2019

Spatial Genetic Structure and Local Adaptation within and among Foxtail Pine (Pinus balfouriana subsp. balfouriana) Populations Located in the Klamath Mountains, California

Rebecca D. Piri

Spatial Genetic Structure and Local Adaptation within and among Foxtail Pine (*Pinus balfouriana* subsp. *balfouriana*) Populations Located in the Klamath Mountains, California

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science at Virginia Commonwealth University.

by

Rebecca Dahlberg Piri

Director: Andrew J. Eckert, Ph.D.
Associate Professor, Department of Biology

Virginia Commonwealth University
Richmond, VA
August 2019

Acknowledgements

I would like to thank to everyone who helped me through this process. To my advisor, Dr. Andrew Eckert, thank you for your guidance and support and for holding me to a high standard. To my committee, thank you for your valuable feedback. To everyone in the Eckert lab, thank you for your patience, reassurance, and friendship. To Mitra Menon, thank you for your invaluable help with all of my technical issues. To my former boss, Dr. Eugene Maurakis, thank you for giving me the confidence to pursue this degree.

And finally, thank you to my family and my husband, Sepehr. I would not have been able to do this without your unwavering love and belief in me.

Table of Contents

List of Figures

List of Tables

ABSTRACT

SPATIAL GENETIC STRUCTURE AND LOCAL ADAPTATION WITHIN AND AMONG FOXTAIL PINE (*PINUS BALFOURIANA* SUBSP. *BALFOURIANA*) POPULATIONS LOCATED IN THE KLAMATH MOUNTAINS, CALIFORNIA
By Rebecca Dahlberg Piri

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science at Virginia Commonwealth University.

Virginia Commonwealth University, 2019

Major Director: Andrew J. Eckert, Ph.D., Assistant Professor, VCU Department of Biology

Foxtail pine (*Pinus balfouriana*) is a subalpine conifer endemic to California, notably separated into two disjunct subspecies. Previous studies have described the northern subspecies, *Pinus balfouriana* subsp. *balfouriana,* as having an uncommonly high level of genetic differentiation and no discernible spatial patterns in phenotypic variation. This study seeks to characterize the spatial genetic structure and patterns of selection of the northern subspecies (*Pinus balfouriana* subsp. *balfouriana*) using genome-wide data and to identify the influence of ecology and environment on the unique genetic patterns. I show that genetic differentiation among populations is much less than previously estimated (F_{ST} = 0.000644) and there is weak isolation-by-distance structure, but ongoing gene flow is unlikely. Within populations, stand density and competitor effects contribute to inbreeding. I also show that previously measured traits are predominantly determined by genetics. Analyzing by sliding window in the genome, I show that connectivity patterns vary widely throughout the genome and identify several areas that are important to the genetic architecture of the phenotypic traits and plasticity (GxE). Overall, there is high connectivity, genetic similarity, and genetically based trait variation among and within populations of the northern subspecies of foxtail pine due to historical processes, despite biotic interactions driving inbreeding. Persistent genetic isolation, however, may make adaptation to future climate a challenge for the subspecies.

INTRODUCTION

Subalpine montane ecosystems are cool-climate, conifer-rich forests at the highest altitude that trees can establish on mountaintops and act as a transition between the alpine tundra and forest ecoregions (Smith et al., 2008). Conifers found in this region are keystone species in their habitats and are ecologically important to lower-elevation montane habitats, as they can stabilize snow pack on mountaintops, which affects water availability during the growing season (Lapp et al., 2005). Their status is threatened by climate change, though, which modifies global habitats, selection pressures, and species distributions. The fast pace of environmental change and altered distribution of genetic variants leaves many species unable to keep pace, which can lead to extinction (Forister et al., 2010; Hoffmann & Sgro, 2011; Parmesan et al., 1999; Chen et al., 2011; Dawson et al., 2011). These effects are magnified in montane ecosystems, with temperatures rising quicker than most other habitats and many lower-altitude species shifting upwards to escape unfavorable conditions (Pauli et al., 2012; Pepin et al., 2015). These ecological changes leave native species to deal with both shifting climate and higher competition (Steinbauer et al., 2018). Subalpine conifers are further strained by long life cycles that make adaptation very slow (Pauli et al., 1996; Leitch & Leitch, 2012). To understand resiliency of these ecosystems and how best to approach forest management to mitigate climate effects, we must understand the patterns of genetic structure and evolutionary potential of conifers in the subalpine range.

Genetic diversity is key to a species' resiliency and longevity, providing variation upon which selection can act (Charlesworth & Charlesworth, 1987). Spatial genetic structure (SGS) is the non-random distribution of that variation among populations due to a variety of processes, including current or historical gene flow, colonization, and drift (Wright, 1931; Slatkin 1985; Hewitt 2001). These distribution patterns can influence adaptation and demography shifts (Sork et al. 2001). In conifers, genetic diversity within populations is generally high and differentiation among populations low due to long generation times and large effective population sizes (Achere et al., 2005; Aitken et al., 2008; Namroud et al., 2008). Wind-dispersed pollen and out-breeding mating systems also promote gene flow (Achere et al., 2005, Namroud et al., 2008), while fluctuating selection pressures, spatially variable selection pressures, and low levels of linkage disequilibrium maintain phenotypic diversity (Aitken et al., 2008). Still, dispersal limitation can cause isolation by distance (IBD), reducing connectivity between geographically isolated populations, or decrease pollen pool diversity within populations around barriers (Slatkin, 1993; Dyer & Sork, 2001; Savolainen et al., 2007). Demographic clines may also be present from sequential, founder-effect bottlenecks due to range expansion (Holliday et al., 2010). Diversity among populations due to differentiation can facilitate local adaptation, which is important for species resiliency (Aitken et al., 2008; Alberto et al., 2013; Lind et al. 2018), although persistent gene flow limitations exacerbated by drift and varying selection pressures can increase genetic load and reduce diversity within populations (Wright, 1951; Lowe, 2004).

Though processes like gene flow affect all loci, selection pressures and effective population sizes vary across the genome, resulting in uneven effects of drift and selection across loci within the genome (Cavalli-Sforza, 1966; Slatkin, 1985). Metrics used to measure patterns of population genetic structure are calculated as a function all alleles available, both genic and non-genic, resulting in an average value (Wright, 1951; Nei 1972, 1978; Excoffier et al., 1992), so that use of these global statistics can be uninformative about evolutionary processes related to drift and selection. To uncover genome-wide patterns of selection and drift, identifying locus-specific effects that vary significantly from the genome-wide patterns is instrumental (Eveno et al. 2008; Namroud et al. 2008). Several past studies have used this approach in conifers to identify selection using F_{ST} outlier analysis, and identified candidate genes for adaptive traits (Eveno et al. 2008; Namroud et al. 2008; Eckert et al., 2010).

Foxtail pine, the focal species of this study, is a long-lived, subalpine conifer endemic to the mountains of California, separated into two distinct subspecies by a 500km range disjunction. *Pinus balfouriana* subsp. *austrina,* the southern subspecies, is found in the Sierra Nevada Mountains, while *Pinus balfouriana* subsp. *balfouriana*, the northern subspecies, is found in the Klamath Mountains. The subspecies diverged during the early to mid-Pleistocene and currently have virtually no gene flow between them (Eckert et al., 2008). Divergent characteristics are evident that further distinguish the subspecies, including morphology and habitat differences (Mastrogiuseppe & Mastrogiuseppe, 1980). Genetic structure among local populations in each region also differs. While genetic diversity overall is higher

in the southern subspecies, differentiation among populations is higher in the northern subspecies (northern subspecies F_{st} = 0.242, southern subspecies F_{st} = 0.075) (Oline et al., 2000; Eckert et al., 2008). The distinctive population structure and uncommonly high differentiation of foxtail pine in the Klamath Mountains is ideal for a study seeking to understand the relative contributions of different evolutionary processes to the generation of unique spatial genetic patterns.

The Klamath Mountains are a region of high botanical diversity and environmental heterogeneity, with stark habitat shifts over short spatial distances (Sawyer & Thornburgh, 1977). Unlike the southern subspecies in the Sierra Nevada, which is the dominant tree species in its range (Cheng, 2004) and often occurs in pure stands (Lloyd et al., 1997), the northern subspecies experiences high competition in mixed stands that are common in this region (Cheng, 2004; Eckert, 2006). When found with co-occurring species, such as western white pine (*Pinus monticola* Dougl. ex D. Don) and red fir (*Abies magnifica* A Murray), the northern subspecies exhibits reduced population abundance and signs of shade stress (Eckert, 2006). At fine spatial scales, the northern subspecies can escape these competitors by establishing in microsites, created by boulders breaking up the landscape into many small, isolated sites. At broader spatial scales, they can utilize less-nutrient, ultramafic soils (Eckert, 2006). Competition influences the distribution of the northern subspecies of foxtail pine and likely provides additional selection pressures (Antonovics & Levin, 1980).

Abundant competitor species may also impact gene flow in the northern subspecies of foxtail pine populations. Barriers on the landscape, such as tree

density and rock formations, can disrupt pollen movement and decrease genetic diversity of pollen pools (Dyer & Sork, 2001). In northern foxtail populations, where the density of co-occurring conifer species also negatively impacts the population density, this effect is likely magnified (Eckert & Sawyer, 2002). At the species level, restricted pollen movement within and among populations and low population sizes can cause higher levels of inbreeding and contribute to reduced breeding among populations in the northern subspecies relative to the southern subspecies. High inbreeding, and therefore low genetic diversity, could put the subspecies at risk of an inbreeding depression (Charlesworth & Charlesworth, 1987; Wright, 1977).

If gene flow is restricted among populations of northern subspecies of foxtail pine, adaptive responses to changing environment may be decreased. Low levels of gene flow among populations, whether due to landscape barriers, pollen desiccation during dispersal, or isolation by distance, can decrease genetic diversity (Wright, 1943; Dyer & Sork, 2001; Bohrerova et al. 2009). For small populations, low levels of gene flow can inhibit adaptation due to limited genetic variation available for selection, which may not contain beneficial alleles (Holt & Gomulkiewicz, 1997). This is especially true for polygenic architectures of fitness, where many genes of small effect determine fitness levels, which is the case for most fitness traits in conifers (Le Corre & Kremer, 2003; Lind et al. 2018).

The phenotypic characteristics used to define the subspecies, including needle and cone morphology and the concentration of terpenes, which are secondary metabolic compounds, are highly variable within the northern subspecies (Mastrogiuseppe & Mastrogiuseppe, 1980; Zacher, 2015). These morphology and terpene concentrations have important fitness implications: terpenes and resin

ducts are key defenses against insects and pathogens (Huber & Bohlmann, 2006; Carmona et al., 2011; Iason et al., 2011), cone morphology is related to reproductive ability and defense against herbivores (Smith, 1970; Niklas & Kyaw, 1983; Coffey et al., 1999) and needle morphology is important for water use efficiency, especially during cold and hot climate shifts (Tranquillini, 1976; Hultine et al., 2000; Mayr et al., 2012). Past analysis found no evidence of a spatial pattern in these traits, which they interpreted as evidence that northern foxtail pine populations are more influenced by drift than selection (Zacher, 2015). However, the genetic and environmental influences on these traits are unknown, and there is no evidence whether or not these traits are adaptive within populations or among different environments.

High genetic differentiation among populations and morphological heterogeneity within populations in the northern subspecies of foxtail pine is well established by previous studies, but the processes affecting these patterns are poorly understood. Previous molecular data included allozymes (Oline et al. 2000), chloroplast, mitochondrial, and candidate gene nuclear DNA (Eckert et al., 2008; Eckert et al., 2010), all of which highlighted high differentiation among populations. Meanwhile, phenotypic heterogeneity exhibits high overall variation, but little to no differentiation among populations using global ANOVA-based tests (Zacher, 2015). The genetic bases of these traits and functional implications have yet to be studied.

The purpose of this study is to uncover mechanisms driving the observed differentiation patterns across the genome, and identify some of the genetic basis for measured traits. I investigated the contributions of drift and selection on the population genetic patterns of the northern subspecies of foxtail pine by evaluating

spatial genetic structure and connectivity among populations, both globally and by sliding window analysis to identify locus-specific effects. My goal is to understand the impact of evolutionary processes and environmental pressures on observed genetic patterns and how those effects vary across the genome. To achieve this goal, I tested the three following hypotheses: 1) Populations are highly differentiated genome-wide, with heterogeneity varying across the genome, and have patterns of connectivity resulting from demographic processes. 2) High levels of inbreeding, due to bottlenecks and reduced pollen and seed movement from high competitor density, influence differentiation among populations. 3) Morphological and chemical traits are correlated to genetic variation and are the result of natural selection pressures from local environments.

METHODS

Data Collection and Preparation

Population Data

Samples for this study include 378 foxtail pine trees sampled from 19 populations located in the Klamath Mountains of California (Figure 1, Appendix A), sampled in the summer of 2014 from seven regions. Trees were sampled at a height of 6 meters for pine needles and geo-referenced along transects. Phenotypic data were also measured for these trees, including 20 total traits, including cone and needle morphology and turpentine chemistry (Appendix B). In addition, 65

additional population observations were collected from GBIF using the *dismo* package in R v 1.1-4 (Hijmans et al., 2017) for niche modeling.

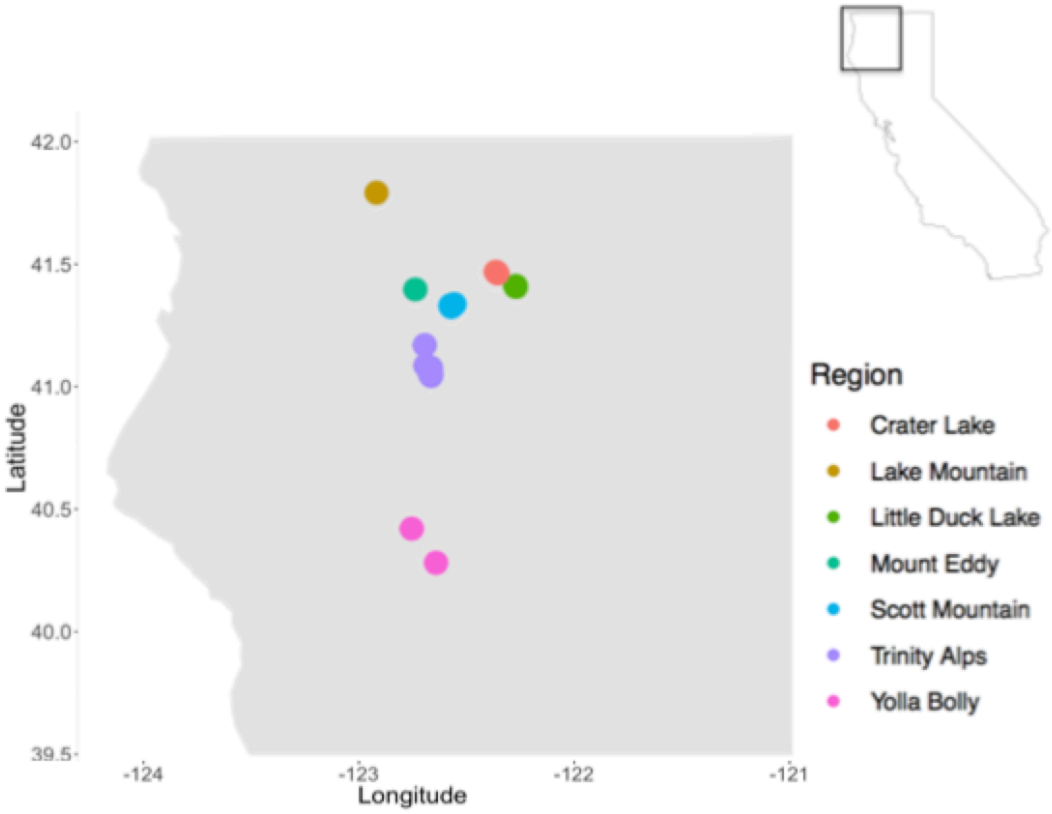


Figure 1: Populations sampled, n = 19

Niche Modeling

Climate data averaged from 1961 to 1990 at a 1km resolution were collected from Climate NA and trimmed to the range of northern foxtail pine. Out of 26 initial variables, 13 were identified as usable after removing variables that were correlated (Pearson's *r*) above 0.90 for the geo-referenced population (Appendix C, D).

The variables with the greatest importance for the climate niche of northern foxtail pine, which were used for later analysis based on their biological significance,

were identified using MaxEnt version 3.4.0 using 1000 bootstrap replicates. All populations, including sampled and unsampled stands, were mapped in ArcMap and thinned to a minimum distance of 1km to match the spatial resolution of the environmental data to avoid over-fitting data to populations reported multiple times. In total, 48 foxtail locations, in addition 19 populations used for genetic analysis, were used (Appendix E). The top variables that jointly explained over 90% of foxtail's niche were selected for later analysis.

Library Preparation and DNA Sequencing

Total DNA was extracted from the needle tissues using Qiagen DNeasy 96 Plant Kits following the manufacturer's protocol. Libraries were prepared for ddRADseq following the Parchman et al. (2012) protocol using EcoR1 and Mse1 restriction enzymes. Ligated, amplified, and pooled PCR products were separated by size using gel electrophoresis in 1% agarose gels, and fragments between 300 and 500 bp were isolated for sequencing. Sequencing was carried out using four multiplexed libraries, each with 96 sampled trees. Novogene performed single-end sequencing of the four multiplexed DNA libraries on the Illumina HiSeq 4000 platform.

Sequenced lanes were demultiplexed into separate fastq files for each individual using GBSX v 1.3-0 (Herten *et al.,* 2015). All reads were then trimmed to a set number of base pairs, excluding primers and barcodes, to help remove inaccuracies at the ends of the sequences, which interfere with *de novo* assembly, using fastp v 0.19.4 (Chen et al*.,* 2018). Length trimming was done to multiple lengths (60, 70, 80, and 90 bp; named Trim 60, Trim 70, Trim 80, and Trim 90) to

ensure genetic patterns were preserved and an optimal number of both single nucleotide polymorphisms (SNPs) and matches to a partial linkage map could be made. Trim lengths excluded adaptors and barcodes at the beginning of the read.

Single nucleotide polymorphisms (SNPs) were identified for all trim replicates of data using a *de novo* assembly in dDocent (Puritz et al., 2014; Puritz et al., 2014) with a CD-HIT similarity threshold of 0.85, and coverage and unique sequence thresholds at 3 and 5, respectively. All data sets were then all filtered using VCFtools (Danecek et al., 2011) using the following criteria: maximum read depth at the 0.50 quantile, minimum phred score of 20, minor allele frequency of 0.005, and maximum 50% missing data. All insertion-deletion (Indel) polymorphisms and SNPs with more than two alleles were also removed. Hierarchical fixation indices were estimated using using *hierfstat* package in R v 0.04-22 (Goudet, 2005). I used values of F_{IS} (>|0.50|) to filter SNPs, so that SNPS with $F_{IS} < -0.5$ were removed from the data set due to possible issues during assembly. I retained SNPs with high positive F_{IS} (>0.50) and $F_{ST} > 0.00$, however, due to their possible significance in the population structure.

The trimmed and filtered data sets were centered and scaled using methods from Patterson et al. (2006), and the normalized genetic patterns of each data set were analyzed with principal components analysis (PCA). The scores for the first two components for each PCA were extracted and compared using pairwise Procrustes rotations using the vegan package in R v 2.5-3. to ensure general patterns were conserved. The best data set, which had the most SNPs and matches to the linkage map, was selected for further analysis.

Alignment to Linkage Map

The contigs from all trimmed *de novo* assemblies were aligned with a previously created linkage map (n ≈ 20,000 contigs) using BLASTN v 2.7.1 (Altschul et al., 2000; Friedline et al., 2015). The linkage map provided information about the relative position of the SNPs in the genome, so patterns of selection and differentiation across the genome could be evaluated. The best two hits for each contig that matched to the linkage map were extracted, and contig matches were retained only if the best hit was at least 85% the same at query contig and the top hit was the same length as the second hit with a higher match quality or was longer than the second hit with at least the same match quality.

Final Data Set Selection

The trimmed dataset with the highest total SNPs and blast hits was used as the final genetic dataset for all following analysis to optimize comparisons between genotype and phenotype and understanding of patterns by region of the genome.

H1: Populations are highly differentiated genome-wide, with heterogeneity varying across the genome, and have patterns of connectivity resulting from demographic processes.

Genetic Structure

Genetic structure among the 19 sampled populations was examined using a suite of standard population genetic methods to assess whether the genome-wide differentiation is, in fact, high and if overall structure follows a demographic model, such as isolation by distance, as hypothesized. The PCA plot for the chosen data set

was analyzed to identify any obvious groups based on all SNPs. Multilocus Wright's F-statistics were then estimated, with uncertainty assessed through bootstrapping across SNPs (n = 1000 replicates with replacement): F_{ST} as a measure of genetic structure and F_{IS} as a measure of fixation using *hierfstat*. The region and population-level organization contributions to F_{ST} were also estimated to assess impact of spatial organization on structure. F_{IS} by population and pairwise F_{ST} comparisons between populations were estimated using mean expected and observed heterozygosity calculated from SNP data to assess structure, possibly due to inbreeding, within populations and pairwise differences between populations. STRUCTURE v 2.3.4 was used to search for genetic patterns among populations, given 2 to 19 groups (Porras-Hurtado et al., 2013). Finally, private alleles, alleles found in only one population, were identified to assess genetic isolation.

F_{ST} values were also estimated for all bins using *hierfstat*, and analyzed for outliers using the 95% confidence interval to identify bins with significantly high or low differentiation. Low values that were below zero were ignored, as they are artifactual.

SNP Binning

To evaluate whether heterogeneity varies across the genome, sliding windows were created. Using only the SNPs that matched the linkage map, groups were selected based on position along the genome, as defined in the previously created linkage map (Friedline et al., 2015). Bins of minimum 40 SNPs were created, with positions that had more SNPs than the sample size randomly sampled to the correct size and positions that had fewer identified SNPs combined with

neighboring positions to increase total sample size. Analysis was repeated for groups of 30 and 50 SNPs to assess whether results are sensitive to sample size. Binned datasets were compared using pairwise t-tests for mean and F tests for variance of F_{ST} by linkage group with Bonferroni corrections at a 95% confidence level (Dunn, 1961).

Population Graphs

Connectivity was evaluated using population graphs to assess whether connectivity patterns are consistent with a demographic model. For all SNPs and then for each sliding window, population were made using the *popgraph* package R 1.5.1 (Dyer & Nason, 2004) to evaluate the spatial structure of genetic patterns using conditional genetic distance (cGD). A genome-wide population graph was created using all identified SNPs from the assembly to evaluate the overall population connectivity structure. A series of subset graphs were also made for each bin to evaluate how the structure of population connectivity varies across the genome and compared to a population graph generated from all the SNPs that match to the linkage map. Graphs were compared using distance congruence tests, which utilize a correlation test, to evaluate how connectivity patterns vary across the genome.

H2: High levels of inbreeding, due to bottlenecks and reduced pollen and seed movement from high competitor density, influence differentiation among populations.

Structural Equation Modeling

A structural equation model (SEM) was built using the *lavaan* package in R v 0.6-3 (Rosseel, 2012) to evaluate and quantify the hypothesized impact of competitor density on inbreeding, both indirectly through effects on foxtail importance and directly by limiting pollen and seed movement. Variables that were normal or with variances 3 orders of magnitude or more greater than other variables were transformed as needed using natural log, square root, or unit conversion. Foxtail population density, competitor population densities (red fir and western white pine), conifer diversity, and substrate heterogeneity data were used from Eckert and Sawyer (2002) and Eckert (2006). In total, there were 15 populations to build the model that had both genetic and environmental data available. Models were evaluated used multiple fit indices, including Chi-squared, RMSEA, CFI, and SRMR (Hu & Bentler, 1999).

H3: Morphological and chemical traits are correlated to genetic variation and are the result of natural selection pressures from local environment.

Redundancy Analysis

The effect of genetics, geography, environment on phenotypic traits were assessed with Redundancy Analysis (RDA) using the *vegan* package in R v 2.5-3 to

determine if multivariate trait variation is correlated with genetic variation, as hypothesized. A whole-genome RDA was made using all SNPs, top environmental variables that determine foxtail's niche, and geographic coordinate data as the predictor matrices and phenotypic trait data as the response matrix. Genetic data were transformed using a Hellinger transformation (Rao, 1995), and trait data, environmental data, and geographic data were centered and scaled. The predictive abilities of the predictor matrices were assessed using variance partitioning, and pure effects of environment and genetics were identified using conditioned models. The predictors for the global RDA model were selected using stepwise model selection with a significance level of 0.05 for adding and removing predictor variables with MASS v 7.3-51.1 (Venables & Ripley, 2002) and evaluated based on fit, variance inflation factors, and statistical significance at an alpha of 0.05. The goal was to identify SNPs associated with environmental adaptation and observed phenotypic traits. Geographic data, latitude and longitude, were included to remove spatial effects.

Then, RDA models were built for each bin and compared to the whole-genome RDA. To avoid overfitting, SNP data were converted to PC axes and only the top axes that jointly explained 80% of the variation in the data were retained for the model. The environmental axes selected for the global model were also included. Bin RDA plots were evaluated with variance partitioning to identify bins with outlier total effect, pure genetic effects, and pure environmental effects.

RESULTS

Data Collection and Preparation

MaxEnt

MaxEnt estimated the niche model shown below, and had an AUC value of 0.946, which indicates the model has strong predictive abilities (Figure 2) (Swets, 1988). The model predicted suitable habitat outside of the natural range of northern foxtail pine, which is habitat with good environmental conditions, but is outside the historic range of foxtail, so it has not established there. Days below zero degrees Celsius, mean annual precipitation, and precipitation as snow had the highest influence on the model. Six variables accounted for over 90% of the estimated niche, as determined by MaxEnt, and will be focused upon in later analysis (Table 1).

Figure 2: MaxEnt niche estimate, red colors represent most suitable habitat

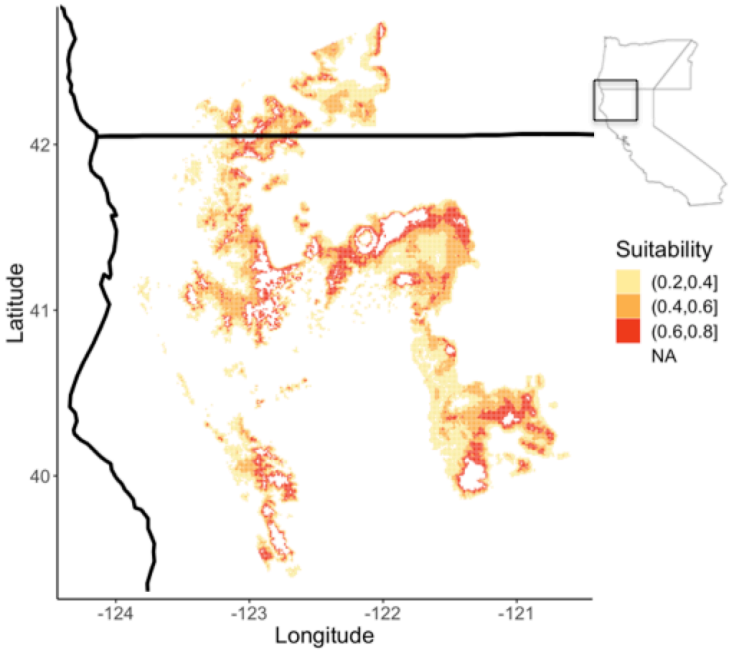


Table 1: Variables that determine over 90% of the northern foxtail's climate niche, as determined by MaxEnt

Variable	Percent Contribution
Degree days below zero	32.9
Mean annual precipitation	25.5
Precipitation as snow	14
Hargreaves climatic moisture deficit	11.9
Summer heat moisture index	4.5
Temperature Difference	4

Genetic Dataset Selection

After all quality filtering, Trim 60 had 92,559 SNPs, Trim 70 had 83,628, Trim 80 had 63,508, and Trim 90 had 47,700. When compared pairwise using Procrustes analysis, the PCA plots for all trimming levels were highly related (Pearson's $r >$ 0.998), indicating trimming levels did not affect interpretations of genetic patterns. The results of blast were as follows: Trim 60 had a total of 3,263 SNPs that could be mapped to the linkage map, Trim 70 had 5,709, and Trim 80 had 5,720. Trim 90 was not matched to the linkage map based on its low over all SNPs and minimal increase in linkage map hits between Trim70 and 80. Trim 70 dataset was ultimately selected to continue in analysis as the primary genetic dataset, due to its high initial SNP identification during assembly and number SNPs able to be located on the existing linkage map.

H1: Populations are highly differentiated genome-wide, with heterogeneity varying across the genome, and have patterns of connectivity resulting from demographic processes.

Genetic Structure

No obvious genetic structure among populations was evident in the PCA, as summarized using the top two PCs, which jointly explain 1.57% of the variance (Figure 3). The global, multilocus F_{ST} was estimated as 0.000644 with a 95% confidence interval of [0.000496, 0.000785]. An insignificant amount of the value of F_{ST} is captured by F_{CT}, which is regional organization, (6.94% with a 95% confidence interval of [-12.1%, 25.5%]). Most of the value of F_{ST} is captured by F_{SC}, population-level organization, (93.1% with a 95% confidence interval of [74.5%, 112%]). The F_{IS} values by population were variable, ranging from -0.03 to 0.21, with an average of 0.07 (Table 2). The pairwise F_{ST} comparisons between populations were low overall and ranged from -0.0014 (between populations 5 and 22) and 0.0078 (between populations 20 and 25), with an average of 0.00056 (Appendix F). Pairwise F_{ST} is significantly correlated with the geographic distance (Pearson's r = 0.163, p value = 0.0335). Structure analysis showed no clear genetic groupings of populations given a number of clusters (K) ranging from 2 to 19 (Figure 4). Only eight private alleles were identified (Appendix G).

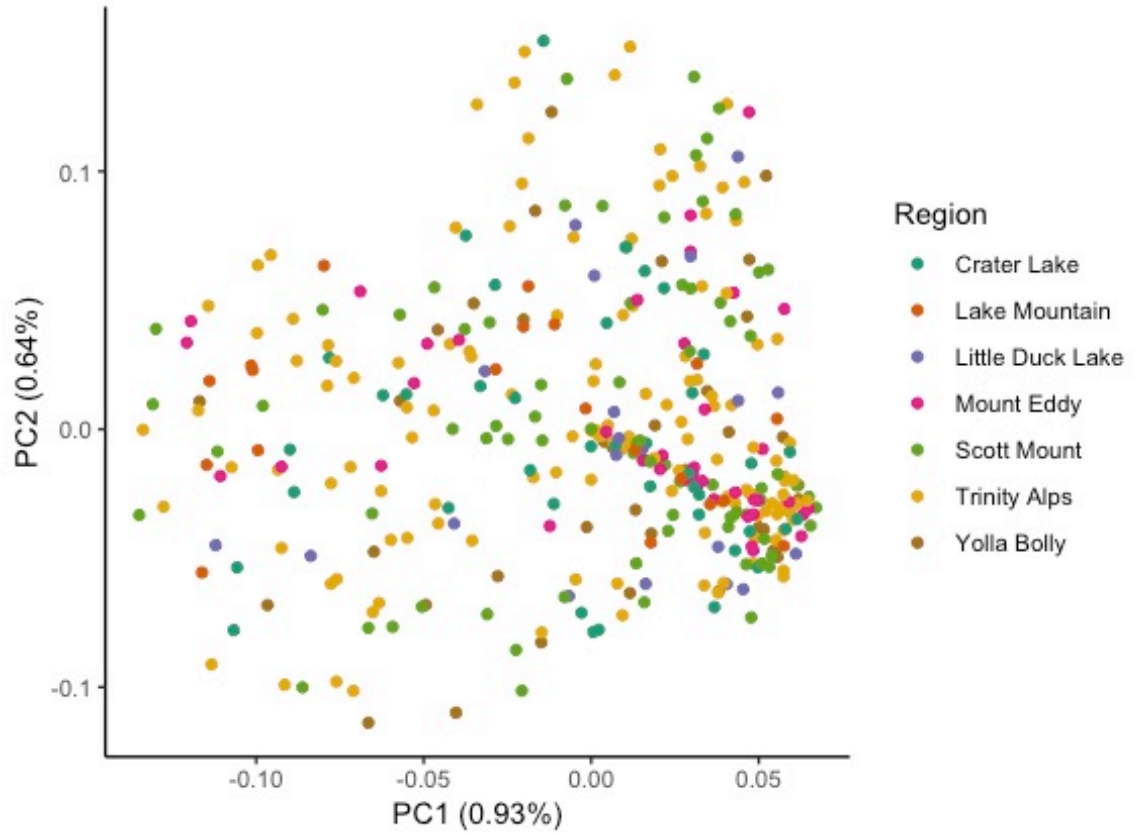
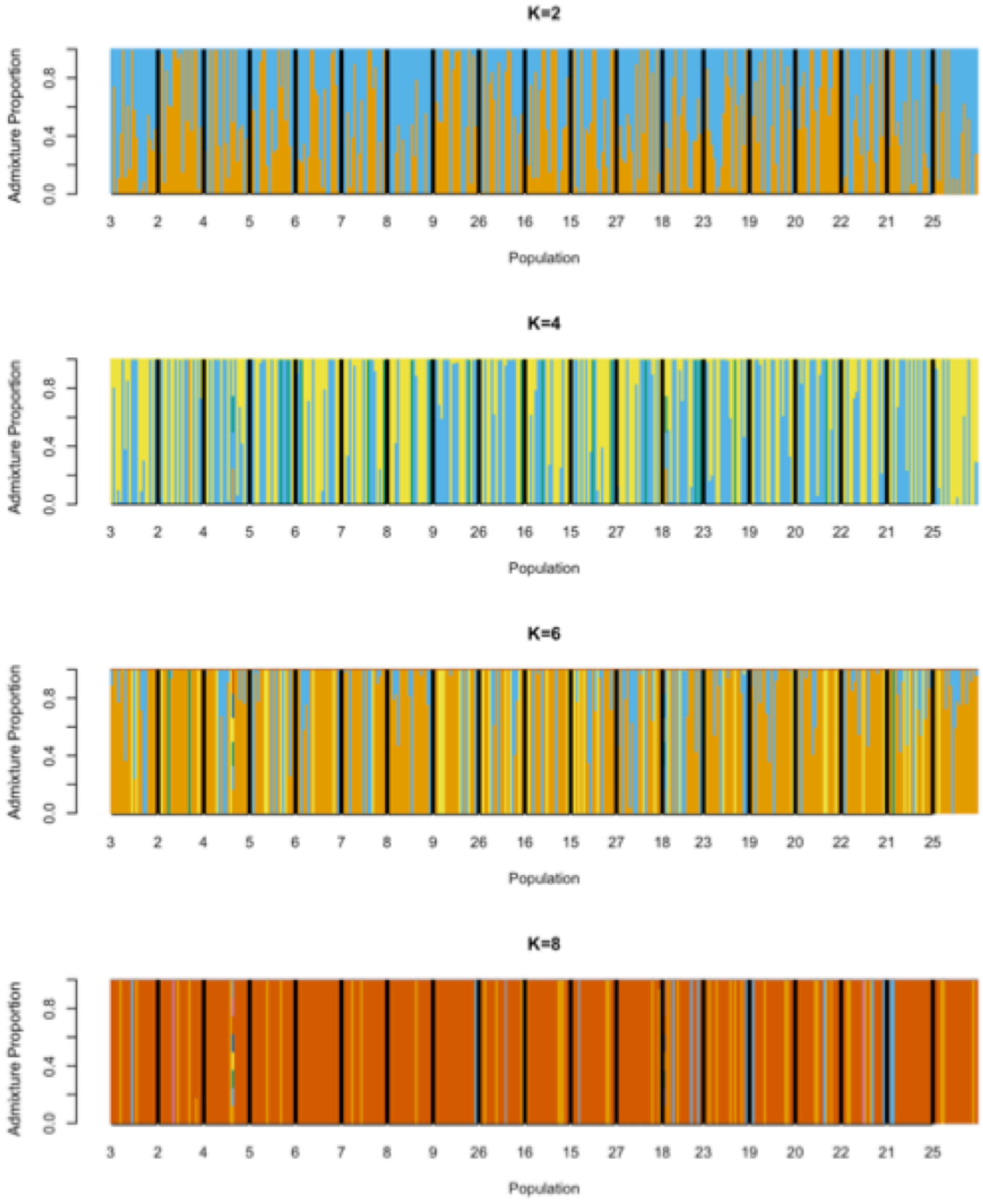


Figure 3: PCA of genetic patterns using all SNPs for the Trim 70 dataset, populations are colored by region.

Table 2: F_{IS} by population

Population	Region	F_{IS}
2	Yolla Bolly	0.187
3	Yolla Bolly	0.041
4	Trinity Alps	0.040
5	Trinity Alps	0.043
6	Trinity Alps	0.078
7	Trinity Alps	0.014
8	Trinity Alps	-0.022
9	Trinity Alps	0.141
15	Scott Mountain	0.025
16	Scott Mountain	0.142
18	Scott Mountain	0.089
19	Little Duck Lake	0.051
20	Little Duck Lake	0.214
21	Crater Lake	0.072
22	Crater Lake	0.066
23	Mount Eddy	0.098
25	Lake Mountain	-0.032
26	Trinity Alps	0.062
27	Scott Mountain	0.071

Figure 4: Select structure analysis results for K = 2, 4, 6, & 8 groupings



SNP Binning

The binned dataset at minimum sample size of 40 SNPs had a total of 90 bins distributed across the 12 linkage groups (Table 3). The mean and variance of F_{ST} for each linkage group was compared to datasets binned for minimum 30 and 50 SNPs using a Student's t test and a two-sided F-test with a Bonferroni correction. None of the comparisons were significant, so the datasets were assumed to capture similar genetic patterns and variation (Table 4). The dataset binned to 40 SNPs was used for all following analysis (Appendix H). At a 95% confidence level, three bins were identified as outliers for F_{ST} (Figure 5, Table 5).

Table 3: Number of bins by linkage group

Linkage Group	Total SNPs	Binned groups
1	474	8
2	588	9
3	420	8
4	204	4
5	499	7
6	837	14
7	497	5
8	478	7
9	561	8
10	287	6
11	297	5
12	567	9
Total	**5,709**	**90**

Table 4: F_{ST} pattern comparison p values between the binning groups (30, 40, and 50)

Linkage Group	Bin 30 / Bin 40 Comparisons		Bin 40 / Bin 50 Comparisons	
	t test for mean p value	Variance test p value	t test for mean p value	Variance test p value
1	0.7470	0.6516	0.8677	0.4707
2	0.6558	0.3473	0.6193	0.3479
3	0.4984	0.6132	0.7569	0.9879
4	0.7146	0.8161	0.9657	0.7691
5	0.5212	0.6828	0.8754	0.2919
6	0.9743	0.6920	0.8453	0.7521
7	0.5269	0.5914	0.9868	0.8990
8	0.5844	0.3705	0.3264	0.6980
9	0.9889	0.6953	0.4548	0.9842
10	0.9482	0.6765	0.7615	0.7007
11	0.4232	0.6001	0.9325	0.8372
12	0.7587	0.8400	0.9201	0.7215

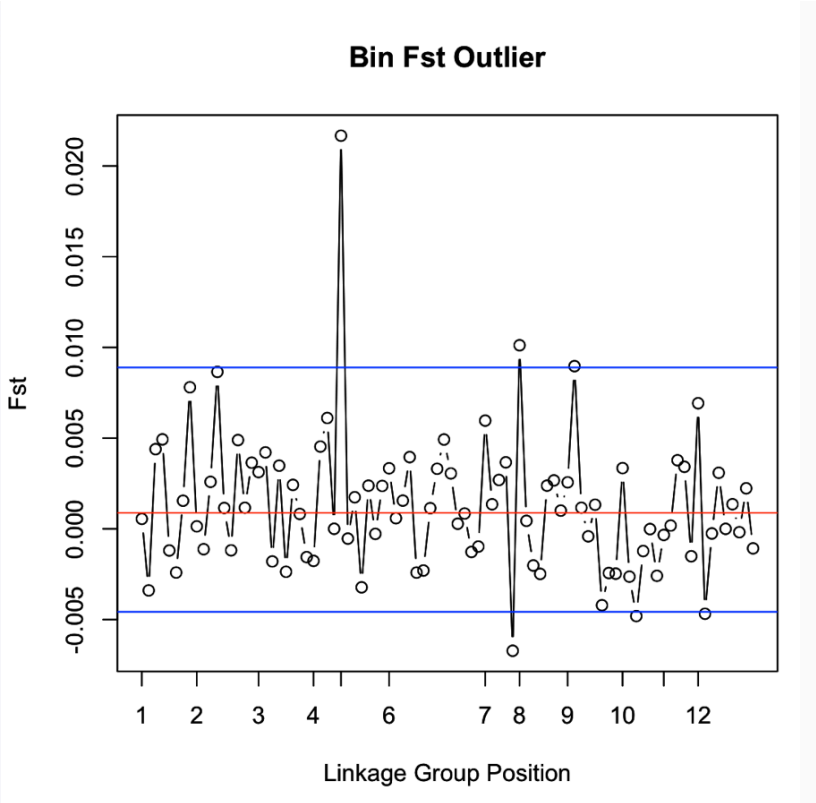


Figure 5: F_{ST} values of all bins plotted with the 95% confidence interval, red line represents the 50% quantile, blue lines represent the 2.5% and 97.5% quantiles.

Table 5: Bins with outlier F_{ST} values

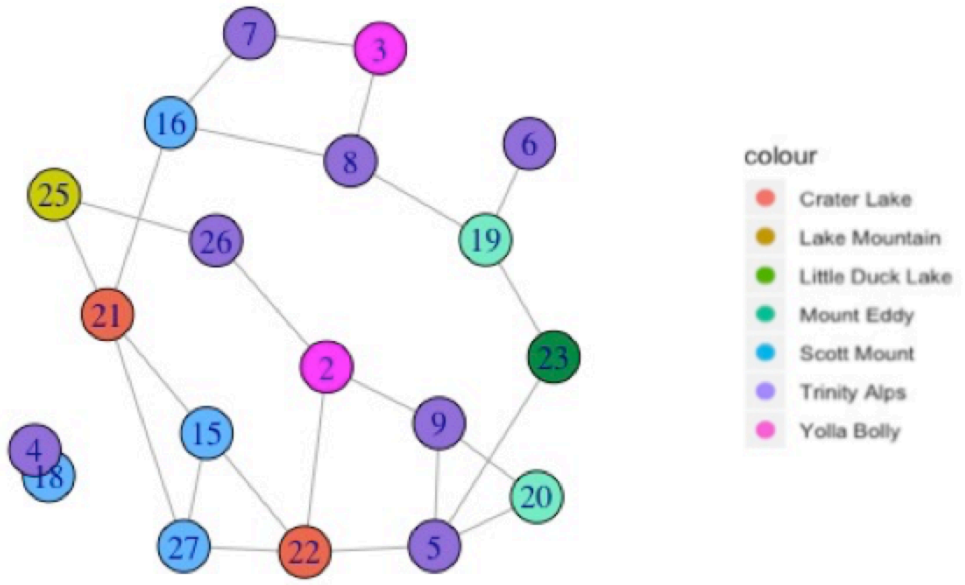
Bin ID (Linkage Group _ Number)	F_{ST}
5_1	0.0217
8_1	0.0101
9_2	0.0090

Population Graph Analysis

The population graph generated based on all SNPs has 24 statistically significant edges between the 19 populations (Figure 6). The global population graph had an average closeness of 0.0214, meaning there was low centrality in the graph. The graph also had an average degree of 2.52, with a maximum degree of 4. Populations with the highest degree belong to the Crater Lake, which are near the in the middle of northern foxtail pine's range (Figure 1). The diameter of the graph was 26.07. The graph has two outlier populations, 4 and 18, which were not connected to other populations in the network, and probably inflate the graph diameter significantly.

The population graphs for the binned data were compared to a global population graph that used all SNPs that matched to the linkage map using a distance congruence test of their conditional genetic distances, with an alternative hypothesis that there is a non-zero correlation between the global graph and the subset bin graph. Of the 90 bins, no bins were significantly correlated with the global graph at the 95% confidence level after a Bonferroni correction (Dunn, 1961), meaning all binned graphs have different connectivity patterns than the global graph.

Figure 6: Population graph using all SNPs



H2: High levels of inbreeding, due to bottlenecks and reduced pollen and seed movement from high competitor density, influence differentiation among populations.

Structural Equation Modeling

Multiple models were created with increasing complexity, but only two models passed all fit indicator cut offs (X^2 p-value > 0.05, RMSEA < 0.1, SRMR < 0.08, CFI > 0.95) (Table 6) (Hu & Bentler, 1999). Model A represents the full hypothesis, including both direct and indirect effects of competitors on F_{IS}, and Model B is the same model without the direct effects of competitor presence on F_{IS}. No models that included environmental variables passed all fit indicators, possibly because the models were too complex for the sample size of the data. Ultimately, Model B was

selected for the final model, which can predict 28.8% of the variation in F_{IS} based on tree density and relative importance of *Pinus balfouriana* (PIBA), both of which decrease F_{IS} (Figure 7). As described in Model B, tree density, PIMO, and ABMA all decrease PIBA (β = -0.67, -0.43, and -0.40, respectively), and tree density and PIBA decrease F_{IS} (β = -0.64 and -0.48). There were also indirect positive effects of tree density, PIMO, and ABMA on F_{IS}, which can be calculated by multiplying the values along the indirect paths, (β = 0.32, 0.21, and 0.18, respectively). Model A, however, had very similar fit scores to Model B, so the direct relationships between competitors and F_{IS} cannot be completely dismissed.

Table 6: SEM models that passed all fit indicator cut offs

Model name	equation	AIC	X^2 p value	RMSEA	SRMR	CFI	R^2 for FIS
Model A	F_{IS} ~ tree density + PIBA + ABMA + PIMO PIBA ~ PIMO + density + ABMA	-51.390	0.478	0.000	0.014	1.000	0.288
Model B	**F_{IS} ~ tree density + PIBA** **PIBA ~ PIMO + density + ABMA**	**-53.45**	**0.379**	**0.000**	**0.055**	**1.000**	**0.190**

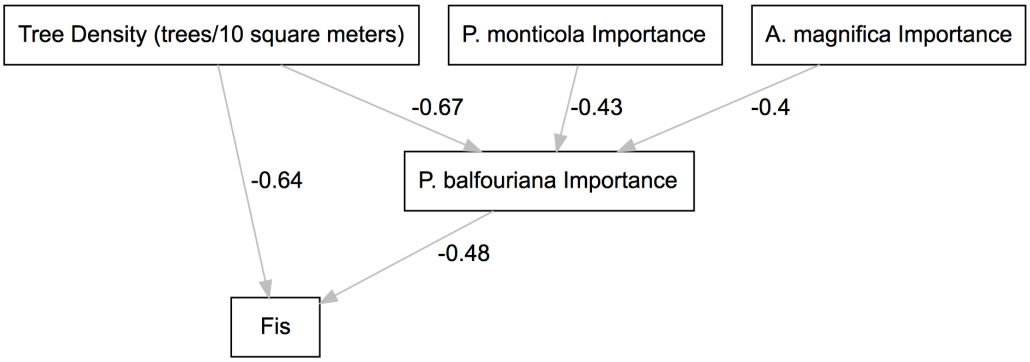


Figure 7: SEM plot for Model B, coefficients are standardized model coefficients

H3: Morphological and chemical traits are correlated to genetic variation and are the result of natural selection pressures from local environment.

Redundancy Analysis

A global, multilocus RDA model was selected using stepwise model selection (Figure 8). To reduce multicollinearity and variance inflation factors, the genetic and environmental data were converted to PC axes. Geographic data greatly increased multicollinearity in the model and reduced predictive ability, so they were removed from the model building process. The final selected model was as follows:

Phenotypic traits ~ GPC 10 + GPC 13 + GPC 16 + EPC 1 + EPC 4

In the model, the GPC 10, GPC 13, and GPC 16 are PC axes that represent the influence of all genetic data (83,628 SNPs) . EPC 1 and EPC 4 are the environmental PC axes (Table 7), and they represent primarily cold temperatures and seasonality, respectively. The model was significant (p-value = 0.001) and the total effect (R^2) is 45.9% (adjusted R^2 = 25.2%). Based on variance partitioning, the genetic axes

independently explained 24.4% of the overall variation (adjusted R^2 =13.2%), while environment explained 15.6% of the variation (adjusted R^2 = 8.7%). The confounded effect of these variables, therefore, was 5.9% (3.3% using adjusted R^2). This represents the portion of the overall effect of these variables that cannot be attributed to either of their independent (i.e., pure) effects. Given these effects, genetic data accounted for 53.2% of the explainable variance in phenotypic traits, while environmental data accounted for 34.0%. The remainder of the explainable variance (12.8%) was due to the confounding between genetic and environmental data. Of the RDA bin models, three models were outliers for both total effect and pure genetic effect, meaning they explained significantly more variation in trait data, (Figure 9 & 10, Table 10), and three models were outliers for pure environmental effect (Figure 11, Table 8). There was no overlap with the outlier bins for F_{ST}.

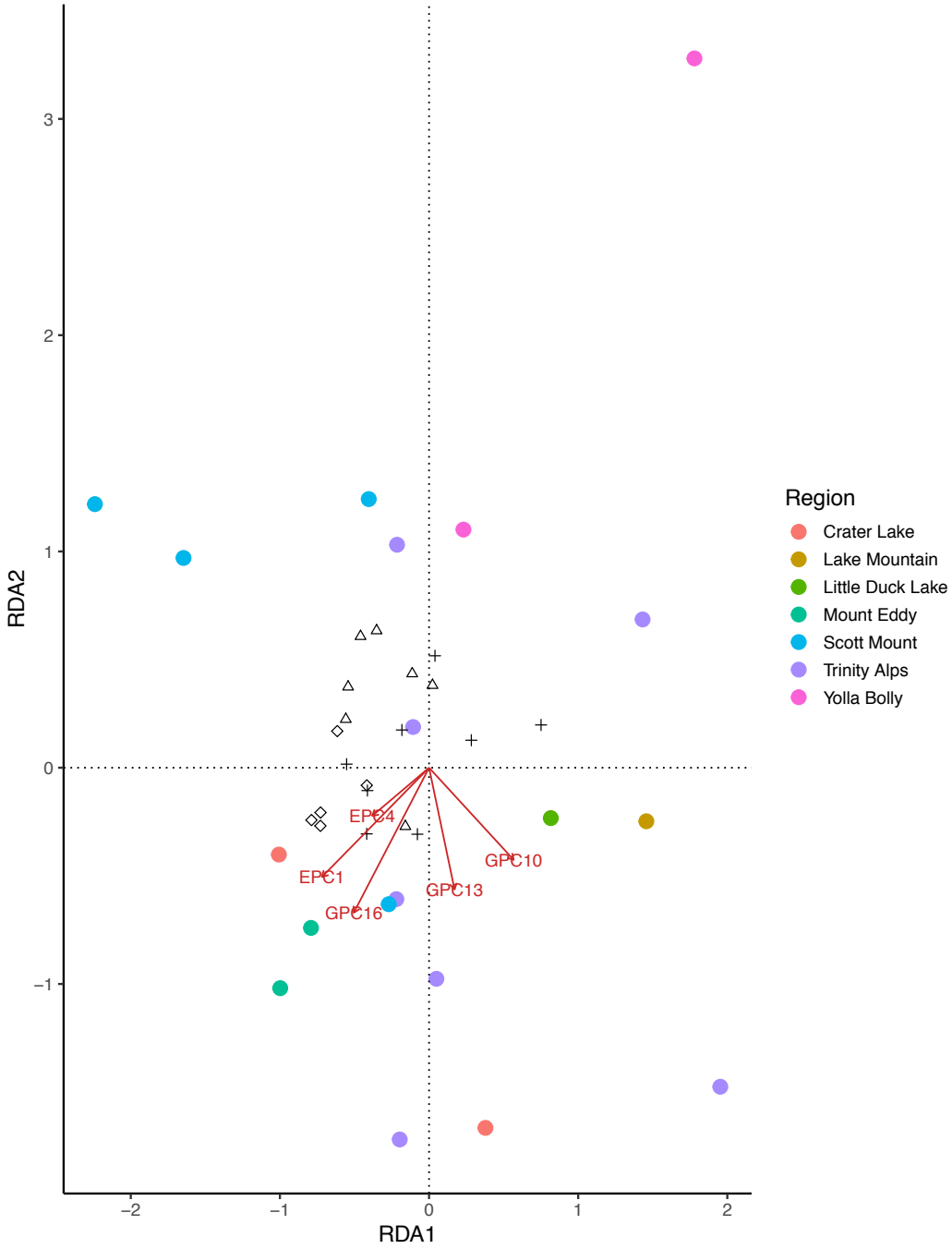


Figure 8: Biplot of selected RDA model, triangles represent chemical traits, pluses represent needle traits, and diamonds represent cone traits

Table 7: PC axes loadings for selected environmental variables

Variable	Environmental PC axis 1	Environmental PC axis 4
MCMT	-0.4339	0.0300
TD	0.1527	-0.9161
MAP	-0.1693	-0.1575
MSP	0.2583	-0.0616
AHM	0.0274	0.1540
SHM	-0.3460	-0.0663
bFFP	0.4120	0.1650
eFFP	-0.3819	-0.2555
PAS	0.3498	-0.0899
CMD	-0.3899	0.0505

Figure 9: Plot of total effect (R^2) for each bin RDA model, the red line represents the 50% quantile, while the blue lines represent the 2.5% and 97.5% quantiles

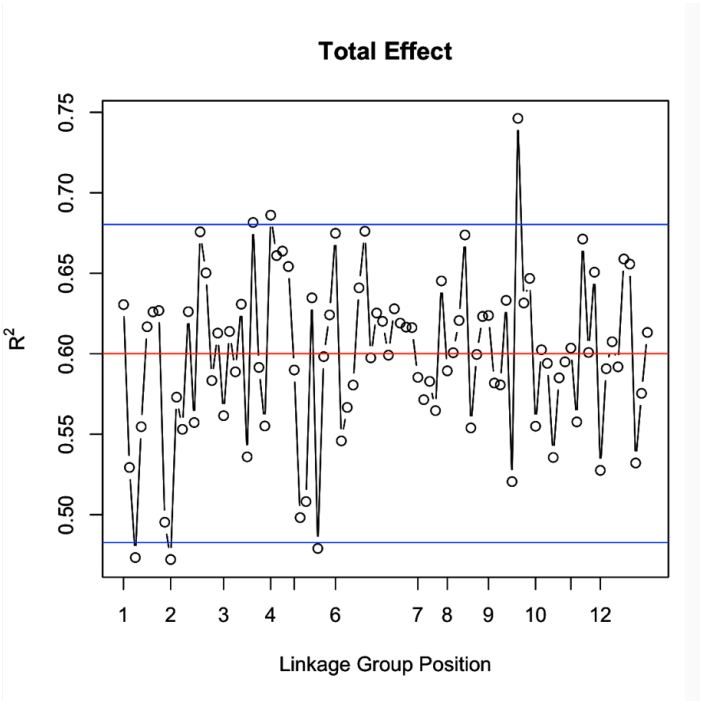


Figure 10: Plot of pure genetic effect (R^2) for each bin RDA model, the red line represents the 50% quantile, while the blue lines represent the 2.5% and 97.5% quantiles

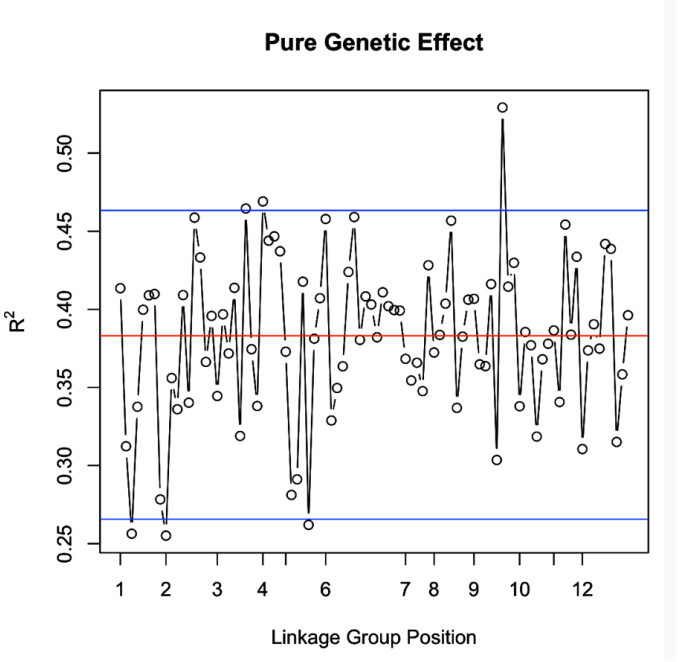


Figure 11: Plot of pure environmental effect (R^2) for each bin RDA model, the red line represents the 50% quantile, while the blue lines represent the 2.5% and 97.5% quantiles

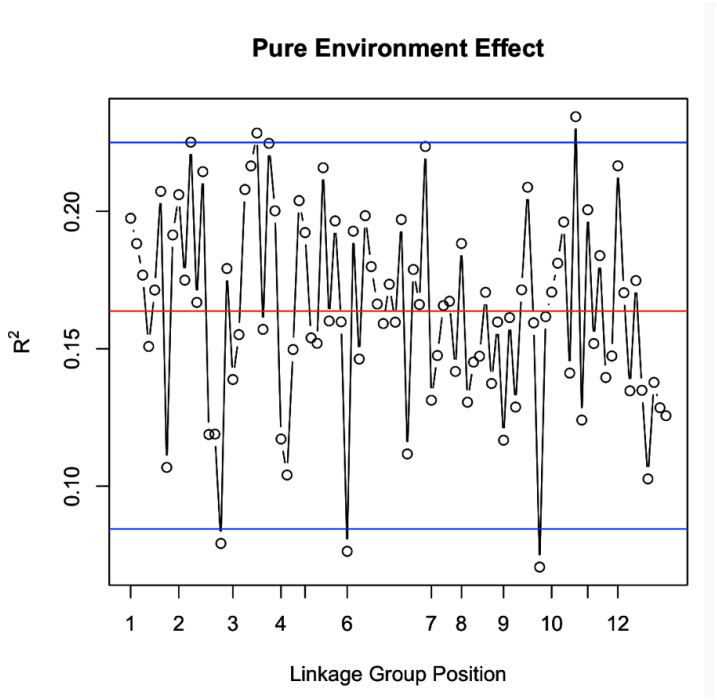


Table 8: Bins that are outliers for partitioned genetic R^2, * indicates value is an outlier

Bin ID	Total Effect R^2	Pure Genetic R^2	Pure Environment R^2
2_11	0.5530	0.3360	0.2252*
3_12	0.5359	0.3189	0.2285*
3_17	0.6816*	0.4645*	0.1570
4_1	0.6861*	0.4691*	0.1172
9_18	0.7462*	0.5292*	0.1594
10_16	0.5850	0.3680	0.2344*

DISCUSSION

The results of this study reveal less population structure in foxtail pine than expected in the first hypothesis based on previous studies, estimating F_{ST} to be 0.000644, compared to F_{ST} = 0.242 (Oline et al., 2000) and F_{ST} = 0.080 (Eckert et al., 2010). Previous studies based predictions on limited genetic data from functional regions, three allozymes and five genic loci respectively, which can overestimate structure due to non-neutrality of the genetic markers. Although differentiation is low, the population structure follows only a weak pattern due to demographic processes as hypothesized, such as isolation-by-distance model, and there was no significant regional structure, as shown by the hierarchical F_{ST} analysis, PCA (Figure 1), Pairwise F_{ST} (Appendix F), and STRUCTURE analysis (Figure 4). If gene flow is ongoing and spatially restricted, one would expect stronger regional structure due to populations that are geographically closer breeding more often than with geographically isolated populations, so patterns of gene flow of this sort is improbable (Slatkin, 1993). Rather, the currently observed similarity is likely due to diversity retained from initial colonization and fragmentation upon the peaks of the Klamath Mountains. The extremely long lifespan of foxtail pine may have drastically

slowed down the loss of genetic diversity among populations, despite lack of active mating among populations in fragmented habitats (Lowe et al., 2005; Petit & Hampe, 2006). The relatively low rate of private alleles identified also supports this conclusion, as populations that have been diverged for many generations would be expected to accumulate private alleles (Appendix G) (Zuckerkandl & Pauling, 1965). Despite differentiation defying expectations of structure, the fragmented populations and evidence for largely independent evolution within stands confirms previous descriptions (Oline et al., 2000; Zacher 2015).

The overall lack of differentiation among populations is consistent with high levels of genetic connectivity, despite the lack of evidence for continued gene flow, as illustrated by the population graph (Figure 6). These connectivity patterns, therefore, likely result from shared variation dating to the original colonization of the Klamath Mountains and fragmentation of foxtail pine across the isolated peaks in this region. Consistent with the lack of spatial genetic structure, the connectivity pattern does not seem to be strongly associated with geographic distance, as regions in the center of the species' range are not more central to the graph, with the exception of Crater Lake populations, which have an average degree of 4 as compared to 2.5 or less for all other regions (Figure 6). The two outlier populations, 4 and 18, are also striking in the graph. Although, these populations do not cluster in any of the genetic structure analyses and their pairwise F_{ST} is not an outlier (-0.0005), therefore their isolation in the network is likely due to equal connections to all other populations that were pruned because they did not contribute significantly to explain connectivity in the network. Still, these populations likely inflate the diameter of the graph.

Although most of my first hypothesis concerning overall patterns was not supported by the data, expectations of variation across the genome were supported by F_{ST} and connectivity patterns. Of the 90 total bins, three bins had outlier F_{ST} values, ranging from 0.00953 to 0.00798, far from the population estimate of 0.000644 (Figure 6, Table 5). F_{ST} outliers can be interpreted as selection pressure acting on specific regions of the genome and may be important for adaptive trait variation (Eveno et al. 2008; Namroud et al. 2008; Eckert et al., 2010). The regions that were outliers for F_{ST} were not outliers for RDA analysis though, so they are not likely associated with the measured traits for this study. The outlier F_{ST} bins also did not capture any of the quantitative trait loci identified on the linkage map for foxtail for water use efficiency (Eckert et al., 2016). The population graph analysis of connectivity was even more variable. All 90 bins were uncorrelated with the global graph, meaning connectivity patterns vary greatly throughout the genome. This result is likely due to the same mechanisms driving varying F_{ST} values: varying effective population sizes, selection pressures, and linkage disequilibrium (Smith & Haigh, 1974; Charlesworth et al., 1993; Charlesworth, 2009; Gossman et al., 2011).

The structural equation modeling largely supported the relationships predicted in my second hypothesis, although the direct relationship between competitors and F_{IS} was not strongly supported. The model identified the connection between foxtail importance, tree density, and competitor importance and a connection between F_{IS}, tree density, and foxtail importance (Figure 7). These results show inbreeding within stands is higher with low foxtail importance and low density, and that competition indirectly affects inbreeding within stands by regulating foxtail pine importance. This confirms past research that showed

increased spatial autocorrelation in stands where the relative importance of foxtail was low (Eckert et al., 2010). Similar patterns of low population densities leading to increasing inbreeding have been documented in other conifers as well (Farris & Mitton, 1984; Morgante et al., 1991; Restoux et al., 2008). The relationship between competitor importance and F_{IS}, as represented in Model A, was not represented in the selected model, but the model measures were very close (AIC difference of 2.06), so there is weak support for direct competitor effects on F_{IS} (Eckert & Sawyer, 2002; Eckert 2006). The power of the SEM analysis in this study, however, was low due to small sample size (n = 15) (Wolf et al., 2013), so that these conclusions should be tempered by the need for further sampling. Models with more complexity, including direct competitor effects and environment, may perform better with a larger dataset.

As expected in my third hypothesis, this study showed evidence for a genetic basis for the measured traits. Although it defied my original expectations, the overall low genetic divergence among populations is congruent with patterns of phenotypic variation previously described for these same individuals, where overall diversity is high, but there is very little differentiation among populations (Zacher, 2015). The similarity of genetic and phenotypic variation patterns supports my expectation that there is a genetic basis for measured phenotypic traits. The redundancy analysis provides further evidence of a genetic basis of the phenotypic traits, and highlights the influence of environmental variables that are important in defining the subspecies' niche on trait plasticity (Figure 8). Based on the global model, needle traits are in line with vectors representing the environmental axes which are mostly associated with seasonality and precipitation, and genetic axis 16, indicating needle

trait variation is affected by genotype by environment plasticity, while cone and chemical traits are most associated with the vectors for genetic axes 10 and 13, indicating cone and chemical trait variation is most associated with genetic variation (Figure 8, Table 7). Further, outlier bins with large pure genetic effects are candidate regions for further investigation into the genetic architecture of these traits, outlier bins with high pure environmental effects are likely regions related to plasticity of the observed traits (i.e., GxE), and outlier bins for total effect may be important local adaptation (Figure 9, Table 8). One outlier bin for environment, bin 3_12, spanned a region with a quantitative trait locus SNP associated with water-use efficiency and foliar nitrogen (Eckert et al., 2016). Bins that were outliers for total effects were the same as outlier bins for pure genetic effects, because most of the explained variance is genetic overall. These bins are candidates for further research into local adaptation, as they provide evidence that measured traits have both genetic and environmental influences related to that section of the genome, which are necessary components for a trait to be adaptive. The bins that were outliers for F_{ST}, however, were not outliers for total, genetic, or environmental effects, likely because outlier F_{ST} regions should represent genes impacting the traits most important for fitness. This implies that while these traits have known links to fitness in other conifers (Wallis, 2007; Garcia et al., 2009; Iason et al., 2011; Mápula-larreta et al., 2011; Zacher, 2015), they may have smaller or negligible effects for fitness of foxtail pine in its current environment.

The connectivity and high genetic similarity among populations due to historical processes was surprising in this study, disproving most of my first hypothesis, but the overall weak spatial genetic structure due to long term

geographic isolation is consistent with past descriptions (Oline et al., 2000). Based on my results, it seems that inbreeding within populations, which can contribute to differentiation among populations, is influenced by foxtail importance, density, and indirect biotic stress from competitors. Finally, this study provides evidence that previously measured traits have a strong genetic basis with some level of environmentally induced plasticity (i.e., effects of environment and genotype-by-environment interactions) and are potentially adaptive, although unlikely to be major components of fitness. Despite high levels of genetic diversity and connectivity, the fragmented population structure without easily detectable levels of current gene flow and poor competitive ability of foxtail pine may impede adaptation to changing climate. Over time, due to changing climate and increased biotic stress due to other species migrating to high altitudes, foxtail may thus be vulnerable to extinction in the Klamath Mountains.

LITERATURE CITED

Achere, V., Favre, M., Besnard, G., Jeandroz, S. (2005). Genomic Organization of Molecular Differentiation in Norway Spruce (*Picea abies*). *Molecular Ecology*, *14*, 3191–3201.

Aitken, S. N., Yeaman, S., Holliday, J. A., Wang, T., Curtis-McLane, S. (2008). Adaptation, Migration or Extirpation: Climate Change Outcomes For Tree Populations. *Evolutionary Applications*, *1*, 95–111.

Alberto, F. J., Aitken, S. N., Alía, R., González-Martínez, S. C., Hänninen, H., Kremer, A., … Savolainen, O. (2013). Potential For Evolutionary Responses to Climate Change - Evidence From Tree Populations. *Global Change Biology*, *19*, 1645–1661.

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., Lipman, D. J. (1997), Gapped BLAST and PSI-BLAST: a New Generation of Protein Database Search Programs, *Nucleic Acids Res.* 25, 3389-3402

Antonovics, J., Levin, D. A. (1980). The Ecological and Genetic Consequences of Density-Dependent Regulation in Plants. *Annual Review of Ecology, Evolution, and Systematics,* 11, 411-52.

Bohrerova, Z., Bohrer, M., Cho, K.D., Bolch, M.A., Linden, K.G. (2009). Determining the Viability Response of Pine Pollen to Atmospheric Conditions During Long-Distance Dispersal. *Ecological Applications*, 19, 656–667.

Carmona D., Lajeunesse M. J., Johnson M. T. (2011) Plant Traits That Predict Resistance to Herbivores. *Functional Ecology,* 25, 358-367.

Cavalli-Sforza, L. L. (1966). Population Structure and Human Evolution. *Proceedings of the Royal Society B*, 164, 362-79.

Charlesworth, D., & Charlesworth, B. (1987). Inbreeding Depression and its Evolutionary Consequences. *Annual Review of Ecology, Evolution, and Systematics,* 18.

Charlesworth, B., Morgan, M. T., Charlesworth, D. (1993). The Effect of Deleterious Mutations on Neutral Molecular Variation. *Genetics* 134, 1289–1303.

Charlesworth ,B. (2009). Fundamental Concepts in Genetics: Effective Population Size and Patterns of Molecular Evolution and Variation. *Nature Reviews Genetics*, 10, 195–205.

Chen I. C., Hill J. K., Ohlemüller, R., Roy D. B., Thomas C. D. (2011). Rapid Range Shifts of Species Associated With High Levels of Climate Warming. *Science,* 333, 1034-1026.

Chen S., Zhou Y., Chen Y., Gu J. (2018). Fastp: An Ultra-Fast All-In-One FASTQ Preprocessor, *Bioinformatics*, 34, 884-890.

Cheng, Sheauchi. tech. ed. 2004. Forest Service Research Natural Areas in California. Gen. Tech. Rep. PSW-GTR-188. Albany, CA: Pacific Southwest Research Station, Forest Service, U.S. Department of Agriculture, 338 p.

Coffey, K., Benkman, C. W., Milligan, B. G. (1999). The Adaptive Significant of Spines on Pine Cones. *Ecology*, 80, 1221-1229.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R., Lunter, G., Marth, G., Sherry, S. T., McVean, G., Durbin R., and 1000 Genomes Project Analysis Group (2011) The Variant Call Format and VCFtools. *Bioinformatics*.

Dawson T. P., Jackson S. T., House J. I., Prentice I. C., Mace G. M. (2011). Beyond Predictions: Biodiversity Conservation In A Changing Climate. *Science*, 332, 53-58.

Dunn, O. J. (1961). Multiple Comparisons Among Means. *Journal of the American Statistical Association*, 56, 52–64

Dyer, R. J., & Nason, J. D. (2004). Population Graphs: The Graph Theoretic Shape of Genetic Structure. *Molecular Ecology*, *13*, 1713–1727.

Dyer, R. J., & Sork, V. L. (2001). Pollen Pool Heterogeneity in Shortleaf Pine, *Pinus echinata* Mill. *Molecular Ecology*, *10*, 859–866.

Eckert, A. J. (2006). Influence of Substrate Type and Microsite Availability on the Persistence of Foxtail Pine (*Pinus balfouriana, Pinaceae*) in the Klamath Mountains, California. *American Journal of Botany*, *93*, 1615–24.

Eckert, A. J., Eckert, M. L., Hall, B. D. (2010). Effects of Historical Demography and Ecological Context on Spatial Patterns of Genetic Diversity Within Foxtail Pine (*Pinus Balfouriana* ; Pinaceae) Stands Located in the Klamath Mountains, California. *American Journal of Botany*, *97*, 650–659.

Eckert, A. J., Harwood, D. E., Lind, B. M., Hobson, E. M., Delfino Mix, A., Maloney, P. E., Friedline, C. J. (2016) The Genetic Architecture of Local Adaptation II: The QTL Landscape of Water-Use Efficiency for Foxtail Pine (*Pinus balfouriana* Grev. & Balf.). Posted to bioRxiv with doi: 10.1101/038240.

Eckert, A. J., van Heerwaarden, J., Wegrzyn, J. L. , Nelson, C. D. , Ross-Ibarra, J., González-Martínez, S. C., Neale, D. B. (2010) Patterns of Population Structure

and Environmental Associations to Aridity Across the Range of Loblolly Pine (*Pinus taeda* L., Pinaceae). *Genetics* ,185, 969-982.

Eckert, A. J., Sawyer, J. O., & Eckert1, A. J. (2002). Foxtail Pine Importance And Conifer Diversity In The Klamath Mountains And Southern Sierra Nevada, California. *Source: Madroño Madroño*, *49*, 33–45.

Eckert, A. J., Tearse, B. R., & Hall, B. D. (2008) A Phylogeographical Analysis of the Range Disjunction for Foxtail Pine (*Pinus Balfouriana* , Pinaceae): The Role of Pleistocene Glaciation. *Molecular Ecology*, *17*, 1983–1997.

Eveno, E., Collada, C., Guevara, M. A., Leger, V., Soto V. A., et al., (2008) Contrasting Patterns of Selection at *Pinus pinaster* Ait. Drought Stress Candidate Genes as Revealed by Genetic Differentiation Analyses. *Molecular Biology and Evolution*, 25, 417–437.

Excoffier L., Smouse P. E., Quattro J. M. (1992) Analysis of Molecular Variance Inferred From Metric Distances Among DNA Haplotypes: Application to Human Mitochondrial DNA Restriction Data. *Genetics*, 131, 479–491.

Farris M. A., Mitton J. B. (1984) Population Density, Outcrossing Rate, and Heterozygote Superiority in Ponderosa Pine. E*volution*. 38, 1151–1154.

Friedline, C. J., Lind, B. M., Hobson, E. M., Harwood, D. E., Mix, A. D., Maloney, P. E., Eckert, A. J. (2015). The Genetic Architecture of Local Adaptation I: The Genomic Landscape of Foxtail Pine (*Pinus balfouriana* Grev. & Balf.) as Revealed From a High-Density Linkage Map. *Tree Genetics & Genomes,* 11.

Forister, M.L., McCall, A. C., Sanders, N. J., Fordyce, N. J., Thorne, J. H., O'Brien, J., Waetjen, D. P., Shapiro, A. M. (2010). Compounded Effects of Climate Change And Habitat Alteration Shift Patterns of Butterfly Diversity. *Proceedings of the National Academy of Sciences*, 107, 2088-2092.

Garcia, R., A.M. Siepielski, and C.W. Benkman. 2009. Cone and Seed Trait Variation in Whitebark Pine (*Pinus albicaulis; Pinaceae*) and the Potential for Phenotypic Selection. *American Journal of Botany,* 96, 1050-1054.

Gossman, T. I., Woolfit, M., Eyre-Walker, A., (2011). Quantifying the Variation in the Effective Population Size Within a Genome. *Genetics*, 189, *1389-1402*.

Goudet, J. (2005). Hierfstat, a Package for R to Compute and Test Hierarchical F-Statistics. *Molecular Ecology Notes*, 5, 184-186.

Hewitt, G. M., (2001) Speciation, Hybrid Zones and Phylogeography — or Seeing Genes in Space and Time. *Molecular Ecology*, 10, 537–550.

Holliday, J. A., Yuen, M., Ritland, K., & Aitken, S. N. (2010). Postglacial History of a Widespread Conifer Produces Inverse Clines in Selective Neutrality Tests. *Molecular Ecology*, *19*, 3857–3864.

Herten K., Hestand M. S., Vermeesch J. R., Houdt J. K. V. (2015) GBSX: a Toolkit for Experimental Design and Demultiplexing Genotyping by Sequencing Experiments. *BMC Bioinformatics*, 16, 73.

Hoffmann, A. A., Sgrò C. M. (2011). Climate Change and Evolutionary Adaptation. *Nature*, 470, 479-85.

Holt, R. D., Gomulkiewicz, R. (1997). How Does Immigration Influence Local Adaptation? A Reexamination of a Familiar Paradigm. *The American Naturalist*. The University of Chicago Press.

Hu, L., Bentler, P. M., (1999) Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria Versus New Alternatives, *Structural Equation Modeling: A Multidisciplinary Journal*, 6, 1-55.

Huber, D., Bohlmann, J. (2006). The Role of Terpene Synthases in the Direct and Indirect Defense of Conifers Against Insect Herbivory and Fungal Pathogens. I n: Tuzun S., Bent E. (eds) *Multigenic and Induced Systemic Resistance in Plants*. Springer, Boston, MA, 296-313.

Hultine, K. R., Marshall, J. D. (2000). Altitude Trends in Conifer Leaf Morphology and Stable Carbon Isotope Composition. *Oecologia*, *123*, 32–40.

Iason, G. R., O'Reilly-Wapstra, J. M., Brewer, M. J., Summers, R. W., Moore, B. D. (2011) Do Multiple Herbivores Maintain Chemical Diversity of Scots Pine Monoterpenes?. *Philos T R Soc B,* 366,1337–1345.

Latter, B. D. H. (1973). The Island Model of Population Differentiation: A General Solution. *Genetics*, *73*, 147–157.

Lapp, S., Byrne, J., Townshend, I., Kienzle, S. (2005). Climate Warming Impacts On Snowpack Accumulation In An Alpine Watershed. *International Journal Of Climatology Int. J. Climatol*, *25*, 521–536.

Le Corre, V., Kremer, A. (2003). Genetic Variability at Neutral Markers, Quantitative Trait Land Trait in a Subdivided Population Under Selection. *Genetics*, *164*, 1205–19.

Leitch, A. R., Leitch, I. J. (2012). Ecological and Genetic Factors Linked to Contrasting Genome Dynamics in Seed Plants. *Source: The New Phytologist*, *194*, 629–646.

Lind, B. M., Menon, M., Bolte, C. E., Faske, T. M. ,Eckert, A. J., 2018. The Genomics of Local Adaptation in Trees: Are We Out of the Woods Yet? *Tree Genetics & Genomes,* 14.

Lloyd, Andrea H.; Graumlich, Lisa J. 1997. Holocene Dynamics of Treeline Forests in the Sierra Nevada. *Ecology.* 78, 1199-1210.

Lowe A.J ., Harris S. A., Ashton P. (2004) *Ecological Genetics: Design, Analysis and Application*. Blackwell: Oxford, UK, 326.

Lowe A. J., Boshier D., Ward M., Bacles C. F. E., Navarro C. (2005). Genetic Resource Loss Following Habitat Fragmentation and Degradation; Reconciling Predicted Theory with Empirical Evidence. *Heredity,* 95, 255–273.

Mápula-larreta, M., López-Upton, J., Vargas-Hernández, J. J., Hernández-Livera, A. (2007). Reproductive Indicators in Natural Populations of Douglas-fir in Mexico. *Biodiversity and Conservation,* 16, 727-742

Mastrogiuseppe, R. J., Mastrogiuseppe, J. D. (1980). A Study of *Pinus balfouriana* Grev. & Balf. (Pinacae). S*yst. Bot*. 5, 86–104.

Mayr S., Schmid P., Beikircher B. (2012) Plant Water Relations in Alpine Winter. In: Lütz C (ed) *Plants in Alpine regions*. Springer, Vienna, 153–162.

Mittermeier, R. A., Robles-Gil, P., Hoffmann, M., Pilgrim, J. D., Brooks, T. B., Mittermeier, C. G., Lamoreux, J. L. & Fonseca, G. A. B. (2004). Hotspots Revisited: Earth's Biologically Richest and Most Endangered Ecoregions. *Cemex*, Mexico City, Mexico 390pp.

Morgante M., Vendramin G. G., Rossi P. (1991) Effects of Stand Density on Outcrossing Rate in Two Norway Spruce (*Picea abies*) Populations. *Can J Bot*, 69, 2704–2708.

Namroud, M. C., Beaulieu, J., Juge, N., Laroche, J., & Bousquet, J. (2008). Scanning the Genome for Gene Single Nucleotide Polymorphisms Involved in Adaptive Population Differentiation in White Spruce. *Molecular Ecology*, *17*, 3599–3613.

Nei M (1972) Genetic Distance Between Populations. *American Naturalist*, 106, 283–292.

Nei M. (1978) Estimation of Average Heterozygosity and Genetic Distance from a Small Number of Individuals. *Genetics*, 89, 583 –590.

Niklas, K., Kyaw Tha Paw, U. (1983). Conifer Ovulate Cone Morphology: Implications on Pollen Impaction Patterns. *American Journal of Botany, 70*, 568-577.

Oline, D. K., Mitton, J. B., & Grant, M. C. (2000). Population And Subspecific Genetic Differentiation In The Foxtail Pine (Pinus Balfouriana). *Evolution*, *54*, 1813–1819.

Parmesan C., Ryrholm N., Stefanescu C. et al. (1999) Poleward Shifts in Geographical Ranges of Butterfly Species Associated with Regional Warming. *Nature, 3*99, 579–583.

Patterson, N., Price, A. L., Reich, D., (2006). Population Structure and Eigenanalysis. *PLOS Genetics*, 2: e190.

Parchman, T. L., Z. Gompert, C. W. Benkman, F. D. Schilkey, J. Mudge, and C. A. Buerkle 2012. Genome Wide Association Mapping Of An Adaptive Trait In Lodgepole Pine. *Molecular Ecology,* 21, 2991-3005.

Pauli, H., Gottfried, M., Grabherr, G. (1996). Effects of Climate Change on Mountain Ecosystems -- Upward Shifting of Alpine Plants. *World Resource Review*, *8*.

Pepin, N., Bradley, R. S., Diaz, H. F., Baraer, M., Caceres, E. B., Forsythe, N., … Yang, D. Q. (2015). Elevation-Dependent Warming in Mountain Regions of the World. *Nature Publishing Group*, *5*.

Petit RJ, Hampe A. (2006). Some Evolutionary Consequences of Being a Tree. *Annu Rev Ecol Evol Syst* 37, 187-214.

Porras-Hurtado, L., Ruiz, Y., Santos, C., Phillips, C., Carracedo, A., Lareu, M. V. (2013). An Overview of STRUCTURE: Applications, Parameter Settings and Supporting Software. *Front Genet,* 4:98.

Puritz, J. B., Hollenbeck, C. M., Gold, J. R. (2014), dDocent: a RADseq, Variant-Calling Pipeline Designed for Population Genomics of Non-Model Organisms. *PeerJ* 2:e431.

Puritz, J. B., Matz, M. V., Toonen, R. J., Weber, J. N., Bolnick, D. I., Bird, C. E. (2014). Demystifying the RAD Fad. *Molecular Ecology* 23, 5937–5942.

Rao, C. R. (1995) A Review of Canonical Coordinates and an Alternative to Correspondence Analysis using Hellinger Cistance. *Qüestiió.* 19, 23–63

Restoux, G., Silva, D. E., Sagnard, F., Torre, F., Klein, E., Fady, B. (2008) Life at the Margin: The Mating System of Mediterranean Conifers. *Web Ecol*., 8, 94–102.

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, 48, 1-36.

Savolainen, O., Pyhäjärvi T., Knürr T. (2007) Gene Flow and Local Adaptation in Trees. *Annual Review of Ecology, Evolution, and Systematics,* 38, 595–619.

Sawyer, J. O., and D. A. Thornburgh. 1977. Montane and sub- alpine vegetation of the Klamath Mountains. P. 699–732 in M. G. Barbour and J. Major (eds.), *Terrestrial Vegetation of California*. John Wiley & Sons, New York.

Slatkin M (1985). Gene Flow in Natural Populations. *Annual Review of Ecology and Systematics*, 16, 393–430.

Slatkin M (1993). Isolation by Distance in Equilibrium and Non- Equilibrium Populations. *Evolution*, 47, 264–279.

Smith, C. C. (1970). The Coevolution of Pine Squirrels (*Tam- iasciurus*) and Conifers. *Ecological Monographs*, 40, 348– 371.

Smith, J. M., Haigh, J.,(1974). The Hitch-Hiking Effect of a Favourable Gene. *Genetics Research*, 23–35.

Smith, W.K., Johnson, D. M., Reinhardt, K., (2008). Encyclopedia of Ecology. *Academic Press*, 144-153.

Sork, V. L., Nason, J., Campbel,l D. R., Fernandez. J. F. (2001) Landscape Approaches to Historical and Contemporary Gene Flow in Plants. *Trends in Ecology and Evolution*, 14, 219-224.

Steinbauer, M. J., Grytnes, J.-A., Jurasinski, G., Kulonen, A., Lenoir, J., Pauli, H., … Wipf, S. (2018). Accelerated Increase in Plant Species Richness on Mountain Summits is Linked to Warming. *Nature*, *556*, 231–234.

Swets, K. (1988) Measuring the Accuracy of Diagnostic Systems. *Science*, 240, 1285–1293.

Tranquillini, W. (1976) Water Relations and Alpine Timberline. In: Lange OL, Kappen L, Schulze E-D (eds) *Water and plant life. Ecological Studies*, 19, 473–491.

Venables, W. N., Ripley, B. D. (2002) *Modern Applied Statistics with S. Fourth Edition*. Springer, New York.

Wakeley, J. (2005). The Limits of Theoretical Population Genetics. *Genetics*, *169*, 1–7.

Wallis, C.M. 2007. Understanding the Roles of Phenolics and Terpinoids in Pine Defense Against Fungal Pathogens. Thesis, The Ohio State University, Columbus, OH, USA.

Whittaker, R. J. (1960). Vegetation of the Siskiyou Mountains, Oregon and California. *Ecologocal Monographs,* 30, 279-338.

Wolf, E. J., Harrington, K. M., Clark, S. L., Miller, M. W (2013). Sample Size Requirements for Structural Equation Models: An Evaluation of Power, Bias, and Solution Propriety. E*ducational and Psychological Measurement*, 76, 913-934.

Wright, S. (1931). Evolution in Mendelian Populations. *Genetics*, *16*, 97–159.

Wright, S. 1943. Isolation by Distance. *Genetics* 28,114–138.

Wright S (1951) The Genetic Structure of Populations. *Annals of Eugenics*, 15, 323–354.

Wright, S. (1977) *Evolution and the Genetics of Populations* vol. 3, *Experimental Results and Evolutionary Deductions*. University of Chicago Press, Chicago.

Zacher, I. (2015). A Morphological And Volatile Terpene Analysis Of Pinus Balfouriana To Test For The Mountain Island Effect In The Klamath Mountains. (Master's Thesis). Retrieved from Humboldt Digital Scholar.

Zhang Z., Schwartz S., Wagner L., Miller W. (2000) A Greedy Algorithm for Aligning DNA Sequences. *Journal of Computational Biology*, 7, 203-214.

Zuckerkandl, E., Pauling, L. 1965. Evolutionary Divergence and Convergence in Proteins. *In*: V. Bryson and H. J. Vogel, eds., *Evolving Genes and Proteins*. Academic Press, New York, 97-166.

APPENDIX

Appendix A: Populations sampled.

Population #	Region	Longitude	Latitude	n
2	Yolla Bolly	-122.97181	40.19563	20
3	Yolla Bolly	-122.85887	40.03787	20
4	Trinity Alps	-122.88764	40.91232	20
5	Trinity Alps	-122.88258	40.91408	20
6	Trinity Alps	-122.88258	40.93994	20
7	Trinity Alps	-122.88221	40.9016	20
8	Trinity Alps	-122.89847	40.94175	20
9	Trinity Alps	-122.90782	40.95049	20
15	Scott Mountain	-122.79199	41.22741	20
16	Scott Mountain	-122.78895	41.22124	20
18	Scott Mountain	-122.77382	41.23386	19
19	Little Duck Lake	-122.48779	41.31433	20
20	Little Duck Lake	-122.49014	41.31688	20
21	Crater Lake	-122.58175	41.38403	19
22	Crater Lake	-122.57407	41.37635	20
23	Mount Eddy	-122.95529	41.30197	20
25	Lake Mountain	-123.13416	41.74908	20
26	Trinity Alps	-122.9108	41.0448	20
27	Scott Mountain	-122.7879	41.2282	20

Appendix B: Measured phenotypic traits.

Measured Trait	Category
A_Pinene8.13	Chemistry
B_Pinene9.10	Chemistry
3-Carene9.67	Chemistry
A_Phell_limone10.15	Chemistry
Methylthymol13.47	Chemistry
Bornyl_Acetate14.32	Chemistry
T_Caryoph16.20	Chemistry
A_Humulene16.7	Chemistry
Large resin duct diameter	Needle
Small resin duct diameter	Needle
Distance between ducts	Needle
Needle length	Needle
Needle width	Needle
Needle thickness	Needle
Scale thickness	Cone
Scale width	Cone
Scale length	Cone
Cone length	Cone
Cone width	Cone
Scale number	Cone

Appendix C: ClimateNA variables. * indicates variables used in MaxEnt model.

Abbreviation	Description
Long	Longitude in Lambert Conformal Conic
Lat	Latitude in Lambert Conformal Conic
AHM	Annual heat-to-moisture index
bFFP	Beginning of frost free period
CMD *	Hargreaves climate moisture deficit (mm)
CMI	Hargreaves climate moisture index (mm)
cmiJJA	Summer (June, July, Aug) moisture index (Hogg 1997 modified Penmnan-Monteith method)
eFFP *	End of frost free period
Eref	Hargreaves reference evaporation (mm)
FFP	Frost-free period
MAP *	Mean annual precipitation (mm)
MAT	Mean annual temperature (°C)
MCMT	Mean coldest month temperature (°C)
MSP *	May to September precipitation (mm)
MWMT	Mean warmest month temperature
NFFD	The number of frost-free days
PAS *	Precipitation as snow (mm) between August in previous year and July in current year
PPT_sm *	Summer precipitation (mm)
PPT_wt	Winter precipitation (mm)
SHM *	Summer heat-moisture index ((MWMT)/(MSP/1000))
Tave_sm	Summer mean temperature (°C)
Tave_wt	Winter mean temperature (°C)
TD *	Temperature difference between MWMT and MCMT, or continentality (°C)
Tmax07 *	Temperature max in July (°C)
Tmin01	Temperature min in January (°C)
DDA *	Degree days above 5°C
DDB *	Degree days below 0°C

Appendix D: Correlations of all climate data for foxtail pine location data.

	long	lat	AHM	bFFP	CMD	CMI	cmiJJA	eFFP	EMT	Eref	FFP	MAP	MAT	MCMT	MSP	MWMT
DDB																
DDA																
Tmin01																
Tmax07																
TD																
Tave_wt																
Tave_sm																
SHM																
PPT_wt																
PPT_sm																
PAS																
NFFFD																
MWMT																1
MSP															1	-0.111
MCMT														1	-0.033	0.860
MAT													1	0.933	-0.098	0.970
MAP												1	0.010	0.155	0.539	0.043
FFP											1	0.231	0.828	0.922	0.026	0.806
Eref										1	0.163	-0.237	0.674	0.410	-0.182	0.680
EMT									1	0.254	0.968	0.136	0.872	0.970	-0.030	0.816
eFFP								1	0.875	-0.126	0.937	0.382	0.623	0.809	0.046	0.595
cmiJJA							1	0.341	-0.020	-0.924	0.059	0.326	-0.460	-0.151	0.295	-0.518
CMI						1	0.605	0.369	0.021	-0.541	0.133	0.942	-0.228	-0.011	0.504	-0.205
CMD					1	-0.760	-0.938	-0.359	0.009	0.897	-0.096	-0.534	0.436	0.148	-0.496	0.452
bFFP				1	-0.194	0.132	0.245	-0.743	-0.934	-0.443	-0.930	-0.042	-0.930	-0.914	-0.001	-0.917
AHM			1	-0.241	0.694	-0.954	-0.498	-0.216	0.136	0.505	0.008	-0.907	0.326	0.171	-0.480	0.266
lat		1	0.247	0.294	0.149	-0.337	-0.190	-0.547	-0.427	0.133	-0.454	-0.317	-0.258	-0.373	0.423	-0.304
long	1	0.049	0.742	-0.277	0.663	-0.745	-0.543	-0.178	0.130	0.469	0.047	-0.675	0.294	0.148	-0.598	0.276

											1
										1	-0.885
									1	0.819	-0.951
								1	0.267	0.743	-0.451
							1	0.734	-0.074	0.434	-0.001
						1	0.171	0.598	0.918	0.958	-0.973
					1	0.914	0.538	0.806	0.744	0.988	-0.830
				1	0.564	0.464	0.295	0.440	0.409	0.516	-0.471
			1	-0.362	-0.044	0.050	-0.172	-0.240	0.123	0.001	-0.115
		1	-0.042	-0.795	-0.364	-0.327	-0.146	-0.219	-0.360	-0.340	0.376
	1	0.401	0.414	-0.654	-0.851	-0.848	-0.307	-0.706	-0.717	-0.850	0.797
1	-0.773	-0.348	0.061	0.469	0.824	0.950	0.050	0.369	0.978	0.886	-0.956
0.850	-0.817	-0.376	0.035	0.526	0.989	0.919	0.528	0.750	0.784	0.987	-0.843
-0.062	0.344	0.742	0.451	-0.885	-0.166	-0.050	-0.163	-0.201	-0.027	-0.095	0.078
0.971	-0.777	-0.355	0.144	0.442	0.840	0.979	0.020	0.442	0.967	0.901	-0.992
0.907	-0.871	-0.348	0.002	0.516	0.974	0.981	0.349	0.717	0.851	0.994	-0.924
0.080	0.412	0.026	0.992	-0.428	-0.041	0.061	-0.173	-0.252	0.142	0.011	-0.122
0.965	-0.647	-0.357	0.205	0.377	0.745	0.875	0.045	0.235	0.963	0.815	-0.896
0.317	-0.678	-0.182	-0.221	0.389	0.750	0.568	0.650	0.981	0.218	0.686	-0.434
0.982	-0.733	-0.360	0.117	0.420	0.777	0.929	-0.016	0.308	0.998	0.846	-0.951
0.852	-0.422	-0.355	0.365	0.294	0.515	0.703	-0.181	-0.062	0.883	0.601	-0.782
-0.072	0.516	0.253	0.300	-0.399	-0.583	-0.320	-0.764	-0.934	0.020	-0.494	0.168
-0.047	0.585	0.075	0.933	-0.489	-0.296	-0.147	-0.383	-0.556	0.038	-0.234	0.043
0.081	-0.587	-0.287	-0.492	0.553	0.547	0.310	0.640	0.886	-0.023	0.453	-0.184
-0.953	0.795	0.311	-0.009	-0.414	-0.885	-0.936	-0.275	-0.514	-0.916	-0.929	0.893
0.201	-0.656	-0.075	-0.904	0.515	0.360	0.278	0.237	0.510	0.126	0.317	-0.206
-0.402	0.088	0.550	-0.370	-0.502	-0.276	-0.277	0.025	0.106	-0.424	-0.261	0.383
0.179	-0.543	-0.442	-0.657	0.606	0.337	0.244	0.295	0.484	0.123	0.290	-0.191
NFFFD	**PAS**	**PPT_sm**	**PPT_wt**	**SHM**	**Tave_sm**	**Tave_wt**	**TD**	**Tmax07**	**Tmin01**	**DDA**	**DDB**

Appendix E: Additional population locations used for MaxEnt niche modeling collected from GBIF.

Specimen ID	Longitude	Latitude
UC1452398	-122.85028	40.03361
UCSB23622	-122.8612	40.0391
UC1037449	-122.9023	40.0417
DS366536	-122.9755	40.1809
HSC39117	-122.9621	40.1847
UCR3733	-122.98139	40.205
UC1134094	-122.9976	40.2092
HSC40495	-122.8809	40.9556
HSC32215	-123.0688	40.9752
CAS584540	-122.9716	41.0185
HSC45575	-122.9159	41.0492
UC1452396	-122.90083	41.05
UC688867	-122.8478	41.0519
UC400297	-122.9078	41.0579
SFV20775	-122.92	41.06
HSC40897	-122.8055	41.1666
JEPS46803	-122.7715	41.239
HSC70593	-122.9476	41.242
HSC31504	-122.9528	41.3026
HSC4807	-122.9389	41.3031
HSC26017	-122.95833	41.305
HSC32576	-122.9546	41.3065
HSC92115	-122.4885	41.3083
UC1452402	-122.4666	41.3142
JEPS57744	-122.49115	41.3184
CHSC55794	-122.50528	41.31917
UC1713756	-122.47674	41.31999
UC1585380	-122.51972	41.3299
JEPS57149	-122.50375	41.33995
GH403119	-122.53367	41.34131
UC1452399	-122.56667	41.35056
UC1452400	-122.5807	41.3548
JEPS82668	-122.58963	41.36097
DS492622	-122.3498	41.3693
UC55641	-123.40542	41.49048
UC1272962	-123.19915	41.55641
UC1452397	-123.13333	41.73472
UC1452397	-123.13333	41.74833
JEPS80562	-123.06306	41.95414
DS160173	-122.8811	40.9104

DS286	-122.781	41.259
A246859	-122.7224	41.2628
UC1713757	-122.55169	41.33154
CAS476512	-123.21133	41.56861
LA86572	-123.206	41.5769
CAS463792	-123.092	41.5787
HSC1416	-122.89251	40.94404
A246860	-122.40139	41.27333

Appendix F: Pairwise F_{ST} comparisons between all populations

	27	26	25	23	22	21	20	19	18	16	15	9	8	7	6	5	4	3	2
2	-7.0E-4	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	2.7E-3	NA
3	1.3E-3	3.0E-4	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA		
4	-1.0E-4	4.0E-5	2.6E-3	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA			
5	4.0E-4	-1.0E-3	4.0E-4	-1.0E-4	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA				
6	-2.0E-4	-1.0E-3	1.1E-3	-5.0E-5	-5.0E-4	NA	NA	NA	NA	NA	NA	NA	NA	NA					
7	2.1E-3	1.5E-3	7.8E-3	1.5E-3	2.3E-3	1.8E-3	NA	NA	NA	NA	NA	NA	NA						
8	2.0E-4	-1.1E-3	2.0E-4	6.0E-4	-7.0E-4	-6.0E-4	2.3E-3	NA	NA	NA	NA	NA							
9	-7.0E-4	-1.2E-3	2.3E-3	1.0E-4	-2.0E-4	-4.0E-4	7.0E-4	-1.0E-4	NA	NA	NA								
15	-1.0E-4	-1.2E-3	5.0E-3	5.0E-4	4.0E-4	1.0E-4	3.0E-4	2.0E-3	3.0E-5	NA									
16	-1.0E-4	-1.2E-3	5.0E-3	5.0E-4	4.0E-4	1.0E-4	3.0E-4	2.0E-3	6.0E-4										
18	-1.0E-4	-1.2E-3	5.0E-3	5.0E-4	4.0E-4	1.0E-4	3.0E-4	2.0E-3											
19	-1.0E-4	-1.2E-3	5.0E-3	5.0E-4	4.0E-4	1.0E-4	3.0E-4												
20	-1.0E-4	-1.2E-3	5.0E-3	5.0E-4	4.0E-4	1.0E-4													
21	-1.0E-4	-1.2E-3	5.0E-3	5.0E-4	4.0E-4														
22	-1.0E-4	-1.2E-3	5.0E-3	5.0E-4															
23	-1.0E-4	-1.2E-3	5.0E-3																
25	-1.0E-4	-1.2E-3																	
26	-1.0E-4																		
27																			

Appendix G: Number of private alleles identified in each population

Population	Number of Private Alleles
2	0
3	0
4	0
5	0
6	1
7	0
8	0
9	0
15	1
16	0
18	1
19	0
20	0
21	0
22	1
23	1
25	1
26	1
27	1

Appendix H: Linkage group positions in each bin for minimum sample size of 40 SNPs.

Linkage Group	Bin ID	# of SNPs	Positions Included
1	1_1	125	0
	1_2	43	1.13, 2.25, 4.84, 2.64
	1_6	44	12.9, 19.36
	1_8	45	20.97, 22.58, 25.81, 30.65, 31.66, 32.26
	1_14	67	33.64, 34.7, 38.71
	1_17	54	46.78, 51.62, 54.85, 59.38, 66.11
	1_23	48	82.22, 85.51, 89.27, 90.24
	1_27	48	92.5, 93.57, 96.8, 101.64
2	2_1	54	0, 1.64, 4.92, 6.55
	2_5	46	8.56, 11.46, 13.1, 18.02, 18.56, 22.93
	2_11	92	26.51
	2_12	53	27.3
	2_13	57	34.4, 37.67, 39.31, 41.49, 42.58
	2_18	92	43.67
	2_19	55	44.22, 52.41, 54.04
	2_22	45	55.68, 56.5, 57.32, 60.6
	2_26	94	70.42, 72.06, 75.33, 78.61, 86.8, 88.44
3	3_1	46	0, 0.9, 3.29, 5.41
	3_5	44	12.32, 14.32, 16.22, 18.93
	3_9	98	21.36, 22.53
	3_11	47	25.81
	3_12	49	30.43, 34.02, 41.82, 44.76, 50.13
	3_17	42	53.71, 55.5, 57.3, 60.88, 65.21, 70.41
	3_23	52	72.97, 73.89, 75.42, 77.22, 80.57
	3_28	42	84.15, 85.94, 91.31, 93.1
4	4_1	46	0, 5.83, 11.67, 30.65, 32.08
	4_6	43	33.47, 37.89, 40.24, 42.28, 45.7, 46.66, 48.12
	4_13	55	49.57, 58.32
	4_15	60	69.99, 90.4, 93.2, 96.23
5	5_1	137	0, 5.18, 6.71
	5_4	40	8.05
	5_5	63	12.07, 17.44, 29.52, 34.22
	5_9	45	37.57, 38.91, 41.6, 48.31, 49.65
	5_14	55	52.33, 55.02, 56.36, 72.46, 75.14
	5_19	75	77.83, 83.2, 90.13, 92.59, 96.99, 100.94, 101.61
	5_26	84	106.01, 107.35
6	6_1	60	0, 1.73, 2.35, 3.47, 7.53, 8.67
	6_7	69	10.4
	6_8	72	14.3, 20.96, 22.45

	6_11	62	29.35, 32.8, 34.62, 38.43, 39.24
	6_16	49	39.86
	6_17	45	40.91, 47.37, 51.19
	6_20	61	51.8, 54.6, 55.25
	6_23	51	57.32, 59.04, 59.72, 64.6
	6_27	47	65.87, 66.38, 67.55, 71.05, 80.36, 81.78
	6_33	48	82.9, 83.612, 85.6, 86.13
	6_37	86	88.28, 89.22
	6_39	84	90.13
	6_40	60	91.87
	6_41	43	92.59
7	7_1	117	0, 1.66, 3.32, 6.64, 8.3, 10.83
	7_7	178	11.61
	7_8	82	13.33, 23.23, 24.89, 29.86, 36.5, 38.16, 43.14, 58.07
	7_16	57	65.59
	7_17	63	76.32, 87.93, 92.91, 96.22
8	8_1	123	0, 8.94
	8_3	62	10.43
	8_4	43	11.92, 17.88, 19.36, 25.32, 26.81
	8_9	62	29.79, 37.24, 38.73, 46.18, 52.14, 56.6, 58.1
	8_16	55	59.58,64.05, 67.03, 75.97
	8_20	48	78.94, 80.44, 83.41, 93.85, 98.31, 99.8, 101.29
	8_28	85	107.16
9	9_1	142	0
	9_2	65	1.56, 4.35, 5.92, 7.32
	9_6	43	10.17, 13.08
	9_8	49	15.98, 17.43, 20.5, 24.7, 26.15
	9_13	54	27.61, 29.05, 30.51, 31.96, 33.2
	9_18	41	34.42, 35.26, 39.15, 43.59, 50.85
	9_23	84	52.3, 55.21, 58.11, 62.47, 65.38, 71.19, 77, 82.82, 91.53
	9_32	83	95.89
10	10_1	49	0, 2.94, 4.42
	10_4	51	6.54, 15.36, 16.2, 19.15
	10_8	42	25.04, 29.46, 30.94, 49.53, 64.81
	10_13	43	66.29, 98.66, 100.07
	10_16	60	104.25
	10_17	42	106.06
11	11_1	46	0
	11_2	46	1.53, 4.47
	11_4	59	6.03, 8.94, 10.43, 14.9, 16.39
	11_9	52	20.86, 26.82, 29.8, 32.78, 41.72, 62.57, 71.51
	11_16	94	78.96, 82.8, 93.86, 108.76
12	12_1	145	0, 3.18, 4.73
	12_4	53	11.12

12_5	42	15.89, 19.07
12_7	43	25.43, 28.61, 31.79, 44.51
12_11	63	46.1, 47.69, 49.28
12_14	47	50.87, 52.46
12_16	41	55.64, 58.82, 65.17, 66.76, 81.07
12_21	60	84.25, 87.43, 89.02, 90.61, 95.37, 96.97
12_27	73	98.56

Appendix I: SEM data dictionary

Variable	Description	Transformations
FIS	Multi-locus F_{IS}	
Density	Trees/10 square meters	
PIBA	*P. balfouriana* relative importance	
ABMA	*A. magnifica* relative importance	Natural log applied for normality
PIMO	*P. monticola* relative importance	

Appendix J: SEM model regression details

Model A				
	Variable	**Estimate**	**P(>|z|)**	**Standardized Estimate**
F_{IS} ~	density	-0.685	0.292	-0.785
	PIBA	-0.292	0.145	-0.750
	ABMA	-0.025	0.025	-0.324
	PIMO	-0.228	0.158	-0.462
PIBA ~	PIMO	-0.544	0.243	-0.430
	density	-1.492	0.347	-0.666
	ABMA	-0.078	0.039	-0.396

Model B				
	Variable	**Estimate**	**P(>|z|)**	**Standardized Estimate**
F_{IS} ~	density	-0.555	0.061	-0.637
	PIBA	-0.188	0.156	-0.482
PIBA ~	PIMO	-0.544	0.025	-0.430
	density	-1.492	0.000	-0.666
	ABMA	-0.078	0.043	-0.396