



Why do Coastal Seeds Fail?

Maria Jose Gomez Quijano, Briana L. Gross, Julie R. Etterson

28.04.2020

Project Number 17-306-09

NOAA Award NA17NOS4190062

This report was prepared by Maria Jose Gomez Quijano, Briana L. Gross and Julie R. Etterson using Federal funds under award NA17NOS4190062 from the Coastal Zone Management Act of 1972, as amended, administered by the Office for Coastal Management, National Oceanic and Atmospheric Administration (NOAA), U.S. Department of Commerce provided to the Minnesota Department of Natural Resources (DNR) for Minnesota's Lake Superior Coastal Program. The statements, findings, conclusions, and recommendations are those of the author(s) and do not necessarily reflect the views of NOAA's Office of Coastal Management, the U.S. Department of Commerce, or the Minnesota DNR.



Final Report

Introduction

In northern Minnesota, the coast of Lake Superior has a significant effect on climate patterns and allows for an extended growing season by producing warmer falls and winters, mild springs, and cooler summers. Coastal forests in Minnesota have both high economic and ecological value because of animal and plant species diversity, tourism, forestry, and the timber industry. Unfortunately, due to rapid climate change, some of the native high value species are declining, leading to the disruption of coastal forests and negatively impacting both the economy and ecosystems. We have undertaken this study to better understand how to preserve and restore Minnesota's coastal forests under the effects of changing climatic conditions. Specifically, our aim is to determine whether coastal Northern red oak populations are genetically differentiated from populations in the rest Minnesota. At present, Minnesota state agencies actively avoid collecting tree seeds from coastal areas for forest regeneration, as observations show that the coastal seeds fail to grow in inland areas, even though there are healthy, seed-producing trees present along the shore. The failure of establishment could be due to local adaptation of the coastal trees, genomic imprinting leading to transgenerational plasticity (i.e., a genetic effect of parental conditions on gene expression in the next generation), or poor provisioning of seeds in the coastal environment.

We used Northern red oak (*Quercus rubra*) as a study system to understand how coastal seed sources might differ from other seed sources in Minnesota because Northern red oak is present along the coast and is predicted to increase in frequency as climate change progresses. Northern red oak is a deciduous tree in the Fagaceae family that it is native to North America and widely distributed through eastern United States and southeastern Canada, inhabiting a diverse range of hardiness zones (3-8) and growing in different soil types such as acidic, loamy, moist, clay, and well-drained. There is evidence from the literature that oak populations can adapt to different environments and potentially to rapid climate change, but population structure of oak populations in coastal vs. inland areas has not yet been studied. In order to study the possible local adaptation of northern red oak to the microclimate by the shore of Lake Superior, we focused on identifying molecular and phenotypic population structure among populations of northern red oak (*Quercus rubra*) from coastal vs. non-coastal regions in Minnesota.

Project Outcomes:

Outcome 1: Genetic Testing – Collect leaf samples from 12-13 individual trees from 30 populations of red oak, 10 from each of the coastal, inland and interior zones. Perform genetic analysis to determine if the tree populations are genetically unique.

Outcome 2: Seed and Germination – Collect twenty mature acorns from each individual tree (approximately 360 trees). Test the acorns for viability. Weigh and plant in the University greenhouse ten acorns per individual in a randomized block design (total of 3,900 seeds). Record germination and juvenile traits such as survival, cotyledon size, leaf number, height and growth rate for six months. Analyze the data to support or refute the hypotheses for why seeds produced by coastal trees differ for germination, survival and other traits.

Outcome 3: Data Share - Develop a technical report presenting the results of both the genetic and the seed/germination trait tests. Share the results with the MNDNR's Silviculture Program, as well as the United States Forest Service, Minnesota Forest Resources Council, and the North Shore Forest Collaborative. Explore other outreach opportunities, including public education in the state parks and presentations to community groups. In addition, the results may be presented at regional and national meetings as well as in research publications. Report must contain acknowledgement of funding.

Work Completed

Outcome 1: Genetic Testing – This grant outcome was met, and we were able to achieve the original goal of determining whether coastal tree populations were genetically unique.

Populations were identified using the Minnesota Department of Natural Resources releve data. We collected samples from three regions within Minnesota, classified as coastal, inland, and interior. Coastal populations were located between 0-10 miles from the coast of Lake Superior, inland populations are located between 11-50 miles from the coast, and interior populations between 51-100 miles from the coast. Leaf tissue was collected from a minimum of 12 individuals in each of 30 different populations, with 10 populations in each region (coastal, inland, interior).

We performed genomic DNA extractions on 358 *Q. rubra* samples and 45 *Q. ellipsoidallis* (used as outgroups) and sent the DNA to the University of Minnesota Genomic Center (UMGC) for RAD-seq. Single nucleotide polymorphisms (SNPs) were identified using two different methods: the [Freebayes](#) pipeline and the [ipyRAD](#) pipeline. We performed all the population genetics analyses using both of the SNP sets to ensure that our results were robust and found no major differences between the results generated from Freebayes vs. the ipyRAD data sets.

Were performed a variety of analyses to determine whether populations were genetically unique. Genetic distance (F_{ST}) was calculated among as well as between populations. Allele frequencies per population were used to construct a neighbor-joining tree of all the *Q. rubra* and *Q. ellipsoidallis* populations. Fine-scale genetic groupings were also examined using a principle component analysis (PCA) and a Bayesian approach to assign each individual to a group based on genetic similarity.

Outcome 2: Seed and Germination – This grant outcome was met in a modified manner, and we were able to use the data we generated to determine whether the coastal populations showed juvenile trait differences compared to non-coastal populations, meeting the original grant goal.

During the fall of 2018, the same *Q. rubra* populations used for leaf samples (outcome 1) were visited for acorn collections. Not all populations were producing acorns, because oak trees generally produce acorns only every 3-5 years. We collected acorns from all productive populations for a total of 1438 acorns across 10 (6 coastal, 3 inland, 1 interior) populations. This was not as many acorns as originally planned in our objective, but it was still a large enough sample size to allow us to compare trait differences in a statistically rigorous manner. We also supplemented this collection during the fall of 2019, by visiting the remaining *Q. rubra* populations, where we collected a total of 2305 acorns across 9 (4 inland, 3 interior, 2 coastal) populations.

Acorns collected in both 2018 and 2019 were used to compare seed mass in coastal vs. non-coastal populations. During the summer of 2019, a fully randomized block design common garden experiment was initiated at the University of Minnesota Duluth greenhouse. The 1438 acorns from 102 maternal lines of the 10 populations collected in fall 2018 were randomly distributed along four blocks. Germination percentage, dates, and seedling height were recorded to quantify juvenile growth. Standard least squares and logistic models were used for seed mass, germination date, growth rate and germination rate to compare populations within each region and between the coastal vs. non-coastal regions.

Outcome 3: Data sharing – This grant outcome was met and we achieved the original grant goal of sharing our data and results.

We developed a technical report presenting the analyses and results of both the genetic and the seed and seedling traits. We shared the results with the MNDNR’s Silviculture Program, as well as the United States Forest Service, Minnesota Forest Resources Council, and the North Shore Forest Collaborative via email. We presented at several outreach events and scientific meetings over the course of the granting period. We made our data publicly available through the XXX database. The technical report and all presentations contained acknowledgement of MLSCP funding.

Outcome Overview – Overall, we were able to complete all the goals in our original grant and meet the outcomes. We did need to modify one of the outcomes (outcome 2) in response to the biology of the species in our study, because the oak populations did not produce as many acorns as we anticipated. Nonetheless, we performed the analyses we needed to evaluate trait differences among populations.

Results

Genetic Testing:

Species genetic assignment - Although our primary interest in this study was the potential for population structure within northern red oak, we also included three northern pin oak (*Q. ellipsoidalis*) populations as reference samples to allow for the detection of hybrids. This was necessary given the overlap between the two species in Minnesota and evidence of *Q. rubra* hybridizing with *Q. ellipsoidalis* in contact zones elsewhere in the Great Lakes region. We identified four *Q. rubra* populations that had *Q. ellipsoidalis* genetic identity. All of our collections were based on the occurrence of *Q. rubra* in MNDNR releve data, and the releve species data did not show any presence of *Q. ellipsoidalis* in any of the four mis-identified populations. One possible explanation for the collection of *Q. ellipsoidalis* at these sites is that trees were misidentified in the original releve survey, but another is that our samples were from a stand of *Q. ellipsoidalis* very close to, but not included in, the original releve. The latter possibility suggests that these would be likely locations of hybrid zones in Minnesota. These four populations were removed from subsequent analyses.

Population differentiation based on neutral molecular markers - Genetic differentiation was low for all of the *Q. rubra* populations regardless of their region; F_{ST} values for all *Q. rubra* populations ranged between 0.02-0.04 (out of a maximum value of 1). All of the principal component analyses (PCA), STRUCTURE analyses, and neighbor joining trees showed two clusters, one of all *Q. ellipsoidalis* and the

second one of all of the *Q. rubra* species, showing no differentiation between coastal, inland, or interior *Q. rubra* populations. When *Q. ellipsoidalis* populations were removed to investigate the clustering of *Q. rubra* populations alone, the PCA and STRUCTURE analyses revealed a large cluster containing most of the coastal, inland and interior populations. The only deviation from this pattern was that most individuals from the HR coastal population and the MG inland populations were slightly diverged from the other populations. Both of these populations are located south of the St. Louis river and worthy of further investigation. Overall, all of the analyses were consistent, showing minimal differentiation between *Q. rubra* populations regardless of their region and no indication that coastal populations were genetically unique based on this measure. The low levels of population differentiation are likely due to the long generation time and extended pollen dispersal range that characterize oaks; both of these factors would ensure the dispersal and maintenance of allelic diversity across populations. The lack of differentiation between regions based on our neutral genetic markers indicates high genetic exchange between populations, but there is still potential for genetically-based phenotypic differences, which we investigated using seed and seedling traits, below.

Seed and Germination:

Genetic differentiation based on phenotypic traits - Despite the low genetic differentiation based on neutral molecular markers between populations and regions, we observed significant genetic differences in seed mass and germination between coastal and non-coastal populations (inland and interior) with our common garden experiment. The coastal region had statistically significantly lower seed mass (p-value= 0.03), earlier germination dates (p-value= 0.0125), and higher germination rates (p-value= 0.0165) than the non-coastal region. These results show that while coastal seeds are smaller, they germinate at a higher rate than non-coastal seeds, indicating that they are unlikely to be poorly provisioned. We also observed 1.5 days earlier germination for coastal populations. It is possible that early germination is adaptive in the coastal environment where seedlings will have low risk of freezing due to the warm spring temperatures produced by the lake, and therefore will have higher establishment success. However, if coastal seeds are planted and grown in inland areas where temperatures are much cooler during the spring months, they could be exposed to freezing temperatures, possibly explaining the failure of establishment that has led foresters to stop accepting these seeds. The observed differences in phenotypic traits provide suggestive evidence that that genes controlling for seed mass and germination might be under natural selection, despite the overall genetic similarities between the populations.

We note that there was also a highly significant variation between populations within each region for every measured trait (p-value < 0.0001), including growth rate. The large variation among populations within each region shows that each individual population had a different response to germination and growth under controlled environmental conditions. For example, the variation in seed mass per population within the coastal region was highly significant. This variation among populations could be due to the wide geographic region across which seeds were collected, spanning several hardiness zones from north to south. Differences in hardiness zones reflects a difference in temperatures during the growing season, and therefore within the coastal region there are *Q. rubra* populations experiencing very different environmental conditions.

Recommendations for coastal management:

Our results show that while coastal red oak populations are not differentiated from inland and interior red oak populations based on neutral genetic markers (genetic testing results), but that they do show unique traits in terms of seed mass, germination percentage, and germination timing (seed and germination results). This combination of results suggest that the coastal oak populations may be uniquely adapted to the coastal climate, although further study is needed to rule out other possible explanations for these patterns, such as transgenerational plasticity. Nonetheless, our results support treating coastal populations as a unique seed collection zone for reforestation efforts, an approach that is being embraced by an increasing number of state and federal agencies. The differences we observed among populations within both the coastal and non-coastal regions also suggest that it is important to sample seeds from a diverse pool of populations to increase the diversity of traits in the seed pool. Similarly, it is important to take into account multiple environmental factors besides seed zones when collecting seeds, such as exact coordinates, temperature, precipitation and other environmental factors across a north-south gradient. These results were made possible entirely by funding from the MLSCP.

Partnerships

In this project, we benefitted from collecting samples in MN State Parks, MN Scientific and Natural Areas, the City of Duluth, and MN State Forests, and one private property, all with permits or permissions from those entities/owners. We regard our access to these properties for sampling one of the keys to this grant's success.

Leveraged Dollars

Additional funds for this project were provided by the UMN-Duluth Swenson College of Science and Engineering and Department of Biology to fund two summer fellowships for undergraduates who worked on this project. Each fellowship was valued at approximately \$4,000 in salary and \$700 for supplies.

Conclusions

The most important lesson from this project was the challenge of incorporating the biological reality of our study system with the timeline of the grant. The first year we went to collect acorns, almost no populations were producing seed. The subsequent year, only 10 populations produced good seed. This is a natural part of oak biology, but it meant that we could not complete our project within the expected timeline. Fortunately, the MLSCP was willing to extend the timeline of the grant and slightly modify one of our outcomes to account for what the oak trees were doing each season. We hope that similar projects in the future will also benefit from such flexibility. We will also plan to incorporate this uncertainty more firmly into our research plans for future grant proposals.

Future Plans

The observed differences in phenotypic traits provide suggestive evidence that there are genetic differences between populations and that genes controlling for seed mass and germination might be under natural selection, despite the homogenizing effect of gene flow. However, in this experiment we were not able to control for transgenerational plasticity, due to the slow growth of *Q. rubra* seedlings. Because of the nature of our experiment, there are other possible explanations for the observed

differences that require further study. In particular, the significant variation between populations within a region for all the traits provides initial evidence that these traits can have a genotype by environment interaction and therefore present an adaptive plastic response. In order to further investigate this interaction, seedlings must be grown under different environmental conditions. If populations respond differently to the treatments but maintain the variation between population, it would indicate an adaptive plastic response of *Q. rubra* seedlings to variable environmental conditions. We were recently awarded a STAR grant funded by the MLSCP to study the growth of the seedling used in this experiment under variable environmental conditions. We will expose the seedlings to different temperatures and water availabilities, to examine growth and survival. This project will allow us to identify genotype by environment interactions, reflecting an adaptive plastic response to these different environmental conditions as well as to reduce environmental carryover effects.

This project fits into a larger effort in the Etterson and Gross labs to understand adaptation of Minnesota plants to different environments. Past work in the Etterson lab has specifically focused on seed-sourcing guidelines, and this work on coastal seed zones will inform this broader research program. Past work in the Gross lab has specifically focused on plants adapted to the coastal environments (arctic relicts), and this work on a more widely distributed species is a complement to that work. In the future all of our results will be submitted as a manuscript to a scientific journal, we will also present our results at conferences in our study field.

Appendices

Appendix 1: The **technical report** (final product) is attached to the submission. The technical report includes the correct acknowledgement and is labeled with project title and number.

Appendix 2: Photos of the greenhouse experiment

Appendix 3: Correspondence from stakeholders in response to sharing our technical report.

Link to data share: <https://conservancy.umn.edu/handle/11299/212844>

Appendix 1: TECHNICAL REPORT

Why do Coastal Seeds Fail?

Population genetic and phenotypic analysis of coastal and non-coastal Northern red oak in Minnesota

Maria Jose Gomez Quijano, Briana L. Gross, Julie R. Etterson

University of Minnesota Duluth
Department of Biology

Project Number 17-306-09
NOAA Award NA17NOS4190062

This report was prepared by Maria Jose Gomez Quijano, Briana L. Gross and Julie R. Etterson using Federal funds under award NA17NOS4190062 from the Coastal Zone Management Act of 1972, as amended, administered by the Office for Coastal Management, National Oceanic and Atmospheric Administration (NOAA), U.S. Department of Commerce provided to the Minnesota Department of Natural Resources (DNR) for Minnesota's Lake Superior Coastal Program. The statements, findings, conclusions, and recommendations are those of the author(s) and do not necessarily reflect the views of NOAA's Office of Coastal Management, the U.S. Department of Commerce, or the Minnesota DNR.



Abstract:

The unique climate of the North Shore region in northern Minnesota is moderated by Lake Superior, resulting in warmer winters, a reduced risk of spring frost, cooler summers, and an extended fall season. Currently, Minnesota state foresters actively avoid collecting tree seeds from this unique climate adjacent to the lake, as observations show that the coastal seeds fail to become established when they are planted inland. Poor performance of coastal seed sources in inland climates suggest that these populations may be genetically differentiated, which could be detected on the basis of molecular genetic population structure or phenotypic divergence. Alternatively, the failure to establish could be driven by environmental carryover from the maternal tree's environment that influences gene regulation in the developing seeds and has maladaptive outcomes in the nonlocal climate. To explore possible explanations for poor performance of coastal seeds, we used neutral molecular markers and a common-garden experiment to examine population differentiation. To examine molecular genetic differentiation, northern red oak (*Quercus rubra*) leaf samples were collected from populations within three categorical distances from Lake Superior (coastal, inland, and interior). We used restriction site associated DNA sequencing (RAD-seq) to identify single nucleotide polymorphisms (SNPs) and examine population structure. To examine phenotypic differentiation, acorns were collected and planted in a fully randomized block design in the greenhouse where germination and juvenile traits were measured for a full season. Both SNP and phenotypic data were analyzed to determine differences among populations corresponding to distance from the lake. Despite the geographical distance between our populations (max = 97.4mi), we found no molecular differentiation among populations. With respect to phenotype, coastal populations had 2% lighter seed that germinated at 7% higher rates, and 1.5 days earlier than non-coastal populations, although 1st year growth rate did not differ. This suggests that while the populations are not differentiated at neutral genetic markers, the populations do differ for key traits which may be associated with genes under natural selection, environmental carryover effects, or genotype by environment interactions.

Introduction:

Within a single species, genetic or environmental variation, as well as biotic and abiotic interactions can lead to differentiation among populations and local adaptation¹. One way to detect differentiation between populations is through molecular genetic analysis of populations structure, which identifies subpopulations or clusters and how these subpopulations are related to each other². With neutral markers, these differences are most strongly related with patterns of gene flow. Genetic differences in phenotypic traits, on the other hand, can be determined by raising populations in a common environment and identifying significant differences among populations³. This type of experiment is especially informative if phenotypic traits are associated with environmental factors, such as a climate gradient^{4,5}. Even more inferential power is obtained by pairing information on gene flow with phenotypic data that identifies important traits for adaptation. Evidence of local adaptation must be obtained experimentally and typically is definitively demonstrated in reciprocal transplant experiments in which the local population has higher fitness in its respective environment compared to other populations growing in the same environment^{4,5}.

Molecular approach to studying population differentiation

Population genetic differentiation can occur because of low gene flow due to geographical distance between populations, or because of adaptation to different environments⁶. Population differentiation can be identified through the study of population structure using molecular markers like single nucleotide polymorphisms (SNPs). High-throughput next generation sequencing techniques such as RAD-seq (restriction associated DNA sequencing) allow for genotyping and identification of SNPs for any species⁷ and does not require prior genomic information for the populations that are being studied⁸. RAD-seq loci can occur in both coding or non-coding regions of the genome, and loci are conserved within or between organisms of closely related species because of the conservation of restriction sites⁸. RAD-seq output can be aligned to a reference genome of the studied or a closely related species, or if a reference genome is not available, a *de novo* assembly can also be performed⁹. RAD-seq data can be used to identify SNPs across the genome, which are used to calculate genomic diversity, F_{ST} values for population differentiation and structure, and evidence of hybridization or introgression between populations or species⁵.

Phenotypic approach to studying population differentiation

Population differentiation can also be observed by raising multiple populations in a common environment and measuring their quantitative traits. If populations are significantly different when environmental effects are equalized, significant phenotypic differences are inferred to be genetically based¹⁰. This classic experimental design has been used to discover population differentiation in a broad range of organisms from plants to lizards¹⁰. These studies have identified phenotypic trait differences that reflected locally adapted populations or populations

with the potential for local adaptation^{11, 12, 13}. Transgenerational plasticity or environmental carryover occurs when the parents can alter the offspring's phenotypic response based on the parental growing environment¹⁴. Transgenerational plasticity can be genetically controlled and has the potential to evolve. However, in some cases transgenerational plasticity can act as “noise” in genetic experiments, giving the impression of genetic differences among populations, while in reality populations are having a response based on their parental environment¹⁴. Removing the effects of different home environments allows genetic differences between populations to be identified and transgenerational plasticity to be ruled out if phenotypic differences are maintained after a long growing period or a refresher generation is performed^{10,15}. Common garden experiments paired with neutral molecular markers can be useful to identify genetic differences and structure between populations grown across an environmental gradient.

Understanding information from these two types of data

Population differentiation can be a consequence of stochastic or adaptive evolutionary processes. Neutral molecular markers can provide insight on stochastic processes such as gene flow or genetic drift. Phenotypic traits on the other hand, provide evidence on how natural selection is acting on adaptive traits in a population. Combining these two techniques allows for the prediction of the ability and potential of species and populations to adapt to a new or changing environment^{16,17}. When these two techniques are combined to identify population structure, there are three possible outcomes that reflect the different effects of the interaction between gene flow and adaptive traits on a population or a species:

- 1) Population structure based on molecular markers and based on phenotypic traits: In this scenario, populations are strongly differentiated because of reduced gene flow either due to environmental constraints or genetic differences that lead to reproductive isolation based on incompatibilities in life-history traits like fecundity or phenology. These genetic differences are the product of strong natural selection on traits that has led to local adaptation^{16,18}.
- 2) Population structure based phenotypic traits but not on molecular markers: In this scenario, populations have high gene flow, usually because of a large dispersal range that is sufficient to overcome genetic drift. However, there can be a genotype by environment interaction allowing individuals to have different phenotypes depending on their growing environment. There can also be cases where natural selection is stronger than gene flow and can act on adaptive traits, giving populations the potential to adapt to a local environment despite high gene flow^{19,20}.
- 3) Population structure based on molecular markers but not on phenotypic traits: In this scenario, populations are differentiated and reproductively isolated due to low gene flow, reduced dispersal, genetic drift such as a founder event, convergent phenotypic evolution, or divergent evolution but phenotypic maintenance. All of these evolutionary processes can lead to what is known as cryptic population structure, where distinct genetic lineages are phenotypically identical^{21,22}.

The importance of climate adaptation

In this study, we are examining the climatic impact of Lake Superior on tree populations that occur on the coast and further inland. Coastal forest in Minnesota have both high economic and ecological value because of animal and plant species diversity, tourism, forestry, and the timber industry. Unfortunately, due to rapid climate change, some of the native high-value species are declining, leading to the disruption of coastal forests and an increase of monocultures from species like red maple (*Acer rubrum*) or invasive species, negatively impacting both the economy and ecosystems. We have undertaken this study to better understand how to preserve and restore Minnesota's coastal forests under the effects of changing climatic conditions. Specifically, our aim is to determine whether populations of coastal plants are genetically differentiated from their inland counterparts, therefore requiring special attention for seed sourcing during restoration.

Large bodies of water influence and moderate the local climate, producing distinct local microclimates^{24,25}. In northern Minnesota, the coast of Lake Superior experiences effects similar to ocean coasts, as the lake itself basically resembles an inland sea. 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 shore²⁶. 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^{25,27,28,29}. For example, on the shore of Lake Superior, there is evidence of arctic-alpine relic plant species such as *Euphrasia hudsoniana* that have persisted in-place since the last glacial maximum, even though the bulk of the species range is distributed in northeast Canada. Populations in Minnesota have been preserved along bedrock and cliff areas of the lake shore due to the unique environmental conditions found there³⁰.

Minnesota State Seed Sourcing Policy near Lake Superior

At present, Minnesota State agencies actively avoid collecting tree seeds from coastal areas for forest regeneration, as observations show that the coastal seeds fail to grow in inland areas of the state (Fig. 1). This is despite the existence of apparently healthy, seed-producing trees along the shore. The failure of establishment could be due to local adaptation of the coastal trees, poor provisioning of seeds in the coastal environment, or genomic imprinting due to transgenerational plasticity. Previous studies have shown adaptive transgenerational plasticity in tree species, where phenotypic traits related to fitness are influenced by the maternal environment and are found to be adaptive when offspring are grown in their maternal environment³¹. In previous years there was a disparity between the seed zones that the Minnesota DNR recognized and the seed zones the USDA Forest Service recommends. The MN DNR suggested that the coast of Lake Superior had two seed zones that span across the north-east and the north-central regions of

the state (Fig. 1a)³². A recent study also showed the need for seed zone flexibility as they observed that plant adaptations align with changing climate conditions rather than with the established fixed seed³³. Currently, the MN DNR is working towards changing their seed zones in the state of Minnesota to resemble the 2019 USDA Forest seed zone map in which the coast of Lake Superior is considered a unique zone (Fig. 1c)³⁴. The effort to bridge the gap between seed zones reinforces the suggestive evidence of differences in coastal and inland environments and organisms, which hints at the possibility for local adaptation. However, until further investigation on these populations is completed, poor seed provisioning and genomic imprinting in coastal areas cannot be ruled out. This study combines population genetic data and phenotypic measurements from a common garden to examine whether coastal populations are differentiated from non-coastal populations.

The study species

Northern red oak (*Quercus rubra*) is a deciduous tree in the Fagaceae family that is native to North America and widely distributed through eastern United States and southeastern Canada³⁵, inhabiting a diverse range of hardiness zones (3-8)³⁶ and growing in different soil types such as acidic, loamy, moist, clay, and well-drained³⁷. 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 mammals³⁸. 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 year³⁸. Like other oak species, *Q. rubra* is prone to interspecific hybridization^{39,40}; it has been shown to hybridize with other species in the subgenus *Erythrobalantus* such as bear oak (*Q. ilicifolia*), willow oak (*Q. phellos*), black oak (*Q. velutina*), swamp red oak (*Q. shumardii*), shingle oak (*Q. imbricaria*), blackjack oak (*Q. marilandica*) and northern pin oak (*Q. ellipsoidalis*)^{39,40}. Of these species, *Q. rubra* and *Q. ellipsoidalis* overlap in the Great Lakes region (including Wisconsin, Michigan, and Minnesota) and several studies using morphometric traits and genetic markers have shown hybridization between northern red oak (*Q. rubra*) and its sister species northern pin oak (*Q. ellipsoidalis*) in the upper peninsula of Michigan and the Apostle Islands^{41,42,43}. The state of Minnesota is part of the natural distribution of both *Q. rubra* and *Q. ellipsoidalis*, which could lead to gene flow between species in areas where the two species overlap.

There is evidence from the literature that oak populations can adapt to different environments and potentially to rapid climate change^{44,11,45,46}, but population structure of oak populations in coastal vs. inland areas has not yet been studied. In order to study the possible local adaptation of northern red oak to the microclimate by the shore of Lake Superior, we focused on identifying molecular and phenotypic population structure among populations of northern red oak (*Quercus rubra*) from three different regions in Minnesota. The three main questions were addressed in this study are:

- (1) Are the Minnesota red oak populations hybridizing with pin oak? Interspecies hybridization can be a source of population differentiation, so it was important to determine whether this was driving differences between coastal and non-coastal populations by including *Q. ellipsoidalis* samples in our molecular genetic analyses.
- (2) Is there population differentiation based on neutral molecular markers between red oak populations of coastal, inland and interior regions of the state of Minnesota? This question was addressed by measuring population structure using RAD-seq data from populations across three different regions. We hypothesized that if trees are adapted to their local environment, then coastal populations will be genetically unique when compared to inland and interior populations.
- (3) Are there genetic differences based on germination and juvenile phenotypic traits of red oak seedlings from coastal vs non-coastal populations? We measured germination rates, germination dates and phenotypic traits such as seed mass and growth rate of oak seedlings in the greenhouse. We hypothesized that if trees are adapted to their local environment, coastal populations will germinate earlier because of higher spring temperatures by the shore due to the moderating effect of Lake Superior.

Materials and Methods:

Population sampling and plant material

Populations for seed and tissue collections were identified using the Minnesota Department of Natural Resources releve data. We collected samples from three regions within Minnesota, classified as coastal, inland, and interior. Coastal populations were located between 0-10 miles from the coast of Lake Superior, inland populations are located between 11-50 miles from the coast, and interior populations between 51-100 miles from the coast. For the molecular marker analysis, leaf tissue from *Q. rubra* trees was collected from 30 different populations, with 10 populations in each region (coastal, inland, interior) and 15-18 individuals in each population during the summer of 2018 (Fig. 2, Table 1). Leaf tissue from three populations of *Q. ellipsoidalis* in Minnesota were also collected to identify misclassified samples and evaluate the potential for hybridization between *Q. rubra* and *Q. ellipsoidalis* (Fig. 2, Table 1). All the collected leaf tissue was stored at -80°C until genomic DNA extractions were performed.

During the fall of 2018, the same *Q. rubra* populations were visited for acorn collections. Like many oak species, *Q. rubra* trees only produce acorns every 3-5 years, and therefore not all the visited populations produced acorns. In 2018, acorns were produced at 10 of the 30 populations (6 coastal, 3 inland, 1 interior) (Table 1). In each population 10-15 trees were sampled, and 20-30 acorns were collected per tree depending on availability, for a total of 1438 acorns across the 10 populations. Collected acorns were floated to distinguish viable from non-viable acorns, and viable acorns were weighed and then stratified for 3 months in a cold room until planting. During the fall of 2019, the remaining *Q. rubra* populations were visited and acorns were collected from

an additional 9 populations (4 inland, 3 interior, 2 coastal) (Table 1). In each population 10-15 trees were sampled, and 20-50 acorns were collected per tree depending on availability, for a total of 2305 acorns across the 9 populations. Collected acorns were floated to distinguish viable from non-viable acorns and viable acorns were weighed.

DNA extractions and Illumina sequencing

A modified CTAB extraction⁴⁷ was performed for at least 12 samples per population using 0.1g of 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 Northern red oak tissue⁴⁸. Extracted DNA was quantified using a Qubit 3.0 fluorometer, any samples with DNA concentrations less than ~10 ng/μl were excluded from sequencing. Genomic DNA of 358 *Q. rubra* samples and 45 *Q. ellipsoidalis* were sent to the University of Minnesota Genomic Center (UMGC) for library preparation and Illumina sequencing, RAD libraries were prepared using BamHI + NsiI restriction enzymes, and 412 dual-indexed 150bp libraries were generated. All generated libraries were combined into a single pool and sequenced on a NextSeq High-output 1x150-bp flow cell. Approximately 270M total reads were generated, with a mean yield of 650K reads per library.

SNP calling

SNPs were identified using two different methods: the Freebayes pipeline and the ipyRAD pipeline. The Freebayes⁴⁹ pipeline was performed by the UMGc with the *Quercus lobata* reference genome used to call variants. The pipeline used by the UMGc has a hardcoded maximum number of reference genome scaffolds or chromosomes and the *Q. lobata* reference genome exceeded that maximum; therefore, scaffolds were concatenated into a lower number of super-scaffolds, generating a stitched genome that was then used to call variants and further downstream analyses. The second method for data analysis was the ipyRAD pipeline⁵⁰. For this method the stitched *Q. lobata* reference genome was used. The ipyRAD workflow consists of sorting and filtering raw reads, mapping the filtered reads and clustering matched reads based on sequence similarity, estimating heterozygosity, and using that estimate to estimate consensus allele sequences as SNPs. Once the allele sequences are estimated, ipyRAD clusters the consensus allele sequences across samples, aligns them and generates multiple output formats for downstream analyses. After the ipyRAD pipeline was used, individual samples that had a low number of consensus reads or an estimated error rate above 0.0036 were removed from further analyses. A second round of filtering was performed on the UMGc and on our ipyRAD generated data using VCFtools⁵¹ to remove sites with a phred quality score lower than 20 and fewer than 95% genotypes called. Genotypes with a minimum depth of below 5 and a mean depth below 15, individuals with more than 25% missing data, and all indels in the dataset were also filtered out and removed from further analyses. We performed all the analyses using both of the SNP sets and found no major differences between the results generated from Freebayes vs. the ipyRAD data sets, indicating that the method of SNP identification did not influence our

results. Below, we present and discuss the results from the SNPs generated by ipyRAD pipeline; results from the Freebayes pipeline are included as an appendix (Fig. S1-4).

Molecular population structure analyses

Genetic variation within and among populations was characterized using the R package Adegenet⁵². Allele frequencies for each population were calculated and used to estimate observed heterozygosity (H_o) and expected heterozygosity (H_e). Pairwise genetic differentiation (F_{ST}) among as well as between populations was calculated using the R package Higherfstat⁵³ and Adegenet⁵². Allele frequencies per population were used to construct a neighbor joining tree of all the *Q. rubra* and *Q. ellipsoidalis* populations using the R package Ape⁵⁴. Genetic groupings were also analyzed using Identity-By-Descent measures of the SNP data to construct principle component analyses with the R package SNPrelate⁵⁵. Population structure analysis was performed using a Bayesian approach on the multilocus genotype data to assign each individual to a cluster group using the program STRUCTURE 2.3.4⁵⁶. To determine the numbers of clusters (K) that best fitted the data, four iterations of each K (2-6) were performed with burn-in period of 10,000 and 20,000 Markov chain Monte Carlo reps after the burn-in period. The software Structure Harvester⁵⁷ was used to visualize likelihood values and (ΔK) across the multiple values of K using the Evanno method and determine the number of clusters that best fit the data. The shiny web app pophelper⁵⁸ was used to graphically display population clustering.

Phenotypic population structure analyses

During the summer of 2019, a fully randomized block design common garden experiment was initiated at the University of Minnesota Duluth greenhouse (46°49'00.7"N 92°05'12.1"W). 1438 acorns from 102 maternal lines of the 10 populations collected in fall 2018 were randomly distributed along four blocks. Each of the experimental blocks contained acorns from each maternal line and each population in order to have a balanced fully randomized design. All the seedlings were kept at the same temperature (53°C min – 68°C max) and were watered every other day for the duration of the experiment. Germination dates were recorded every other day until 80% of the seedlings had emerged and then weekly until the end of the experiment. Height was measured to quantify juvenile growth and was measured weekly for the first five weeks of the experiment and then every other week until the end of the experiment due to the slow growth of the seedlings. Height measurements were used to calculate growth rate for all the seedlings using the formula $\left(\frac{\text{Final height (cm)} - \text{Initial height (cm)}}{N_{weeks}} \right)$ to estimate the total increase in growth during the length of the experiment.

Standard least squares models were used for seed mass, germination date and growth rate to estimate the variation in each trait within and between each region (coastal vs. non-coastal). Seed mass was log transformed in order to achieve normality of the data and follow the statistical model assumptions. For the germination dates and growth rates analysis, seed mass (untransformed) was used as a covariate, with block, region and population nested within region

were used as fixed factors. For germination rates, seeds that germinated were coded as 1 and seeds that did not germinate were coded as 0; the data was used for a logistic regression model using weight as a covariate, with block, region and population nested within region used as fixed factors. Two coastal populations, “AA” (n = 2) and “ET” (n = 3), were removed as they destabilized the analysis because of insufficient sample size. We observed high variation between populations within each region, which was accounted for when building graphs using the least square mean and standard errors for every model rather than with the arithmetic mean. The least square means takes into account the variation explained by the covariate, the random factor and the nested fixed factors in the model. All statistical models were done in JMP Pro 14 software⁵⁹ and all graphs were constructed in R version 1.1.456⁶⁰.

Results:

Molecular population genetic differentiation

After filtering all the SNPs with ipyRAD, we retained 11,325 loci, 22,854 alleles. We performed a preliminary STRUCTURE analysis (K =2) to confirm the genetic identity of all the *Q. rubra* and *Q. ellipsoidalis* populations. Our results showed four *Q. rubra* populations (SP, BL, GW and II) that had *Q. ellipsoidalis* (pin oak) genetic identity (Fig. S5). These four populations were removed from subsequent analyses, as they were likely *Q. ellipsoidalis* populations that were misidentified as *Q. rubra* during tissue collection.

After the removal of the four populations there were 341 individuals remaining. Population-wide genetic differentiation estimates including *Q. rubra* and *Q. ellipsoidalis* were low, with $F_{ST} = 0.0510$. Genetic differentiation was highest for *Q. ellipsoidalis* populations when compared to the *Q. rubra* populations (Fig. 3). Pairwise F_{ST} was also calculated comparing each *Q. rubra* population with one another. Pairwise genetic differentiation was low for all of the *Q. rubra* populations regardless of their region. F_{ST} values for all *Q. rubra* populations ranged between 0.02 -0.04, where the TG, MG, HR and FF populations that had the higher pairwise F_{ST} values (Fig. 4). The F_{ST} values show low genetic differentiation within species but high genetic differentiation between species. The principal component analyses (PCA) showed two clusters, one of all *Q. ellipsoidalis* and the second one of all of the *Q. rubra* species, showing no differentiation between coastal, inland, or interior *Q. rubra* populations (Fig. 5A). Next, *Q. ellipsoidalis* populations were removed to investigate the clustering of *Q. rubra* populations alone. The PCA revealed a large cluster that contained most of the coastal, inland and interior populations. However, most individuals from the HR coastal population and the MG inland populations were distributed across PC1 and PC2 diverging from the main observed cluster (Fig. 5B). The HH and LE interior populations also diverged to a smaller extent from the large cluster. The principal component analysis was consistent with the F_{ST} statistics, showing minimal population differentiation between *Q. rubra* populations regardless of their region.

In accordance with the two species and the three hypothesized regions used for this study, our population structure analysis was performed using ranging values of $K = 2-4$. The $K = 2$ clustering was used to confirm species genetic assignment; at this level of K all the *Q. rubra* individuals grouped in one cluster with all the *Q. ellipsoidalis* individuals in the second cluster. Shared ancestry but no evidence of hybridization between the two species was observed, as the *Q. rubra* populations individuals all shared a slight, uniform level of *Q. ellipsoidalis* identity and vice versa (Fig. 6A). For the higher levels of division ($K = 3$ and $K = 4$), we observed no differentiation between populations from coastal, inland or interior regions (Fig. 6B - C). However, for $K = 3$, we observed that the HR and the MG populations diverged slightly from the other populations. The coastal population HR has a higher number of individuals that have identity of the third cluster, while the MG inland population had a higher number of individuals that little to no identity in the third cluster (Fig. 6B). This slight differentiation of HR and MG is consistent with the patterns we observed in the PCA analysis (Figure 5B). The neighbor-joining tree based on allele frequencies of *Q. rubra* and *Q. ellipsoidalis* resulted in two distinct clades, one containing all the *Q. rubra* populations regardless of their region, and another one containing all the *Q. ellipsoidalis* populations (Fig. 7).

Phenotypic population differentiation

The seedlings from coastal and non-coastal populations showed differences in mass and germination when grown in the common garden. The standard least squares analysis showed a significant difference in seed mass between regions and between populations nested within each region for the seeds that we collected in 2018 and planted in the common garden during the Summer of 2019 (Table 2). The coastal region seed mass (LS $\mu = 4.27\text{g}$) was lower and significantly different than for the non-coastal region seed mass (LS $\mu = 4.32\text{g}$). Within the coastal region, population JC had lower mass (LS $\mu = 3.42\text{g}$) than the MS (LS $\mu = 4.35\text{g}$) and TG (LS $\mu = 5.04\text{g}$) populations (Fig. 8A). For the seeds collected during the fall of 2019 that were not used for the common garden experiment, we also observed significant difference between the coastal and non-coastal regions as well as significant differences between the populations within each region (Table 2). The coastal region (LS $\mu = 3.57\text{g}$) again had significantly lower seed mass than the non-coastal region (LS $\mu = 3.73\text{g}$). However, in this case most of the population variation was observed between populations within the non-coastal region. Here, the EV population had lower seed mass (LS $\mu = 3.04\text{g}$) than all the other non-coastal populations (Fig. 8B). 2018 seed mass was used as a covariate and was significantly different for all the further analyses.

The germination phenology analysis showed that coastal and non-coastal regions had significantly different germination dates. Coastal populations (LS $\mu = 148.7$ JDN) germinated on average two days earlier than non-coastal populations (LS $\mu = 150.2$ JDN). Populations nested within region were also significantly different. Within the coastal region, the TG population germinated two days earlier (LS $\mu = 145.9$ JDN) than the coastal average, having the earliest

germination date (May 25, 2019) in the common garden experiment. Within the non-coastal region, the SC population germinated two days later (LS $\mu = 152.8$ JDN) than the non-coastal average, having the latest germination date (June 1, 2019) in the common garden experiment (Fig. 8C). Block was also a significantly different factor.

Percent germination analysis showed a significant difference between regions. The coastal region germination rate was 79% while the non-coastal region germination rate was 72%. Percent germination for population nested within region was also a significant factor. The non-coastal region had significant variation between populations, with a 30% difference between the lowest and the highest germination rates. Two non-coastal populations, NJ and SC, had the lowest germination rates of the experiment with a 54% and 61% percent germination respectively. Percent germination between populations within the coastal region only had a 10% difference between the lowest and the highest germination rates (Fig. 8D).

Growth rate as a measure of juvenile phenotypic trait showed a non-significant difference between coastal and non-coastal regions. However, there was a significant difference in growth rate for populations nested within region. Populations from the coastal region had more uniform growth rate (LS $\mu = 1.148$ cm/week), with a difference in height of 0.23 cm/week between the lowest and the highest coastal population. Populations from the non-coastal region had more variation in growth (LS $\mu = 1.152$ cm/week), with a difference of 0.64 cm/week between the lowest and the highest non-coastal population (Fig. 8E). Block was also a significantly different factor.

Discussion:

Anecdotal accounts from foresters suggest that some tree species used for restoration could be locally adapted to the microclimate near Lake Superior. Minnesota state agencies have observed that coastal seeds fail to grow in inland areas, which has led to the avoidance of coastal seed collections in the past. Currently, new seed sourcing policies are being considered that recognize this region as a unique seed zone, which is parallel to the new U.S. Forest Service seed transfer guidelines³⁴. In addition, there is mounting evidence in the literature from studies that combine molecular marker and phenotypic data that reinforce the possibility of local adaptation in oak species based on one or both sources of information^{11-13,44-46}. For example, studies on *Q. oleoides*, found no molecular population structure and low F_{ST} values, but also showed significant phenotypic differentiation in functional traits related to drought tolerance and environmental variability^{12,13}. Other studies performed on *Q. lobata* have found high F_{ST} values at some candidate gene loci related to flowering and temperature stress genes, indicating potential loci under diversifying selection between populations, and indicating *Q. lobata*'s ability to respond to changing climate conditions⁶¹.

Population differentiation based on neutral molecular markers

Although our primary interest in this study was the potential for population structure within northern red oak, we also included three northern pin oak (*Q. ellipsoidalis*) populations as reference samples to allow for the detection of hybrids. This was necessary given the overlap between the two species in Minnesota and evidence of *Q. rubra* hybridizing with *Q. ellipsoidalis* in contact zones elsewhere in the Great Lakes region^{41,42,43}. The inclusion of these reference populations allowed us to identify four putative *Q. rubra* populations in our original collections (SP, BL, GW, II) that were genetically *Q. ellipsoidalis* in identity. We were also able to confirm the species identity of all of the remaining *Q. rubra* populations. All of our collections were based on the occurrence of *Q. rubra* in MN DNR releve data, and the releve species data did not show any presence of *Q. ellipsoidalis* in any of the four misidentified populations. For the SP (coastal) and II (interior) populations there was evidence of bur oak (*Q. macrocarpa*) presence at the sites. However, there are vast morphological differences between *Q. rubra* and *Q. macrocarpa* which eliminates the possibility that the collected samples were *Q. macrocarpa* individuals. One possible explanation for the collection of *Q. ellipsoidalis* at these sites is that trees were misidentified in the original releve survey, 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 these would be likely locations of oak hybrid zones in Minnesota. In either case, the genetic assignment of these populations to *Q. ellipsoidalis* rather than *Q. rubra* requires further investigation in future studies through fine-scale mapping and genetic identification of the trees in these populations.

Outside of these four mis-identified populations, we did not detect evidence of hybrid zones between the *Q. rubra* and *Q. ellipsoidalis* in the remaining populations, despite evidence for hybridization in the literature. The STRUCTURE analysis did reveal shared ancestry between the two species, which is manifest as a slight reciprocal genomic contribution from each species group to the other that stays relatively constant across the populations. This shared ancestry is in accordance with the close taxonomic relationship and likely long-term gene flow between the species, which is common in oaks. However, we note that we collected samples from mature trees for this study, and thus we cannot rule out the potential for the presence of hybrids at the seedling stage that might later be selected against.

We did not detect genetic differentiation among *Q. rubra* populations according to either distance from the lake or other factors based on our clustering analyses (STRUCTURE, PCA, neighbor-joining tree) or F_{ST} values. The low levels of population differentiation are likely due to the long generation time and extended pollen dispersal range (recorded to be up to 100 km⁶²) that characterize oaks; both of these factors would ensure the dispersal and maintenance of allelic diversity across populations. Indeed, our calculated pairwise F_{ST} values are similar to the ones found for *Quercus* and other tree species such as *Q. oleoides* ($F_{ST} = 0.09$)¹³, *Q. robur* ($F_{ST} = 0.07$)⁶³, *Eucalyptus albens* ($F_{ST} = 0.018$)⁶⁴ and *Pinus taeda* ($F_{ST} = 0.04$)⁶⁵. Such low levels of F_{ST}

indicates low genetic differentiation between populations consistent with panmixia, a sign of high levels of gene flow between populations despite the geographical distance. We did identify two populations MG (interior) and HR (coastal), located south of the St. Louis river, that showed slightly higher population genetic differentiation based on all of our analyses. The observed differences in these two populations could be signs of inbreeding or early reproductive isolation stages due to geographical location, which could be reducing the amount of gene flow between populations. Other than these minor differences, we did not detect population structure based on molecular markers. However (as noted in the introduction) that does not mean the populations do not have the potential to adapt to different environments, as strong natural selection can outcompete the homogenizing effect of gene flow.

Genetic differentiation based on phenotypic traits

Despite the low genetic differentiation based on neutral molecular markers between populations and regions, we observed significant genetic differences in seed mass and germination between coastal and non-coastal populations (inland and interior) with our common garden experiment. The coastal region had significantly lower seed mass, earlier germination dates and higher germination rates than the non-coastal region, but there was also a highly significant variation between populations within each region. The large variation between populations within each region shows that each individual population had a different response to germination and growth under controlled environmental conditions.

The observed significant differences between regions and populations suggest that traits related to germination success and establishment might be under natural selection given the unique environmental conditions exhibited by the coast of Lake Superior. The moderating effect of Lake Superior produces warmer spring temperatures (March-May) by the shore and cooler temperatures in inland areas. Our results showed overall high germination success under controlled environmental conditions, with over 70% germination success for both coastal and non-coastal regions. Moreover, coastal populations had a significantly higher germination rate than inland populations, indicating that coastal populations are unlikely to be poorly provisioned. We also observed earlier germination dates for coastal populations, and later germination dates for non-coastal populations. This early germination date may give *Q. rubra* coastal seedlings an advantage in coastal environments where spring temperatures are mild and there is low risk of freezing, therefore increasing their establishment success. However, if coastal seeds are planted and grown in inland areas where temperatures are much cooler during the spring months, they could be exposed to freezing temperatures after germinating early, possibly explaining the failure of establishment that has led to the avoidance of coastal seeds in the past.

Multiple studies have shown the relationship between seed mass and seedling performance, where larger seeds generally produce larger seedlings, have higher germination rates and are more stress resistant^{66,67}. For this reason, seed mass was included as a covariate in all our

analyses. We observed 2% lower seed mass in 2018 and 4% lower seed mass in 2019 for coastal populations than for non-coastal populations and found that the variation in seed mass per population within the coastal region was highly significant. This variation among populations could be due to population location; despite the fact that all the coastal populations are located within 10 miles of the lake shore, the populations were located in northern or southern coastal areas, and therefore fell in different hardiness zones. The difference in hardiness zones reflects a difference in temperatures during the growing season, and therefore within the coastal region there are *Q. rubra* trees experiencing different environmental conditions, which can affect the resource allocation and production of acorns. Therefore, it is important to acknowledge that some of the observed phenotypic differences are not driven by the region these populations belong to but rather by other environmental factors, like temperature or annual precipitation.

The observed differences in phenotypic traits provide suggestive evidence that there are genetic differences between populations and that genes controlling seed mass and germination might be under natural selection, despite the homogenizing effect of gene flow. However, in this experiment we were not able to control for transgenerational plasticity, due to the slow growth of *Q. rubra* seedlings, which makes a refresher generation practically impossible. Transgenerational plasticity (also known as environmental carryover) can affect the expressed phenotype of an individual grown in a common garden, because the effect of maternal growing environment might persist even when the seed is grown in different conditions⁶⁸. Environmental carryovers can last up to the late seedling stage, which in *Q. rubra* individuals can take up to five years⁶⁸. In order to eliminate environmental carryovers while taking into account the impracticality of a refresher generation, the seedlings could be grown for multiple seasons to provide a clearer evaluation of juvenile growth traits and fitness. If, after multiple years of growth in a common garden the observed genetic differences are maintained, it can be concluded that coastal populations are locally adapted to the lake shore environment.

Our primary interest in this experiment was to identify differences in germination rates and juvenile traits in *Q. rubra* seedlings. Because of the nature of our experiment and the long lifespan of the species, there are other possible explanations for the observed differences that cannot be explained by our experimental design and require further study. The significant variation between populations within a region for all the traits provides initial evidence that these traits can have a genotype by environment interaction, and therefore present an adaptive plastic response. In order to further investigate this interaction, seedlings must be grown under different environmental conditions. If populations respond differently to the treatments but maintain the variation between population, it would indicate an adaptive plastic response of *Q. rubra* seedlings to variable environmental conditions. On the other hand, the observed phenotypic differences could be due to gene interactions such as epistasis or gene regulation that leads to the differential expression of genes related to seed mass, germination and early seedling establishment.

Conclusion

In this study, we evaluated population differentiation of *Q. rubra* populations across the state of Minnesota using molecular markers and quantitative traits. We based our study hypotheses on state agencies observations about the failure of establishment of coastal seeds when planted in more inland areas within its same seed zone. We observed no molecular population structure but significantly phenotypic differences between regions and populations. Our results suggest that while the populations are connected by high levels of gene flow, they still differ for key traits regarding germination success, which could be under natural selection due to the moderating effect of Lake Superior. As a suggestion for future seed collections, it is important to sample seeds from a diverse pool of populations to increase the germination rates. Similarly, it is important to take into account multiple environmental factors besides seed zones when collecting seeds, such as exact coordinates, temperature, precipitation and other environmental factors. These environmental variables can better explain the seed sourcing when paired with the seed zone where collections are performed, given the fact that climate is rapidly changing and that the lake effect can create microclimates along the shore affecting growth and survival of young *Q. rubra* individuals.

Data share: <https://conservancy.umn.edu/handle/11299/212844>

Acknowledgements:

This report was prepared by Maria Jose Gomez Quijano, Briana L. Gross and Julie R. Etterson using Federal funds under award NA17NOS4190062 from the Coastal Zone Management Act of 1972, as amended, administered by the Office for Coastal Management, National Oceanic and Atmospheric Administration (NOAA), U.S. Department of Commerce provided to the Minnesota Department of Natural Resources (DNR) for Minnesota's Lake Superior Coastal Program. The statements, findings, conclusions, and recommendations are those of the author(s) and do not necessarily reflect the views of NOAA's Office of Coastal Management, the U.S. Department of Commerce, or the Minnesota DNR.

Tables and figures:

Table 1. Northern red oak and Northern pin oak collection sites and tissues collected.

Species name	Population Name	Region	Latitude	Longitude	distance from the lake (mi)	Tissue Collected
<i>Q. rubra</i>	ET	Coastal	46.776132	-92.12431	1.28	Leaf / Acorns 2018
<i>Q. rubra</i>	MM	Coastal	46.8843774	-91.975458	1.65	Leaf / Acorns 2019
<i>Q. rubra</i>	HY	Coastal	46.8303967	-92.08472	1.91	Leaf
<i>Q. rubra</i>	AA	Coastal	47.428242	-91.20655	2.60	Leaf
<i>Q. rubra</i>	TG	Coastal	47.3368634	-91.270446	2.66	Leaf / Acorns 2018
<i>Q. rubra</i>	GG	Coastal	47.44995	-91.2339	6.10	Leaf / Acorns 2019
<i>Q. rubra</i>	MS	Coastal	46.6952234	-92.231578	8.73	Leaf / Acorns 2018
<i>Q. rubra</i>	ST	Coastal	46.673576	-92.273644	10.3	Leaf
<i>Q. rubra</i>	JC	Coastal	46.642718	-92.322142	10.5	Leaf / Acorns 2018
<i>Q. rubra</i>	HR	Coastal	46.668554	-92.351586	10.8	Leaf
<i>Q. rubra</i>	EE	Inland	47.31909	-91.48048	11.7	Leaf
<i>Q. rubra</i>	MG	Inland	46.617996	-92.316025	14.3	Leaf / Acorns 2019
<i>Q. rubra</i>	CC	Inland	46.702444	-92.638807	26.0	Leaf / Acorns 2019
<i>Q. rubra</i>	NJ	Inland	46.4077174	-92.371857	27.0	Leaf / Acorns 2018
<i>Q. rubra</i>	SL	Inland	46.8388372	-92.719922	30.8	Leaf / Acorns 2019
<i>Q. rubra</i>	FF	Inland	46.247724	-92.3913	36.6	Leaf / Acorns 2018
<i>Q. rubra</i>	SC	Inland	46.1041104	-92.478812	47.4	Leaf / Acorns 2018
<i>Q. rubra</i>	BS	Inland	46.2071725	-92.819646	51.0	Leaf / Acorns 2018
<i>Q. rubra</i> *	SP	Inland	46.7785583	-93.228834	51.0	Leaf / Acorns 2019
<i>Q. rubra</i>	EV	Inland	47.479658	-92.517111	51.4	Leaf / Acorns 2019
<i>Q. rubra</i> *	BL	Interior	47.96372	-92.000742	56.8	Leaf
<i>Q. rubra</i>	HH	Interior	46.8718874	-93.325495	59.0	Leaf
<i>Q. rubra</i>	GL	Interior	46.1243969	-92.997152	61.7	Leaf / Acorns 2018
<i>Q. rubra</i>	KW	Interior	46.541621	-93.373846	62.8	Leaf
<i>Q. rubra</i>	LE	Interior	47.6920825	-92.629149	65.1	Leaf
<i>Q. rubra</i>	CS	Interior	45.853561	-92.748816	68.5	Leaf
<i>Q. rubra</i>	SS	Interior	46.31162	-93.352731	68.7	Leaf / Acorns 2019
<i>Q. rubra</i> *	GW	Interior	47.694641	-92.81142	69.6	Leaf
<i>Q. rubra</i>	ML	Interior	46.16718	-93.778303	91.7	Leaf / Acorns 2019
<i>Q. rubra</i> *	II	Interior	45.951059	-92.662983	97.4	Leaf
<i>Q. ellipsoidalis</i>	HA	Pin oak	45.38362	-93.16834	108.0	Leaf
<i>Q. ellipsoidalis</i>	WO	Pin oak	45.21747	-92.77972	111.0	Leaf
<i>Q. ellipsoidalis</i>	AR	Pin oak	44.86216	-93.61626	158.0	Leaf

*Populations removed from analyses

Table 2. Least square analyses and logistic regression analyses of phenotypic traits measured from the fully randomized block design common garden experiment. **Accessibility note:** There are merged cells in this table. The heading “Block” refers to columns 2-4, the heading “Region” refers to columns 5-7, the heading “Population(region)” refers to columns 8-10, and the heading “Seed mass” refers to columns 11-13.

	Block			Region			Population(region)			Seed mass		
	df	F / χ^2	p	df	F / χ^2	p	df	F / χ^2	p	df	F / χ^2	P
Seed mass 2018 (log)	-	-	-	1	4.52	0.03	6	24.99	<0.0001	-	-	-
Seed mass 2019 (log)	-	-	-	1	8.87	0.0029	7	43.42	<0.0001	-	-	-
Germination phenology	3	6.81	0.0002	1	6.25	0.0125	6	11.37	<0.0001	1430	10.39	0.0013
% germination	3	5.74	0.124	1	5.74	0.0165	6	72.15	<0.0001	1430	30.22	<0.0001
Growth rate	3	22.06	<0.0001	1	0.0045	0.9463	6	12.91	<0.0001	1430	61.01	<0.0001

Bolded values represent statistically significant <0.05 p-values.

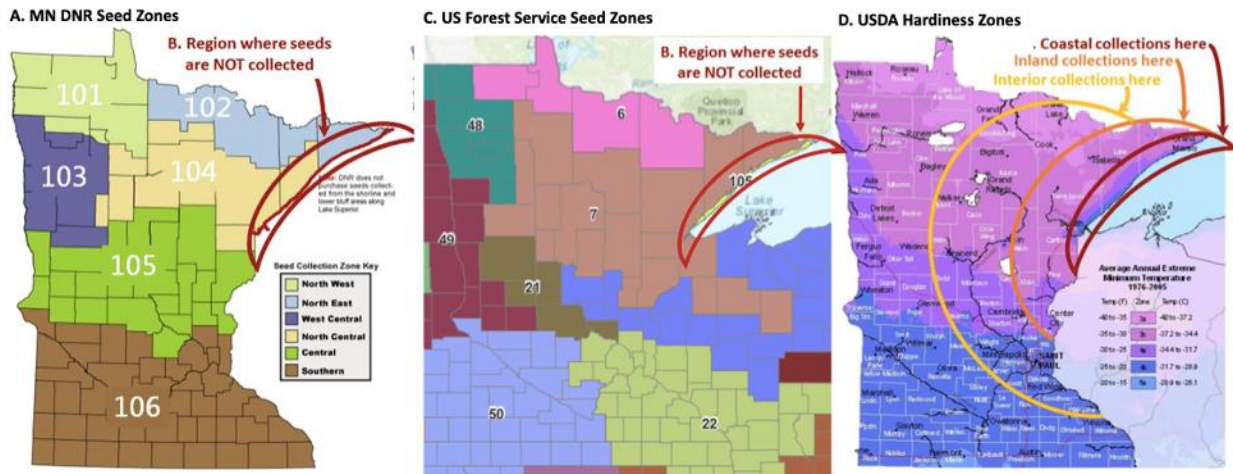


Figure 1. MN seed transfer zones and hardiness zones. A) Seed transfer zones for the state of Minnesota according to the Minnesota DNR. B) coastal region where seeds are not collected because of poor establishment. C) Seed transfer zones for the state of Minnesota according to the USDA Forest Service Eastern Seed Zone Forum. D) Hardiness zones for the state of Minnesota according to the USDA. Zones where leaf and seed samples are collected for this study. Coastal (red), Inland (orange) and Interior (yellow). Adapted from: MDNR Seed collection zones <https://www.dnr.state.mn.us/forestry/nursery/collection-map.html>, USDA Eastern Seed Zone Forum seed collection zones <http://www.easternseedzones.com> and USDA Plant hardiness zones map <https://planthardiness.ars.usda.gov/PHZMWeb/> adapted from Gross and Etterson.

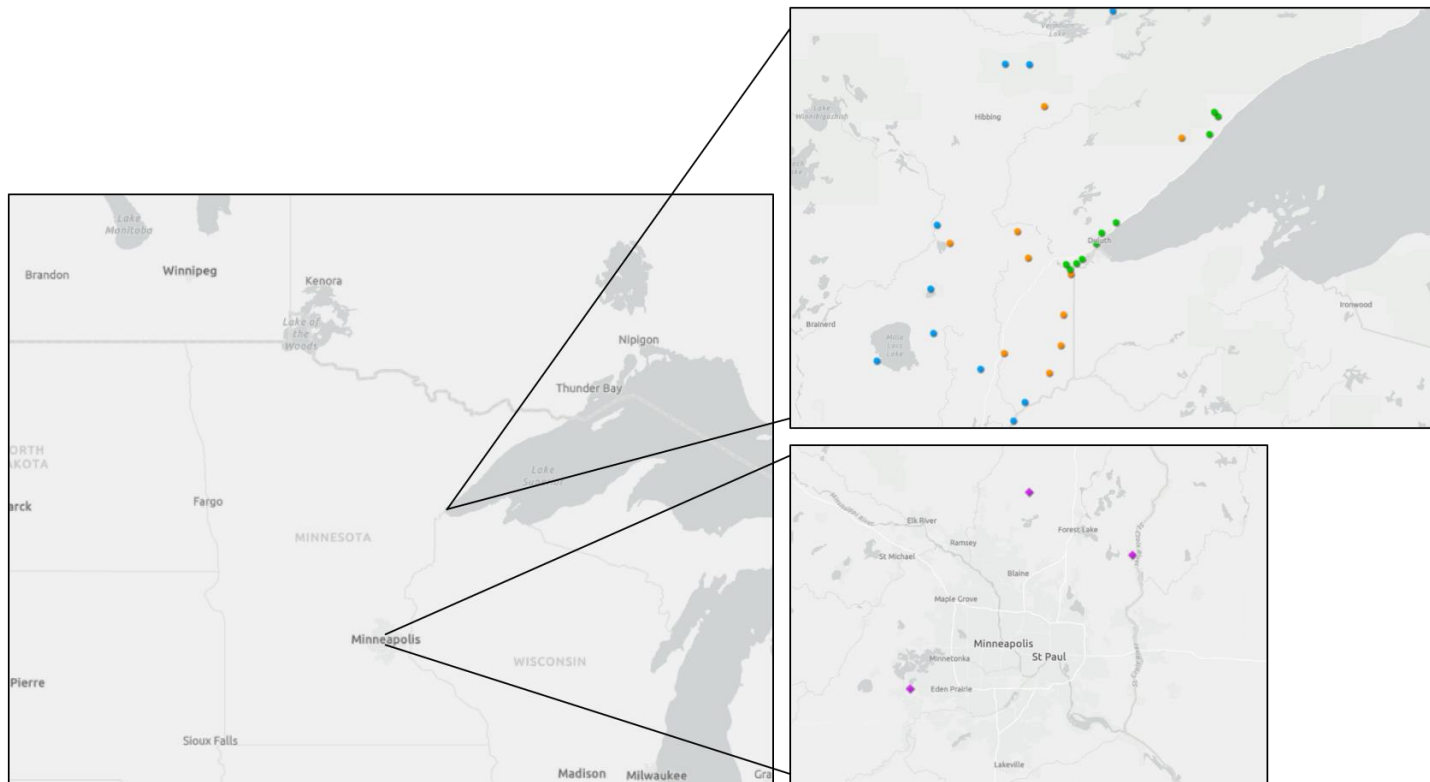


Figure 2. Northern red oak (circles) and northern pin oak (diamonds) collection sites. Coastal populations (green) are between 0-10 miles of the coast. Inland populations (orange) are between 11-50 miles of the coast. Interior populations (blue) are between 51-100 miles of the coast. Leave tissue and acorns were collected in northern red oak populations, however only leaf tissue was collected from northern pink oak sites (purple).

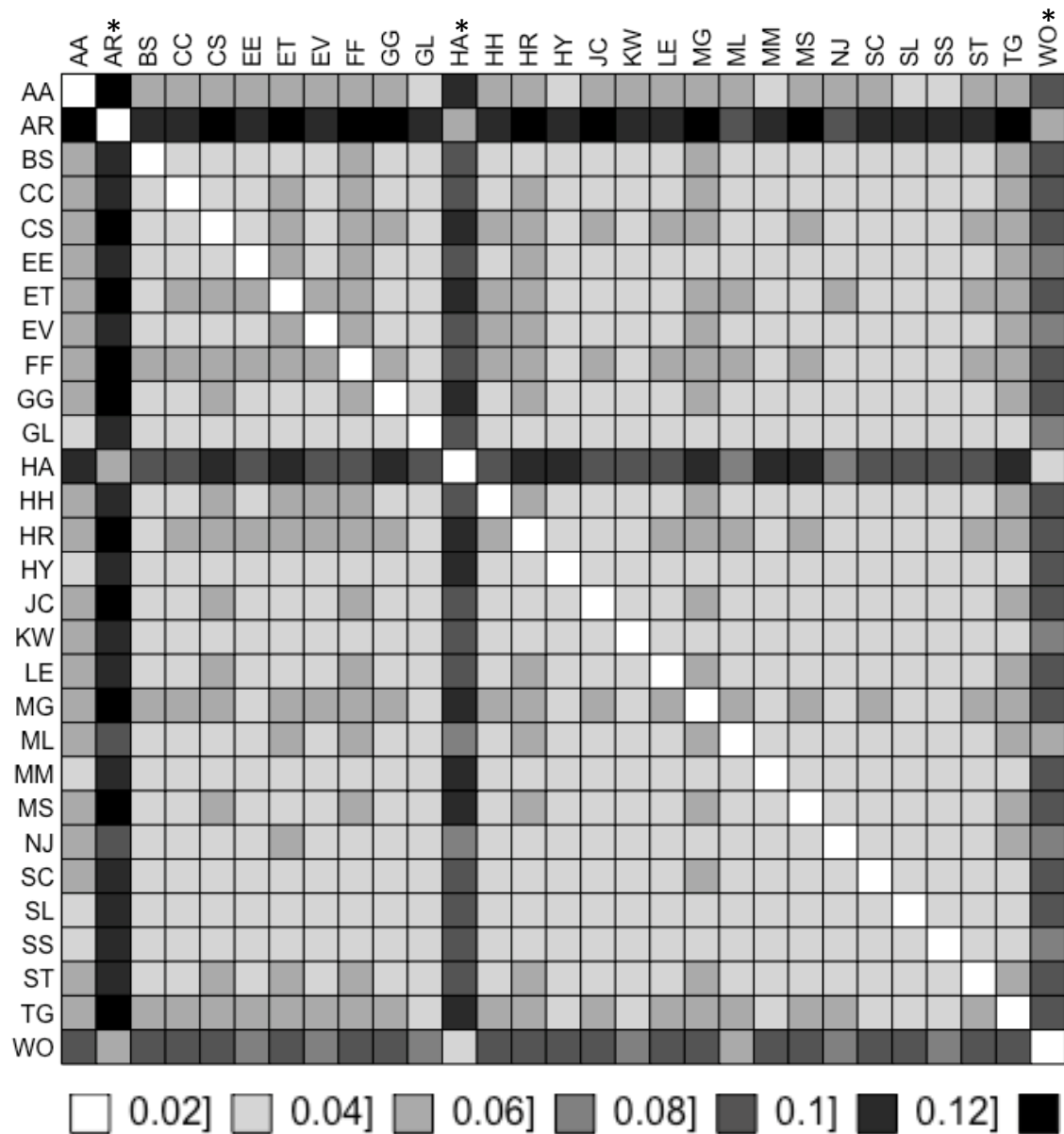


Figure 3. F_{ST} matrix for *Q. rubra* and *Q. ellipsoidalis* populations. Gray squares represent the pairwise F_{ST} value for each population, populations are ordered alphabetically and not by region. Lighter squares represent lower F_{ST} values that represent lower genetic differentiation between populations, darker squares represent higher F_{ST} values that show higher genetic differentiation between populations. Populations marked with an asterisk are the *Q. ellipsoidalis* used in the analysis as outgroups.

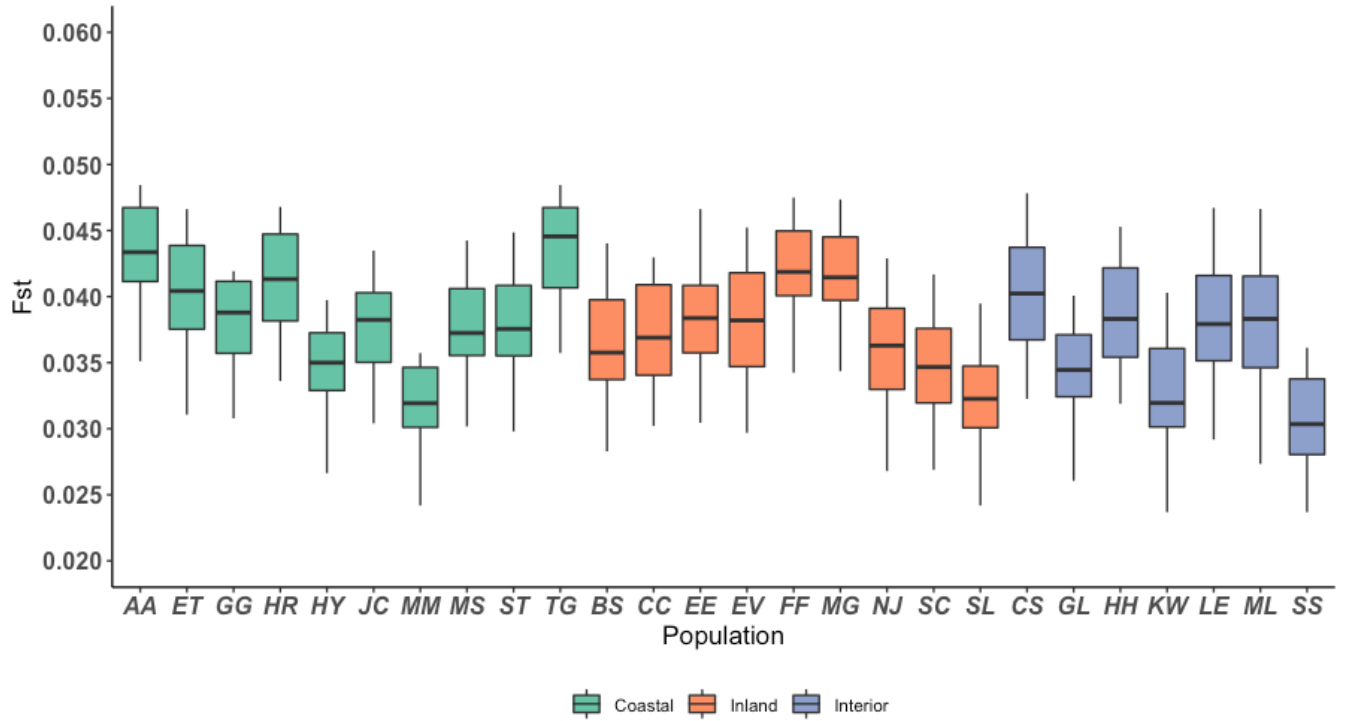


Figure 4. *Q. rubra* individual population F_{ST} values compared to all the other populations in the study. Boxplots represent the distribution of the F_{ST} values of all the coastal (green), inland (orange), interior (blue) populations compared to one another.

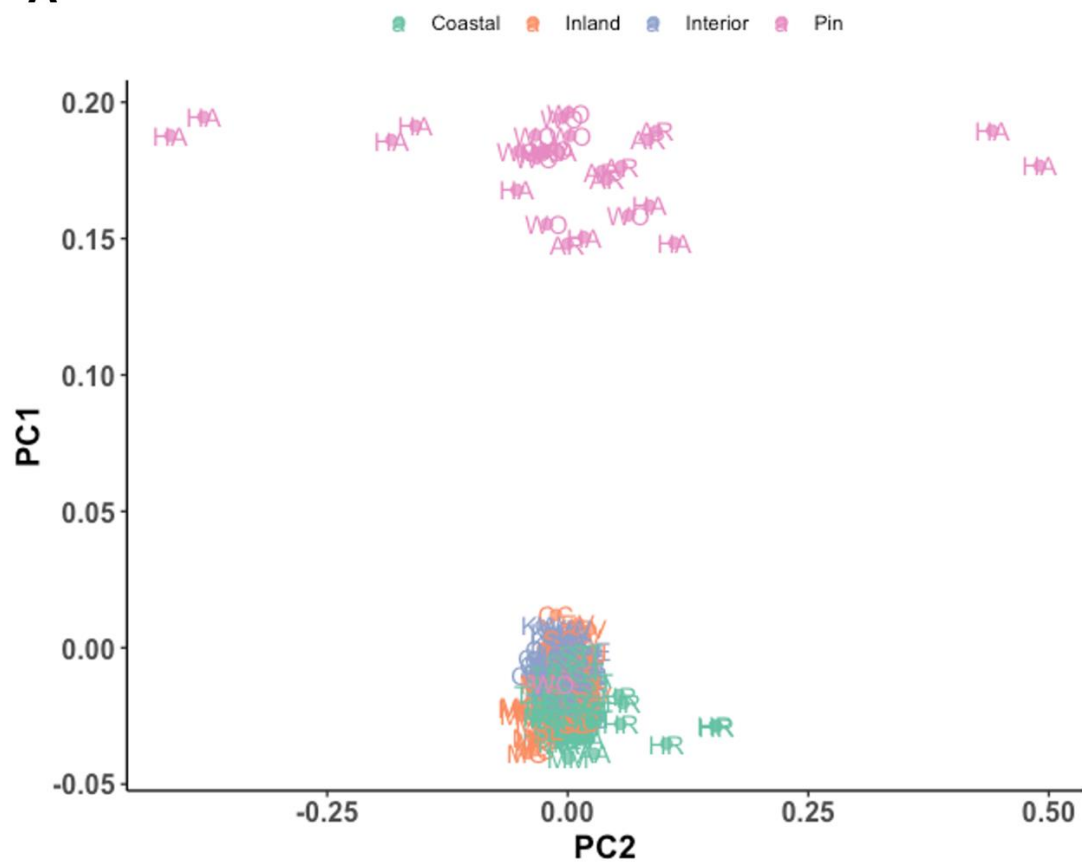
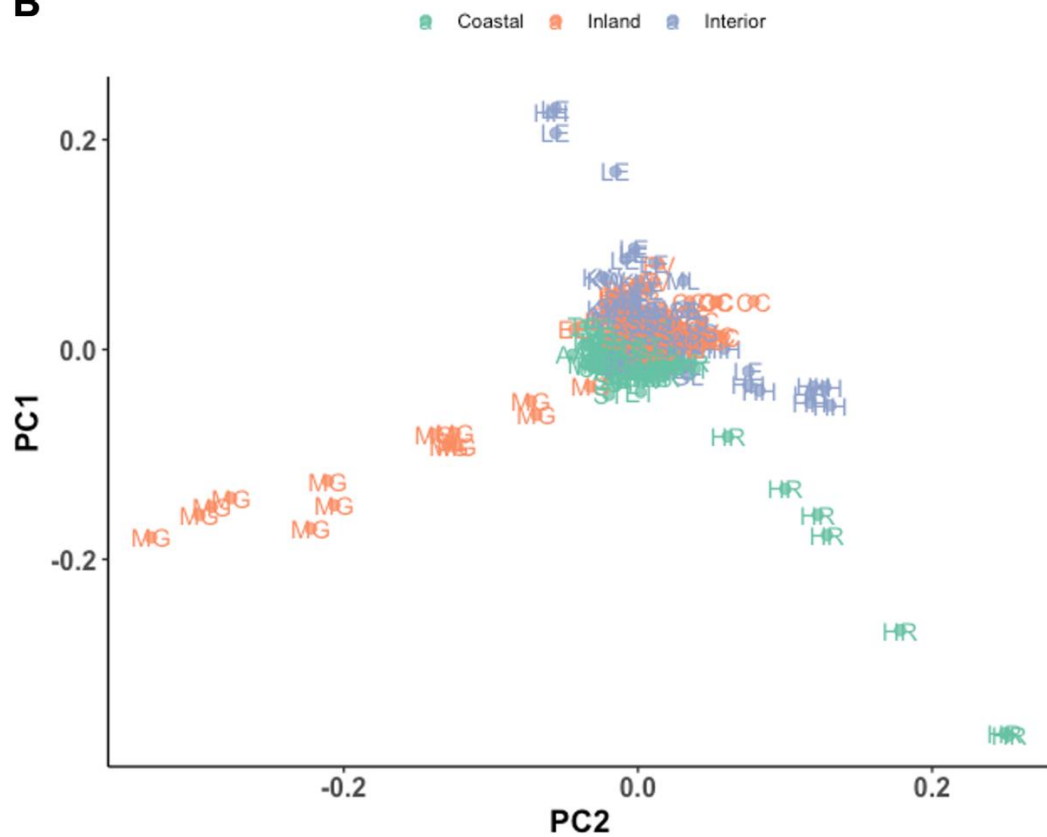
A**B**

Figure 5. Principal component analysis of coastal (green), inland (orange), interior (blue) *Q. rubra* populations and *Q. ellipsoidalis* populations (pink). PCA including all the *Q. rubra* and *Q. ellipsoidalis* populations (a). *Q. ellipsoidalis* populations were removed to identify *Q. rubra* population clustering (b). For each principal component analysis populations are coded by their region color and by the population two letter codes (Table 1) as individual points.

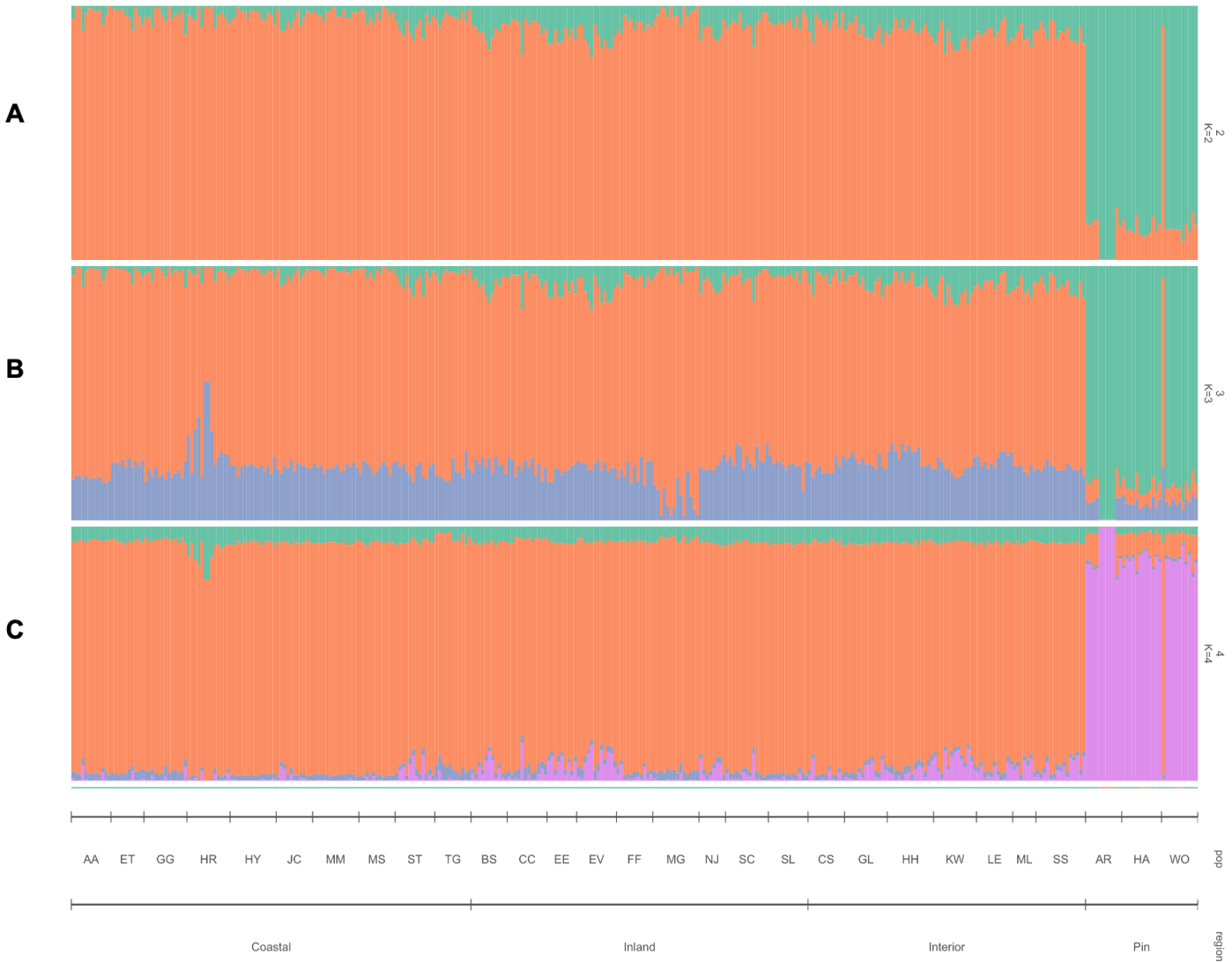


Figure 6. STRUCTURE plot analysis for genetic assignment of *Q. rubra* and *Q. ellipsoidalis* populations for three values of K. Genetic assignment using two clusters K=2 to confirm species identity (a). Values of K are increased up to four as we hypothesize three regions for our analyses (b-c). The proportion of each cluster is shown in different colors for all the individuals in the analysis.

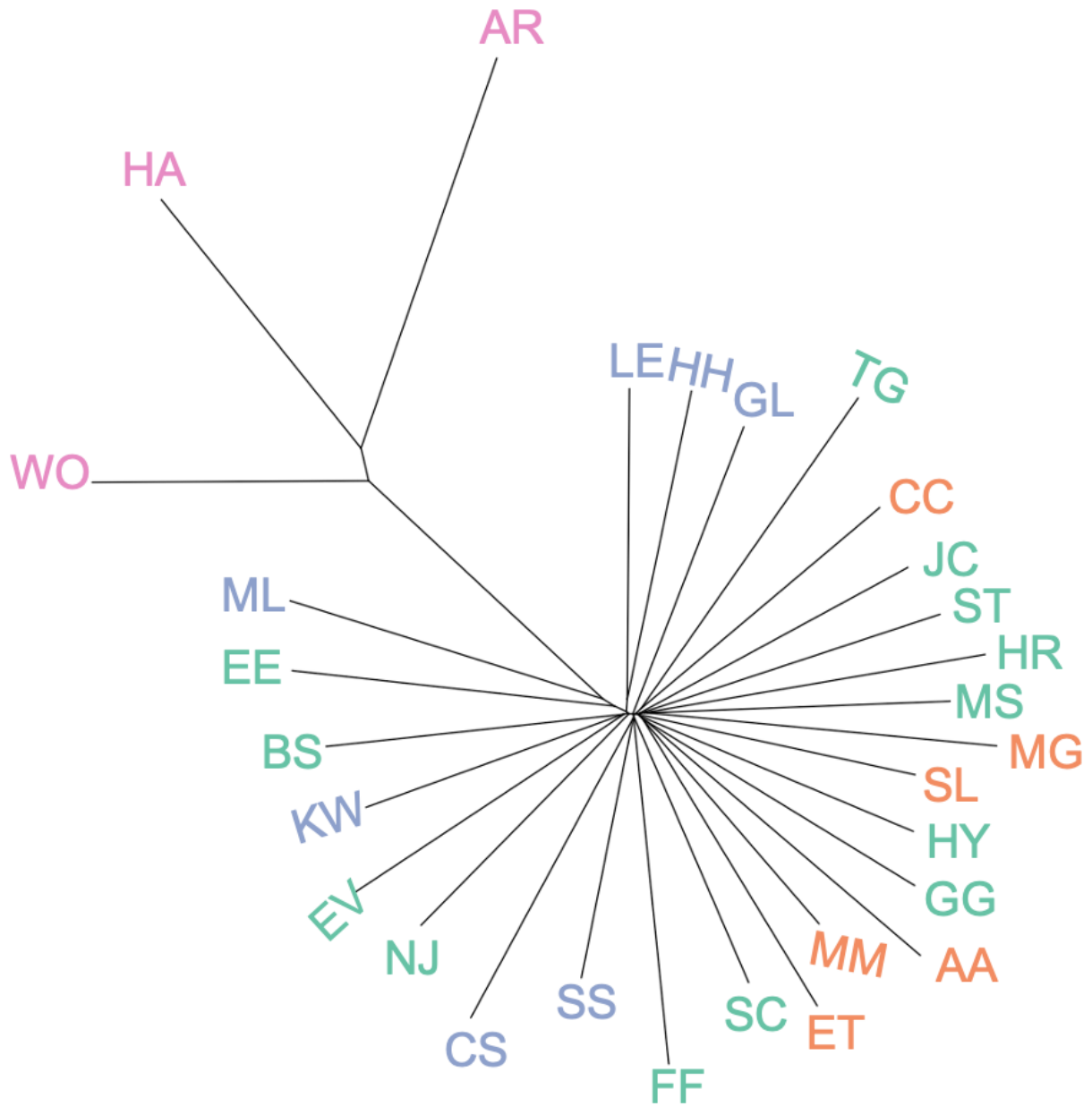


Figure 7. Neighbor joining tree based on population allele frequencies of coastal (green), inland (orange), interior (blue) *Q. rubra* populations and *Q. ellipsoidalis* populations (pink). The distance between both HR and MG compared to the remaining populations, was slightly larger than when other populations compared to each other. However, these differences were minimal and do not reflect a large allele frequency difference from the rest of the *Q. rubra* populations.

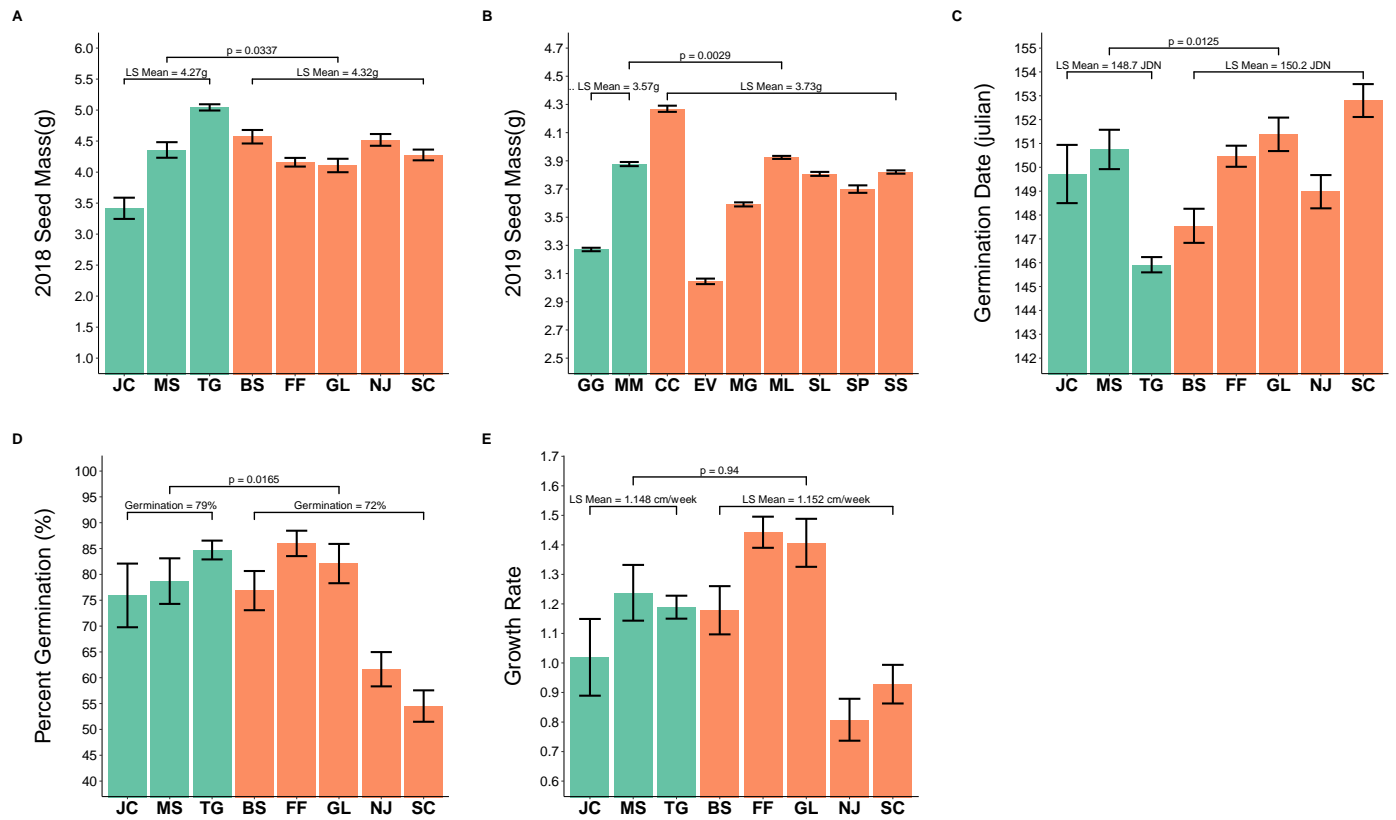


Figure 8. Bar graphs of the least square means of all the measured traits in the common garden experiment. 2018 and 2019 Seed mass values were logarithmically transformed to achieve normality. Each bar in plot represents either a coastal (green) or a non-coastal (orange) population, the brackets that surrounds the populations contain the least square mean value for the region (coastal or non-coastal) all together. The top bracket contains the p-value for the analysis of significance between regions, p-values < 0.05 are considered statistically significant and p-values > 0.05 are considered non statistically significant. The traits measured in the common garden experiment where seed mass before planting in 2018 (a) and in 2019 (b) (seeds not planted in common garden), germination Julian dates (c), percent germination (d) and growth rate (e) measured as the $\left(\frac{Final\ height\ (cm) - Initial\ height\ (cm)}{N_{weeks}}\right)$.

Literature Cited:

1. 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., Pereira, J. S., Prinz, A. , Scherer-Lorenzen, M. , Siamantziouras, A. D., Terry, A. C., Troumbis, A. Y. and Lawton, J. H. (2001), Local adaptation enhances performance of common plant species. *Ecology Letters*, 4: 536-544. doi:10.1046/j.1461-0248.2001.00262.x
2. Greenbaum, G., Templeton, A. R. & Bar-David, S. (2016) Inference and analysis of population structure using genetic data and network theory. *Genetics*, 202, 1299.
3. 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>
4. Lascoux, M. , Glémin, S. and Savolainen, O. (2016). Local adaptation in plants. In eLS, John Wiley & Sons, Ltd (Ed.). doi:10.1002/9780470015902.a0025270
5. Savolainen, O., Lascoux, M. & Merilä, J. (2013) Ecological genomics of local adaptation. *Nat. Rev. Genet.* 14, 807.
6. Kawecki, T. J. and Ebert, D. (2004), Conceptual issues in local adaptation. *Ecology Letters*, 7: 1225-1241. doi:10.1111/j.1461-0248.2004.00684.x
7. 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. *Nat. Rev. Genet.* 17, 81–92.
8. 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. doi:10.1002/ece3.512
9. Davey, J. W., & Blaxter, M. L. (2010). RADSeq: next-generation population genetics. *Briefings in functional genomics*, 9(5-6), 416–423. doi:10.1093/bfpg/elq031
10. 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, 249–254.
11. Sork, V., Stowe, K., & 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. Retrieved from <http://www.jstor.org/stable/2462691>
12. José A Ramírez-Valiente, Julie R Etterson, Nicholas J Deacon, Jeannine Cavender-Bares, Evolutionary potential varies across populations and traits in the neotropical oak *Quercus oleoides*, *Tree Physiology*, Volume 39, Issue 3, March 2019, Pages 427–439, Retrieved from <https://doi.org/10.1093/treephys/tpy108>
13. Ramírez-Valiente, JA, Deacon, NJ, Etterson, J, et al. Natural selection and neutral evolutionary processes contribute to genetic divergence in leaf traits across a precipitation

- gradient in the tropical oak *Quercus oleoides*. *Mol Ecol.* 2018; 27: 2176–2192. <https://doi.org/10.1111/mec.14566>
14. Donelson, J, Salinas, S, Munday, PL, Shama, LNS. Transgenerational plasticity and climate change experiments: Where do we go from here? *Glob Change Biol.* 2018; 24: 13–34. <https://doi.org/10.1111/gcb.13903>
 15. Huang X, Liu D, Wang D, Shi X, Simon J-C (2015) Molecular and Quantitative Genetic Differentiation in *Sitobion avenae* Populations from Both Sides of the Qinling Mountains. *PLoS ONE* 10 (3): e0122343. doi:10.1371/journal.pone.0122343
 16. Räsänen, K. and Hendry, A.P. (2008), Disentangling interactions between adaptive divergence and gene flow when ecology drives diversification. *Ecology Letters*, 11: 624–636. doi:[10.1111/j.1461-0248.2008.01176.x](https://doi.org/10.1111/j.1461-0248.2008.01176.x)
 17. Berlin, S., Trybush, S.O., Fogelqvist, J. *et al.* (2014). Genetic diversity, population structure and phenotypic variation in European *Salix viminalis* L. (Salicaceae). *Tree Genetics & Genomes* **10**, 1595–1610. <https://doi.org/10.1007/s11295-014-0782-5>
 18. Jordan, M.A., Snell, H.L., Snell, H.M. and Jordan, W.C. (2005), Phenotypic divergence despite high levels of gene flow in Galápagos lava lizards (*Microlophus albemarlensis*). *Molecular Ecology*, 14: 859–867. doi:[10.1111/j.1365-294X.2005.02452.x](https://doi.org/10.1111/j.1365-294X.2005.02452.x)
 19. Petit, R. J., and A. Hampe. 2006. Some evolutionary consequences of being a tree. *Annual Review of Ecology Evolution and Systematics* 37: 187–214.
 20. Zanella, C.M., Bruxel, M., Paggi, G.M., Goetze, M., Buttow, M.V., Cidade, F.W. and Bered, F. (2011), Genetic structure and phenotypic variation in wild populations of the medicinal tetraploid species *Bromelia antiacantha* (Bromeliaceae). *American Journal of Botany*, 98: 1511–1519. doi:[10.3732/ajb.1000458](https://doi.org/10.3732/ajb.1000458)
 21. 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 of the United States of America*, 92(18), 8254–8258. <https://doi.org/10.1073/pnas.92.18.8254>
 22. Ribeiro, Â.M., Lloyd, P. and Bowie, R.C.K. (2011), A tight balance between natural selection and gene flow in a southern African arid-zone endemic bird. *Evolution*, 65: 3499–3514. doi:10.1111/j.1558-5646.2011.01397.x
 23. Anderson, J. T., Willis, J. H., & Mitchell-Olds, T. (2011). Evolutionary genetics of plant adaptation. *Trends in genetics : TIG*, 27(7), 258–266. doi:10.1016/j.tig.2011.04.001
 24. Busoms, S., Teres, J., Huang, X. Y., Bomblies, K., Danku, J., Douglas, A., ... Salt, D. E. (2015). Salinity is an agent of divergent selection driving local adaptation of *Arabidopsis* to coastal habitats. *plant physiology*, 168(3), 915–929. doi:10.1104/pp.15.00427
 25. Chen, G., & Schemske, D. (2015). Ecological differentiation and local adaptation in two sister species of Neotropical *Costus* (Costaceae). *Ecology*, 96(2), 440–449.
 26. Moen, S. Minnesota Sea Grant - Lake Superior's Natural process. (2018) Available at: <http://www.seagrant.umn.edu/superior/processes>.
 27. Vanwalleghe, T., & Meentemeyer, R. (2009). Predicting forest microclimate in

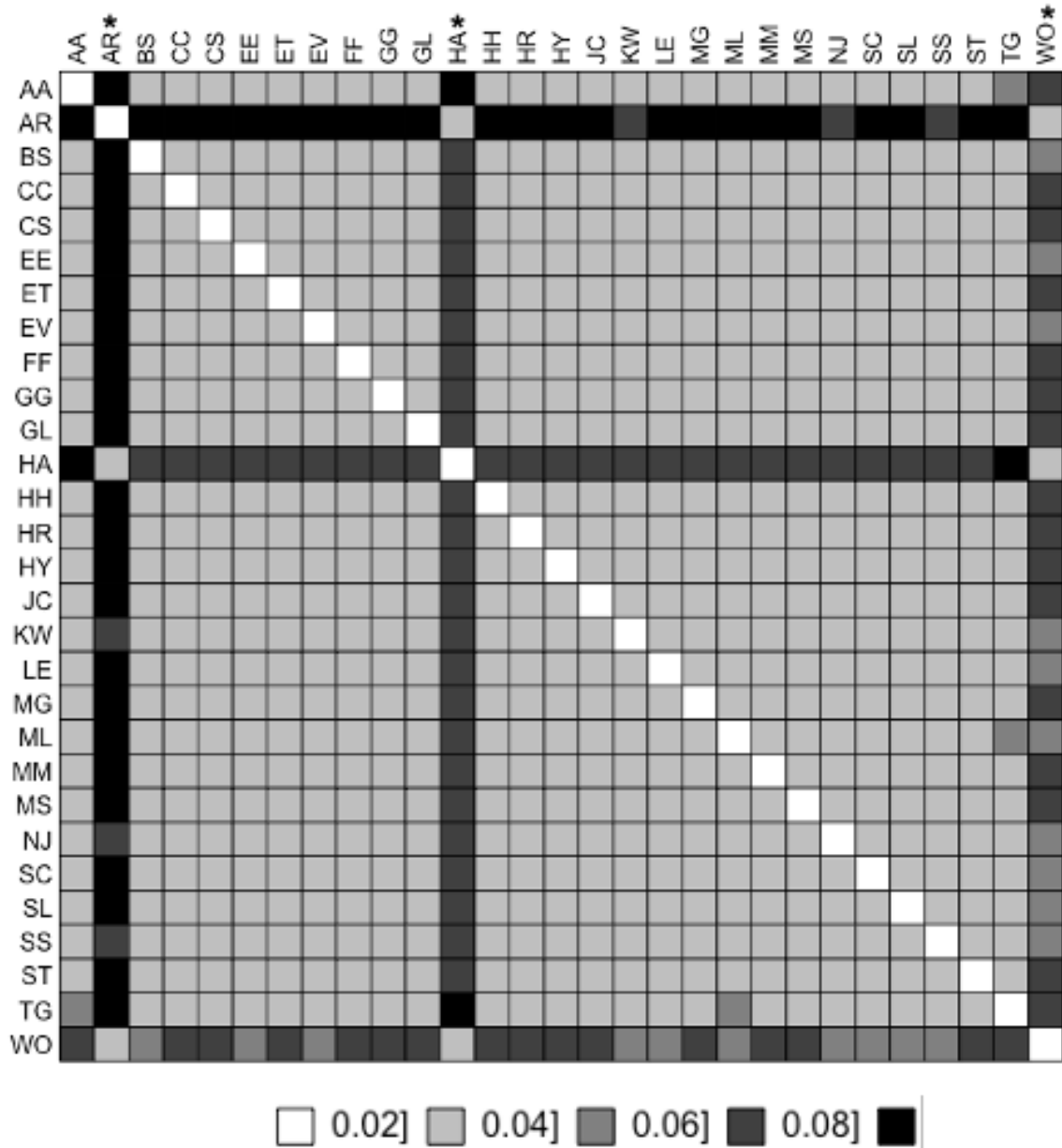
- heterogeneous landscapes. *Ecosystems*, 12(7), 1158-1172.
28. Vallés, S., Fernández, J., Dellafiore, C., & Cambrollé, J. (2011). Effects on soil, microclimate and vegetation of the native-invasive *Retama monosperma* (L.) in coastal dunes. *Plant Ecology*, 212(2), 169-179.
 29. Nevo, E. (1995). Asian, 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.
 30. Zlonis, K.J., Gross, B.L. Genetic structure, diversity, and hybridization in populations of the rare arctic relict *Euphrasia hudsoniana* (Orobanchaceae) and its invasive congener *Euphrasia stricta* . *Conserv Genet* 19, 43–55 (2018). <https://doi.org/10.1007/s10592-017-0995-x>
 31. Galloway, L. F., & Etterson, J. R. (2007). Transgenerational Plasticity Is Adaptive in the Wild. *Science*, 318(5853), 1134–1136. doi: 10.1126/science.1148766
 32. Resources, M. D. of N. Seed Collection Map. (2019). Available at: <https://www.dnr.state.mn.us/forestry/nursery/collection-map.html>.
 33. Etterson, J. R., Cornett, M. W., White, M. A., and Kavajecz, L. C.. 2020. Assisted migration across fixed seed zones detects adaptation lags in two major North American tree species. *Ecological Applications* 00(00):e02092. 10.1002/eap.2092
 34. Service, U. F. Eastern Seed Zone Forum. (2019). Available at: <http://www.easternseedzones.com>.
 35. MinnesotaWildflowers.info. Minnesota Wildflowers (*Quercus Rubra*). (2012).
 36. USDA. USDA Hardiness Zones. (2019) Available at: <https://planthardiness.ars.usda.gov/PHZMWeb/>.
 37. Day, A. Northern Red Oak (*Quercus rubra*) (2019). Available at: <https://www.arborday.org/trees/treeguide/treedetail.cfm?itemID=877>.
 38. Woods, M. M. No Title. (2018). Available at: <http://www.myminnesotawoods.umn.edu/2018/08/why-are-there-so-many-acorns-this-year/>.
 39. Moran, E.V., Willis, J. and Clark, J.S. (2012), Genetic evidence for hybridization in red oaks (*Quercus* sect. *Lobatae*, Fagaceae). *American Journal of Botany*, 99: 92-100. doi:10.3732/ajb.1100023
 40. Sander, I. Northern Red Oak (*Quercus rubra*), Fagaceae Beech family. Available at: https://www.srs.fs.usda.gov/pubs/misc/ag_654/volume_2/quercus/rubra
 41. Lind, J.F., Gailing, O. Genetic structure of *Quercus rubra* L. and *Quercus ellipsoidalis* E. J. Hill populations at gene-based EST-SSR and nuclear SSR markers. *Tree Genetics & Genomes* 9, 707–722 (2013). <https://doi.org/10.1007/s11295-012-0586-4>
 42. Jensen, R.J., Hokanson, S.C., Isebrands, J.G. and Hancock, J.F. (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: 1358-1366. doi:10.1002/j.1537-2197.1993.tb15375.x

43. Gailing, O., Kostick, S., Caré, O., Khodwekar, S. 2018. Leaf morphological and genetic variation between *Quercus rubra* and *Quercus ellipsoidalis*: comparison of sympatric and parapatric populations. *Annals of Forest Research* DOI:10.15
44. Sork, V. L. (2016), Gene flow and natural selection shape spatial patterns of genes in tree populations: implications for evolutionary processes and applications. *Evol Appl*, 9: 291-310. doi:10.1111/eva.12316
45. Gugger, P. F., Fitz-Gibbon, S. , PellEgrini, M. and Sork, V. L. (2016),.Species-wide patterns of DNA methylation variation in *Quercus lobata* and their association with climate gradients. *Mol Ecol*, 25: 1665-1680. doi:10.1111/mec.13563
46. Rellstab, C. , Zoller, S. , Walthert, L. , Lesur, I. , Pluess, A. R., Graf, R. , Bodénès, C. , Sperisen, C. , Kremer, A. and Gugerli, F. (2016). Signatures of local adaptation in candidate genes of oaks (*Quercus* spp.) with respect to present and future climatic conditions. *Mol Ecol*, 25: 5907-5924. doi:10.1111/mec.13889
47. Sork Lab:Protocols. (2018). *OpenWetWare*, Available at: https://openwetware.org/mediawiki/index.php?title=Sork_Lab:Protocols&oldid=1037048.
48. Barta, C. E., Bolander, B., Bilby, S. R., Brown, J. H., Brown, R. N., Duryee, A. M., Edelman, D. R., Gray, C. E., Gossett, C., Haddock, A. G., Helsel, M. M., Jones, A. D., Klingseis, M. E., Leslie, K., Miles, E. W., & Prawitz, R. A. (2017). In Situ Dark Adaptation Enhances the Efficiency of DNA Extraction from Mature Pin Oak (*Quercus palustris*) Leaves, Facilitating the Identification of Partial Sequences of the 18S rRNA and Isoprene Synthase (IspS) Genes. *Plants* (Basel, Switzerland), 6(4), 52. <https://doi.org/10.3390/plants6040052>
49. Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. arXiv preprint arXiv:1207.3907 [q-bio.GN] 2012
50. Eaton DAR & Overcast I. “ipyrad: Interactive assembly and analysis of RADseq datasets.” *Bioinformatics* (2020).
51. Petr Danecek, Adam Auton, Goncalo Abecasis, Cornelis A. Albers, Eric Banks, Mark A. DePristo, Robert E. Handsaker, Gerton Lunter, Gabor T. Marth, Stephen T. Sherry, Gilean McVean, Richard Durbin, 1000 Genomes Project Analysis Group, The variant call format and VCFtools, *Bioinformatics*, Volume 27, Issue 15, 1 August 2011, Pages 2156–2158, <https://doi.org/10.1093/bioinformatics/btr330>
52. Jombart T, Ahmed I (2011). “adegenet 1.3-1: new tools for the analysis of genome-wide SNP data.” *Bioinformatics*. doi: [10.1093/bioinformatics/btr521](https://doi.org/10.1093/bioinformatics/btr521).
53. GOUDET, J. (2005), hierfstat, a package for r to compute and test hierarchical *F*-statistics. *Molecular Ecology Notes*, 5: 184-186. doi:[10.1111/j.1471-8286.2004.00828.x](https://doi.org/10.1111/j.1471-8286.2004.00828.x)
54. Paradis E. & Schliep K. 2018. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35: 526-528.
55. Zheng X, Levine D, Shen J, Gogarten S, Laurie C, Weir B (2012). “A High-performance Computing Toolset for Relatedness and Principal Component Analysis of SNP Data.” *Bioinformatics*, 28(24), 3326-3328. doi: [10.1093/bioinformatics/bts606](https://doi.org/10.1093/bioinformatics/bts606).

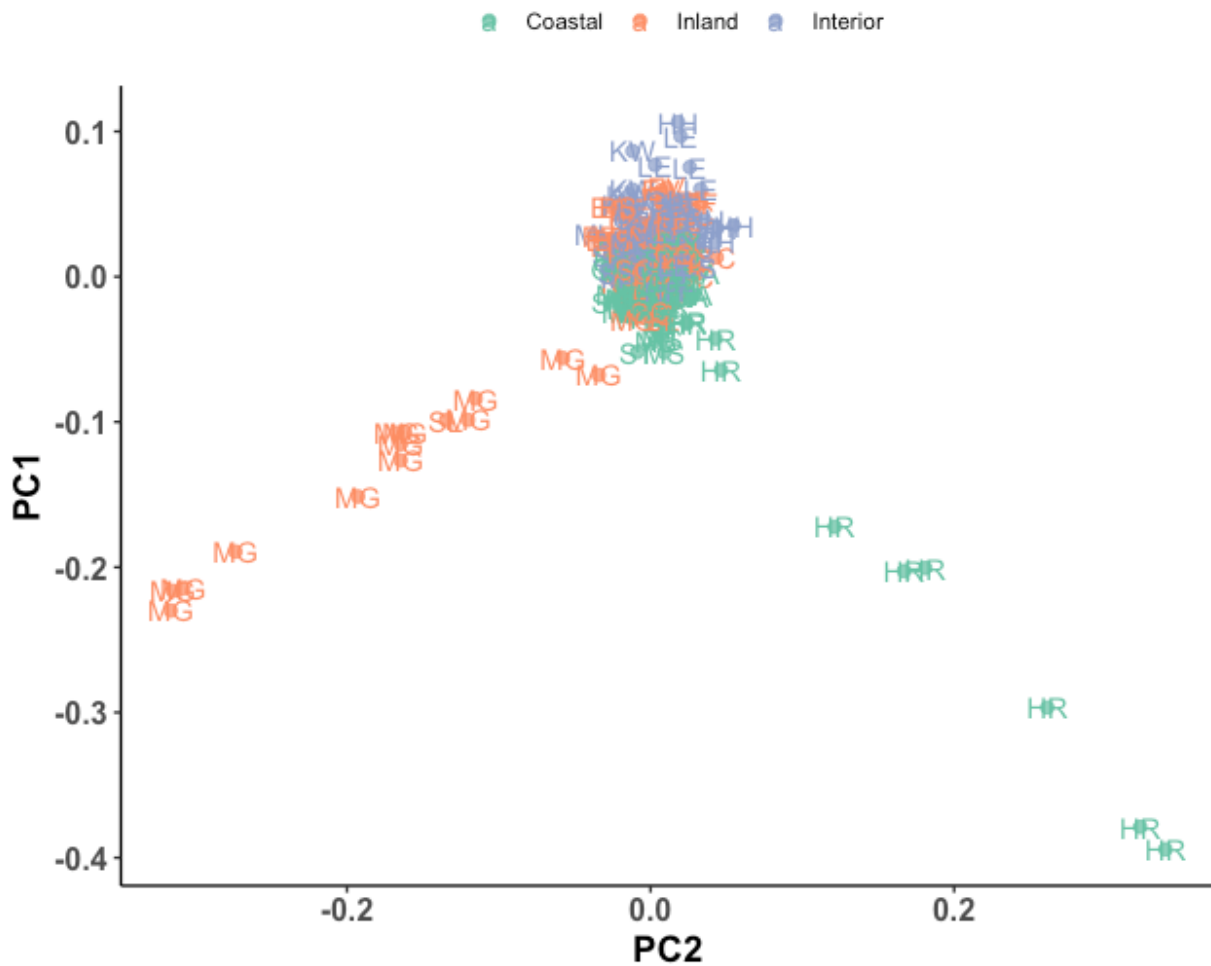
56. Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155(2), 945–959.
57. Earl, Dent A. and vonHoldt, Bridgett M. (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* vol. 4 (2) pp. 359-361. doi: 10.1007/s12686-011-9548-7 <http://www.springerlink.com/content/jnn011511h415358/>
58. Francis RM (2017). “pophelper: An R package and web app to analyse and visualise population structure.” *Molecular Ecology Resources*, 17(1), 27-32. doi: 10.1111/1755-0998.12509, pophelper R package version 2.3.0, <https://github.com/royfrancis/pophelper>.
59. JMP®, Version 14. SAS Institute Inc., Cary, NC, 1989-2019.
60. R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>.
61. Sork, V. L., Squire, K. , Gugger, P. F., Steele, S. E., Levy, E. D. and 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: 33-46. doi:10.3732/ajb.1500162
62. Schueler, S., Schlünzen, K.H. Modeling of oak pollen dispersal on the landscape level with a mesoscale atmospheric model. *Environ Model Assess* 11, 179 (2006). <https://doi.org/10.1007/s10666-006-9044-8>
63. 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, 231–241. <https://doi.org/10.1007/s10709-005-4014-7>
64. Murray, KD, Janes, JK, Jones, A, Bothwell, HM, Andrew, RL, Borevitz, JO. Landscape drivers of genomic diversity and divergence in woodland Eucalyptus. *Mol Ecol*. 2019; 28: 5232– 5247. <https://doi.org/10.1111/mec.15287>
65. Eckert, A. J., Bower, A. D., González-Martínez, S. C., Wegrzyn, J. L., Coop, G., & Neale, D. B. (2010). Back to nature: Ecological genomics of lob- lolly pine (*Pinus taeda*, Pinaceae). *Molecular Ecology*, 19, 3789–3805. <https://doi.org/10.1111/j.1365-294X.2010.04698.x>
66. Bonfil, C. (1998), The effects of seed size, cotyledon reserves, and herbivory on seedling survival and growth in *Quercus rugosa* and *Q. laurina* (Fagaceae). *Am. J. Bot.*, 85: 79-87. doi:10.2307/2446557
67. Quero, J.L., Villar, R., Marañón, T., Zamora, R. and Poorter, L. (2007), Seed-mass effects in four Mediterranean *Quercus* species (Fagaceae) growing in contrasting light environments. *Am. J. Bot.*, 94: 1795-1803. doi:10.3732/ajb.94.11.1795
68. Roach, D. A., & Wulff, R. D. (1987). Maternal effects in plants. *Annual review of ecology and systematics*, 18(1), 209-235.

61. Bonfil, C. (1998), The effects of seed size, cotyledon reserves, and herbivory on seedling survival and growth in *Quercus rugosa* and *Q. laurina* (Fagaceae). *Am. J. Bot.*, 85: 79-87. doi:10.2307/2446557
61. Quero, J.L., Villar, R., Marañón, T., Zamora, R. and Poorter, L. (2007), Seed-mass effects in four Mediterranean *Quercus* species (Fagaceae) growing in contrasting light environments. *Am. J. Bot.*, 94: 1795-1803. doi:10.3732/ajb.94.11.1795
62. Roach, D. A., & Wulff, R. D. (1987). Maternal effects in plants. *Annual review of ecology and systematics*, 18(1), 209-235.

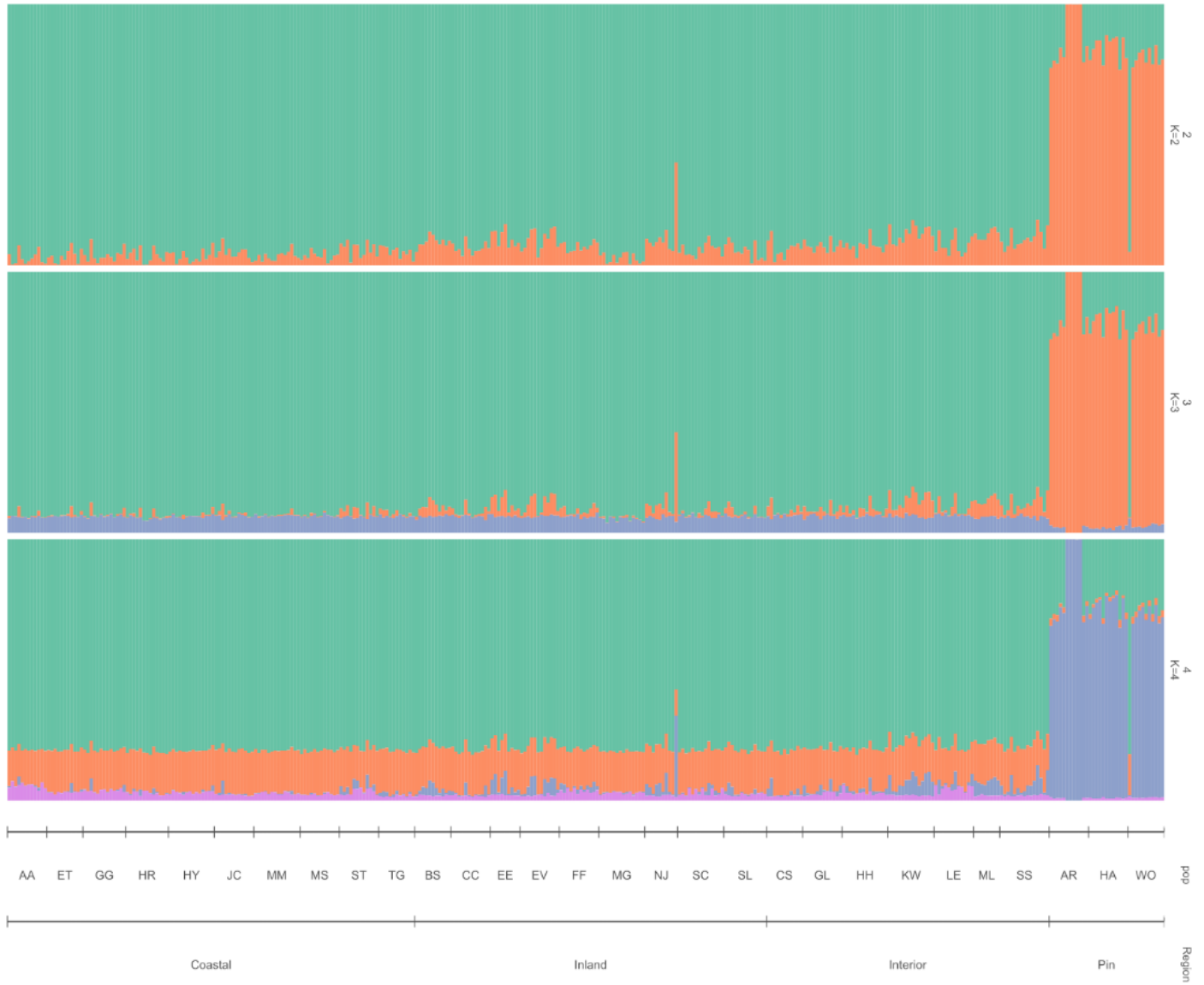
Supplementary Figures:



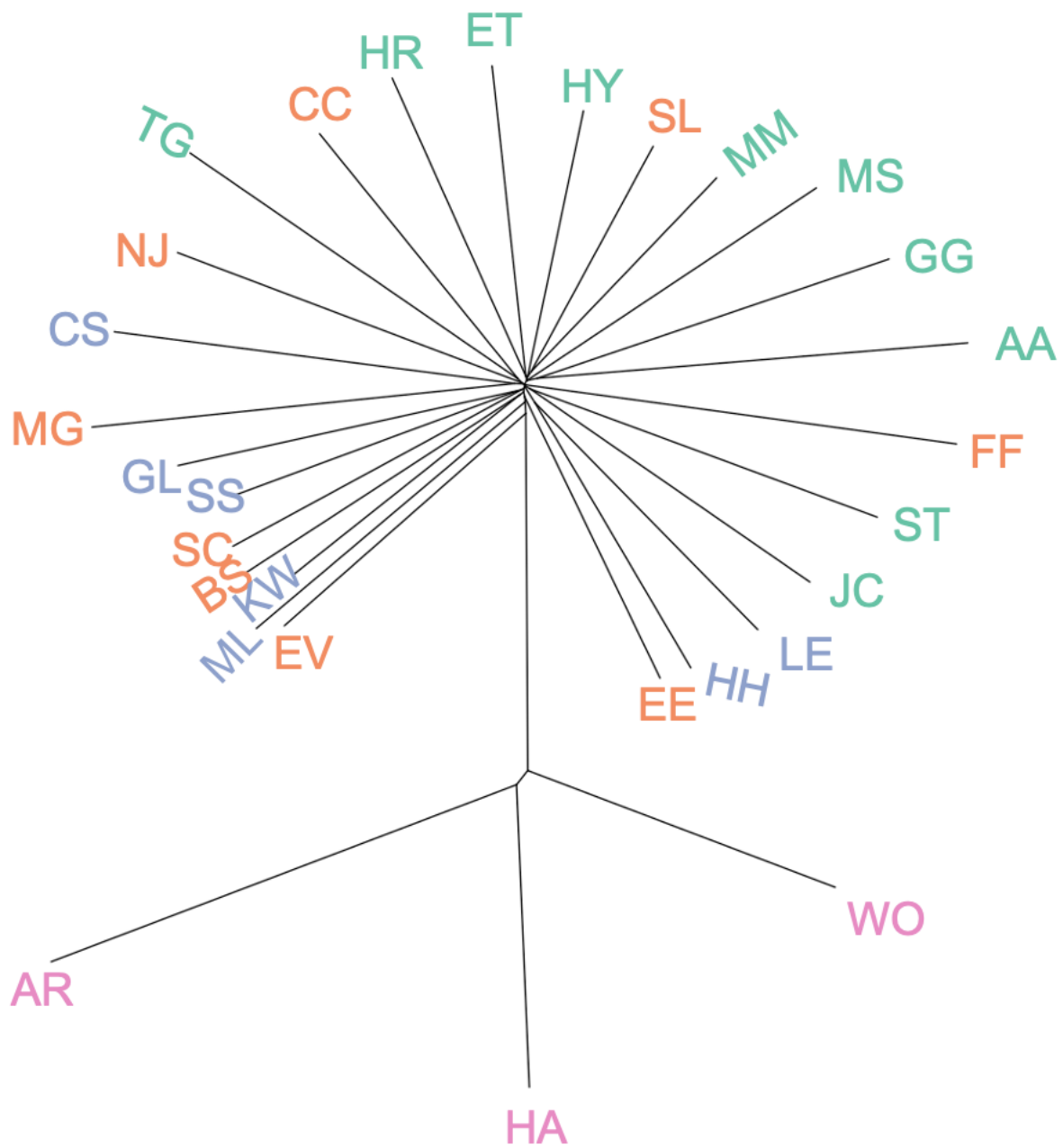
Supplementary Figure 1. F_{ST} matrix for *Q. rubra* and *Q. ellipsoidalis* populations using the Freebayes SNPs. Gray squares represent the pairwise F_{ST} value for each population, populations are ordered alphabetically and not by region. Lighter squares represent lower F_{ST} values that represent lower genetic differentiation between populations, darker squares represent higher F_{ST} values that show higher genetic differentiation between populations. Populations marked with an asterisk are the *Q. ellipsoidalis* used in the analysis as outgroups.



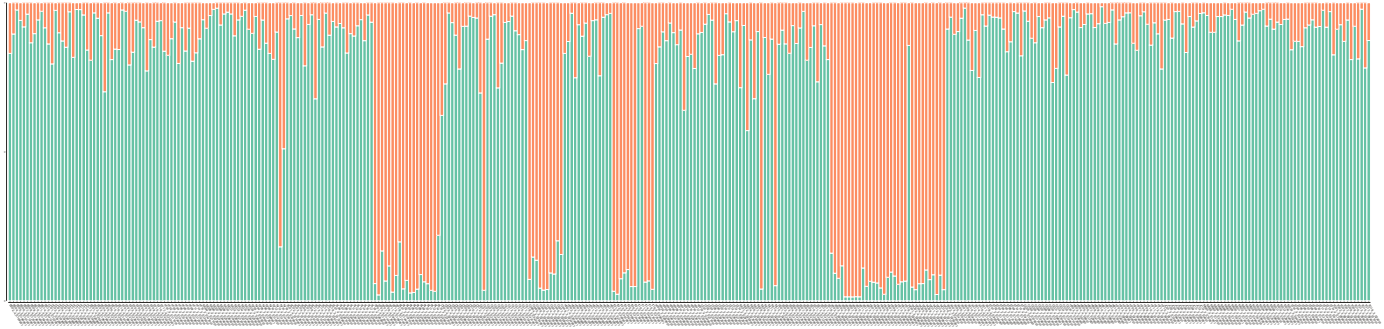
Supplementary Figure 2. Principal component analysis using of coastal (green), inland (orange), interior (blue) *Q. rubra* populations using Freebayes SNPs. Populations are coded by their region color and by the population two letter codes (Table 1) as individual points.



Supplementary Figure 3. STRUCTURE plot analysis for genetic assignment of *Q. rubra* and *Q. ellipsoidalis* populations for three values of K using Freebayes SNPs. Genetic assignment using two clusters K=2 to confirm species identity (a). Values of K are increased up to four as we hypothesize three regions for our analyses (b-c). The proportion of each cluster is shown in different colors for all the individuals in the analysis



Supplementary Figure 4. Neighbor joining tree based on population allele frequencies from Freebayes SNPs of coastal (green), inland (orange), interior (blue) *Q. rubra* populations and *Q. ellipsoidalis* populations (pink).

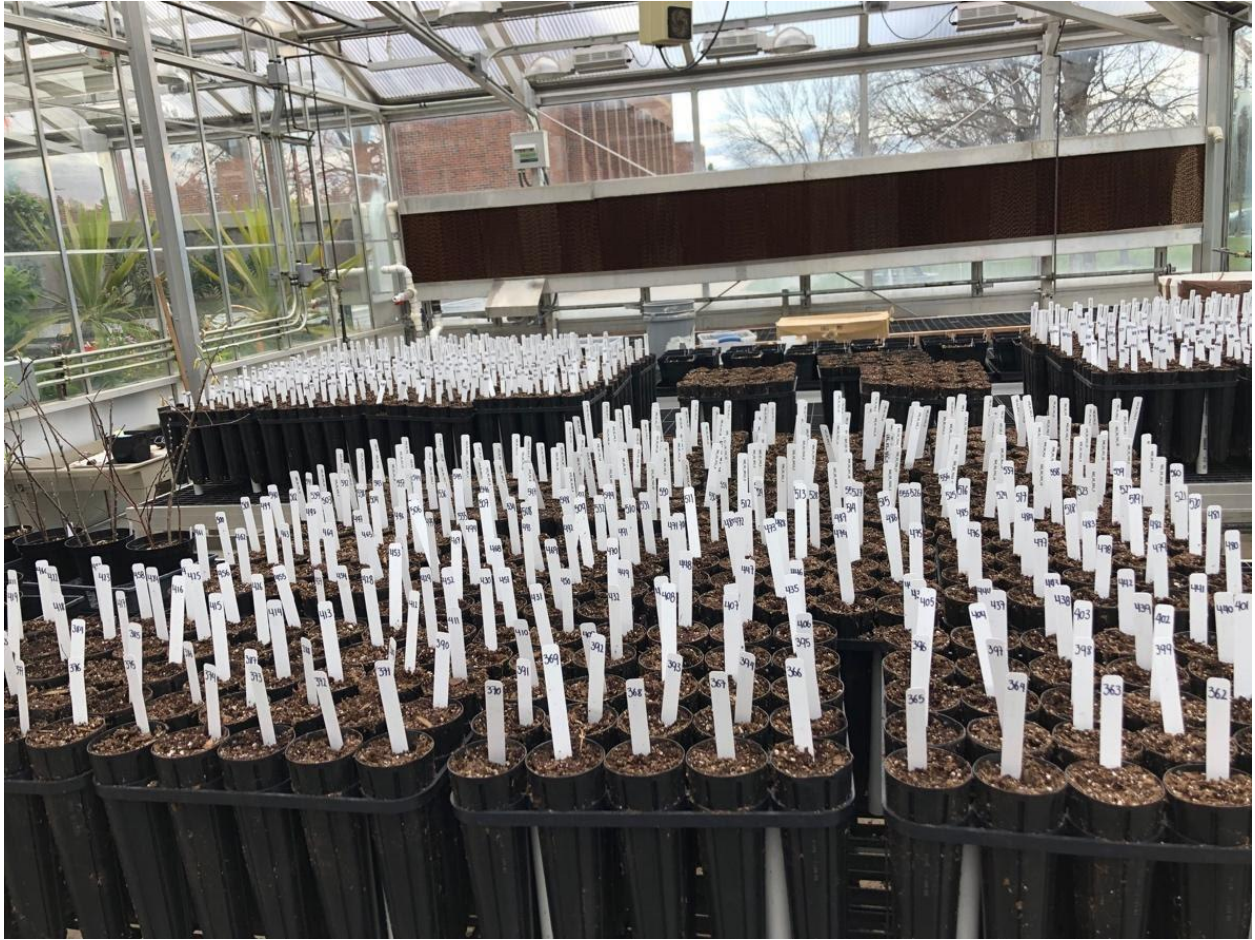


Supplementary Figure 5. STRUCTURE plot analysis for genetic assignment of *Q. rubra* and *Q. ellipsoidalis* populations for $K = 2$ to confirm species identity. *Q. rubra* populations (SP, BL, GW and II) were clustered with the *Q. ellipsoidalis* populations, resembling *Q. ellipsoidalis* identity. SP, BL, GW and II populations were removed from all the subsequent analyses.

Appendix 2: Photos of the greenhouse experiment



04.29.2019 – red oak acorns from 3 different populations before being planted in the common garden



05.09.2019 - Common Garden experimental set-up at the UMD Greenhouse



05.23.2019 - Northern red oak seedling two weeks after planting



05.31.2019 - Northern red oak seedling three weeks after planting



06.07.2019 – Northern red oak seedlings one month after planting



07.08.2019 – Height measurements of seedlings



07.22.2019 – Overview of northern red oak seedlings at UMD Greenhouse

Appendix 3: Correspondence from stakeholders in response to sharing our technical report.

From: Palik, Brian -FS brian.palik@usda.gov
Subject: RE: Northern red oak study
Date: April 30, 2020 at 7:23 AM
To: Briana Gross blgross@d.umn.edu, Pike, Carolyn - FS carolyn.c.pike@usda.gov, Sweeney, Carrie -FS carrie.sweeney@usda.gov
Cc: Maria Jose Gomez gomez312@d.umn.edu, Julie Etterson jetterso@d.umn.edu



Briana, thanks for sharing this. Interesting work!
Brian

Brian Palik, PhD
Science Leader for Applied Forest Ecology
USDA Forest Service-Northern Research Station
p: 218-326-7116
brian.palik@usda.gov

1831 E Hwy 169
Grand Rapids, MN 55744
www.nrs.fs.fed.us/people/Palik
www.adaptivesilviculture.org/node/954



Caring for the land and serving people


From: Briana Gross [mailto:blgross@d.umn.edu]
Sent: Wednesday, April 29, 2020 8:58 PM
To: Pike, Carolyn - FS <carolyn.c.pike@usda.gov>; Sweeney, Carrie -FS <carrie.sweeney@usda.gov>; Palik, Brian -FS <brian.palik@usda.gov>
Cc: Maria Jose Gomez <gomez312@d.umn.edu>; Julie Etterson <jetterso@d.umn.edu>
Subject: Northern red oak study

Dear Carolyn, Carrie, and Brian,

Over the past several years, Julie Etterson and I have been working on an MLSCP-funded project focused on whether Northern red oak populations along the coast of Lake Superior are genetically differentiated from non-coastal populations. We have recently closed the grant and wanted to share the results with you in case it is useful in your seed sourcing efforts for reforestation.

In short, we found that while the coastal populations are not differentiated based on neutral genetic markers, they do show significant differences from non-coastal populations for traits that might influence seedling establishment success (percent germination and germination date). We have attached the technical report detailing these and other findings, and are happy to answer any questions you might have about this study. We will also continue to work on this project over the coming year and will keep you updated as we delve deeper into the data we've generated.

Sincerely,
Briana and Julie

From: Pike, Carolyn - FS carolyn.c.pike@usda.gov 
Subject: RE: Northern red oak study

Date: April 30, 2020 at 7:27 AM

To: Palik, Brian -FS brian.palik@usda.gov, Briana Gross blgross@d.umn.edu, Sweeney, Carrie -FS carrie.sweeney@usda.gov
Cc: Maria Jose Gomez gomez312@d.umn.edu, Julie Etterson jetterso@d.umn.edu



Thanks Briana! That was my suspicion, but it is very cool to see the work completed. So the question remains: should the north shore be a separate seed collection zone for trees and plants? Let me know your thoughts!

Sincerely, Carrie Pike

From: Palik, Brian -FS
Sent: Thursday, April 30, 2020 8:24 AM
To: Briana Gross <blgross@d.umn.edu>; Pike, Carolyn - FS <carolyn.c.pike@usda.gov>; Sweeney, Carrie -FS <carrie.sweeney@usda.gov>
Cc: Maria Jose Gomez <gomez312@d.umn.edu>; Julie Etterson <jetterso@d.umn.edu>
Subject: RE: Northern red oak study

Briana, thanks for sharing this. Interesting work!
Brian

Brian Palik, PhD
Science Leader for Applied Forest Ecology
USDA Forest Service-Northern Research Station
p: 218-326-7116
brian.palik@usda.gov

1831 E Hwy 169
Grand Rapids, MN 55744
www.nrs.fs.fed.us/people/Palik
www.adaptivesilviculture.org/node/954



Caring for the land and serving people

From: Briana Gross [<mailto:blgross@d.umn.edu>]
Sent: Wednesday, April 29, 2020 8:58 PM
To: Pike, Carolyn - FS <carolyn.c.pike@usda.gov>; Sweeney, Carrie -FS <carrie.sweeney@usda.gov>; Palik, Brian -FS <brian.palik@usda.gov>
Cc: Maria Jose Gomez <gomez312@d.umn.edu>; Julie Etterson <jetterso@d.umn.edu>
Subject: Northern red oak study

Dear Carolyn, Carrie, and Brian,

Over the past several years, Julie Etterson and I have been working on an MLSCP-funded project focused on whether Northern red oak populations along the coast of Lake Superior are genetically differentiated from non-coastal populations. We have recently closed the grant and wanted to share the results with you in case it is useful in your seed sourcing

From: Larson, Dave (DNR) <dave.larson@state.mn.us>
Subject: RE: Northern red oak study
Date: April 30, 2020 at 9:10 AM
To: Briana Gross <blgross@d.umn.edu>



Briana and Julie,

Thanks for your work on this research project and thanks for sending the report. I will share it with local DNR staff along the Shore.

Dave Larson, Region Silviculturist

Minnesota Department of Natural Resources

Northeast Region Headquarters
1201 East Hwy 2
Grand Rapids, MN 55744
Office Phone 218-328-8898
Cell Phone 218-398-3815
Dave.Larson@state.mn.us

Silviculture - The Essence of Forest Management

From: Briana Gross <blgross@d.umn.edu>
Sent: Wednesday, April 29, 2020 8:49 PM
To: Schuller, Dave (DNR) <david.schuller@state.mn.us>; Dubuque, Paul A (DNR) <paul.dubuque@state.mn.us>; Reinikainen, Mike (DNR) <mike.reinikainen@state.mn.us>; Larson, Dave (DNR) <dave.larson@state.mn.us>; Baysal, Cassandra (DNR) <cassandra.baysal@state.mn.us>
Cc: Julie Etersson <jetterso@d.umn.edu>; Maria Jose Gomez <gomez312@d.umn.edu>
Subject: Northern red oak study

This message may be from an external email source.

Do not select links or open attachments unless verified. Report all suspicious emails to Minnesota IT Services Security Operations Center.

Dear all,

Some years ago, Dave Schuller wrote a letter of support as Julie Etersson and I applied for an MLSCP grant to study whether Northern red oak populations along the coast of Lake Superior were genetically differentiated from non-coastal populations. We have completed this study, and found that while the coastal populations are not differentiated based on neutral genetic markers, they do show significant differences for traits that might influence establishment success (percent germination and germination date).

We have attached the technical report detailing these and other findings, and are happy to answer any questions you might have about this study. We will also continue to analyze the data over the coming year and will keep you updated as we delve deeper into the data we've