

UVM ScholarWorks

Effects of range position and inbreeding on early life fitness traits in red spruce, *Picea rubens* (Sarg.)

Item Type	undergraduate thesis
Authors	Munson, Helena V.
Download date	2026-05-13 22:13:47
Item License	http://creativecommons.org/licenses/by-nc-nd/3.0/
Link to Item	https://hdl.handle.net/20.500.14849/5741

**Effects of range position and inbreeding on early life
fitness traits in red spruce, *Picea rubens* (Sarg.)**

Helena Munson, Environmental Science (BS)

College of Arts and Sciences Honors College Thesis

Department of Plant Biology

College of Agricultural and Life Science

Thesis Supervisor: Stephen Keller, Ph.D.

Committee Chair: Charlotte Mehrrens, Ph.D.

Third Committee Member: Donald Stratton, Ph.D.

Abstract

Red spruce (*Picea rubens* Sarg.) is a boreal coniferous tree native to the eastern United States and Canada. As a result of post-glacial range expansion and more recent land-use change, red spruce populations have become fragmented, creating small isolated populations throughout the southeastern part of its range. A major question then becomes whether this range fragmentation has led to inbreeding and susceptibility of young seedlings to the effects of inbreeding depression. This study investigates variation in early-life fitness traits in relation to population size and inbreeding history in 340 open-pollinated seed families of red spruce trees sampled from populations located in three different source regions: the core, margin, and southeastern edge of the range. Four measures of fitness were used to calculate overall seedling early life fitness: average seed mass, germination proportion, survivorship, and growth (seedling height) after 12 weeks. To estimate inbreeding history for use in predicting seedling fitness, we used whole exome sequences obtained for the mother tree for each of the 340 families, and used these sequences to estimate population and individual-level homozygosity. Early life traits and a composite measure of overall seedling fitness were then predicted by genetic inbreeding history and source region using linear mixed-effect models with backwards selection and model ranking based on Akaike information criterion (AIC). There was a significantly higher level of population and individual-level homozygosity in the edge region when compared to the core and margin. Overall seedling fitness was also significantly lower in the edge region, consistent with an association between range fragmentation, inbreeding, and early life fitness. However, mixed models that accounted for random effects of population and family level variation did not show evidence of statistically significant associations between fitness traits with source region and inbreeding history. One possible explanation could be a counter-balancing that is occurring between the isolated, inbred

population structure and the historic genetic variation in the southern part of the range, making some traits appear more significant than others. My results suggest that there appears to be a higher level of inbreeding and a lower level of overall seedling fitness in southern edge populations that could be caused by population inbreeding depression. These findings may have important implications for current and future restoration efforts targeted at red spruce as local environments continue to change.

Introduction

Mating among genetically related individuals (consanguineous mating) or through self-reproduction (selfing) is a frequent consequence of small population size, and greatly increases the rate of population genetic drift and may incur individual and/or population level fitness effects (Angeloni, Ouborg, & Leimu, 2011; Mosseler et al., 2000). The frequency of homozygous loci increases and conversely heterozygosity decreases in an inbred population as there is less genetic mixing among unrelated individuals (Wang, Caballero, & Hill, 1998). Levels of inbreeding therefore can be quantitatively measured based on the observed versus expected heterozygosity within a population based on Hardy-Weinberg equilibrium (Hendrick & Cockerham, 1986). A reduction in fitness because of increased homozygosity due to inbreeding is termed *inbreeding depression*. Inbreeding depression can be caused by either an increase in the homozygosity of lethal alleles or by an accumulation of mildly deleterious (i.e., non-lethal) homozygous alleles across the genome that are otherwise masked when in the heterozygous state (Charlesworth & Willis, 2009).

Inbreeding caused by limited gene flow could also put pressure on populations that favor a specific environment, lowering the potential for local adaptation in inbred populations as local

environments change (García-Fernández, Iriondo, & Escudero, 2012). If gene loci that contribute to local adaptation also contribute to inbreeding depression, then inbreeding depression could potentially vary with time (Abu Awad & Billiard, 2017). However, these fluctuations in inbreeding depression and inability to locally adapt can lower effective (reproducing) population sizes, putting small populations at a great risk for extinction (Cheptou & Donohue, 2011).

Inbreeding and genetic variation are strongly dependent on current patterns of gene flow and population connectivity, as well as historical processes such as migration of the distributional range in response to past climate changes. Evidence from fossil pollen suggests that migration of many tree species occurred towards higher elevations and more northerly latitudes that supported favorable environments conditions as the climate changed at the end of the last ice age (Boisvert-Marsh, Périé, & De Blois, 2014; Koo, Madden, & Patten, 2014; Schaffler & Jacobson, 2002). These migration patterns often resulted in the establishment of “leading” and “trailing (rear) edge” populations within a species range that can possess distinct yet complex demographic and genetic histories (Hampe & Petit, 2005). While the rear range edge may become more fragmented as the range shifts towards more northerly climatic regions, rear edge populations may also retain more ancestral genetic variability that can be reduced at the leading edge through founder effects and genetic drift due to successive migration events (Mosseler, Major, & Rajora, 2003; Pluess, 2011). For example, a study on holm oak provided evidence that successive long-distance dispersal events have produced reductions in genetic diversity at the (northern) leading edge due to founder effects (Hampe, Pemonge, & Petit, 2013). Thus, both current population connectivity and historic range expansion are likely to affect the genetic diversity of populations in different parts of the range. While the trailing edge of species’ ranges might increase the rate of inbreeding due to

fragmentation (Levin, 2011), it might not necessarily mean a reduction in fitness due to lower genetic variance within a population.

Coniferous trees offer an excellent study system to investigate the importance of demographic and genetic history on inbreeding and inbreeding depression. Fossil conifer pollen preserved in sediments of lakes and bogs show that climate change at the end of the last ice age resulted in dramatic latitudinal and elevational shifts in abundance, degree of population fragmentation, and genetic diversity (Davis and Shaw 2001). Further, despite being wind-pollinated, conifers and other gymnosperm trees are reproductively self-compatible, and thus susceptible to both inbreeding and inbreeding depression. Evidence for inbreeding depression has been found in black spruce (*Picea mariana*) in which inbreeding depression was suggested based on segregation distortion of genetic markers in offspring (Scoles, Kang, Major, & Rajora, 2011). In a related study on Douglas Fir (*Pseudotsuga menziesii*) that used experimental crosses to generate a range of inbreeding levels, individuals that were self-mated had the highest rate of offspring mortality, suggesting susceptibility to inbreeding depression (Stoehr, Ott, & Woods, 2015).

This thesis aims to look at the effects of inbreeding depression in the conifer species *Picea rubens* (Sarg.), red spruce, across its range in the eastern United States. The range of *P. rubens* is fragmented into non-continuous isolated patches at its higher elevation southern edge while forming more continuous stands in the northeastern core of its range (Koo, Patten, Teskey, & Creed, 2014; Major, Mosseler, Johnsen, Campbell, & Malcolm, 2015). There is also evidence to suggest that historically, red spruce has expanded its range northward over the Holocene period in response to climatic changes (Schauffler & Jacobson, 2002). Like other conifers, red spruce is reproductively monocious meaning it is self-compatible, having both male and female flowers that

will produce seed cones upon fertilization after reaching maturity (Govindaraju, 1988; Hart, 1959). The reproductive self-compatibility and range dynamics of *P. rubens* make it a good candidate for understanding the relationship between population size, demogenetic history, and inbreeding depression. By using early life traits as measures of fitness, the relative fitness of spruce offspring can be compared to molecular-based estimates of inbreeding at the individual and population-level across its range. Specifically, I aim to test the following questions:

- *How do early life traits of seedlings, grown under common garden conditions, differ depending on source regions?*
- *Are there higher individual and population levels of inbreeding in the fragmented southern edge of the range?*
- *What is the relationship between genotypic levels of inbreeding in mother trees with the phenotypic early life fitness traits of their seedlings?*

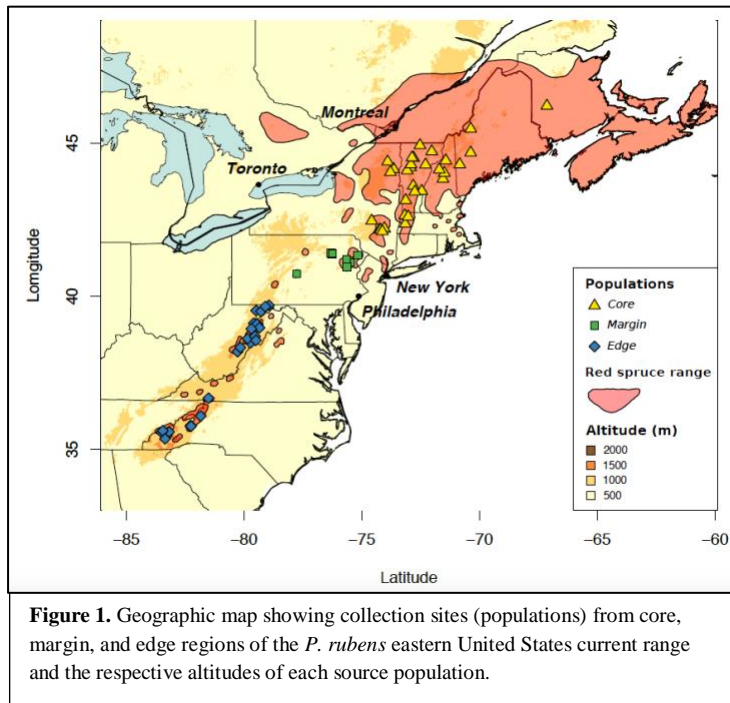
Methods

Red spruce is a conifer present throughout the eastern United States and Canada, with a geographical range from the Appalachian mountains northwards to Prince Edward Island (Mosseler et al., 2000). *P. rubens* is best adapted to cool, temperate climates, with optimal annual temperatures ranging from -18°C to -1°C in the winter months and 11°C to 27°C in the summer months. Across its range, source elevation varies from sea level in the northeast up to 1900 meters in the Appalachian Mountains (Blum, 1990). Red spruce has been well studied with respect to local adaptation, climate variability, genomics, and inbreeding (Scoles et al., 2011). Red spruce is known to have a large genome, between 0.85×10^{10} and 2.4×10^{10} base pairs in length, but possess relatively low genetic diversity (Koo, Patten, & Creed, 2011; Koo, Patten, et al., 2014; Scoles et

al., 2011).

Phenotypic Seedling Trait Collection

In order to determine early life fitness, seeds from 340 open-pollinated maternal families representing 64 collection sites of red spruce were collected from across its range. Of the 340 families, 123 families (23 sites) were from range core, 98 (18 sites) were from the range margin, and 123 (23 sites) were from the



and 123 (23 sites) were from the southern range edge (see Figure 1). Assignment of families to range core/margin/edge were done based on genetic analysis of population substructure (T. Capblancq and S. Keller, unpublished). From each family, 50 seeds were weighed to the nearest 0.01 milligram. Weighed seeds were put into petri dishes

containing wetted filter paper and coarse sand and starting on March 19, 2018 were put into a germination chamber set to 20°C for 16 hour dark periods and 30°C for 8 hour photoperiods. Petri dishes were checked and watered as needed.

Germination was scored once the radicle of a seedling was visible outside of the seed coat, at which time the seedling was transplanted into a container labeled with a barcode specific to the individual plant. For each of the 340 families, 15 seedlings were planted as they germinated. The seedlings were randomly divided among 5 blocks (A, B, C, D, & E) and assigned to a common

garden location in which they would eventually be planted: Vermont (VT), Maryland (MD), or North Carolina (NC). The container racks were put into the UVM greenhouse to grow in the spring of 2018 and kept between 21 and 24°C during the 16 hour photoperiod and between 15 and 18°C during the 8 hour dark period for the summer months. At the end of the seedling planting, petri dishes were retained to count the number of ungerminated seeds left in the dishes to determine the overall germination proportion out of the initial 50 seeds for each family. After 12 weeks of growth in the greenhouse, determined by the planting date recorded for each rack, I measured the height of the seedlings from the base of growth to the top to 0.1 mm using calipers. Mortality was also recorded on a periodic basis for each of seedling. Dead seedlings were assigned a zero value for height.

To look at overall multiplicative fitness for each family based on the phenotypic traits of the seedlings, I multiplied the 50-seed weight, germination proportion, and the mean height value to yield an overall early life fitness value. Mortality was accounted for by weighting the average height per family by the frequency of mortality (contributing height values of 0.0 mm). A second round of seed germination and planting was required for the larger common garden experiment this study is part of, but seed mass and germination were not measured in the second round. Therefore, a total of 330 out of the original 340 families were included in the multiplicative fitness estimations.

DNA Extraction and Sequencing

As part of a larger NSF grant in the Keller lab, DNA was extracted from the 340 maternal trees for genomic DNA sequencing of the red spruce exome. The resulting DNA sequences were used in the current study to estimate inbreeding levels of the maternal trees and their populations

for comparison to the fitness trait data of the seedlings. For DNA extraction, frozen (-80°C) needle tissue was arrayed from each family and extracted using a Qiagen DNeasy 96 Plant Kit. The DNA concentration of each extraction sample was quantified using a fluorometric-based DNA binding assay (Qubit; Invitrogen, USA) read on a BioTek Synergy H4 Multi-plate Reader. Samples were also run on an agarose gel to ensure high quality, high molecular weight DNA. Samples were then normalized to a working volume of 35 µL using an Eppendorf epMotion 5070 and sent to RAPiD Genomics at the University of Florida for exome capture and Illumina sequencing.

Complete details of exome sequence capture probe design, library construction, and Illumina sequencing are available from Dr. Thibault Capblancq (tcapblanc@uvm.edu) in the Keller Lab by request. In brief, capture baits were designed from two different reference transcriptomes for a close relative of red spruce (*Picea glauca*) based on previously published sequences (Rigault et al., 2011)(Yeaman et al., 2014). In total there were 80,000 120 base pair long probes generated from the protein coding region of the *P. glauca* genome. Red spruce DNA was mechanically sheared to 400 base pair long fragments and captured using biotinylated beads. Any sequences in the red spruce fragments that complemented the exomic probes would stick to the probes while uncomplemented DNA would wash out. The resulting capture sequences were then ligated with sample-specific barcodes and sequenced on 1 lane of Illumina HiSeq X. This process generated 150 base pair long *P. rubens* DNA fragments that were mapped back to the exomic reference genome for *P. glauca* using the genome aligner BWA-MEM.

Genotype and Inbreeding Calculations

Working with Dr. Capblancq, genotypes across the reconstructed red spruce exome were estimated using the program ANGSD (Analysis of Next Generation Sequencing Data)

(Korneliussen, Albrechtsen & Nielsen 2014). ANGSD was used to estimate and assign the probability of different genotypes based on the number of overlapping mapped sequences at different polymorphic sites across all the individuals.

Dr. Capblanq used the allele frequencies resulting from ANGSD analyses to calculate the number of expected heterozygous individuals under Hardy-Weinberg equilibrium (i.e., random mating) within each population. By comparing the expected and observed frequency of heterozygous genotypes, he identified the proportion of polymorphic sites showing an excess of homozygosity. I used this proportion as a metric for the magnitude of inbreeding at the population level. In a similar method, Dr. Capblanq estimated an individual-level inbreeding coefficient by calculating the proportion of homozygous genotypic sites within the exomic data of each individual family (i.e., per maternal tree). Only the polymorphic sites for which there were data for at least two maternal individuals within each of the 64 populations were used.

Statistical Analyses

To test for source region and inbreeding effects on early life fitness traits, I used linear models (lm), generalized linear mixed effect models (glmer), and linear mixed effect models (lmer) in R (R core team 2018). Separate models were run for each of the four fitness traits (50 seed weight, germination proportion out of 50, seedling mortality, and seedling height after 12 weeks) as well as the overall multiplicative fitness values. I designated source regions (core, margin, edge), population-homozygosity and individual-homozygosity (and their interactions) as fixed effects, and population and family as random effects.

Full models incorporating fixed effects, interactions between fixed effects, and random effects were run before backwards selection criteria was used to eliminate non-significant

interactions one by one to yield minimally adequate models. Akaike information criterion (AIC) was used to rank models based on how well models fit the data, after being penalized for the number of parameters (complexity). The mixed-effect model with the lowest AIC value, meaning it had the best fit, was then retained for inference regarding each individual fitness trait.

Results

Variation in Seedling Fitness Traits

Early life fitness traits showed considerable variation within my experiment, despite seeds and seedlings being measured in a common environment. Looking at the individual components of fitness (seed weight, germination proportion, seedling mortality, and seedling height), 50-seed weight was significantly different with respect to source region ($p < 0.05$; Figure 2), with seed weight of the edge and margin regions being significantly higher than the core region. Seed germination proportion was also significantly different among the three regions ($p < 0.0001$; Figure 3), being significantly lower in the edge region. There was a significantly positive relationship between seed weight and germination proportion ($p < 0.0001$; Figure 4), with higher 50-seed weights corresponding with an increase

in the germination proportion for each family, with a significant amount of variation ($R^2 = 0.06$).

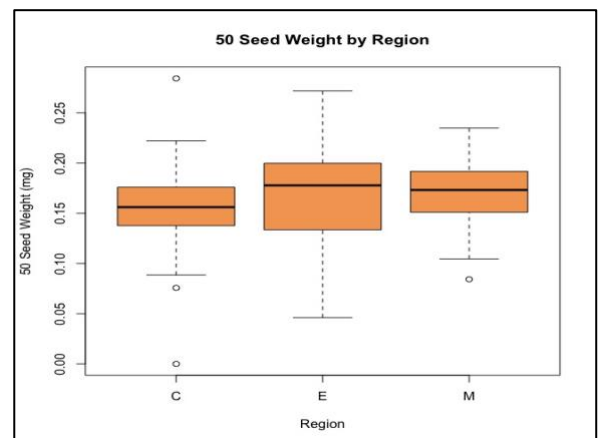


Figure 2. Boxplot distribution of 50-seed weight with respect to the three regions: C= core, E= edge, M= margin. For core region, min= 0.0, Q1= 0.14, mean= 0.16, Q3= 0.18, max= 0.28. For edge region, min= 0.05, Q1= 0.13, mean= 0.17, Q3= 0.20, max= 0.27. For marginal region, min= 0.08, Q1= 0.15, mean= 0.17, Q3= 0.19, max= 0.23. Summary of linear model of seed weight with respect to region yielded a p-value for edge region of 0.031* and a p-value of 0.01* for marginal region.

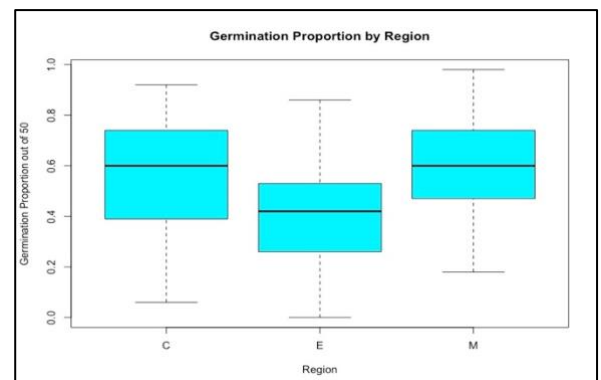


Figure 3. Boxplot distribution of germination proportion with respect to the three regions: C= core, E= edge, M= margin. For core region, min= 0.06, Q1= 0.40, mean= 0.57, Q3= 0.74, max= 0.92. For edge region, min= 0.00, Q1= 0.26, mean= 0.41, Q3= 0.53, max= 0.86. For marginal region, min= 0.18, Q1= 0.47, mean= 0.60, Q3= 0.74, max= 0.98. Summary of linear model of germination proportion with respect to region yielded a p-value for edge region of $2.21e-09^{***}$.

Seedling mortality, was significantly different with respect to region, showing a significantly lower proportion of dead seedlings in the edge region compared to the other regions ($p < 0.0001$;

Figure 5). Seedling height at 12 weeks was not

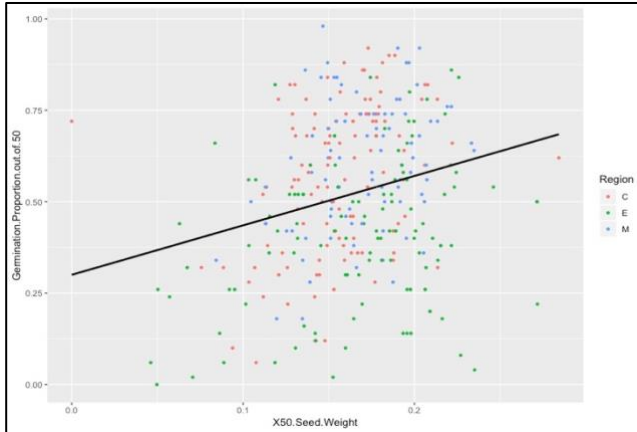


Figure 4. Linear regression of effect of 50 seed weight on germination proportion out of 50, regions are distinguished by different colors: C= core, E= edge, M= margin. R^2 for the regression line is 0.062. ANOVA test of significance yielded a p-value of $7.04e-06^{***}$.

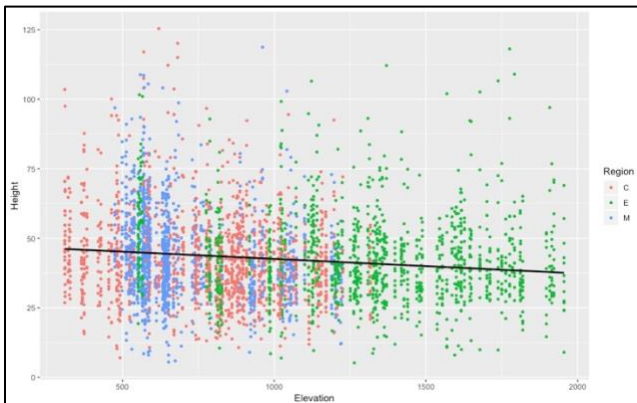


Figure 6. Linear regression of fixed effect of elevation on seedling height after 12 weeks, incorporating the random effect of region: C= core, E= edge, M= margin. R^2 for the regression line is 0.02. ANOVA test of significance yielded a p-value of $9.434e-16^{***}$.

significantly differentiated among source

regions. However, source elevation showed a strong association with height ($p < 0.0001$; Figure 6). Seedlings from mother trees of the edge region made up the majority of the highest elevation values, and overall there was a negative trend between elevation and seedling height, although considerable variation was still present ($R^2 = 0.02$). Seedling height was also significantly

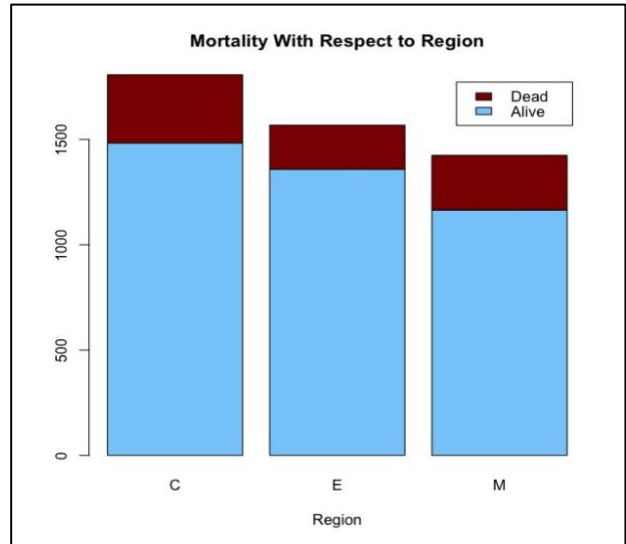


Figure 5. Barplot showing the proportion of dead and alive individual seedlings with respect to source region. The percent of dead seedlings for each region is 17.9%, 13.3%, and 18.2% for core, edge, and margin respectively. Summary of generalized linear model yielded a p-value for edge region of $p = 0.0003^{***}$.

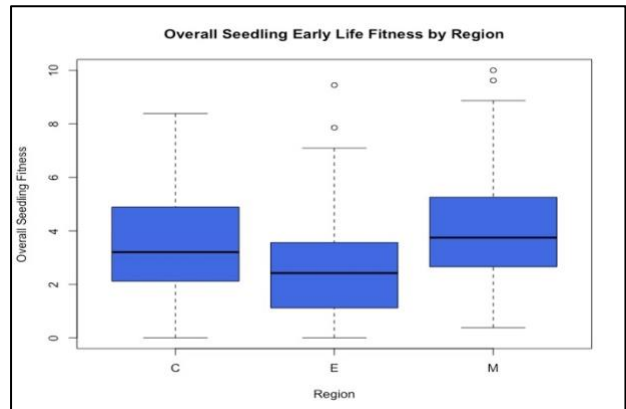


Figure 7. Boxplot distribution of relative fitness values categorized by source region: C= core, E= edge, M= margin. For core region, min= 0.0, 1st Q= 2.18, mean= 3.46, 3rd Q= 4.90, max= 8.40. For marginal region, min= 0.38, 1st Q= 2.66, mean=4.01, 3rd Q=5.26, max=10.00. For edge region, min=0.0, 1st Q=1.19, mean=2.61, 3rd Q=3.56, max=9.45. ANOVA statistical test yielded a p value for edge region of $p = 0.001^{**}$ and a p value for margin region of $p = 0.045^*$.

related to 50-seed weight when the effects of both elevation and seed weight were incorporated into a linear model ($p < 0.001$). Overall multiplicative fitness values showed an effect of region, with significantly ($p = 0.001$) lower mean fitness for families from the edge and significantly higher ($p < 0.05$) mean fitness for families from the marginal region compared to the range core (Figure 7).

Variation in Genomic Inbreeding at Population and Individual Levels

Population-level proportion of excess homozygous sites was significantly different among

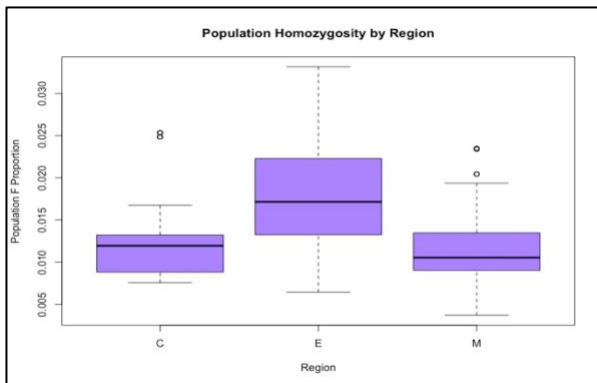


Figure 8. Boxplot distribution of proportion of population level excess homozygous sites with respect to the three source regions, C= core, E= edge, M=margin. For core region, min= 0.0076, Q1= 0.0088, mean= 0.012, Q3= 0.013, max= 0.025. For edge region, min= 0.0064, Q1= 0.013, mean= 0.017, Q3= 0.022, max= 0.033. For marginal region, min= 0.0037, Q1= 0.0091, mean= 0.012, Q3= 0.013, max= 0.023. Summary of linear model of population homozygosity with respect to region yielded a p-value for the edge region of 2.1×10^{-12} ***.

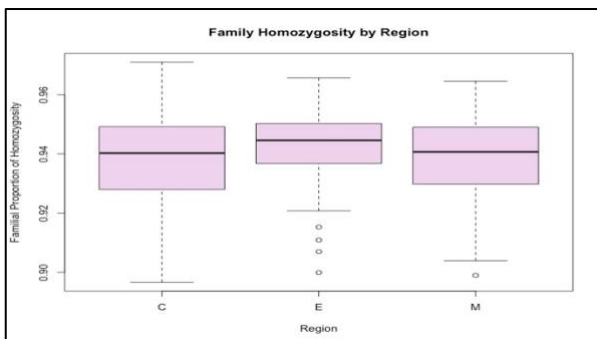


Figure 10. Boxplot distribution of proportion of family level excess homozygous sites with respect to the three regions. C= core, E= edge, M= margin. For core region, min= 0.90, Q1= 0.93, mean= 0.94, Q3= 0.95, max= 0.97. For edge region, min= 0.90, Q1= 0.93, mean= 0.94, Q3= 0.95, max= 0.97. For marginal region, min= 0.90, Q1= 0.93, mean= 0.94, Q3= 0.95, max= 0.96. Summary of linear model of family homozygosity with respect to region yielded a p-value for edge region of 0.046*.

the three source regions, with the greatest population homozygosity evident in the edge region ($p < 0.001$; Figure 8). This pattern can also be seen in Figure 9., which shows the geographical

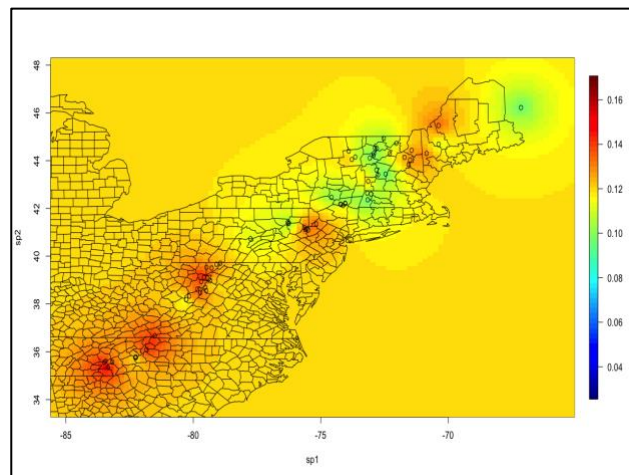


Figure 9. Geographical range map of 64 collection sites used for common garden planting from three source regions: Core, Margin, and Edge shown from northern to southern latitudes. Proportion of excess homozygous sites for each population is shown along a color gradient for the entire range.

distribution of the 64 populations displayed with a color gradient based on proportion of population level excess homozygous sites. The darker red

color indicates a higher proportion of homozygous sites. Variation in individual-level homozygosity was also significant among regions (Figure 10), with families from the edge region showing significantly higher proportions of individual homozygosity than families from the core region ($p < 0.05$).

Relative Importance of Source Range and Inbreeding for Early Life Fitness

Correlations between traits and inbreeding were evident for several life stages (Figure 11) phenotypic and genotypic traits analyzed in the mixed effect models. In order to better understand the factors contributing to the individual phenotypic traits, linear mixed effect models were used to analyze the effects of source region and inbreeding metrics while accounting for random effects of population and family (Table 1).

Starting with the earliest life fitness trait, 50-seed weight was a family level trait and included only a random effect of population. The model with the lowest AIC ranking, indicating the best fit for the data, included the individual effects of region, population and family level homozygosity, and the interactions between factors.

Like seed weight, germination proportion was a family level trait. The initial model included 50 seed weight as a fixed effect due to their highly significant relationship (see Figure 4). The model with the lowest AIC score for germination included interactions between region and both levels of homozygosity and the random effect of population (Table 1).

Looking at individual seedling mortality, the full model was unable to converge so it had to be simplified to only include one interaction, region*population homozygosity (determined by relative significance for both interactions), and the random effect of family (which was more variable than population). After backwards model selection, the simplified model with the lowest

AIC ranking included the fixed effects of region and individual homozygosity and the random effect of family. The model showed a negative trend with respect to edge region, consistent with the results shown in Figure 5. and a significantly positive trend with respect to family level homozygosity ($p < 0.05$).

For height values of the individual seedlings, the initial model that included 50-seed weight, due to their significant relationship ($p < 0.001$), and retained all of the fixed effects and interactions between region and both levels of homozygosity had the lowest AIC score indicating that, although it was the most complex model, it was the best fit model for the seedling height data.

Looking at overall early life fitness, like the other family level traits, population was the only random effect while region and homozygosity were fixed effects. The model for early life fitness that best fit the data based on AIC ranking was the most complex model that included all of the individual factors and the interactions between them. While none of the individual predictors in the model were statistically significant (Table 1), a simple linear regression between fitness and population inbreeding was highly significant ($p < 0.0001$), with a negative trend between overall fitness and population homozygosity as predicted under the hypothesis of inbreeding depression (Figure 12).

Table 1. Table of trend estimates for each phenotypic trait (columns) with respect to factors (rows) of the models for each trait with the lowest AIC score, signifying the best fit based on the data. Significance is noted with asterisk(s) next to the respective estimation values.					
	50 SeedWeight	Germination Proportion	Seedling Mortality	Seedling Height	Seedling Early Life Fitness
50 Seed Weight		estimate= 1.64***		estimate= 23.58*	
Edge Region	estimate= -0.094	estimate= 1.15	estimate= -0.61	estimate= 33.35	estimate= 6.72
Margin Region	estimate= -0.65*	estimate= 2.05	estimate= 0.06	estimate= 84.31	estimate= 1.61
Population Homozygosity	estimate= -1.32	estimate= -5.16		estimate= 5.61	estimate= -18.12
Family Homozygosity	estimate= -0.15	estimate= 0.83	estimate= 20.76*	estimate= 18.22	estimate= 1.39
Edge Region*PopHomozygosity	estimate= -0.50	estimate= -4.46		estimate= -20.32	estimate= -13.22
Margin Region*PopHomozygosity	estimate= 1.56	estimate= 10.32		estimate= 144.41	estimate= 33.37
Edge Region*FamHomozygosity	estimate= 0.12	estimate= -1.29		estimate= -36.11	estimate= -7.11
Margin Region*FamHomozygosity	estimate= 0.69*	estimate= -2.31		estimate= -91.48	estimate= -1.98

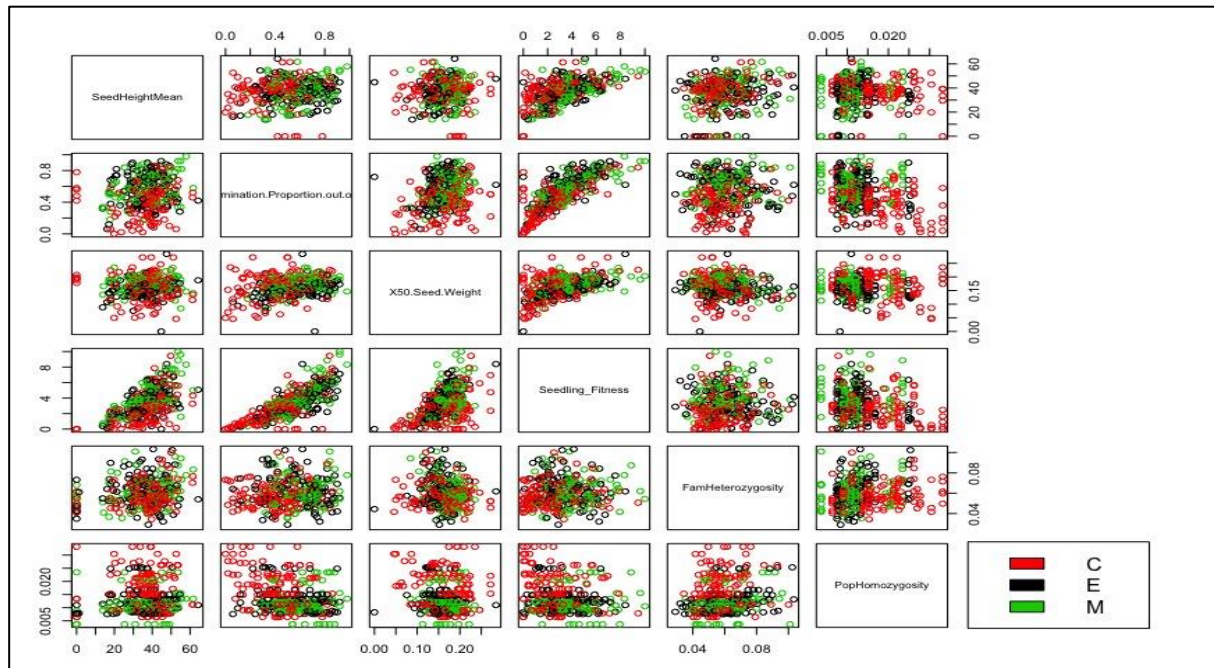


Figure 11. Pair wise plot showing the correlation between the family level phenotypic traits (mean seedling height, germination proportion out of 50, and 50 seed weight), overall family relative fitness values, and genotypic values for population level and family level homozygosity. Plot is colored by region: C= core (red), E= edge (black), M= margin (green).

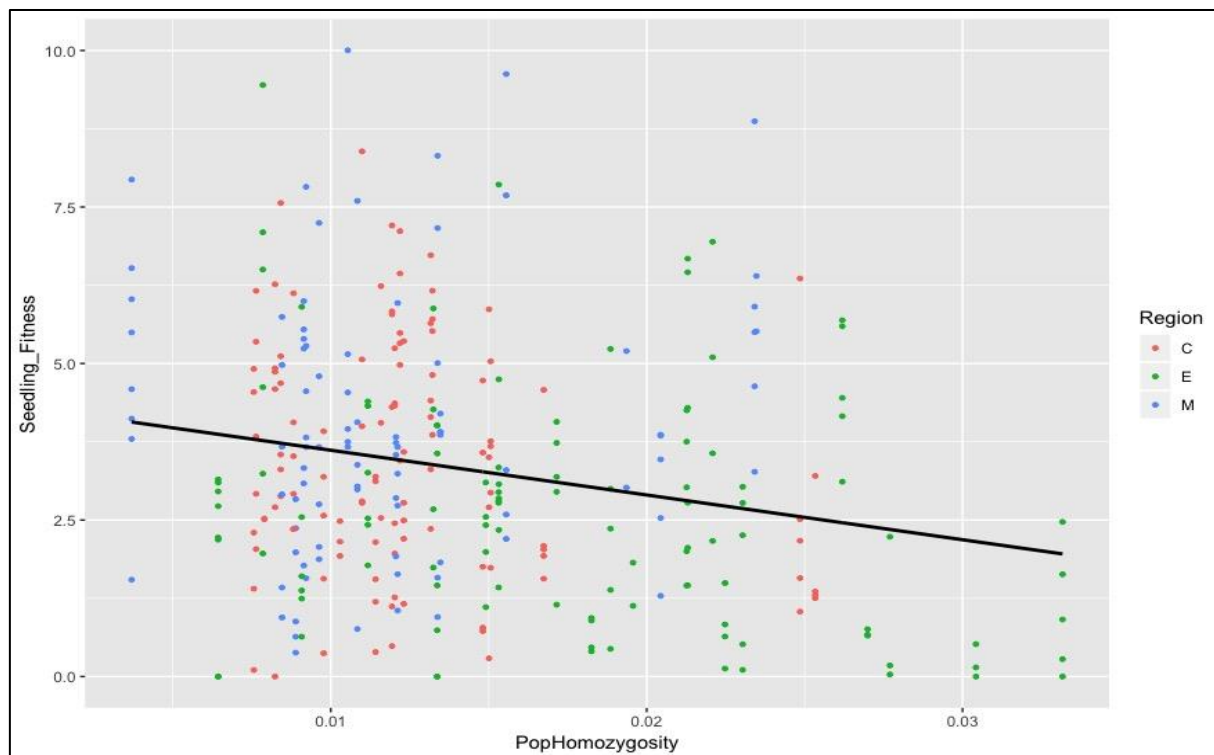


Figure 12. Mixed effect model of source population homozygosity proportion effect on overall relative fitness for each family. Source regions are shown in different colors: C=core, E=edge, M=margin. ANOVA test of significance for population homozygosity effect on relative fitness yielded a p-value of $7.68e-05^{***}$. R^2 for the regression is 0.049.

Discussion

The objective of this study was to look at early life traits as measures of fitness for *Picea rubens* seedlings and compare those traits with their source region and genotypic levels of inbreeding calculated from exome sequencing of the mother trees. When looking at the phenotypic traits individually, there appeared to be some relationship between source environment and fitness. Height was significant with respect to elevation and despite being grown under the same conditions in the greenhouse, there appeared to be a negative relationship between source elevation and overall seedling height, with the highest source elevation and shortest seedlings being from the edge region. Seed weight also had a significant effect on seedling height, based on the linear models. This could imply that climatic conditions from the source environment of the seedlings could be contributing to potential resource allocation from the mother tree to the seeds, affecting early life traits and seedling relative fitness. The study of *Calluna vulgaris*, *Erica cinerea*, and *Erica vagans* done by (Vera, 1997), suggested that seed weight has a direct correlation with overall seedling survival and growth, with larger, more nutrient substantive seeds generally outcompeting smaller seeds in germination and growth.

Overall there was a significantly higher proportion of population level homozygosity and individual level homozygosity in the edge region when compared to the core or margin. The higher proportion of homozygosity is indicative of a higher level of inbreeding due to the deviation away from expected levels of population and individual heterozygosity based on Hardy Weinberg equilibrium (Vieira, Fumagalli, Albrechtsen, & Nielsen, 2013). This supports the idea that smaller, more fragmented populations, like those in the southern edge, have generally low to null levels of gene flow and are more susceptible to loss of heterozygosity due to genetic drift and inbreeding (Kardos, Taylor, Ellegren, Luikart, & Allendorf, 2016; Nei & Tajima, 1981).

Despite significant differences among regions in early life fitness traits and in genome-wide inbreeding, most traits did not show significant associations in mixed-models that accounted for the variability among levels of population and family random effects (Table 1). Only seedling mortality showed a significant relationship with family level homozygosity that was robust to the inclusion of random effects in the mixed-effect models. Surprisingly, multiplicative fitness was not significantly predicted by region or inbreeding, despite mortality contributing as a component of fitness. However, there was significantly lower overall fitness for families from the edge region when compared to the core or margin, and also a significantly negative relationship between overall fitness and population level homozygosity, with fitness decreasing as homozygosity increased. Both of these results are inline with my predictions that early life fitness should be lowest in areas where range fragmentation and inbreeding rates are highest. However, when incorporating those factors into the best fit mixed effect model for relative fitness, the effects of region and homozygosity were non-significant indicating that there may be a significant relationship between relative fitness and population homozygosity, but it could be masked by the combined effects of region, population and family homozygosity, and their interactions.

There appeared to be a very strong correlation between overall seedling fitness and germination proportion (Figure 11), which might have meant that the significant relationship between germination proportion and region was heavily influencing the significant relationship between overall relative fitness and region. To account for this, mean seedling height, that had been weighted with zeros to account for seedling mortality, was used as a proxy for early life fitness to see whether or not the linear mixed effect model for mean seedling height would be significant. The model was not significant for any of the fixed effects included in the best fit model for fitness, indicating that although germination proportion might have an influence on fitness, it

was not alone driving the significant relationship between fitness and region.

Other explanations for the relatively weak correspondence between fitness traits and genomic estimates of inbreeding could be due to statistical factors associated with the sequence data. The metrics we used for population and family level homozygosity were non-parametric proxies to F_{IS} and inbreeding coefficients directly calculated from Hardy-Weinberg equations. These proxies were necessary because of low coverage of sequence reads at polymorphic sites which required the application of genotype probabilities instead of assigning genotypes with “hard calls”. By calculating the proportion of homozygous sites from the exome sequences of the individual trees and the population wide homozygosity explicitly, the genomic data were able to provide proxy estimates of inbreeding levels, but do not correspond strictly to Hardy-Weinberg estimation of population and family levels of inbreeding, which could have had some effects on the model estimations.

Another future research need to better understand the causes of variation in early life is to estimate evolutionary models of the demogenetic history of red spruce throughout its range. There is some evidence to suggest that there was a northward expansion of spruce during the Holocene towards cooler, wetter environments that spruce is better adapted to (Blum, 1990; Schaffler & Jacobson, 2002). If there was a historic southern refugia that migrated northward as the climate warmed, that could have long standing implications for the genetic diversity of the entire red spruce range. If red spruce had migrated northward away from the source pool refugia in the south, that could have generated a leading edge, trailing edge dynamic of the range (Hampe & Petit, 2005). Through a succession of founder effects, the genetic variation at the leading edge of the range could have become diluted compared to the trailing source edge, leading to potentially higher level of genetic variation in the trailing edge, or a non-significant difference between the phenotypic

fitness of northern and southern regions of the range.

This dynamic has been suggested in other studies such as the one done on *Capsella bursa-pastoris*, shepherd's purse (Orsucci, Milesi, Hansen, Girodolle, & Lascoux, 2019), in which individuals ascending from an Asian colonization front or from the more genetically rich European core of its range were compared based on life history traits. The authors found that individuals ascending from the genetically rich core did better than those from the colonization front for most life history traits including: germination rate, vegetative traits, phenology, and reproductive traits, but those from the colonization front actually did better in competition than those from the core. In terms of red spruce, my results could indicate a potential counter-balancing of the negative effects of the current fragmented, isolated population structure with the historic genetic diversity present in the southern range. This could provide an explanation for the variably significant individual early life traits in the edge region (i.e. higher seed weight but lower germination proportion and lower mortality) but overall lower seedling fitness consistent with the predictions of inbreeding depression in the southern edge.

Even if the unknown demographic history across red spruce's range introduces potential complexity to interpreting my findings, there is still abundant evidence to suggest that genetic drift and increasing levels of inbreeding in small isolated populations can greatly reduce adaptive potential (García-Fernández et al., 2012; Willi, Van Buskirk, & Hoffmann, 2006). As the climate continues to change, the smaller effective population sizes in the edge region could make those populations more susceptible to loss of fitness and local extinction. This study provides evidence to support there being a higher level of inbreeding at a population level and family level in the edge region, meaning that even if the early life phenotypic traits are not dramatically reduced now, there is still potential for consequences of inbreeding depression in the future with continuing

shrinkage of suitable habitat due to climate change (Boisvert-Marsh et al., 2014).

The adaptive potential of current populations of *Picea rubens* throughout its range may be something to consider in future restoration efforts, especially in the southern edge. Genetic diversity is a major factor that should be considered when looking at the potential success of short and long term restoration efforts as it can play a role in the short term plasticity of a population with respect to stress response as well as the long term adaptive potential as local environments will continue to change (Sujii et al., 2017; Thomas et al., 2014; Yeaman et al., 2014). This study provides evidence that the highest levels of inbreeding throughout the range occur in the southern edge, as well as reduced overall relative early life fitness when compared to the more continuous core and marginal regions. There are still some caveats of this study to consider when looking at the current interactions between environment, phenotypic traits, and levels of inbreeding, but all of these factors together are necessary to understand the future trajectory of small, potentially inbred populations with respect to climate change and what that means for future restoration efforts.

Acknowledgements

I would like to thank Dr. Stephen Keller for all of his contributions to my project and for the time and resources he has put into being my thesis advisor. I would like to thank Dr. Thibaut Capblanq for his contributions to my data analyses and thesis process. I would like to thank Sonia DeYoung and Ethan Thibault for their contributions to the field and lab work portions of my project. I would like to acknowledge the entire Keller lab for all of their time and support with my thesis and I would finally like to thank my other committee members, Dr. Charlotte Mehrtens and Dr. Donald Stratton for their time and effort to review my project and be on my thesis defense committee.

Funding for this project came from the Keller Lab NSF Spruce grant as well as through a SURF provided by the Carl Reidel Scholarship Fund and the UVM Office of Undergraduate Research (FOUR).

Literature Cited

- Abu Awad, D., & Billiard, S. (2017). The double edged sword: The demographic consequences of the evolution of self-fertilization. *Evolution*, *71*(5), 1178–1190. <https://doi.org/10.1111/evo.13222>
- Angeloni, F., Ouborg, N. J., & Leimu, R. (2011). Meta-analysis on the association of population size and life history with inbreeding depression in plants. *Biological Conservation*, *144*(1), 35–43. <https://doi.org/10.1016/j.biocon.2010.08.016>
- Blum, B. M. (1990). Red Spruce, 250–259. Retrieved from https://plants.usda.gov/factsheet/pdf/fs_piru.pdf
- Boisvert-Marsh, L., Périé, C., & De Blois, S. (2014). Shifting with climate? Evidence for recent changes in tree species distribution at high latitudes. *Ecosphere*, *5*(7), 1–33. <https://doi.org/10.1890/ES14-00111.1>
- Charlesworth, D., & Willis, J. H. (2009). The genetics of inbreeding depression. *Nature Reviews. Genetics*, *10*(11), 783–796. <https://doi.org/10.1038/nrg2664>
- Cheptou, P.-O., & Donohue, K. (2011). Environment-dependent inbreeding depression: its ecological and evolutionary significance. *The New Phytologist*, *189*(2), 395–407. <https://doi.org/10.1111/j.1469-8137.2010.03541.x>

Davis, M.B. and Shaw, R.G., 2001. Range shifts and adaptive responses to Quaternary climate change. *Science*, 292(5517), pp.673-679.

García-Fernández, A., Iriondo, J. M., & Escudero, A. (2012). Inbreeding at the edge: Does inbreeding depression increase under more stressful conditions? *Oikos*, 121(9), 1435–1445. <https://doi.org/10.1111/j.1600-0706.2011.20219.x>

Govindaraju, D. R. (1988). Life histories, neighbourhood sizes, and variance structure in some North American conifers. *Biological Journal of the Linnean Society*, 35(1), 69–78. <https://doi.org/10.1111/j.1095-8312.1988.tb00459.x>

Hampe, A., Pemonge, M. H., & Petit, R. J. (2013). Efficient mitigation of founder effects during the establishment of a leadingedge oak population. *Proceedings of the Royal Society B: Biological Sciences*, 280(1764). <https://doi.org/10.1098/rspb.2013.1070>

Hampe, A., & Petit, R. J. (2005). Conserving biodiversity under climate change: The rear edge matters. *Ecology Letters*, 8(5), 461–467. <https://doi.org/10.1111/j.1461-0248.2005.00739.x>

Hart, A. C. (1959). Silvical Characteristics of Red Spruce. *USDA Forest Service*, 18.

Hendrick, P. W., & Cockerham, C. C. (1986). Partial inbreeding: equilibrium heterozygosity and the heterozygosity paradox. *Evolution*, 40(4), 856–861.

Kardos, M., Taylor, H. R., Ellegren, H., Luikart, G., & Allendorf, F. W. (2016). Genomics advances the study of inbreeding depression in the wild. *Evolutionary Applications*, 9(10), 1205–1218. <https://doi.org/10.1111/eva.12414>

Koo, K. A., Madden, M., & Patten, B. C. (2014). Projection of red spruce (*Picea rubens* Sargent) habitat suitability and distribution in the Southern Appalachian Mountains, USA. *Ecological*

Modelling, 293, 91–101. <https://doi.org/10.1016/j.ecolmodel.2014.06.005>

Koo, K. A., Patten, B. C., & Creed, I. F. (2011). *Picea rubens* growth at high versus low elevations in the Great Smoky Mountains National Park: evaluation by systems modeling. *Canadian Journal of Forest Research*, 41(5), 945–962. <https://doi.org/10.1139/x10-243>

Koo, K. A., Patten, B. C., Teskey, R. O., & Creed, I. F. (2014). Climate change effects on red spruce decline mitigated by reduction in air pollution within its shrinking habitat range. *Ecological Modelling*, 293, 81–90. <https://doi.org/10.1016/j.ecolmodel.2014.07.017>

Korneliussen, T.S., Albrechtsen, A. and Nielsen, R., 2014. ANGSD: analysis of next generation sequencing data. *BMC bioinformatics*, 15(1), p.356.

Levin, D. A. (2011). Mating system shifts on the trailing edge, *109*(3), 613–620. <https://doi.org/10.1093/aob/mcr159>

Major, J. E., Mosseler, A., Johnsen, K. H., Campbell, M., & Malcolm, J. (2015). Growth and allocation of *Picea rubens*, *Picea mariana*, and their hybrids under ambient and elevated CO₂. *Canadian Journal of Forest Research*, 45(7), 877–887. <https://doi.org/10.1139/cjfr-2014-0525>

Mosseler, A., Major, J. E., & Rajora, O. P. (2003). Old-growth red spruce forests as reservoirs of genetic diversity and reproductive fitness. *Theoretical and Applied Genetics*, 106(5), 931–937. <https://doi.org/10.1007/s00122-002-1156-1>

Mosseler, A., Major, J. E., Simpson, J. D., Daigle, B., Lange, K., Park, Y. S., ... Rajora, O. P. (2000). Indicators of population viability in red spruce, *Picea rubens*. I. Reproductive traits and fecundity. *Canadian Journal of Botany*, 78(7), 928–940. <https://doi.org/10.1139/b00-065>

- Nei, M., & Tajima, F. (1981). GENETIC DRIFT AND ESTIMATION OF EFFECTIVE POPULATION SIZE. *Genetics*, 98, 625–640. Retrieved from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1214463/pdf/625.pdf>
- Orsucci, M., Milesi, P., Hansen, J., Girodolle, J., & Lascoux, M. (2019). Ecological strategy and genetic load in the shepherd ' s purse (*Capsella bursa-pastoris*) from the core and the limit of its natural range .
- Pluess, A. R. (2011). Pursuing glacier retreat: Genetic structure of a rapidly expanding *Larix decidua* population. *Molecular Ecology*, 20(3), 473–485. <https://doi.org/10.1111/j.1365-294X.2010.04972.x>
- Porcher, E., & Lande, R. (2016). Inbreeding depression under mixed outcrossing, self-fertilization and sib-mating. *BMC Evolutionary Biology*, 16(1), 1–14. <https://doi.org/10.1186/s12862-016-0668-2>
- R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Rigault, P., Boyle, B., Lepage, P., Cooke, J. E. K., Bousquet, J., & Mackay, J. J. (2011). Genome Analysis A White Spruce Gene Catalog for Conifer Genome Analyses. *Plant Physiology*, 157, 14–28. <https://doi.org/10.1104/pp.111.179663>
- Schauffler, M., & Jacobson, G. L. (2002). Persistence of coastal spruce refugia during the Holocene in northern New England, USA, detected by stand-scale pollen stratigraphies. *Journal of Ecology*, 90(2), 235–250. <https://doi.org/10.1046/j.1365-2745.2001.00656.x>
- Scoles, G., Kang, B.-Y., Major, J. E., & Rajora, O. P. (2011). A high-density genetic linkage map

of a black spruce (*Picea mariana*) × red spruce (*Picea rubens*) interspecific hybrid. *Genome*, 54(2), 128–143. <https://doi.org/10.1139/G10-099>

Stoehr, M., Ott, P., & Woods, J. (2015). Inbreeding in mid-rotation coastal Douglas-fir: implications for breeding. *Annals of Forest Science*, 72(2), 195–204. <https://doi.org/10.1007/s13595-014-0414-0>

Sujii, P. S., Schwarcz, K. D., Grando, C., de Aguiar Silvestre, E., Mori, G. M., Brancalion, P. H. S., & Zucchi, M. I. (2017). Recovery of genetic diversity levels of a Neotropical tree in Atlantic Forest restoration plantations. *Biological Conservation*, 211(April), 110–116. <https://doi.org/10.1016/j.biocon.2017.05.006>

Thomas, E., Jalonen, R., Loo, J., Boshier, D., Gallo, L., Cavers, S., ... Bozzano, M. (2014). Genetic considerations in ecosystem restoration using native tree species. *Forest Ecology and Management*, 333, 66–75. <https://doi.org/10.1016/j.foreco.2014.07.015>

Vera, M. L. (1997). *Effects of Altitude and Seed Size on Germination and Seedling Survival of Heathland Plants in North. aPlant vuEcology* (Vol. 133). Retrieved from <https://www-jstor-org.ezproxy.uvm.edu/stable/pdf/20050545.pdf?refreqid=excelsior%3Aa7c68a40576a4c44da7fef5c83338>

Vieira, F. G., Fumagalli, M., Albrechtsen, A., & Nielsen, R. (2013). Estimating inbreeding coefficients from NGS data: Impact on genotype calling and allele frequency estimation. *Genome Research*, 23(11), 1852–1861. <https://doi.org/10.1101/gr.157388.113>

Wang, J., Caballero, A., & Hill, W. G. (1998). The effect of linkage disequilibrium and deviation from Hardy-Weinberg proportions on the changes in genetic variance with bottlenecking.

Heredity, 81(2), 174–186. <https://doi.org/10.1038/sj.hdy.6883900>

Warren, R. L., Keeling, C. I., Yuen, M. M. Saint, Raymond, A., Taylor, G. A., Vandervalk, B. P., ... Bohlmann, J. (2015). Improved white spruce (*Picea glauca*) genome assemblies and annotation of large gene families of conifer terpenoid and phenolic defense metabolism. *Plant Journal*, 83(2), 189–212. <https://doi.org/10.1111/tpj.12886>

Willi, Y., Van Buskirk, J., & Hoffmann, A. A. (2006). Limits to the Adaptive Potential of Small Populations. *Annual Review of Ecology, Evolution, and Systematics*, 37(1), 433–458. <https://doi.org/10.1146/annurev.ecolsys.37.091305.110145>

Yeaman, S., Hodgins, K., Suren, H., Nurkowski, K., Rieseberg, L., Holliday, J., & Aitken, S. (2014). Conservation and divergence of gene expression plasticity following c. 140 million years of evolution in lodgepole pine (*Pinus contorta*) and interior spruce (*Picea glauca* x *Picea engelmannii*). *New Phytologist*, 203, 578–591.