



VCU

Virginia Commonwealth University
VCU Scholars Compass

Theses and Dissertations

Graduate School

2018

NATURAL AND ANTHROPOGENIC DRIVERS OF TREE EVOLUTIONARY DYNAMICS

Brandon M. Lind
Virginia Commonwealth University

Follow this and additional works at: <https://scholarscompass.vcu.edu/etd>



Part of the [Botany Commons](#), [Evolution Commons](#), [Forest Management Commons](#), [Genetics Commons](#), [Genomics Commons](#), [Integrative Biology Commons](#), [Other Ecology and Evolutionary Biology Commons](#), [Other Forestry and Forest Sciences Commons](#), [Plant Biology Commons](#), [Plant Breeding and Genetics Commons](#), and the [Population Biology Commons](#)

© Brandon M. Lind

Downloaded from

<https://scholarscompass.vcu.edu/etd/5359>

This Dissertation is brought to you for free and open access by the Graduate School at VCU Scholars Compass. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of VCU Scholars Compass. For more information, please contact libcompass@vcu.edu.

© Brandon M. Lind 2018

All Rights Reserved

NATURAL AND ANTHROPOGENIC DRIVERS OF TREE EVOLUTIONARY DYNAMICS

A dissertation submitted in partial fulfillment of the requirements of the degree of
Doctor of Philosophy at Virginia Commonwealth University

by

BRANDON MICHAEL LIND

BACHELOR'S DEGREE IN CHEMISTRY, UNIVERSITY OF NORTHERN IOWA 2010

BACHELOR'S DEGREE IN BIOLOGY, UNIVERSITY OF NORTHERN IOWA 2010

DIRECTOR: ANDREW J. ECKERT, Ph.D.

ASSOCIATE PROFESSOR, DEPARTMENT OF BIOLOGY

Virginia Commonwealth University

Richmond, Virginia

April 27, 2018

Acknowledgement

“What is a tree? A tree is a big plant with a stick up the middle. Everybody knows that.”

– Colin Tudge

The Tree: A natural history of what trees are, how they live, and why they matter

I wish to thank my amazing wife, Sevima, and daughter, Ella Marie; without them none of this would be possible. To Andrew Eckert, my advisor, who was a fantastic mentor and friend. To all of my committee members and many collaborators that helped produce great science. And to the students and faculty of VCU who created such a great environment to work, learn, and grow. Thank you!

Table of Contents

	Page
Acknowledgement.....	ii
Table of Contents.....	iii
List of Tables.....	vi
List of Figures.....	ix
List of Abbreviations.....	xvii
Abstract.....	xix
Chapter 1. Introduction.....	1
Factors affecting tree evolutionary dynamics.....	2
Gene Flow, Mating System, and Dispersal.....	2
Population size and density.....	5
Selection and Genetic Architecture.....	6
Summary.....	8
Objectives.....	8
Chapter 2 – Effects of forest management on fine-scale gene flow in <i>Pinus lambertiana</i> Dougl.	9
Chapter 3 – Local adaptation of <i>Pinus albicaulis</i> Engelm. across fine spatial scales.....	9
Chapter 4 - Local adaptation and linkage maps I: A first step towards describing genetic architectures of complex traits in natural populations.....	10
Chapter 5 - The genetic architecture of local adaptation II: The QTL landscape of water-use efficiency for foxtail pine (<i>Pinus balfouriana</i> Grev. & Balf.).....	11
Chapter 6 – The current state and future directions of forest tree genomics.....	11
Chapter 2. Effects of forest management on fine-scale gene flow in <i>Pinus lambertiana</i>	12
Abstract.....	12
Introduction.....	13
Methods.....	17
Study area, sampling, and focal species.....	17
Analysis of tree spatial structure.....	19
DNA extraction, microsatellite amplification.....	19
Genetic diversity measures.....	21
Analysis of spatial genetic structure.....	21
Parentage analysis.....	22
Using parentage analysis to further quantify fine-scale geneflow.....	25
Results.....	26
Analysis of tree spatial structure.....	26
Diversity measures.....	33
Analysis of spatial genetic structure.....	34
Quantifying fine-scale gene flow.....	36

Discussion.....	42
Conclusion.....	49
Chapter 3. Local adaptation of <i>Pinus albicaulis</i> Engelm. across fine spatial scales of the Lake Tahoe Basin, USA	50
Abstract	50
Introduction	51
Methods	55
Focal species, study area, and sampling	55
Common gardens and phenotypic measurements	57
DNA extraction, sequencing, and analysis	57
Identifying focal sets of loci	59
Inferring signatures of local adaptation	60
Results	63
SNP filtering and characterization	63
Genotype-environment analysis	63
Genotype-phenotype analysis	66
F_{ST} outlier analysis	69
Assessing possible artefactual signals of high $\hat{D}_{a(ij)}$	70
Intersection of SNPs within and across methods	71
Discussion.....	72
Standing genetic variation for fitness related traits.....	73
Water availability as a driver of local adaptation.....	74
Implications to whitebark management.....	75
Limitations and concluding remarks	77
Acknowledgments.....	79
Chapter 4. Local adaptation and linkage maps I: A first step towards describing genetic architectures of complex traits in natural populations	80
Abstract	80
Introduction	81
Materials and Methods	84
Focal species	84
Sampling	86
Library preparation and sequencing	86
DNA sequence analysis	87
Linkage analysis.....	89
Consensus map construction and biological interpretation	92
Results	93
DNA sequence analysis	93
Linkage mapping.....	96
Consensus map construction and biological interpretation	98
Discussion.....	101
Acknowledgments.....	105
Chapter 5. Local adaptation and linkage maps II: The QTL landscape of water-use efficiency in <i>Pinus balfouriana</i>	106
Abstract	106
Introduction	107

Materials and methods	110
Focal species	110
Common garden	111
Phenotypic determination	112
Sequence analysis and genotype determination	112
Species distribution modeling	114
Quantitative genetic analysis	116
Results	118
Sequence analysis and genotype determination	118
Species distribution modelling	119
Quantitative genetic analysis	123
Discussion	128
Climate drivers of the current geographical distribution and WUE	129
Genetic architecture of water-use efficiency	131
Conclusions	136
Chapter 6. Current state and future directions of forest tree genomics	137
Abstract	137
Introduction	138
Identifying heritable phenotypic variation	141
Expectations for the loci underlying quantitative traits	148
Gene action	150
Negative selection	156
Positive selection	157
Gene flow	161
Summary	163
The genomics of local adaptation in trees	163
Common approaches used to identify adaptive loci	163
Current progress in trees	164
Are we out of the woods yet?	166
The path forward	168
Box 6.1: A step in the right direction: Synergism between GWAS and Genomic Selection	169
Stepping off the pack – what’s in our pack?	171
At the trail junction – where to next?	175
Needed data	176
Standardized data reporting	179
Empirical tests of theory	182
Concluding remarks	183
Acknowledgements	184
Chapter 7. Conclusions	185
References	Error! Bookmark not defined.
Vita	Error! Bookmark not defined.

List of Tables

- Table 2.1** Genetic diversity measures (standard deviation) by treatment. N : number of individuals [adults, seedlings]; A_T : total number of alleles; A : mean number of alleles per locus; A_e : effective number of alleles (harmonic mean across loci); H_o , H_e : respectfully observed and expected heterozygosity for nuclear markers; A_p : average number of private alleles. For A , A_e , H_o , and H_e , values indicate averages across loci, where values for each locus were calculated across all three treatment replicates simultaneously. H_o and H_e used only nuclear markers, whereas other genetic diversity columns considered all loci.34
- Table 3.1.** Population location and associated attributes. Population size – total (maternal trees with seedlings in common garden). Climatic values were ascertained from data spanning 1971-2000. Ann. precipitation – annual precipitation; AWC – available water capacity at 25cm or 50cm soil depth; CEC – cation exchange capacity; GDD – growing degree days above 5°C; Max solar rad input – maximum solar radiation input; WC-15bar – water capacity at -15bar (wilting point); WC-1/3bar – water capacity at -1/3bar (field capacity). Asterisks indicates populations from which seeds sampled from cones were planted in a common garden. Environmental variables are averaged across subplots.56
- Table 3.2.** Signatures of allele frequency shifts associated with environmental distance. Significant Mantel tests (9999 permutations) from comparisons among $pwD_{a(ij)}$ matrices from SNPs associated with environment (first column) against environmental Euclidian distance (second column). Environmental variables as in Table 3.1. † indicates a comparison

in which at least one variable is water-related, or was associated with annual precipitation.	58
Table 3.3. Signatures of allele frequency shifts associated with environmental distance. Significant Mantel tests (9999 permutations) from comparisons among allele frequency shifts ($pwD_{a(ij)}$) of SNPs associated with phenotype (second column) against environmental Euclidian distance (third column). Selection Criterion refers to the process used to identify SNPs associated with phenotype. † indicates a comparison in which at least one variable is water-related, or was associated to annual precipitation.....	70
Table 4.1. Attributes of the data structure related to maternal tree.....	95
Table 4.2. Attributes of single-tree and the consensus linkage maps. Values for ratio variables are totals and are not averaged across linkage groups (see Tables S4.1–S4.5).	97
Table 5.1 Summary of the families ($n=5$) used for QTL mapping. Sibling counts represent the numbers of siblings genotyped and phenotyped for each family. Additional siblings for each family are still growing within the common garden (see Materials & Methods).....	111
Table 5.2 Mean (standard deviation) of read metrics by family	118
Table 5.3 Attributes of linear mixed models used to estimate familial and regional effects for each phenotypic trait. Values in parentheses are 95% parametric bootstrap confidence intervals (see Materials and Methods).....	123
Table 5.4 Comparisons of linear mixed models using the Akaike Information Criterion (AIC) by trait were used to select the best model (bolded text). In these models, the intercept was a fixed effect, while families nested within regions and regions were random effects. The AIC weight is calculated using the standardized relative likelihoods, where the relative likelihood is given as $e^{(-0.5 \times \Delta AIC)}$. For this calculation, ΔAIC is the difference between the AIC for each model and the AIC for the best model (bolded text), where the best model is the one with the lowest AIC. The weights are then calculated as each of relative likelihoods over the sum of	

the relative likelihoods, thus making the sum of the weights equal to 1. Akaike weights can also be considered as the conditional probabilities for each model..... 124

Table 5.5 Summary of QTLs for each trait that survive multiple test corrected significance thresholds at either the level of the whole genome ($\alpha = 0.05$ for $G_{0.05}$) or a chromosome ($\alpha = 0.01$ for $C_{0.01}$, $\alpha = 0.05$ for $C_{0.05}$). ^aLG, Linkage group, ^bPVE, percent variance explained; PVE_c , corrected percent variance explained, ^cThe threshold value for the F-statistic under the null model as determined using the listed value of α (0.05 or 0.01) and permutations following Churchill and Doerge (1994) for either individual linkage groups (C) or the entire genome (G). ^d95% CI, 95% confidence interval determined through bootstrap analysis (n = 1,000 replicates)..... 126

Table 5.6 Summary of two QTL models fit to each significant QTL from Table 4. Bolded P-values are less than 0.05. ^aLG, Linkage Group; ^bPVE, percent variance explained by both QTLs 129

Table 6.1. Where to next? The Path Forward identifies meaningful ways in which we can progress our understanding of the architecture underlying complex traits in trees. Here we outline some questions that can be used to guide future inquiry as the number of markers and sequence length increase, and annotation becomes more precise and specific to tree biology..... 170

List of Figures

Figure 2.1 Teakettle Experimental Forest, California (Latitude: 36.9606, Longitude: -119.0258).

(A) Spatial arrangement of treatments (BC = burned understory thin; BN = burned no-thin; BS = burned shelterwood thin; UC = unburned understory thin; UN = unburned no-thin; US = unburned shelterwood thin). Replicates for each treatment are numbered one through three from south to north. (B) Mapped coordinates (Universal Transverse Mercator) and elevation (meters) of pre-treatment adults ≥ 5 cm diameter at breast height. (green: *P. lambertiana*, red: *P. jeffreyi*, gray: *A. concolor*, blue: *A. magnifica*, orange: *C. decurrens*, black: *Quercus*, *Salix*, and remaining species.) 18

Figure 2.2 Univariate analysis of adult *P. lambertiana* (PiLa) spatial structure, $g_{inhom}(r)$, by treatment replicate for each distance class, r . Gray : null confidence envelope; Solid black line : observed $g_{inhom}(r)$. Red dashed line : null expectation of complete spatial randomness, $g_{inhom}(r) = 1$. Individuals are aggregated if $g_{inhom}(r) > 1$, inhibited if $g_{inhom}(r) < 1$. BC1 was below the threshold sample size allowed by spatstat.....27

Figure 2.3 Univariate analysis of adult shade tolerant (*A. magnifica* and *A. concolor*; ShadeTol) spatial structure, $g_{inhom}(r)$, by treatment replicate for each distance class, r . Gray : null confidence envelope; Solid black line : observed $g_{inhom}(r)$. Red dashed line : null expectation of complete spatial randomness, $g_{inhom}(r) = 1$. Individuals are aggregated if $g_{inhom}(r) > 1$, inhibited if $g_{inhom}(r) < 1$28

Figure 2.4 Univariate analysis of *P. lambertiana* (PiLa) seedling spatial structure, $g_{inhom}(r)$, by treatment replicate for each distance class, r . Gray : null confidence envelope; Solid black

line : observed $g_{inhom}(r)$. Red dashed line : null expectation of complete spatial randomness, $g_{inhom}(r) = 1$. Individuals are aggregated if $g_{inhom}(r) > 1$, inhibited if $g_{inhom}(r) < 1$29

Figure 2.5 Bivariate analysis of spatial structure between *P. lambertiana* seedlings (seed) and *P.*

lambertiana adults, $g_{inhom,seedling,adult}(r)$, by treatment replicate for each distance class, r .

Gray : null confidence envelope; Solid black line : observed $g_{inhom,seedling,adult}(r)$. Red

dashed line : null expectation of complete spatial random-ness, $g_{inhom,seedling,adult}(r) = 1$.

Individuals are aggregated if $g_{inhom,seedling,adult}(r) > 1$, inhibited if

$g_{inhom,seedling,adult}(r) < 1$30

Figure 2.6 Bivariate analysis of spatial structure between *P. lambertiana*

seedlings (seedling) and adult *Abies* spp. (shadetol), $g_{inhom,seed,shadetol}(r)$, by treatment

replicate for each distance class, r . Gray : null confidence envelope; Solid black line :

observed $g_{inhom,seed,shadetol}(r)$. Red dashed line : null expectation of complete spatial

random-ness, $g_{inhom,seed,shadetol}(r) = 1$. Individuals are aggregated if $g_{inhom,seed,shadetol}(r) >$

1, inhibited if $g_{inhom,seed,shadetol}(r) < 1$31

Figure 2.7 Bivariate analysis of spatial structure between *P. lambertiana* adults

(PiLa-Adult) and adult *Abies* spp. (shadetol), $g_{inhom,PiLa-adult,shadetol}(r)$, by treatment

replicate for each distance class, r . Gray : null confidence envelope; Solid black line :

observed $g_{inhom,PiLa-adult,shadetol}(r)$. Red dashed line : null expectation of complete spatial

random-ness, $g_{inhom,PiLa-adult,shadetol}(r) = 1$. Individuals are aggregated if

$g_{inhom,PiLa-adult,shadetol}(r) > 1$, inhibited if $g_{inhom,PiLa-adult,shadetol}(r) < 1$32

Figure 2.8 Analysis of spatial genetic structure (sensu Smouse & Peakall 1999) between *P.*

lambertiana adults (first row), *P. lambertiana* seedlings (second row), and between *P.*

lambertiana adults and seedlings (third row) by treatment (columns) across distance classes

within plots (main panel) or across TEF (insets). Values of $r_{gh} = 0$ indicate random spatial

patterns of genotypes, $rgh > 0$ indicate clustering of alike genotypes, and $rgh < 0$ indicate spatial inhibition of alike genotypes.35

Figure 2.9 Mother-offspring dispersal events by treatment for (A) dispersal between in-plot individuals, (B) dispersal into plot from an out-of-plot mother, and (C) the ratio of these values. There were no events in which a known mother dispersed seed to another plot, therefore B is utilizing information from parentage analysis that indicated the mother of a given seedling was not sampled. Orange letters within each plot show significant differences between medians, as inferred from separate Kruskal-Wallis tests (see main text of Results). Vertical lines indicate standard deviations.37

Figure 2.10 Father-mother dispersal events by treatment for (A) dispersal between in-plot individuals, (B) dispersal into plot from an out-of-plot mother, and (C) the ratio of these values. Plot-level tallies were those of in-plot mothers receiving pollen from either an in-plot father (A) or an out-of-plot (sampled or unsampled) father (B). Orange letters within each plot show significant differences between medians, as inferred from separate Kruskal-Wallis tests (see main text of Results). Vertical lines indicate standard deviations.39

Figure 2.11 Dispersal distances for seed (A) and pollen (B) calculated from the most probable pedigree from parentage analysis, considering only pedigrees with known mothers (A) or known parents (B). Orange letters within each plot show significant differences between medians, as inferred from separate Kruskal-Wallis tests for mother-offspring and father-mother dispersal distances (see main text of Results).40

Figure 2.12 Fitted 2D-t dispersal kernel for seed (red) and pollen (black) using shape parameters inferred from parentage analysis (Moran & Clark 2011). Dashed lines show the 95% credible interval. This figure is truncated at the maximum distance within plots ($200\sqrt{2}$ m) to focus on differences at short distances.40

Figure 2.13 Fractional parentage across parentage analysis cycles for (A) maternal assignment and (B) paternal assignment (see Methods) with adult individuals along x-axes and seedling

individuals along y-axes. Each cell represents the fraction of the cycles a particular seedling was assigned to a given adult (black ~ 0 to red to orange to yellow to white ~1).41

Figure 2.14 Dispersal distances between mothers and offspring (first column) and between fathers and mothers (second column) using assigned adults from any location (A-B) and for only in-plot individuals (C-D). Orange letters within each plot show significant differences between medians, as inferred from separate Kruskal-Wallis tests (see main text of Results).43

Figure 3.1. Populations used for sampling *P. albicaulis* within the Lake Tahoe Basin (dark outline). Annual precipitation is given for each population to demonstrate the west-east rain shadow experienced across fine spatial scales. Asterisks indicate populations in the common garden study.54

Figure 3.2. Allele frequency covariance ($D_{a(ij)}$) among loci associated to environment by bayenv2. In white are the median values from $D_{a(ij)}$ calculated among focal SNPs associated to environment. Black bars display the 95th percentile of the null distribution of median $D_{a(ij)}$. Environmental variables are grouped by those related to soil (CEC through WC-1/3bar) and those related to climate or geography (Annual precipitation through T-min Jan), with variables related to water availability grouped together in the center of the figure (AWC-25cm through Annual precipitation). All sets of focal loci had median $D_{a(ij)}$ greater than the 100th percentile of the null distribution, as indicated by two stars (**). Environmental variables as in Table 3.1.65

Figure 3.3. Violin plots for the kernel density estimator of the posterior distributions (light grey) taken from Bayesian sparse linear mixed models (BSLMM) executed in GEMMA for (A) the proportion of variance explained by SNPs included in the model (PVE) and (B) the number of SNPs underlying the phenotypic trait (N_{SNP}). Priors for N_{SNP} and PVE were [1,300] and

[0.01,0.9], respectively. Dark grey vertical bars display the first through third interquartile range, with the median represented by the white dot.....68

Figure 3.4. Allele frequency covariance ($D_{a(ij)}$) among loci associated to phenotype by GEMMA.

In white are the median values from $D_{a(ij)}$ calculated among focal SNPs associated to phenotype. Black bars display the 95th percentile of the null distribution of median $D_{a(ij)}$. (A) SNPs identified in the top 99.9th percentile of *PIP*, (B) SNPs identified in the top 99.8th percentile of *PIP*. One star (*) indicates that the median focal $D_{a(ij)}$ was greater than the 95th percentile of the null distribution, whereas two stars (**) indicate that the focal median $D_{a(ij)}$ was greater than the 100th percentile of the null distribution.69

Figure 4.1. Geographical locations of foxtail pine samples used to construct a common garden

located in Placerville, CA. Circles denote the 15 unique locations from which 4 to 17 maternal trees were sampled. Circles enclosed in squares denote locations from which maternal trees used in linkage mapping were sampled. Photo credits: lower: T. Burt; upper: A. Delfino Mix.85

Figure 4.2. Sharing of contigs across maternal tree maps from which the consensus map was

constructed. Counts in each cell represent the number of unique contigs appearing on the final consensus map. Unique contigs for the yellow and green maternal trees were largely discarded to make estimation of pairwise recombination fractions computationally feasible (see Materials & Methods section). KM Klamath Mountains, SN Sierra Nevada.....94

Figure 4.3. Consensus linkage map of 12 linkage groups, derived from SNPs among individuals

of four populations. Working inward from the outermost section of the figure, for each linkage group: (1) the solid black bars represent the span of recombination distances (in cM) for markers; (2) the individual tick marks show the locations of the markers, and the colors represent the density of annotation of the SNPs at that position ($\geq 50\%$ = green, $\geq 25\%$ = red, $< 25\%$ = black) to homologous locations in lobolly pine); (3) The black density plot

represents the counts of SNPs from all four families mapping to a specific position in the linkage group; (4) the colored density plots show the contribution SNPs from the individual families to the markers on the map at each position and are shown in order by total read count in the library, with yellow having the most and red having the least amount of reads. Linear plots of linkage groups comprising the consensus map are given in Figure S4.9..100

Figure 5.1 Species distribution models (SDMs) created using MaxEnt are good predictors of the current geographical range of foxtail pine (inlaid maps; AUC = area under the receiver operating characteristic curve). Precipitation and temperature-related variables are differentially important, as measured by variable contributions to each model, to the SDM of each regional population of foxtail pine, with precipitation-related variables more important for the Klamath Region and temperature-related variables more important for the southern Sierra Nevada. Variable contribution scores (± 1 standard deviation derived from 10 replicated runs of MaxEnt per SDM) are uncorrelated (Spearman's $\rho = -0.065$). For symbols without apparent error bars, the diameter of the circle was greater than the standard deviation.....119

Figure 5.2 Ranks of variable importance (low rank = more important) based on variable contribution (VC) scores and permutation importance (PI) scores to the SDM for each regional population are moderately correlated ($r = \text{Spearman's } \rho$). Variable types are denoted using filled circles, with black used for temperature-related variables, white for precipitation-related variables, and gray for variables related to both temperature and precipitation.....121

Figure 5.3 Familial and regional level means (± 1 standard error) by trait (left: $\delta^{13}\text{C}$, right: $\text{N}_{\mu\text{g}}$) are differentiated across families and regions relative to the global mean. Dashed gray lines give global means across all families for each trait. Estimates for the Klamath Mountains (KM) are given as filled circles, while estimates for the southern Sierra Nevada (SN) are

given as filled triangles. Familial names are given as colors (see Materials and Methods).

..... 125

Figure 5.4 The relationship between traits based on family means (± 1 standard error) is positive (Pearson's $r = 0.415$), although statistically non-significant at $\alpha = 0.05$ ($P = 0.487$). Dashed gray lines give global means across all families for each trait. 127

Figure 5.5 The distributions of the F-statistic derived from single QTL models across each linkage group for carbon isotope discrimination and nitrogen content of needles reveals the isolated nature of QTLs. The dashed horizontal line in each panel is the genome-wide significance threshold ($\alpha = 0.05$) for the F-statistic based on the permutation scheme ($n = 1,000$ permutations) suggested by Churchill and Doerge (1994). Significant QTLs are denoted with filled circles ($\alpha = 0.05$, genome-wide), filled triangles ($\alpha = 0.01$, chromosome-wide) or filled squares ($\alpha = 0.05$, chromosome-wide). 128

Figure 6.1. Average Q_{ST} for each of 14 trait categories from literature review calculated by weighting each estimate by the number of families used in the estimation. Error bars represent the standard deviation of the weighted averages. Numbers above error bars represent total number of estimates, with total number of unique species in parentheses. Asterisks indicate median values of the unweighted Q_{ST} distribution. ColdH = cold hardiness, DisRes = disease resistance, DroughtH = drought hardiness, HaIR = herbivore and insect resistance, LaNP = leaf and needle properties, Phen = phenology, PSM = plant secondary metabolites, Reprod = reproduction, ResAllo = resource allocation, SaSP = seed and seedling properties, WoodProp = wood properties. 147

Figure 6.2. Relevant quantitative genetic concepts are needed to understand the evolution of polygenic traits. (A) Additive and non-additive effects at a single locus, where a is defined as the additive effect (also known as the average effect of allelic substitution [α] when there is no dominance) and d is defined as the dominance deviation. With dominance, $\alpha = a[1 + k(p - q)]$, where k is the degree of dominance ($k = 0$: additive, $k = 1$: dominance, $k > 1$: over-

dominance, see Lynch & Walsh 1998). (B) Polygenic traits are determined by multiple genes, each with additive (shown) and non-additive (not shown) effects. The total additive effect is the sum of the additive effects at all causative loci. (C) Additive-by-additive epistasis, where the additive effect of an allele at the PHY_A SNP depends on what allele it is paired with at the RPL13 SNP. In this case, the effects can be thought of as dependent in the following manner using the four possible haplotypes at the PHY_A (A/T SNP) and RPL13 (C/T SNP) SNPs – AC: +5, AT: -2, TC: -1, TT: 4. (D) The effect of genetic drift on the additive genetic variance as determined by 100 independent, causative loci. Each line represents a simulation of genetic drift in a constant sized population ($n = 500$ diploids) conditioned on initial allele frequencies across loci (\mathbf{p}_1) and effect sizes ($\boldsymbol{\alpha}$). The expected mean across all 100 simulations is given by the dashed black line. Any given simulation can deviate strongly from this expectation (solid black line). Thus, when the elements of \mathbf{p} change over time, in this case due to genetic drift, so does the additive genetic variance. See also Supplemental Box S6.1. 152

Figure 6.3. Insights from genotype-phenotype literature review. (A) Counts of per-locus percent variance explained (r^2) estimates from single-locus genotype-phenotype associations from literature review. Note logarithmic x-axis. (B) Distribution of per-locus r^2 values for individual trait groups within genotype-phenotype literature review. Values along x-axis are total number of estimates and number of species across estimates. Not shown are nine outliers for disease resistance to *Cronartium ribicola* in *Pinus monticola* (range = [0.402, 1.0]) from Lui et al. 2017. Abbreviations as in Figure 6.1. 166

List of Abbreviations

AIC – Akaike Information Criterion
Ann. precipitation – annual precipitation
CEC – cation exchange capacity
AUC – area under the curve
AWC – available water capacity
bp – Base-pairs of DNA
BSLMM – Bayesian sparse linear mixed model
CART – classification and regression tree model
cM – centiMorgan
cpDNA – chloroplast DNA
ddRADseq – double digest restriction-site associated DNA sequencing
dbRDA – distance-based redundancy analysis
DNA – deoxyribonucleic acid
Gbp – Giga base-pairs of DNA
GDD – growing degree days above 5°C
GEA – genotype-environment association
GPA – genotype-phenotype association
GS – genomic selection
GWAS – genome-wide association study
GxE – genotype-by-environment interaction
IBD – isolation-by-distance
LD – linkage disequilibrium
LFMM – latent factor mixed model
LTBMU – Lake Tahoe Basin Management Unit
Max solar rad input – maximum solar radiation input
MCMC – Markov chain Monte Carlo
MSE – mean squared error

mtDNA – mitochondrial DNA

MYA – million years ago

QTL – quantitative trait locus/loci

PVE – percent variance explained

RADseq – restriction-site associated DNA sequencing

RAPTURE – restriction-site associated DNA sequence capture

PCR – polymerase chain reaction

PI – permutation importance

RDA – redundancy analysis

ROC – receiver operating characteristic curves

SDM – species distribution model

SGS – spatial genetic structure

SNP – single nucleotide polymorphism

TEF – Teakettle Experimental Forest

USDA NRCS – United States Department of Agriculture Natural Resource Conservation Service

USFS – United States Forest Service

VC – variable contribution

WC- $\frac{1}{3}$ bar – water capacity at $-\frac{1}{3}$ bar (field capacity)

WC-15bar – water capacity at -15bar (wilting point)

WUE – water-use efficiency

Abstract

NATURAL AND ANTHROPOGENIC DRIVERS OF TREE EVOLUTIONARY DYNAMICS

Brandon Michael Lind – Doctor of Philosophy – Integrative Life Sciences

A dissertation submitted in partial fulfillment of the requirements of the degree of Doctor of Philosophy at Virginia Commonwealth University

Virginia Commonwealth University, 2018

Major Director: Andrew Eckert, Ph.D. – Associate Professor of Biology

Species of trees inhabit diverse and heterogeneous environments, and often play important ecological roles in such communities. As a result of their vast ecological breadth, trees have become adapted to various environmental pressures. In this dissertation I examine various environmental factors that drive evolutionary dynamics in three *Pinus* species in California and Nevada, USA. In chapter two, I assess the role of management influence of thinning, fire, and their interaction on fine-scale gene flow within fire-suppressed populations of *Pinus lambertiana*, a historically dominant and ecologically important member of mixed-conifer forests of the Sierra Nevada, California. Here, I find evidence that treatment prescription differentially affects fine-scale genetic structure and effective gene flow in this species. In my third chapter, I describe the development of a dense linkage map for *Pinus balfouriana* which I use in chapter four to assess the quantitative trait locus (QTL) landscape of water-use efficiency across two isolated ranges of the species. I find evidence that precipitation-related variables structure the geographical range of *P. balfouriana*, that traits related to water-use efficiency are heritable and differentiated across populations, and associated QTLs underlying this phenotypic variation explain large proportions

of total variation. In chapter five, I assess evidence for local adaptation to the eastern Sierra Nevada rain shadow within *P. albicaulis* across fine spatial scales of the Lake Tahoe Basin, USA. Here, genetic variation of traits related to water availability were structured more so across populations than neutral variation, and loci identified by genome-wide association methods show elevated signals of local adaptation that track soil water availability. In chapter six, I review theory related to polygenic local adaptation and literature of genotype-phenotype associations in trees. I find that evidence suggests a polygenic basis for many traits important to conservation and industry, and I suggest paths forward to best describing such genetic bases in tree species. Overall, my results show that spatial and genetic structure of trees are often driven by their environment, and that ongoing selective pressures driven by environmental change will continue to be important in these systems.

Chapter 1.

Introduction

Forested systems cover roughly 31% of Earth's total land area (Costanza et al. 1997; FAO 2010), accounting for approximately 90% of terrestrial biomass (Whitaker 1975), and provide habitat to the vast majority of Earth's terrestrial biodiversity. Additionally, many valuable ecosystem services are carried out in forests such as the production of seeds, leaves, and bark for wildlife forage, forest cover, the production of oxygen, carbon sequestration, the mitigation of droughts and floods, air and water filtration, timber for building and other materials, as well as the reduction of erosion, protracting snowmelt, and desertification. Thus, identifying the factors that shape evolutionary dynamics of tree species is of the utmost importance to conservation, management, and industrial agencies. However, the large size and long generation times of many tree species have limited the study of a number of aspects of their evolution. Some of the earliest studies employed common gardens and reciprocal transplants to understand how aspects of the environment shape properties of survival and growth, most often with the goal of delineating populations for conservation management or for optimizing yield from industrial plantations. Up until the last two decades, however, evolutionary insight at the DNA sequence level have been largely out of reach, due in part to the cost of producing molecular markers and the large genome size of many trees. Even so, recent reductions in the cost of molecular data, alongside an increasing number of molecular resources, have provided novel opportunity for evolutionary insight within and among tree species. In this dissertation, I aim to take advantage of such resources to better understand how environmental change (either from anthropogenic or natural sources) influences the evolutionary dynamics of tree populations, and in particular to identify the underlying genetic components conferring adaptive responses to the environment. To do so, I take advantage of three different study systems, *Pinus lambertiana* Dougl. (Pinaceae) in a fire-

suppressed mixed-conifer forest in the Sierra Nevada mountains of California, *P. albicaulis* Engelm. (Pinaceae) in upper montane forests of the Lake Tahoe Basin overlapping California and Nevada, and *P. balfouriana* Grev. & Balf. (Pinaceae) of high-elevation forests of the Klamath and Sierra Nevada mountains of California.

Factors affecting tree evolutionary dynamics

Gene Flow, Mating System, and Dispersal

Gene flow is a considerable factor shaping the evolutionary dynamics of populations, and of trees in particular. Gene flow, in the context of trees, can be defined as the sharing of genetic material between populations through the dispersal of seed or pollen that survive to contribute genetic material to subsequent generations. The extent to which populations are genetically differentiated across the landscape, due to a number of causes (discussed below), is ultimately influenced by gene flow, which, when sufficiently strong, or without other acting factors, will homogenize allele frequencies across populations resulting in little to no detectable genetic differentiation (Haldane 1930; Wright 1931; Slatkin 1987). However, the extent of gene flow between two given populations of trees will often depend, in part, on their degree of geographic isolation, as the dispersal distances of seed and pollen are often limited (Savolainen et al. 2007). When this is the case, spatially varying degrees of isolation can lead to isolation-by-distance (IBD) where populations with lesser degrees of isolation will be more genetically similar than populations with greater degrees of isolation where allele and genotype frequencies across populations experience higher degrees of independence. In these cases, geographically based genetic differentiation is referred to as spatial genetic structure (SGS). As hinted at previously, SGS is not only influenced by the degree of gene flow between populations. In many cases, instances of SGS can be due to serial colonization, genetic drift (random allele frequency fluctuations due to the finite number of meiotic products and individuals or offspring), geographically based differences in selection pressures or strengths, nonrandom mating, as well as with differences in mutation rate, allele frequencies among sexes, mating system (i.e., the

degree of outcrossing), reproductive output or success, patterns of genetic recombination, violations to the independent assortment of multiple loci during gamete formation, and, more often than not, some combination of these factors.

Ultimately, SGS will impact the probability of persistence of populations. At local scales this is particularly true, especially in circumstances in which dispersal is limited, where the degree of relatedness among individuals can impact aspects of the mating system and the effective levels of consanguineous breeding (i.e., inbreeding) which can often be deleterious due to the union of rare recessive alleles. In some cases, gene flow can rescue populations by introducing genetic variation and can further facilitate the rate of adaptation through the gain of adaptive alleles from other populations, but this is not always the case as gene flow may in some cases swamp selected alleles from contributing to local adaptation (Haldane 1930), particularly at range margins (García-Ramos & Kirkpatrick 1997). At local spatial scales, however, dispersal ability is the greatest factor influencing fine-scale SGS within populations. For wind pollinated trees, gene flow resulting from the movement of seeds and pollen can be estimated through analysis of maternally-inherited mitochondrial DNA (mtDNA), paternally-inherited chloroplast DNA (cpDNA), or from biparentally-inherited nuclear DNA. Using such data has led to findings that suggest that dispersal distances of pollen are often orders of magnitude higher than that of seeds (e.g., Ennos 1994; see Table 2 in Petit & Hampe 2006). Further, it is often the case that tree species exhibit little genetic differentiation between populations (Howe et al. 2003; Alberto et al. 2013) due to extensive gene flow among populations (Slavov et al. 2004; Savolainen et al. 2007) generally to a higher degree than most herbaceous plants (Hamrick et al. 1992). Alongside parentage analysis, which can estimate fine-scale gene flow (e.g., Moran & Clark 2011), quantification of pollen pools from local trees can also be used to estimate aspects of pollen movement (e.g., Smouse et al. 2001).

The mating system of a given species can have profound effects on the degree of SGS within and across populations. The mating system not only defines the mode of genetic transmission between generations, and thus levels of inbreeding, but in concert with levels of

dispersal can influence effective population sizes (the number of individuals from a Wright-Fisher population that would result in similar patterns of drift and/or genetic variation as that within a realized population), and the degree of subdivision due to genetic drift and selection (Schoen and Brown 1991; Holsinger 2000). Conifers often exhibit mixed mating systems and while outcrossing rates within coniferous species varies between populations, individuals, and between years (Mitton 1992), the majority of those species studied are outcrossed and have low levels of selfing (<10%, O'Connell 2003). At fine spatial scales, the degree of correlated paternity (i.e., the proportion of full sibs among maternal progeny) will further influence patterns of SGS, potential levels of kin competition, and inbreeding depression of subsequent generations (Hardy et al. 2004). For example, increased inbreeding rates within *Pinaceae* have been associated with reduced seed set and germination (Kärkkäinen et al. 1999). Deleterious alleles of large-effect have caused inbreeding depression in early life stages (e.g., embryo abortion) while those alleles of lesser effect tend to interact with quantitative characters such as growth and fecundity during maturity (Sorensen 2001).

Various aspects of the biology of a given species will ultimately impact the (effective) mating system as well. For instance, the density of mature conspecifics (in some cases influenced by silvicultural practices), manner of pollination (e.g., insect- or wind-dispersed, or iso- versus anisotropic dispersal patterns), phenological partitioning (e.g., flowering synchrony, seed drop), and phenotype (e.g., height, reproductive effort) will also affect mating patterns and SGS of trees across the landscape (El-Kassaby et al. 2003; Robledo-Arnuncio et al. 2004; Geremew et al. 2018). As pointed out by Kärkkäinen & Savolainen (1993), the degree of selfing in conifers can in some cases be prevented (though not completely) through spatial or temporal isolation of male and female function. Further, selfing should be more common in isolated trees where higher degrees of selfing are due to decreased concentration of pollen from other trees, and less common for trees with lower reproductive output (Kärkkäinen & Savolainen 1993). The findings from past studies are mixed, however, as there have been instances of sparse populations

exhibiting higher degrees of selfing (e.g., Farris & Mitton 1984; Knowles et al. 1987) as well as other instances where no pattern was found despite large differences in densities (e.g., Neale & Adams, 1985; Furnier & Adams, 1986) while relationships between outcrossing rates and pollen output have been weak (Shea 1987; Denti & Schoen 1988). In any case, such demographic patterns and life history consequences will ultimately influence standing genetic variation, and thus the potential for adaptation, or extirpation, of constituent populations.

Population size and density

The number of individuals in a population (i.e., the population size) will determine many aspects of evolutionary dynamics of a given species. Specifically, the effective population size (defined above) was introduced by Sewall Wright (1931; 1933; 1938; 1969) and extended by others to facilitate calculations for evolutionary rates of change caused by genetic drift (defined above; Charlesworth 2009). This distinction is important, as it was found very early on that census population sizes often differ, sometimes dramatically, from effective population sizes (Frankham 1995; Hough et al. 2013) such that evolutionary expectations will differ from reality if not taking these disparities into account. A number of factors will influence the effective population size, such as a disproportionate number of individuals between sexes, variability in reproductive output, non-random mating, the degree of overlap between generations, temporal changes in population size, unequal contributions from populations to the migrant pool, as well as both spatial and genetic structure of individuals across the landscape (Charlesworth 2009; Laporte & Charlesworth 2002). While estimates differ by species, many clades of trees exhibit large effective population sizes, which ultimately impacts the efficacy of selection (see Figure 2 in Hough et al. 2013).

With respect to tree species, the density of conspecifics and continuity of populations will also impact evolutionary outcomes. Indeed, many aspects of conifer biology are affected by their surrounding environment as well as the density of hetero- and conspecifics. For instance, outcrossing rates of conifer species are often tied to population density (Farris & Mitton 1984) and surrounding tree heights (O'Connell et al. 2003), while removal of proximal individuals can

increase pollen and gene flow distances by reducing potential mates and removing once impeding vegetation (Dyer & Sork 2001). Thus, disturbance, *sensu lato*, has the potential to alter contemporary demographic and reproductive dynamics through both direct (population-level) and indirect (ecological-level) impacts (Mouillot et al. 2013). For instance, stand density is positively correlated with outcrossing rate of ponderosa pine (*Pinus ponderosa*, Farris & Mitton 1984) and Scots pine (*P. sylvestris*, Robledo-Anuncio et al. 2004). Conversely, Sork & Smouse (2006) demonstrate that for many plant species habitat fragmentation and physical isolation do not impede (and can occasionally increase) pollen flow. Because pollen and seed dispersal decrease exponentially with distance, and the outcrossing rate is dependent upon genotypic patterns of conspecifics and the spatial density of community structure as a whole, there is strong evidence to suggest that alteration of the community through disturbance may change existing mating patterns.

Selection and Genetic Architecture

When selection pressure is strong, populations must adapt or migrate to avoid extirpation. The spatial extent and strength of selection will likely covary with both biotic and abiotic environmental variation, in some cases considerably. In cases where spatially varying selection causes local adaptation, local populations will have higher fitness than genotypes introduced from elsewhere (see Fig. 1a in Savolainen et al. 2013). While phenotypic differences can be indicative of spatially heterogeneous selection, in these cases serial colonization and other demographic causes (e.g., IBD) must be ruled out. For the genera considered for this dissertation, dissecting fitness into aspects of survival and reproduction can in some cases be difficult, given the long-lived nature of such species, as well as the representativeness of seedling phenotypes for total lifetime fitness. However, for many plant taxa, selection pressures are expected to be strongest for variation in survival during the juvenile stages of development (Donohue et al. 2010), particularly for those taxa with high reproductive output, as is the case for many tree species. As such, juvenile stages in plants have been found to contribute substantially to total lifetime fitness

(Postma and Ågren 2016). Phenotypic traits associated with juvenile survival have thus received the majority of genetic research focus in trees. Such studies have led to insights gained through a long history of common garden experimentation (Langlet 1971; Morgenstern 1996). For example, traits such as growth (e.g., height and diameter), form (e.g., specific gravity, straightness), phenology (e.g., bud flush, bud set), juvenile performance (e.g., germination rate, seed traits), and physiology (e.g., cold hardiness, water use efficiency) have all been shown to be under moderate to high genetic control (reviewed in Cornelius 1994; Howe et al. 2003; Alberto et al. 2013). Despite the vast majority of neutral variation remaining within populations (Howe et al. 2003; Neale and Savolainen 2004), quantitative genetic variation for these traits is often partitioned more so among populations than these neutral expectations (Lind et al. 2018; Chapter 6). Thus there exists ubiquitous evidence that many populations of trees are adapted to their environments (Savolainen et al. 2007; Alberto et al. 2013; Sork et al. 2013; Boshier et al. 2015; Prunier et al. 2015; Holliday et al. 2017) even across fine spatial scales where gene flow is expected to be a particularly prevalent driving force (e.g., Mitton 1989; 1999; Budde *et al.* 2014; Csilléry et al. 2014; Vizcaíno *et al.* 2014; Eckert *et al.* 2015; Holliday *et al.* 2016; Roschanski *et al.* 2016) providing further support that fine spatial scales are relevant to adaptation (Richardson *et al.* 2014). However, uncovering the genetic basis for these adaptive traits will often depend not only on available resources, but upon the neutral and non-neutral evolutionary history of the populations under consideration as well as the underlying genetic architectures of the dissected traits (i.e., the multi-dimensional relationship between genotype and phenotype through causative variants, their relative location within a genome, expression, pleiotropic effect, environmental influence, and degree of dominance, epistasis, and additivity). In Chapter 6, I review current theory regarding polygenic local adaptation and how past evolutionary history can have a profound effect on underlying genetic architectures. I use this synthesis to remark on the current state of the genomics of local adaptation in trees and identify available resources and the path forward for this field.

Summary

Throughout their evolutionary history, species of trees have occupied vast and heterogeneous landscapes (e.g., as evidenced from pollen and fossil records; Hewitt 1999; Willis & van Andel 2004). As such, heterogeneous selection has led to many instances of local adaptation (Hereford 2009; Boshier et al. 2015; Lind et al. 2018), where recent selection occurring over periods of post-glacial colonization has been a predominate influence of current spatial patterns of quantitative genetic differentiation (Savolainen et al. 2013). With this in mind, management of tree species across fine and broad spatial scales will need to consider the past evolutionary history of constituent species, and how natural and anthropogenic influences can drive future evolutionary dynamics. In this dissertation, I aim to take advantage of such circumstances to understand how natural and anthropogenic influence affects evolutionary processes in *Pinus* species native to California and Nevada, USA.

Objectives

The main aims of this dissertation were to understand how natural and anthropogenic forces can influence tree evolutionary dynamics, and gear results and conclusions to be used towards economic and conservation applications. With this in mind, three study systems were chosen, 1) *Pinus lambertiana* in the southern Sierra Nevada mountains of California, 2) *P. balfouriana* in the southern Sierra Nevada and Klamath Mountains of northern California, and, 3) *P. albicaulis* in the central Sierra Nevada mountains. The species used in this dissertation, while all with range extents within California and small portions of Nevada, represent a variety of the myriad forest ecosystems of the western United States, particularly those that receive considerable management attention, and thus have broad application of the results and conclusions to management considerations elsewhere. Further, the publications resulting from these studies exemplify the utility of modern genomic and analytical approaches to non-model organisms.

Chapter 2 – Effects of forest management on fine-scale gene flow in *Pinus lambertiana* Dougl.

Chapter 2 presents an assessment of how forest management affects evolutionary processes impacting fitness within fire-suppressed populations of a historically dominant pine species, sugar pine (*Pinus lambertiana* Dougl.). In order to understand impacts on the mating system and fine scale gene flow in this ecologically important species, we leverage the infrastructure of Teakettle Experimental Forest, a United States Forest Service site that applies treatments of prescription fire, a common fuel reduction technique, as well as their interaction to a fire-suppressed forest in the southern Sierra Nevada. In this study we addressed the following objectives:

- 1) Understand how forest management (fire, thinning, and their interaction) affected pollen and seed dispersal from within and out of sampled plots;
- 2) Describe how forest management affected spatial patterns of constituent trees;
- 3) Describe how forest management affected spatial genetic structure of constituent trees.

Chapter 3 – Local adaptation of *Pinus albicaulis* Engelm. across fine spatial scales

Past studies quantifying the spatial extent of adaptive variation in trees have historically been carried out at broad spatial scales that often represent the geographic ranges of study species. However, management agencies are often limited to a subset of the species' range. There is thus a disconnect between the information gained from such studies and the scale at which implications of their findings can be addressed and put into practice. In this chapter, and building from results from a common garden study, I leverage genetic, climatic, and edaphic data from natural stands of whitebark pine (*Pinus albicaulis* Engelm.) to understand how the environment can drive fine-scale local adaptation within the Lake Tahoe Basin Management Unit. The following questions were addressed in Chapter 3:

- 1) Is there evidence that a long-lived, outcrossing plant species exhibiting high levels of gene flow can be locally adapted across fine spatial scales?

- 2) What is the genetic basis and relationship among loci underlying adaptation in such a species?
- 3) How similar are the genetic bases of fitness-related phenotypes to the loci putatively under selection from the environment?

Chapter 4 - Local adaptation and linkage maps I: A first step towards describing genetic architectures of complex traits in natural populations

For non-model organisms, genomic resources needed to fully describe the genetic architecture of complex traits can face many hurdles. Because of their large genome size, conifers face many challenges in this regard, particularly with respect to the production of physical sequence maps with ordered genetic markers. Such resources can be used to test fundamental theory regarding the expected genetic architecture underlying adaptive traits under various demographic and evolutionary scenarios. To produce such resources, linkage maps are often used to both order and elongate blocks of sequence data. Across taxa, such maps were predominately made from well understood familial relationships or mating designs, but maps of single trees began increasing in numbers in the 1990s. In conifers, maternally derived haploid tissue (i.e., the megagametophyte) can be used to order markers within individual trees within a fairly straightforward framework. This chapter is a first in a series that aims to understand and test fundamental theories of local adaptation in a non-model conifer, *Pinus balfouriana* Grev. & Balf. Using individuals collected from populations that historically exhibited gene flow but have subsequently become isolated from one another since the last glacial maximum, we address the following objective:

- 1) Construct a dense linkage map using next-generation sequencing technology in a non-model organism across both Klamath and Sierra Nevada populations

Chapter 5 - The genetic architecture of local adaptation II: The QTL landscape of water-use efficiency for foxtail pine (*Pinus balfouriana* Grev. & Balf.)

This chapter builds upon Chapter 4 and begins to dissect the genetic architecture of water-use efficiency (WUE) in foxtail pine. Using niche modeling and indirect measures of WUE, the ratio of carbon isotopes fixed during photosynthesis ($\delta^{13}\text{C}$), as well as nutrient utilization, the ratio of nitrogen isotopes fixed during leaf-level resource utilization (μgN), we address the following objectives:

- 1) Explore the importance of water availability as a determinant of geographical range of foxtail pine,
- 2) Investigate the degree to which variation in $\delta^{13}\text{C}$ and μgN is genetically based, differentiated among populations, and underlain by a genetic architecture with large effect loci

Chapter 6 – The current state and future directions of forest tree genomics

Knowledge of the loci underlying quantitative traits will benefit both conservation and industrial endeavors. Specifically, quantifying geographic variation of adaptive loci will inform planting and assisted migration efforts while the identification of these loci can be used in predictive breeding techniques. As such, there is great interest within the forest genetics community to identify the molecular underpinnings of traits important to adaptation and commercial enterprise. In this chapter I address several objectives: First I review the common garden approach and how it can be used to build evidence for local adaptation, as well as both heritable and quantitative genetic variation. Next, I review evolutionary theory and summarize the expectations of allele frequencies, effect size, and genomic organization of adaptive traits as it relates to gene action, negative selection, positive selection, and gene flow. I then contextualize the methods and results from genotype-phenotype associations with these expectations and conclude with recommendations for the field to most efficiently advance our understanding of the genomics of quantitative traits in trees.

Chapter 2.

Effects of forest management on fine-scale gene flow in *Pinus lambertiana*

Abstract

Frequent, low-severity fires in historic western North American forests were once a major driver of ecological patterns and processes, creating resilient systems dominated by widely spaced pine species. However, a century of fire-suppression has caused overcrowding, altering forest composition to shade-tolerant species while increasing competition, leaving trees stressed and susceptible to pathogens, insects, and high-severity fire. Exacerbating the issue, the probability of fire incidence is expected to increase with changing climate, while fire season has been observed to begin earlier and last longer than historic trends. Consequently, forest thinning has been identified as the primary management tool to mitigate these risks. Yet little is known of how thinning, fire, or their inevitable interaction affect evolutionary processes of constituent pine species, specifically regarding processes that influence fitness and play an important role in the opportunity for selection and population persistence. Here we assess management impact on fine-scale gene flow in a historically dominant and ecologically important shade-intolerant pine species in the Sierra Nevada, *Pinus lambertiana* Dougl. We find that treatment prescription differentially affects fine-scale genetic structure and effective gene flow in this species. Specifically, thinning prescriptions in isolation increases relative genetic structure between adults and seedlings, and that seed and pollen dispersal respectively increase and decrease as a function of increasing disturbance intensity. It is likely that these disequilibrated systems will continue to develop with time since a disturbance event, and that future dynamics of such systems should be continually monitored to ensure effective management.

Introduction

Many aspects of conifer biology are affected by their surrounding environment as well as the density of hetero- and conspecifics. For instance, outcrossing rates of conifer species are often tied to population density (Farris & Mitton 1984) and surrounding tree heights (O'Connell *et al.* 2003), while removal of proximal individuals can increase pollen and gene flow distances by reducing potential mates and removing once impeding vegetation. Thus, disturbance, *sensu lato*, has the potential to alter contemporary demographic and reproductive dynamics through both direct (population-level) and indirect (ecological-level) impacts (Mouillot *et al.* 2013).

For many historic forests of the western United States, natural disturbances such as fire were commonplace and equilibrated many ecosystem functions and processes (Covington *et al.* 1994). Fire regimes in these regions had return intervals on decadal scales (10-17 years), in contrast to wetter climates where fire return intervals were (sub)centennial (50+ years; (North *et al.* 2005). Resultantly, these ecosystems experienced frequent, low-severity burns and were populated by fire-adapted genera, equilibrating forests dominated by resilient, widely spaced trees of various pine species. Yet over the past 150 years, anthropogenic influence has resulted in forests that are now fire-suppressed and overgrown by shade-tolerant species, causing increased competition, leaving trees stressed and susceptible to fungal and bark beetle attacks (Bonello *et al.* 2006).

Stand densification has also increased the frequency and probability of contemporary, high-severity fires. Between 2012 and 2014 in California alone, 14,340 fires burned 1.1 million acres and injured or killed nearly 300 individuals (NIFC 2014). Collectively, fires across California, the Great Basin, Southwest, and Rocky Mountain territories have burned a combined 8.8 million acres between 2014 and 2015 (NIFC 2015), while Forest Service scientists predict future fires to reach unprecedented levels, covering over 12-15 million acres annually (USDA Forest Service 2016a) requiring the United States Forest Service (USFS) to budget \$2,300,000,000 on wildfire management, suppression, and preparedness for the 2016 fiscal year (USDA Forest Service

2016b). Exacerbating the issue, analyses of fire season length and onset have shown that seasons are beginning earlier and lasting longer than historic trends (Westerling 2006) while climate models predict extreme weather favorable to fire to become more frequent, and ignited fires to increase in severity, size, and required suppression efforts (Miller *et al.* 2009).

Because of these contemporaneous trends, large-scale legislative efforts of forest thinning have been implemented to simultaneously restore fire-frequent ecosystems to their pre-settlement resilience as well as to protect urban development and human life, as fuel reduction treatments of this type have been shown to be an effective tool in decreasing fire severity and ignition probability (Agee & Skinner 2005; Schwilk *et al.* 2009; Safford *et al.* 2009). For example, the Sierra Nevada Forest Plan Amendment (USDA Forest Service 2004) mandates that 50% of initial thinning treatments take place near urban populations, while the remaining thinning take place in natural wildland stands. To encourage fire resiliency the USFS has implemented fuel reduction treatments across 6.1 million acres of western, fire-suppressed forestland in 2014 (USDA Forest Service 2016a) while foresters are calling for an overhaul of management policy to implement these thinning treatments to a far greater extent (North *et al.* 2015). While congruent with historic forest structure, these actions will orient these already disequibrated systems on trajectories of unknown evolutionary consequence.

Through timber harvests, land use conversion, and fire suppression, forests have undergone systemic shifts in composition, structure, and disturbance regimes that are incongruous to the natural and evolutionary histories of the inhabitant species (Collins *et al.* 2011; Larson & Churchill 2012). Consequentially, anthropogenic forest disturbance has been at the forefront of conservation attention for decades (Ledig 1988; 1992). The extent of human impact on forested land has received particular attention as a result of the empirical expectations developed from population genetic theory. Specifically, because of the reduction in individual tree density overall, and in particular for larger trees that asymmetrically contribute gametes to reproduction (Richardson *et al.* 2014), harvested forests are thought to be specifically subjected

to population bottlenecks, potentially altering existing mating systems or available gene pools while decreasing genetic variability within populations and increasing differentiation from native stands (Cloutier *et al.* 2006; Kramer *et al.* 2008; Lowe *et al.* 2015). These consequences can have considerable influence on the fitness of affected populations, as inbreeding depression can have deleterious effects on growth and reproductive potential while drastic changes in gene pool availability or mating system can alter a population's potential to adapt to local conditions.

Past studies investigating the genetic effects of North American forest management show mixed evidence of harvest influence. These studies sub-sample populations and primarily focus on diversity consequences across a range of molecular markers (often microsatellites). Nearly ubiquitously, management studies of North American conifers compare genotypic diversity indices (e.g., H_E , H_O , allelic richness, etc.) between treatments to detect management influence (Cheliak *et al.* 1988; Gömöry 1992; Buchert *et al.* 1997; Adams *et al.* 1998; Rajora *et al.* 2000; Macdonald *et al.* 2001; Perry & Bousquet 2001; Rajora & Pluhar 2003; El-Kassaby *et al.* 2003; Marquardt *et al.* 2007; Fageria & Rajora 2013a; b). However, the same diversity values can manifest under completely different scenarios and tests of significance between population values for a small number of markers may therefore be under-informative, particularly for sub-sampled populations, as these differences can result from sampling bias or neutral evolutionary processes unrelated to management. Additionally, these investigations also often employ F_{ST} analyses to assess statistical significance between treated and untreated stands (Thomas *et al.* 1999; Perry & Bousquet 2001; Marquardt *et al.* 2007; Fageria & Rajora 2013a; b). Though when used in this context, this test is simply signifying whether the allelic frequencies in (sub)populations under study are likely to have been sampled from the same ancestral population (Holsinger & Weir 2009). Very often, the treated and untreated stands are physically adjacent (derived of a common ancestral population) and only under extreme perturbation should significance be expected. In cases where significance is detected, and other than to assess relative diversity between stands, such differentiation does little to inform how management is affecting ongoing evolutionary

processes affecting fitness as such processes may ameliorate bottlenecks due to management. It would therefore be difficult to draw such conclusions without assessing other processes.

Very seldom in North American studies of forest management are evolutionary processes influencing fitness specifically examined (but see Neale & Adams 1985). As such, and despite generally nonsignificant findings, authors caution interpretation (Finkeldey & Ziehe 2004; Namroud *et al.* 2012). Very often the scale of sampling (both in terms of numbers of individuals and the degree of temporal variation), as well as the lack of investigation into evolutionary dynamics have been offered as inadequate, and that further investigation into evolutionary consequences of natural and anthropogenic disturbance could give valuable insight to forest managers and fill a vital knowledge gap in this regard (Namroud *et al.* 2012). Indeed, incongruence between theoretical predictions and empirical results from studies evaluating genetic consequences of forest disturbance has created a paradox within the literature (Kramer *et al.* 2008). Yet as Lowe *et al.* (2015) point out, we may have been looking in the wrong place. They argue that instead of simply assaying mature cohorts to understand the genetic consequences of disturbance, future attention should include progeny arrays as well as the relative regenerative success across a wide range of influences. Additionally, they contend that the type and magnitude of the genetic response itself may be better understood through the variation in mating and breeding systems of studied species. Of particular importance, Lowe *et al.* (2015) advise scientists that the most fruitful research endeavors will incorporate quantitative approaches to understanding evolutionary mechanisms, specifically those connecting changes in pollination to mating systems and evolutionary fitness, and that these efforts will likely generate critical knowledge regarding the mechanisms driving the dynamics we observe.

Given the evidence for the efficacy of thinning treatments' effect in reducing fire severity and ignition potential (Agee & Skinner 2005; Schwilk *et al.* 2009; Safford *et al.* 2009), the USFS's continued commitment to implementing these thinning treatments to a far greater extent (North *et al.* 2015; USDA Forest Service 2016a), as well as forecasted inevitability of widespread, large-

scale future fire incidence in the western United States (Westerling 2006), interactions between fire and forest thinning management are certain. To ensure forests are resilient to frequent fire and disturbance, and to ensure habitat for public recreation and native wildlife, the interactive impact of management and fire must be understood in an evolutionary framework. Here, we investigated the evolutionary impact of forest management on fire-suppressed populations of sugar pine (*Pinus lambertiana* Dougl.) within Teakettle Experimental Forest (TEF), a USFS site located in the central Sierra Nevada of California. Using microsatellite markers, we employ parentage analysis and assess impact upon various processes known to affect fitness such as the mating patterns, effective dispersal distances, and fine-scale genetic structure.

Methods

Study area, sampling, and focal species

Teakettle Experimental Forest (TEF) is a fire-suppressed, old-growth forest watershed in the central Sierra Nevada mountains of California. The 1300-ha forest ranges from 1900–2600m in elevation and consists of five conifer species representative of fire-suppressed forests of the Sierra range: white fir (*Abies concolor* [Gordon] Lindley ex Hildebrand), red fir (*A. magnifica* A. Murray), incense cedar (*Calocedrus decurrens* [Torr] Florin), Jeffrey pine (*Pinus jeffreyi* Balf.), and sugar pine (*P. lambertiana*). Historically, fire burned the area every 11-17 years, but had been suppressed for 135 years (North et al. 2005) while logging had been completely absent within the watershed (North 2002). Six treatments were applied to neighboring 4-ha plots at TEF (each 200m x 200m, Figure 2.1a) by crossing two levels of burn (no-fire and fire) with three levels of thinning (no-thinning, overstory-thinning, and understory-thinning). The understory thinning prescription followed guidelines in the California spotted owl (CASPO) report (Verner et al. 1992) which is now widely used for fuel management in California (SNFPA 2004). Each treatment was replicated three times for a total of 18 plots covering 72ha. Understory-thinning removes all trees with a diameter at breast height (DBH) ≥ 76 cm and ≤ 25 cm, while overstory-thinning removes all trees > 30 cm DBH except 18-22 of the largest trees per hectare. Treatments were applied to the

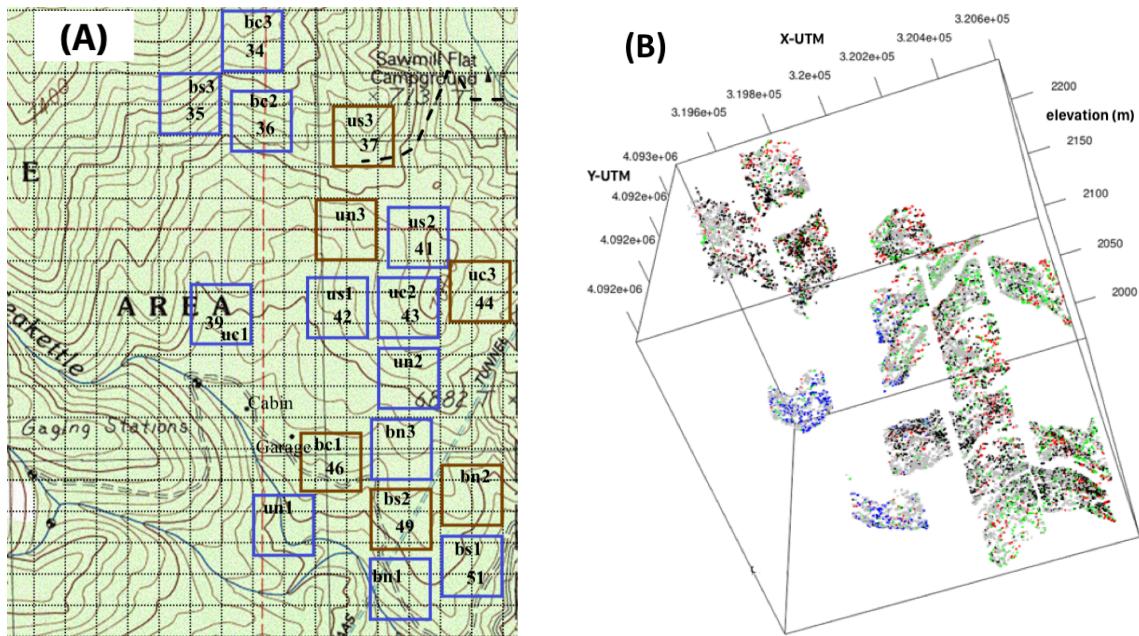


Figure 2.1 Teakettle Experimental Forest, California (Latitude: 36.9606, Longitude: -119.0258). (A) Spatial arrangement of treatments (BC = burned understory thin; BN = burned no-thin; BS = burned shelterwood thin; UC = unburned understory thin; UN = unburned no-thin; US = unburned shelterwood thin). Replicates for each treatment are numbered one through three from south to north. (B) Mapped coordinates (Universal Transverse Mercator) and elevation (meters) of pre-treatment adults ≥ 5 cm diameter at breast height. (green: *P. lambertiana*, red: *P. jeffreyi*, gray: *A. concolor*, blue: *A. magnifica*, orange: *C. decurrens*, black: *Quercus*, *Salix*, and remaining species.).

watershed over 2000 and 2001. Plot inventories of pre-treatment (1999), and post-treatment (2001, 2004, and 2011) conditions mapped individual trees on a 3D coordinate system that was then translated into Universal Transverse Mercator (colored dots, Figure 2.1b). Only standing boles ≥ 5 cm DBH were included in plot inventories, which recorded species, DBH, spatial coordinates, and decay class. Post-treatment inventories updated DBH, decay class, and added individuals to the dataset once they reached 5cm DBH. Here, seedling and saplings are all pine stems < 5 cm DBH. For these, basal diameter and spatial coordinates were recorded over the summers of 2012 and 2013 while collecting needle tissue samples from the full census of all live *P. lambertiana* within plots across all treatments ($N = 3,135$). *Pinus lambertiana* is a historically dominant member of mixed-conifer forests of the Sierra Nevada, and continues to play important ecological roles for numerous taxa. This species is shade-intolerant and is an important focus of restoration of forest resilience to pre-settlement conditions of frequent fire.

Analysis of tree spatial structure

Using plot-level *P. lambertiana* individuals, we estimated spatial structure of seedlings and adults across 10-meter distance classes, r , separately using univariate inhomogeneous pair correlation functions ($g_{inhom}(r)$) from the `spatstat` library in R (Baddeley et al. 2015) using an isotropic edge correction. This statistic was chosen over Ripley's K, or its linearized version (L) because of advocacy for $g_{inhom}(r)$ over these statistics (see `spatstat` manual). This analysis tests the null hypothesis that the 2D spatial arrangement of points (adults or seedlings) is not significantly different from complete spatial randomness (CSR; i.e., a Poisson distribution of inter-point distances with inhomogeneous intensities of points). We calculated null confidence envelopes for each test using 199 null simulations of CSR using the same intensity of the pattern of individuals analyzed. For trees that coincide with the null model of CSR $g_{inhom}(r) = 1$, with spatial aggregation resulting in $g_{inhom}(r) > 1$, and with spatial inhibition resulting in $g_{inhom}(r) < 1$ (Baddeley et al. 2015); significance was judged using the null confidence envelopes described above. We also repeated this analysis for shade-tolerant individuals (i.e., *A. concolor* and *A. magnifica* grouped together). Further, we extended the univariate inhomogeneous pair correlation function to its bivariate equivalent, $g_{inhom,i,j}(r)$, to test for spatial affinity between two groups i and j , using similar methods as above for edge correction and null confidence envelopes. We calculated $g_{inhom,i,j}(r)$ between unique combinations of *P. lambertiana* adults, *P. lambertiana* seedlings, and shade-tolerant individuals. Hypothesis testing and interpretation of bivariate $g_{inhom,i,j}(r)$ was carried out as with univariate $g_{inhom}(r)$. Results from these analyses will allow us to compare standing spatial structure of trees against spatial genetic autocorrelation (see below) and make inferences about ecology of these species as well as how treatments at TEF are affecting ongoing evolutionary dynamics.

DNA extraction, microsatellite amplification

Total genomic DNA was extracted according to manufacturer's protocol using the DNeasy 96 Plant Kit (Qiagen, Germantown, MD) from finely ground *P. lambertiana* samples within a

subset of the factorial treatments at TEF: unburned-no-thin control plots (hereafter UN), understory-thin (CASPO) plots without burn application (hereafter UC), and burned understory-thin plots (hereafter BC) for a total of 1,348 individuals. For each individual, three chloroplast (paternally inherited, Wofford *et al.* 2014: pt71936, pt87268, pc10) and four nuclear (biparental inheritance, Echt *et al.* 1996: rps50, rps02, rps12, rps39) microsatellite markers were amplified (using fluorescent dyes NED, PET, VIC, and FAM) per the original publications with minor modifications using BIO-RAD iProof high fidelity DNA polymerase. Specifically, for all chloroplast markers (Wofford *et al.* 2014) we ran an initial denaturation step of 98°C for 30s, where the next 36 cycles consisting of a denaturation step of 10s at 98°C, an annealing step at 58°C for 30s, and an extension step at 72°C for 30s, followed by a final extension step at 72°C for 10mins. For nuclear markers RPS12 and RPS39 (Echt *et al.* 1996), we ran an initial denaturation step of 98°C for 30s followed by eight cycles of a denaturation (98°C for 10s), annealing (60°C for 30s) and an extension (72°C for 30s), which were subsequently followed by six cycles of denaturation (98°C for 10s), annealing (60°C for 30s; which decreased 0.5°C every step until reaching 57°C), and extension (72°C for 30s), with a final extension of 72°C for 10mins. For nuclear marker RPS50 (Echt *et al.* 1996), this procedure was similar to that of RPS12 and RPS50 except that the second cycles cycled 20 times and the annealing step decreased 0.5°C until reaching 50°C. For RPS02 (Echt *et al.* 1996) we ran the second set of cycles for 10 cycles where the annealing step decreased 0.5°C until reaching 55°C. The chloroplast markers were chosen for their primer conservation across *Pinus*, *Trifoliae*, *Parrya*, and *Quinquifolia* subsections of the *Pinus* genus (Wofford *et al.* 2014) while the chosen nuclear markers have been amplified in eastern white pine (*P. strobus* L., Echt *et al.* 1996) and both sets successfully amplified on a subset of individuals at TEF as a proof-of-concept judged by gel electrophoresis. Multiplexed individuals (one fluorescent dye per well) were analyzed using the Applied Biosystems 3730xl fragment analyzer at Cornell University (<http://www.biotech.cornell.edu/brc/genomics-facility>) and genotypes were called using

GeneMaker v2.6.7 (see Supplemental Information XXXX; <http://www.softgenetics.com/GeneMarker.php>).

Genetic diversity measures

Treatment-specific diversity measures (total number of alleles, A_T ; mean number of alleles per locus, A ; effective number of alleles per locus, A_e ; observed and expected heterozygosity for nuclear markers, respectfully H_o , H_e ; average number of private alleles, A_P ; and overall means for each category) were calculated for each treatment and averaged across loci in order to compare dynamics at TEF to studies within the literature. For estimates of H_o and H_e , only nuclear markers were used. To quantify variation in these measures we also report standard deviation. We also calculated hierarchical multi-locus F_{ST} according to Weir & Cockerham (1984) for nuclear markers using the `hierfstat` package in R (Goudet & Jombart 2015) and calculated treatment-specific F_{ST} in a similar manner in order to compare fixation indices across treatments. Further, single- and multi-locus exclusion probabilities for parentage analysis (see below) were calculated using python scripts modified from `gstudio` (v1.5.0; Dyer 2016).

Analysis of spatial genetic structure

To quantify spatial genetic autocorrelation at a distance class (lag) h (hereafter r_g^h), we used functions from the `PopGenReport` package in R (Adamack & Gruber 2014). Specifically, using multi-locus genetic distances (Smouse & Peakall 1999) and Euclidean geographic distances among spatial coordinates of individuals, we calculated r_g^h across distances classes h corresponding to approximately 10-meter bins for *P. lambertiana* seedlings, *P. lambertiana* adults, as well as a bivariate approximation for the clustering of *P. lambertiana* adult genotypes to those of seedlings. For a given distance class, h , spatial patterning of multi-locus genotypes are unrelated to (i.e., random relative to) the spatial patterns of individuals if $r_g^h = 0$, aggregated if $r_g^h > 0$, and dispersed if $r_g^h < 0$. We estimated null confidence intervals by taking the 2.5th and 97.5th quantiles of $M = 1000$ estimates of $r_{g,m}^h$, where 999 of these estimates were computed by randomly

permuting individual genotypes across empirical spatial coordinates, with the M^{th} permutation being the empirical estimate of r_g^h itself (Smouse & Peakall 1999). We created correlograms for nuclear and chloroplast markers both in isolation and in combination, but present only those using full genotypes as correlograms by marker type showed similar patterns as full genotypes. We used these correlograms to quantify spatial aggregation of genotypes so that conclusions based on treatment effects could be compared and contextualized with ongoing evolutionary dynamics at TEF.

Parentage analysis

To quantify fine-scale gene flow at TEF, we conducted parentage analysis using our genetic markers. Joint estimation of parentage and dispersal parameters (i.e., mean dispersal distances of seed and pollen) were achieved by expanding methods of Moran & Clark (2011). This method simultaneously estimates parentage and mean dispersal distances for seed and pollen within a Bayesian framework, taking into account genotyping error and variation in individual fecundity while treating dispersal processes inside and outside of the mapped areas in a coherent manner, which is critical if the dispersal kernel is to reflect both long- and short-distance movement. Here, all sampled adults are characterized by a multi-locus genotype and a mapped coordinate. Additionally, there exists a sample of seedlings, each of which has not only a genotype and location, but an estimated pedigree as well, which can consider any adult as either mother or father, or of a selfing event (though we excluded possible selfing events from analyses). The probability of the pedigree considering two in-plot parents, before incorporating information regarding genotype, is estimated from the probability of pollen to mother movement over the given distance and of seed movement over the distance between mother and seedling, as well as the parental prior distribution for fecundity and pollen production. For the study here, pollen production was considered proportional to fecundity (as in Moran & Clark 2011) and was estimated by fitting a 2nd-order power polynomial regression to data from Figure 6 in Fowells & Schubert (1956) where Cone Count = $0.0098(\text{dbh}^2) - 0.4811(\text{dbh}) + 10.651$. After calculating cone

counts using this regression, we set fecundity for all adults <25cm DBH to zero given observed cone counts from Fowells & Schubert (1956). For dispersal priors, we set the seed dispersal kernel shape parameter, u_s , to 253.31, corresponding to a mean dispersal distance of 25m (Millar et al. 1992; Fowells & Schubert 1956) while the pollen dispersal kernel shape parameter prior, u_p , was set to 2279.72, corresponding to a mean pollen distance of 75m (Wright 1976; Neale 1983; Millar et al. 1992). For priors to the standard deviation of mean dispersal we set seed (pollen) to 1013.21 (9118.90) corresponding to standard deviations of 50m (75m).

Given that either parent could have produced the offspring the likelihood that this pair is the true parents relative to all other possible parent pairs depends on the dispersal kernel priors for seed and pollen, and the seed and pollen production of all trees both inside and outside of the plot (the fraction of all possibilities; Moran & Clark 2011). To evaluate the probability of an offspring having one parent in the plot and the other outside of the plot, a set of potential out-of-plot parent-densities, dp_1, \dots, dp_{20} , each 10m progressively outside of the plot is considered (see supplemental figure S3.1 in Moran & Clark 2011). Pollen and seed movement into the plot is approximated by assuming first that all seed/pollen produced within each quarter-polygon, v , originates from a tree located dp_v meters from the midpoint of each side outside of the plot. The expected out-of-plot pollen (seeds) reaching an in-plot mother (a seedling's location) from each quarter-polygon outside of the plot is calculated based on the average density and average fecundities of trees outside of the plot and then multiplied by the probability of dispersal to the point within the plot. Summing over each distance class over each side gives the total expected out-of-plot pollen/seed dispersal to points inside of the plot. However, to calculate the probability of an in-plot versus an out-of-plot father, the expected pollen arriving at an out-of-plot mother from another out-of-plot father must first be calculated using the concentric circles around the sampled plot and the distance classes described above. The fraction of rings falling outside the plot determines the fraction of pollen received from each distance class, dp_v , expected to come from

outside trees. Once error rates (e_1) and dropout rates (e_2) of genotyping are calculated through regenotyping individuals (see supplemental information), the probability of a pedigree, seed and dispersal parameters given the offspring genotype, distances, error rates, and pollen/seed production can be estimated (Moran & Clark 2011). Very rarely have previous studies investigating effects of forest management (or using parentage analysis towards such goals) incorporated error and dropout rates into subsequent inferences.

For the current study, out-of-plot densities were extrapolated for each side of the six plots used at Teakettle from densities and DBH distributions (our proxy for fecundity) revealed in pre-treatment surveys (North 2002). Due to the proximity of the treated plots, all adult trees and seedlings across UC, BC, and UN treatments were considered simultaneously for parentage assignment. Our methods therefore extend Moran & Clark (2011) from a single plot of sampled individuals to multiple plots across the landscape by accounting for out-of-plot polygonal boundaries (distance classes) that would have overlapped existing plots, instead using the standing structure of neighboring plots in estimating probabilities of parentage and excluding area outside of a given plot that overlaps any other plot (and is thus probabilistically considered via existing genotypes). Additionally, instead of considering any given pedigree as symmetrical (i.e., with no consideration for which tree was the pollen or seed donor) we utilize genotyped markers separately to consider whether a given pedigree is for a mother-father pair, or for a father-mother pair (i.e., we only considered nuclear markers for a potential mother, and all markers for a potential father). The most probable pedigree for each seedling was identified by assessing the proportion of the proposed pedigree across chains in the Gibbs sampler (as in Moran & Clark 2011). This method was further modified to improve computational efficiency by multiprocessing appropriate elements of the script by utilizing custom python scripts run on the VCU Center for High Performance Computing cluster (CHiP) and the `SNOW` library (v0.4-2; Tierney et al. 2016) in R (v3.3.3; R Core Team 2017).

Using parentage analysis to further quantify fine-scale gene flow

In addition to estimates of the mean seed and pollen dispersal from parentage analysis (see above), we used these parentage assignments to further classify fine-scale gene flow at TEF. Using the full set of most probable pedigrees identified from parentage analysis, we first quantified the number of in-plot vs. out-of-plot dispersal events averaged across each treatment replicate. Then, using the most probable parentage assignment for each offspring, we quantified mean dispersal distances from sampled mothers to seedlings, as well as between sampled fathers to sampled mothers. To better account for uncertainty in parentage assignment (i.e., to account for fractional parentage assignment), we also calculated mean dispersal distance by treatment by considering all pedigrees with known individuals weighted by the probability of assignment. Specifically, for mean seed dispersal, for each seedling we calculated the weighted average of mother-offspring distances across pedigrees of non-zero probability that included known mothers. Here, each weight was the probability of assignment, $p_{seed,pedigree}$, divided by the probability of assignment of this seedling to a known mother ($1 - U_M$) where U_M is the sum of the probabilities across all non-zero pedigrees that included an unsampled mother. Treatment-level averages were then calculated across these weighted distances. For pollen dispersal, for each seedling we considered only pedigrees of non-zero probability where both the mother and father were known, weighting each distance by the probability of assignment, $p_{seed,pedigree}$, divided by the probability of assignment to known parents ($1 - U_{seed,pedigree}$) where $U_{seed,pedigree}$ is the sum of the probabilities across all non-zero pedigrees that included at least one unsampled parent. Treatment-level averages were then calculated from these weighted distances.

Scripts used in analyses described above can be found in IPython notebook format (Pérez & Granger 2007) at <https://github.com/brandonlind/teakettle>.

Results

Analysis of tree spatial structure

Univariate Analysis

Across treatments, *P. lambertiana* adults generally exhibited spatial aggregation at distance classes less than 20 meters, where this signal decreased with increasing disturbance intensity with UN plots showing the greater magnitudes of $g_{\text{inhom}}(r)$ than UC or BC plots at these small distance classes (Figure 2.2). For adult shade-tolerant species (*A. magnifica* and *A. concolor* combined), the extent of spatial aggregation at large distance classes decayed with increasing disturbance intensity (Figure 2.3) where UC treatments generally exhibited greater magnitudes of $g_{\text{inhom}}(r)$ than BC treatments in small distance classes. For *P. lambertiana* seedlings, UN treatments generally had significant aggregation and much larger magnitudes of $g_{\text{inhom}}(r)$ at larger distance classes than other treatments, while seedlings in BC treatments exhibited greater magnitudes of $g_{\text{inhom}}(r)$ across small distance classes than either UC or UN plots (Figure 2.4).

Bivariate Analysis

The spatial affinity of *P. lambertiana* seedlings to *P. lambertiana* adults, $g_{\text{inhom,seedling,adult}}(r)$, generally tended from randomness ($g_{\text{inhom,seedling,adult}}(r) = 1$) to spatial inhibition ($g_{\text{inhom,seedling,adult}}(r) < 1$) with decreasing intensity of disturbance (i.e., from undisturbed UN plots, to thin-only UC plots, to thinned-and-burned BC plots). UN plots tended to show consistent inhibition across distance classes greater than about 15m, whereas observed $g_{\text{inhom,seedling,adult}}(r)$ for UC plots tended to align with the lower extent of the confidence interval with fewer instances of significant inhibition between adult and seedlings (Figure 2.5). A similar trend for increasing spatial inhibition between *P. lambertiana* seedlings and shade-tolerant adults ($g_{\text{inhom,seedling,adult}}(r)$), as well as for *P. lambertiana* adults and shade-tolerant adults

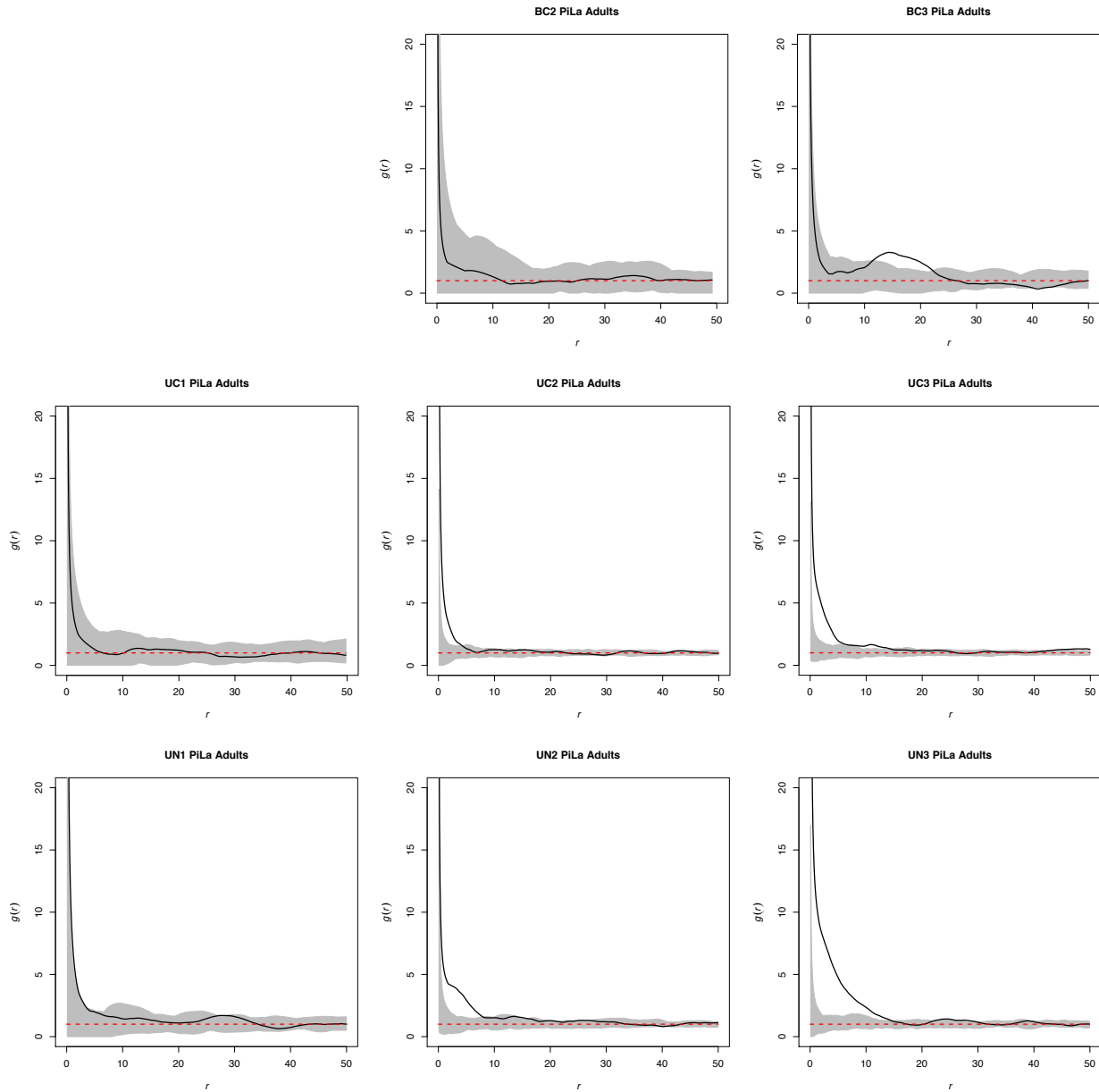


Figure 2.2 Univariate analysis of adult *P. lambertiana* (PiLa) spatial structure, $g_{inhom}(r)$, by treatment replicate for each distance class, r . Gray : null confidence envelope; Solid black line : observed $g_{inhom}(r)$. Red dashed line : null expectation of complete spatial randomness, $g_{inhom}(r) = 1$. Individuals are aggregated if $g_{inhom}(r) > 1$, inhibited if $g_{inhom}(r) < 1$. BC1 was below the threshold sample size allowed by spatstat.

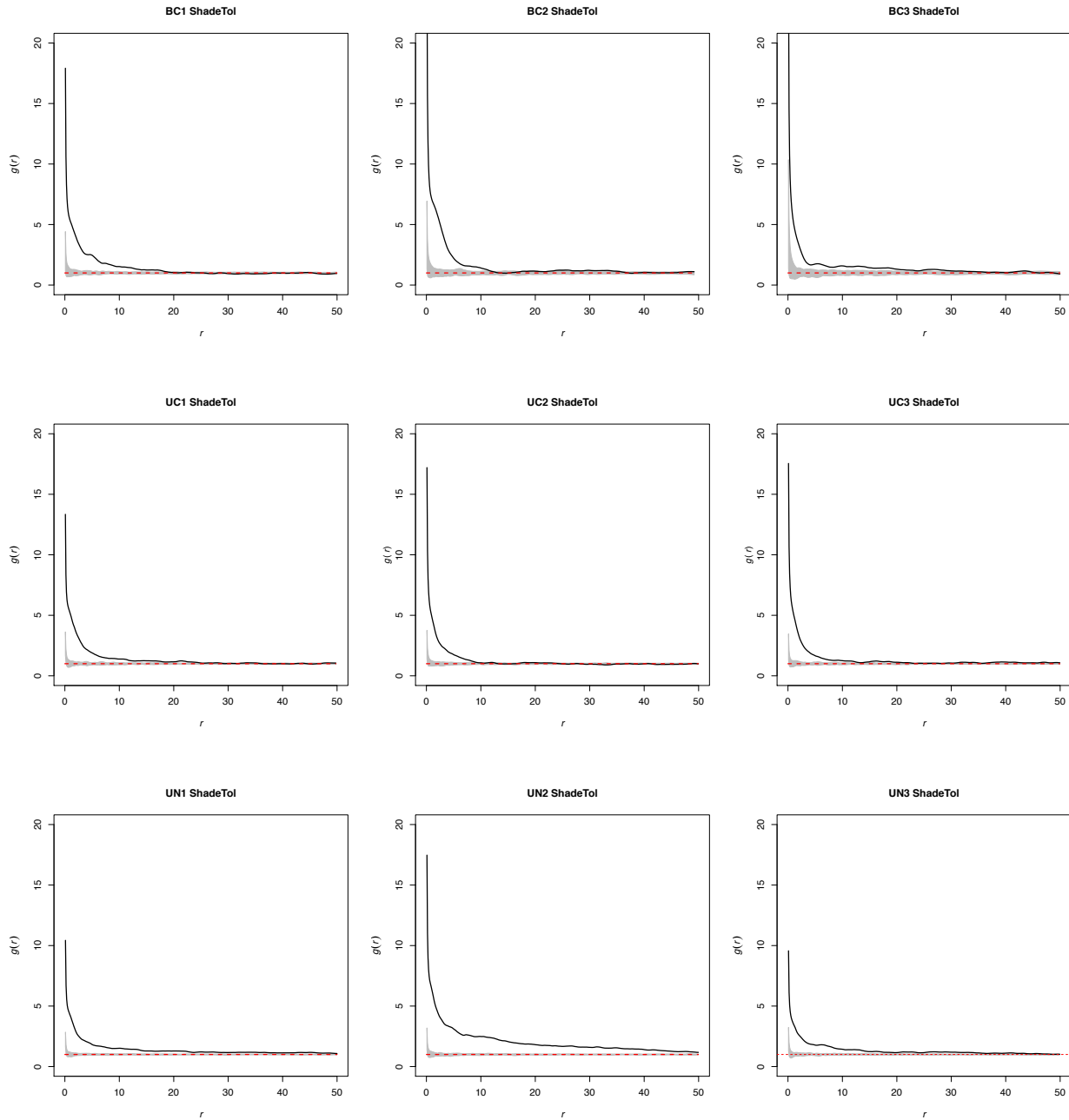


Figure 2.3 Univariate analysis of adult shade tolerant (*A. magnifica* and *A. concolor*; ShadeTol) spatial structure, $g_{inhom}(r)$, by treatment replicate for each distance class, r . Gray : null confidence envelope; Solid black line : observed $g_{inhom}(r)$. Red dashed line : null expectation of complete spatial randomness, $g_{inhom}(r) = 1$. Individuals are aggregated if $g_{inhom}(r) > 1$, inhibited if $g_{inhom}(r) < 1$.

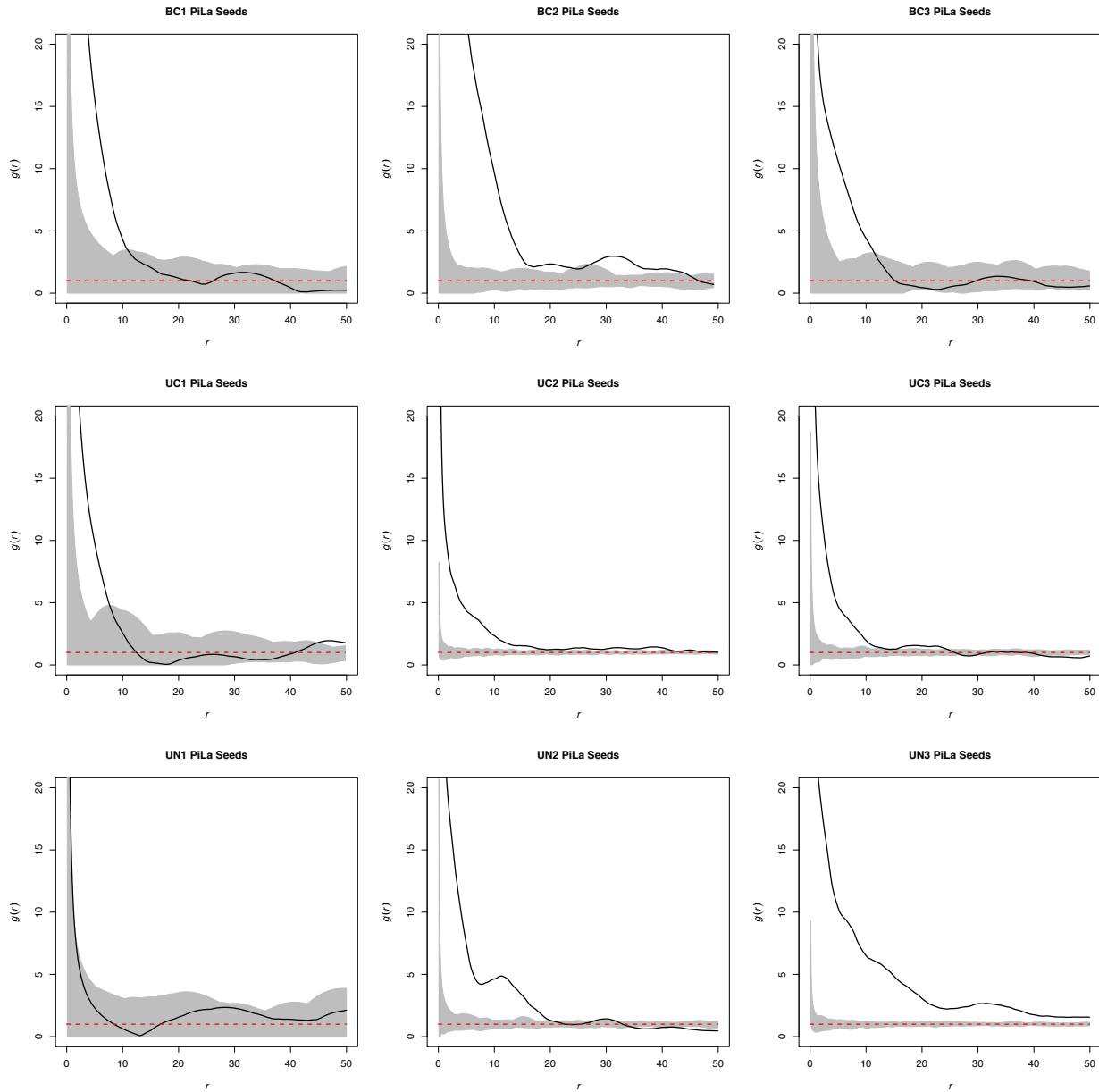


Figure 2.4 Univariate analysis of *P. lambertiana* (PiLa) seedling spatial structure, $g_{inhom}(r)$, by treatment replicate for each distance class, r . Gray : null confidence envelope; Solid black line : observed $g_{inhom}(r)$. Red dashed line : null expectation of complete spatial randomness, $g_{inhom}(r) = 1$. Individuals are aggregated if $g_{inhom}(r) > 1$, inhibited if $g_{inhom}(r) < 1$.

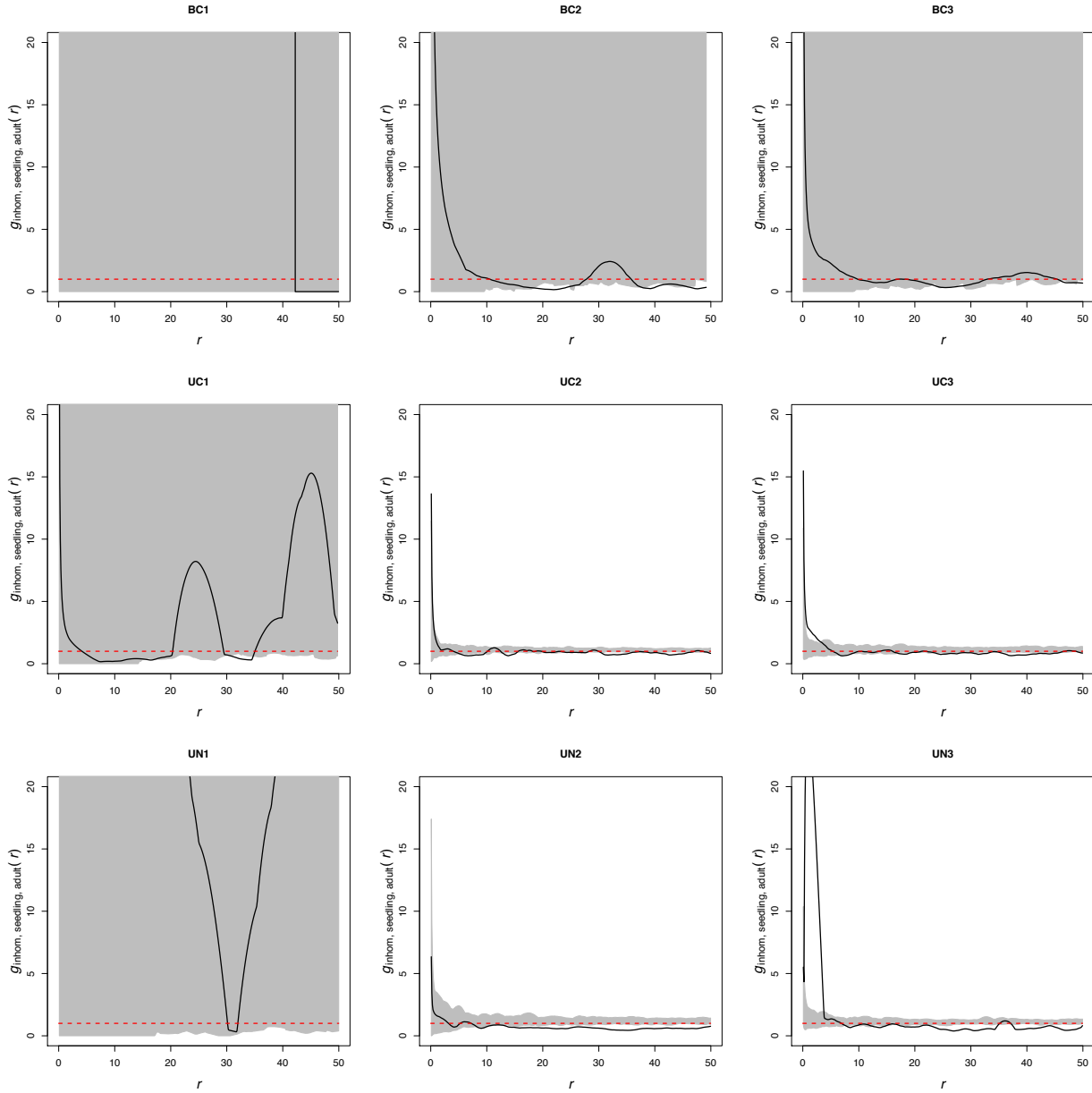


Figure 2.5 Bivariate analysis of spatial structure between *P. lambertiana* seedlings (seed) and *P. lambertiana* adult, $g_{inhom,seedling,adult}(r)$, by treatment replicate for each distance class, r . Gray : null confidence envelope; Solid black line : observed $g_{inhom,seedling,adult}(r)$. Red dashed line : null expectation of complete spatial randomness, $g_{inhom,seedling,adult}(r) = 1$. Individuals are aggregated if $g_{inhom,seedling,adult}(r) > 1$, inhibited if $g_{inhom,seedling,adult}(r) < 1$.

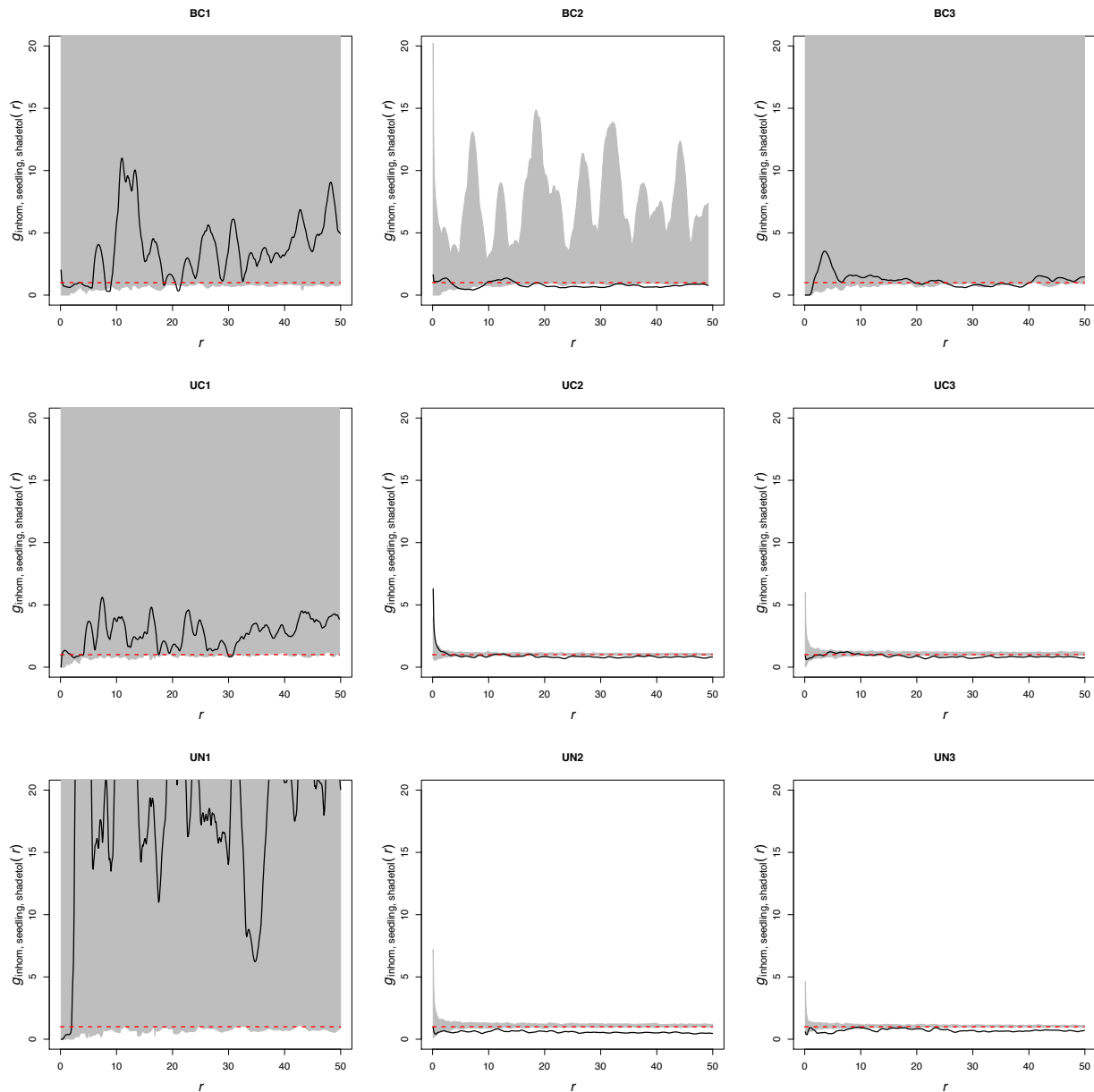


Figure 2.6 Bivariate analysis of spatial structure between *P. lambertiana* seedlings (seedling) and adult *Abies* spp. (shadetol), $g_{inhom,seed,shadetol}(r)$, by treatment replicate for each distance class, r . Gray : null confidence envelope; Solid black line : observed $g_{inhom,seedling,shadetol}(r)$. Red dashed line : null expectation of complete spatial randomness, $g_{inhom,seedling,shadetol}(r) = 1$. Individuals are aggregated if $g_{inhom,seedling,shadetol}(r) > 1$, inhibited if $g_{inhom,seedling,shadetol}(r) < 1$.

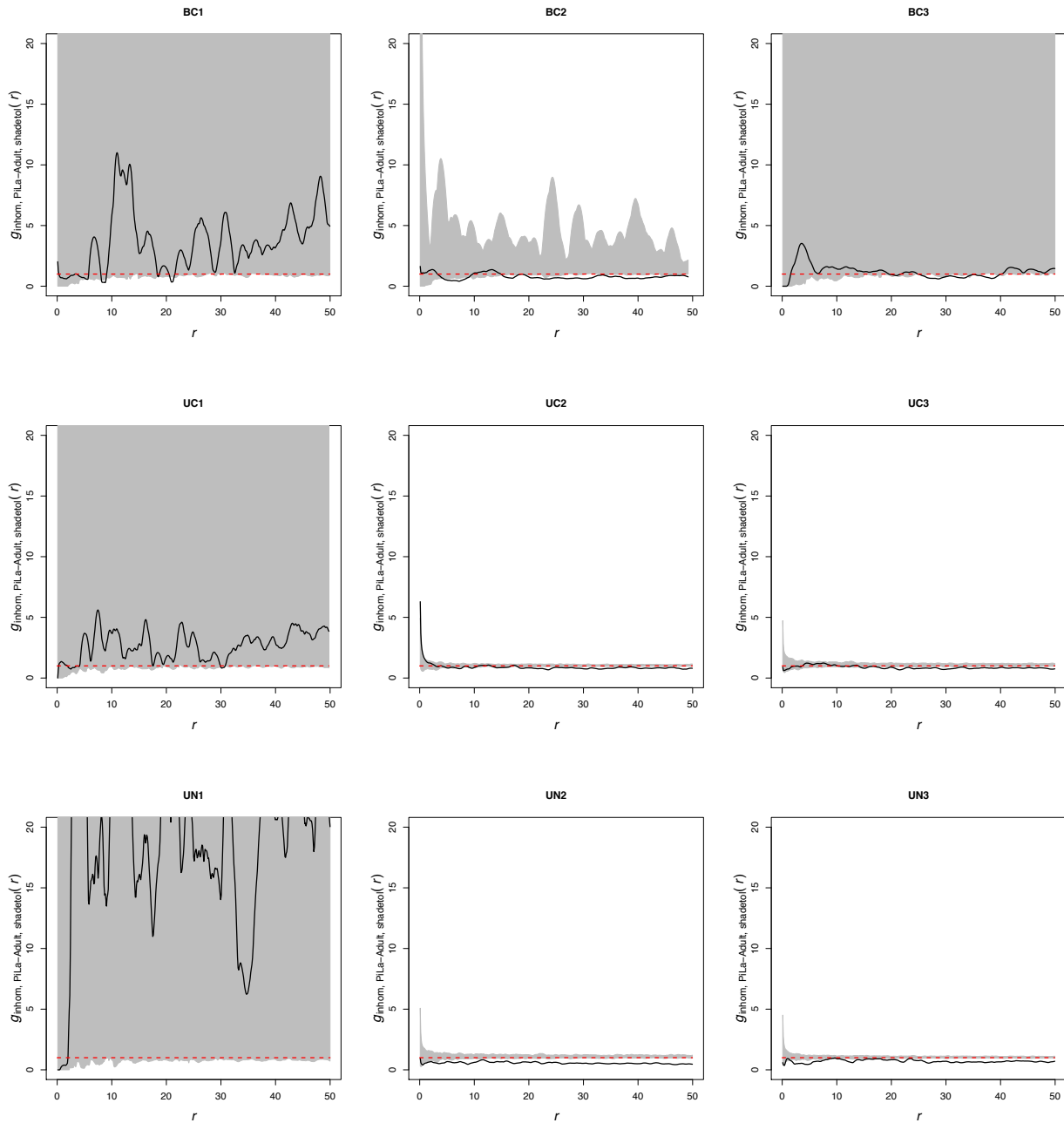


Figure 2.7 Bivariate analysis of spatial structure between *P. lambertiana* adults (*PiLa-Adult*) and adult *Abies* spp. (*shadetol*), $g_{inhom, PiLa-adult, shadetol}(r)$, by treatment replicate for each distance class, r . Gray : null confidence envelope; Solid black line : observed $g_{inhom, PiLa-adult, shadetol}(r)$. Red dashed line : null expectation of complete spatial randomness, $g_{inhom, PiLa-adult, shadetol}(r) = 1$. Individuals are aggregated if $g_{inhom, PiLa-adult, shadetol}(r) > 1$, inhibited if $g_{inhom, PiLa-adult, shadetol}(r) < 1$.

($G_{inhom,seedling,adult}(r)$), was also observed (Figure 2.6-Figure 2.7) where UN treatments generally had a greater inhibition than UC or BC plots, though BC plots also exhibited some evidence of spatial inhibition between groups. The results from the uni- and bivariate analyses of spatial patterns at TEF suggest that pines are generally clustered with other pines, shade-tolerant individuals are clustered with other shade-tolerant individuals, but spatial patterns of shade-tolerant adults generally show spatial inhibition with pine individuals of both classes. Further, together with the univariate spatial clustering of *P. lambertiana* seedlings at small distance classes, these bivariate results suggest there may be ecological drivers influencing realized patterns of seedlings across microenvironments at TEF (e.g., perhaps sites with decreased competition for [or ideal levels of] nutrients or light).

Diversity measures

To compare our results with measures often used across the literature to investigate genetic effects of forest management we calculated various genetic diversity measures (see Methods). Genetic diversity measures (Table 2.1) seemed to be most influenced by census size across the various measures we estimated here. For instance, census size increased from BC (109 individuals) to UN (557 individuals) to UC (682 individuals) where related diversity measures of A_T , A , A_e , and A_p followed this trend. Observed heterozygosity was greatest for UN plots, followed by BC and UC plots, while expected heterozygosity decreased from UC to BC to UN (Table 2.1). Thus, no trend was observed between diversity measures and increasing disturbance treatment at TEF.

Hierarchical F -statistics were calculated with nuclear markers to compare the extent of fixation within and across treatment types, with individuals nested in replicates, replicates nested in treatments, and treatments nested within TEF. The overall multilocus F_{ST} ($F_{rep,TEF}$) at TEF was 0.075, consistent with estimates of many *Pinus* species across various spatial scales (Howe et al. 2003) suggesting that the vast majority of genetic variation was partitioned more so within plots than between plots. The $F_{rep,EF}$ for individual markers varied across markers: rps02 ($F_{rep,TEF} =$

Table 2.1 Genetic diversity measures (standard deviation) by treatment. N : census number of individuals [adults, seedlings]; A_T : total number of alleles; A : mean number of alleles per locus; A_e : effective number of alleles (harmonic mean across loci); H_o , H_e : respectfully observed and expected heterozygosity for nuclear markers; A_p : average number of private alleles. For A , A_e , H_o , and H_e , values indicate averages across loci, where values for each locus were calculated across all three treatment replicates simultaneously. H_o and H_e used only nuclear markers, whereas other genetic diversity columns considered all loci.

Treatment	N	A_T	A	A_e	H_o	H_e	A_p
UN	557 [236,321]	180	25.71 (6.50)	3.23 (1.58)	0.87 (0.06)	0.77 (0.06)	46
UC	682 [307,375]	210	30.00 (7.76)	6.20 (3.07)	0.57 (0.30)	0.84 (0.10)	73
BC	109 [42,67]	107	15.29 (6.80)	4.80 (2.46)	0.82 (0.08)	0.82 (0.07)	5
Mean	449 [195,254]	165.67	23.67	4.74	0.75	0.81	41.3

0.019), rps12 ($F_{rep,TEF} = 0.037$), rps39 ($F_{rep,TEF} = 0.148$), rps50 ($F_{rep,TEF} = 0.103$). Considering only genotypes across replicates of a given treatment, treatment-level estimates of $F_{rep,tx}$ also varied ($F_{rep,UN} = 0.011$, $F_{rep,UC} = 0.109$, $F_{rep,BC} = 0.035$) but showed no pattern with increasing disturbance intensity. Pairwise $F_{rep,tx}$ comparisons between treatments were calculated by considering genotypes across two treatments simultaneously and were used to compare the extent of fixation across disturbance intensity. Here, the three comparisons ranged from 0.050 (UC and UN) to 0.055 (BC and UC) to 0.075 (BC and UN) indicative of increasing relative fixation with increasing disparity for the intensity of disturbance for a given comparison.

Analysis of spatial genetic structure

Analysis of spatial genetic autocorrelation (sensu Smouse & Peakall 1999) was carried out at TEF to better understand how treatment affects standing genetic structure ($P. lambertiana$ adults x $P. lambertiana$ adults), how this standing genetic structure affects genetic structure of seedlings ($P. lambertiana$ seedlings x $P. lambertiana$ seedlings), and the tendency of alike genotypes to be aggregated or inhibited across the treatments as the stands continue to develop after treatment ($P. lambertiana$ adults x $P. lambertiana$ seedlings). In all comparisons, spatial genetic structure in BC treatments did not differ significantly from a random spatial distribution of

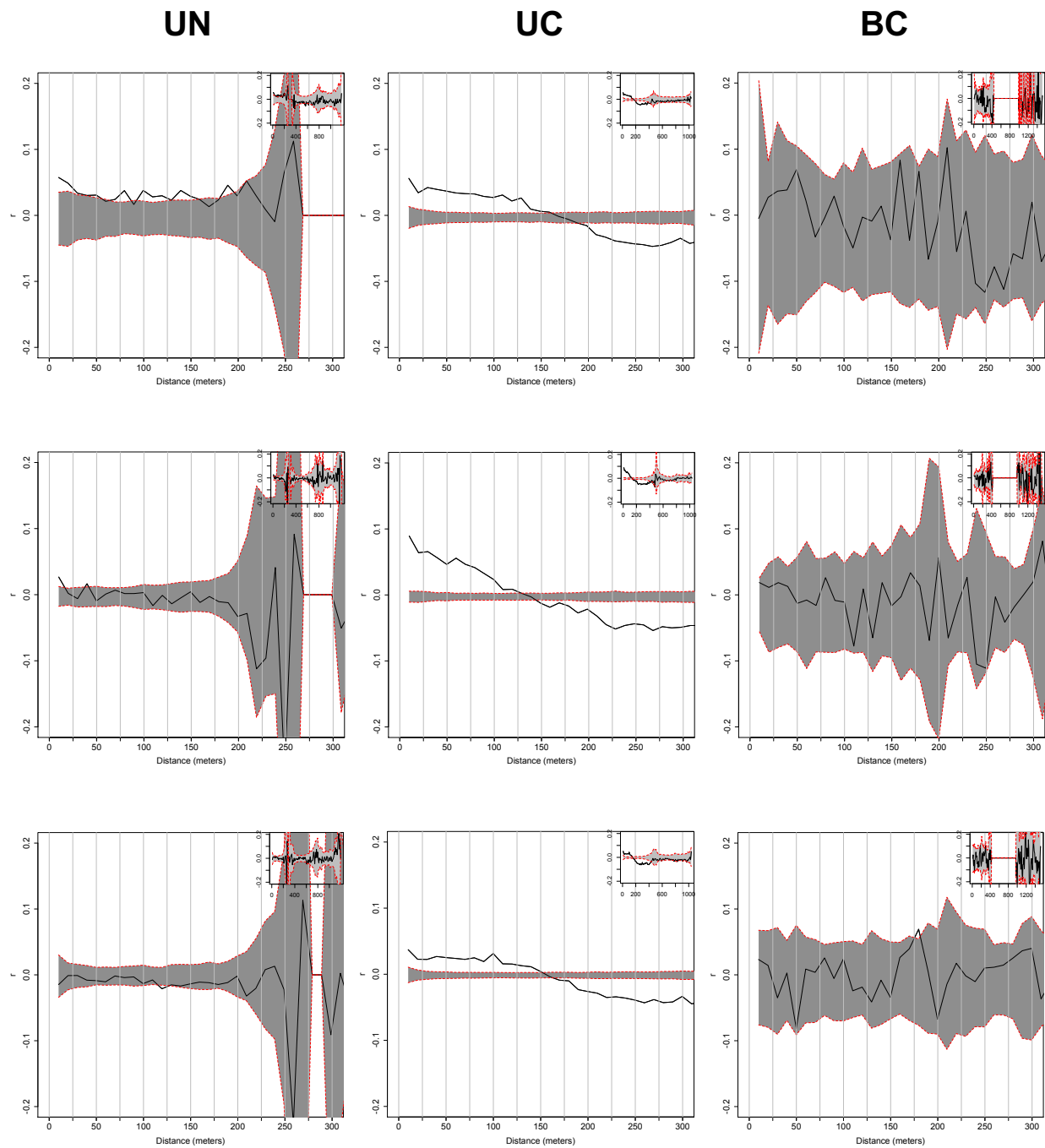


Figure 2.8 Analysis of spatial genetic structure (*sensu* Smouse & Peakall 1999) between *P. lambertiana* adults (first row), *P. lambertiana* seedlings (second row), and between *P. lambertiana* adults and seedlings (third row) by treatment (columns) across distance classes within plots (main panel) or across TEF (insets). Values of $r_g^h = 0$ indicate random spatial patterns of genotypes, $r_g^h > 0$ indicate clustering of alike genotypes, and $r_g^h < 0$ indicate spatial inhibition of alike genotypes.

genotypes (last column Figure 2.8), perhaps due to the relatively small sample sizes (Table 2.1) in distance-class bins. However, there seems to be an effect of treatment on the spatial patterning of genotypes of adults in the UC and UN stands (first row Figure 2.8). While the natural fire-suppressed stands (UN) exhibited small but significant spatial genetic structure for most distance classes up to 200m, UC stands resulted in significant aggregation of adult genotypes at a greater degree than UN up to 150m, where genotypes became spatially inhibited up to the maximum distances in stands ($200(\sqrt{2})$ m; Figure 2.8). These patterns resulted in spatial distributions of seedling genotypes that were randomly distributed except for very short distance classes in UN treatments, and for UC seedlings, resulted in the general pattern observed for UC adults albeit to a higher degree of both aggregation and inhibition (second row Figure 2.8). As a result, alike genotypes between adults and seedlings were aggregated up to 150m in UC plots, whereas this relationship in UN treatments resulted in negative values of $r_{g,adult,seed}^h$ that bordered the confidence envelope for spatial inhibition but were not significantly different from a random spatial distribution of genotypes (third row of Figure 2.8). While the genetic structure of adults is due to the interaction of the effect of treatment on pretreatment conditions, the long-term dynamics of these stands will be influenced by seedling ingrowth. These results suggest that UC treatments may, in the long term, increase the relatedness of individuals across short spatial scales less than 150m relative to either BC or UN treatments.

Quantifying fine-scale gene flow

In-plot vs. out-of-plot dispersal events

To better understand how relative gene flow across plots is influenced by treatment, we quantified the number of in-plot and out of plot dispersal events from the pedigrees identified as most probable from our parentage analysis. To account for sample size differences, we also calculated the ratio of these values. The number of in-plot and out-of-plot dispersal events

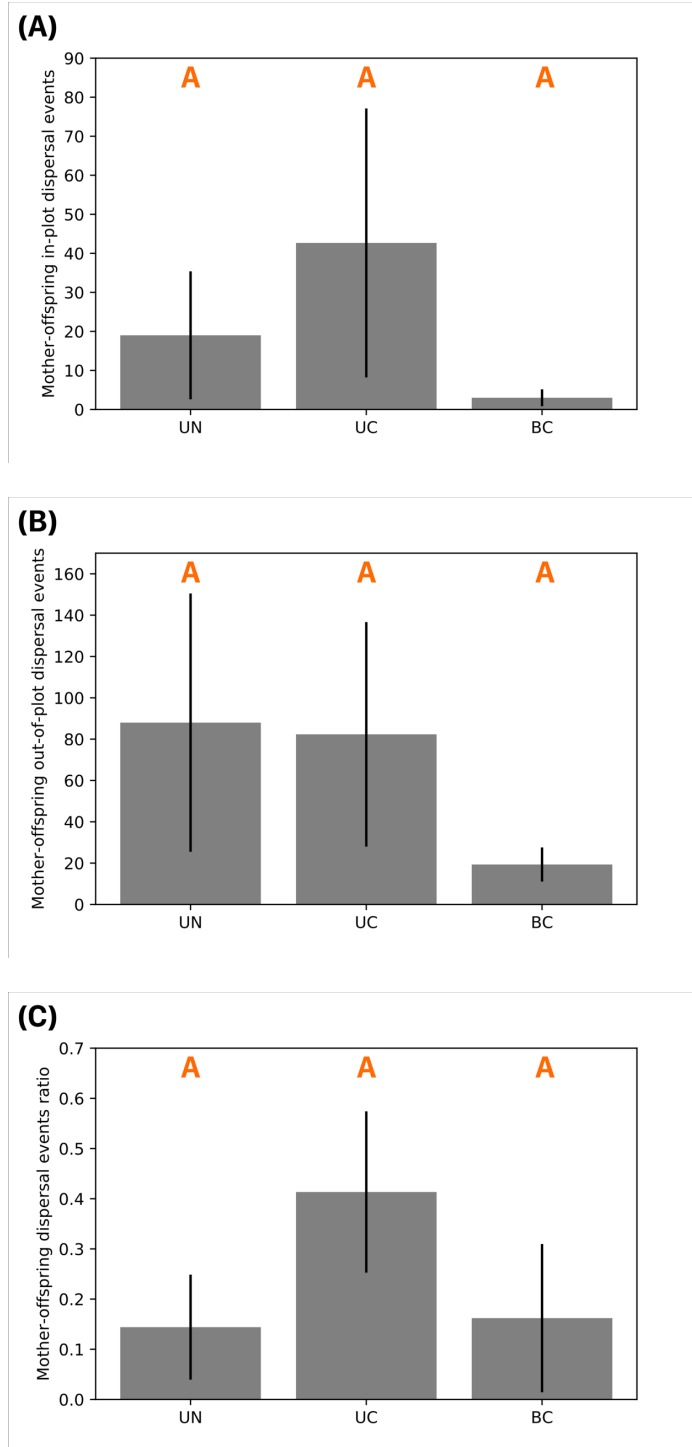


Figure 2.9 Mother-offspring dispersal events by treatment for (A) dispersal between in-plot individuals, (B) dispersal into plot from an out-of-plot mother, and (C) the ratio of these values. There were no events in which a known mother dispersed seed to another plot, therefore B is utilizing information from parentage analysis that indicated the mother of a given seedling was not sampled. Orange letters within each plot show significant differences between medians, as inferred from separate Kruskal-Wallis tests (see main text of Results). Vertical lines indicate standard deviations.

between mother and offspring differed by treatment (Figure 2.9A-B) but not significantly so ($p > 0.4297$). The ratio of these values differed by treatment as well (Figure 2.9C), with UC having the greatest proportion of in-plot dispersal events but overall there were no significant differences among treatments ($p = 0.1926$).

We next quantified the number of in-plot and out-of-plot dispersal events of pollen from the most probable pedigrees identified from parentage analysis. In these cases, out-of-plot pollen dispersal events were tallied as an in-plot mother receiving pollen from an unsampled or out-of-plot father. As with mother-offspring dispersal events we also calculated a ratio of these values. The UC treatment exhibited the most in-plot pollen dispersal events, followed by UN and BC (Figure 2.10a), though these comparisons were not significant ($p = 0.5073$). UN and UC treatments exhibited similar levels of out-of-plot dispersal events (Figure 2.10B), which differed (though not significantly, $p = 0.1376$) from BC out-of-plot events. The ratio of in-plot vs. out-of-plot dispersal events increased with increasing disturbance (Figure 2.10C) but did not differ significantly ($p = 0.1030$).

Median dispersal distances by treatment

Considering the most probable parentage from our model, we calculated the median seed dispersal distances between offspring and known mothers, as well as between the median pollen dispersal between known mothers and fathers (see Methods). Median seed dispersal varied by treatment, being greatest for UN and decreasing with increasing disturbance intensity (Figure 2.11a). Results from a Kruskal-Wallis test indicated significant differences between groups ($p = 0.048$), with *post hoc* tests indicating significant differences between UN and BC ($H = 4.34$, $p = 0.0372$) but not between UN and UC ($H = 2.77$, $p = 0.0959$) or between UC and BC ($H = 2.75$, $p = 0.0970$; Figure 2.11a). Median pollen dispersal also varied by treatment, being greatest for BC treatments, followed by UN and UC treatments, which did not differ significantly ($p = 0.1381$; Figure 2.11b). These realized distances were roughly in line with mean, but smaller than, dispersal distances estimated from dispersal kernel shape parameters in the parentage analysis:

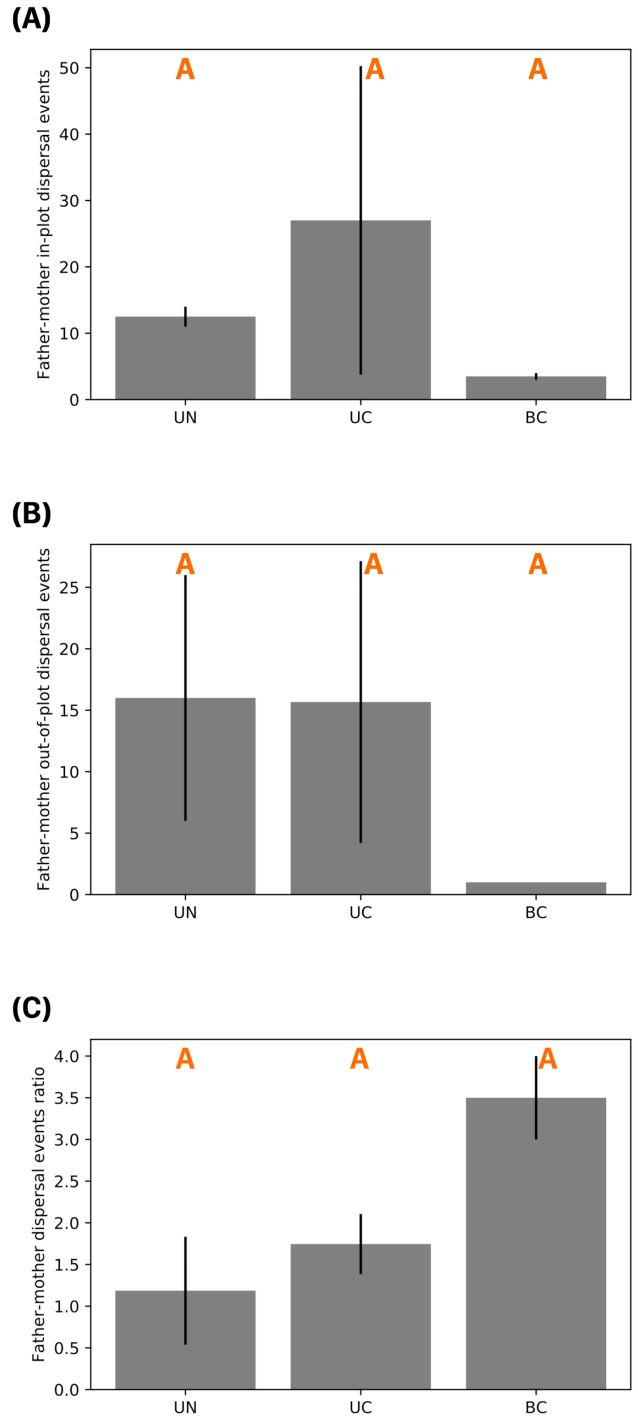


Figure 2.10 Father-mother dispersal events by treatment for (A) dispersal between in-plot individuals, (B) dispersal into plot from an out-of-plot mother, and (C) the ratio of these values. Plot-level tallies were those of in-plot mothers receiving pollen from either an in-plot father (A) or an out-of-plot (sampled or unsampled) father (B). Orange letters within each plot show significant differences between medians, as inferred from separate Kruskal-Wallis tests (see main text of Results). Vertical lines indicate standard deviations.

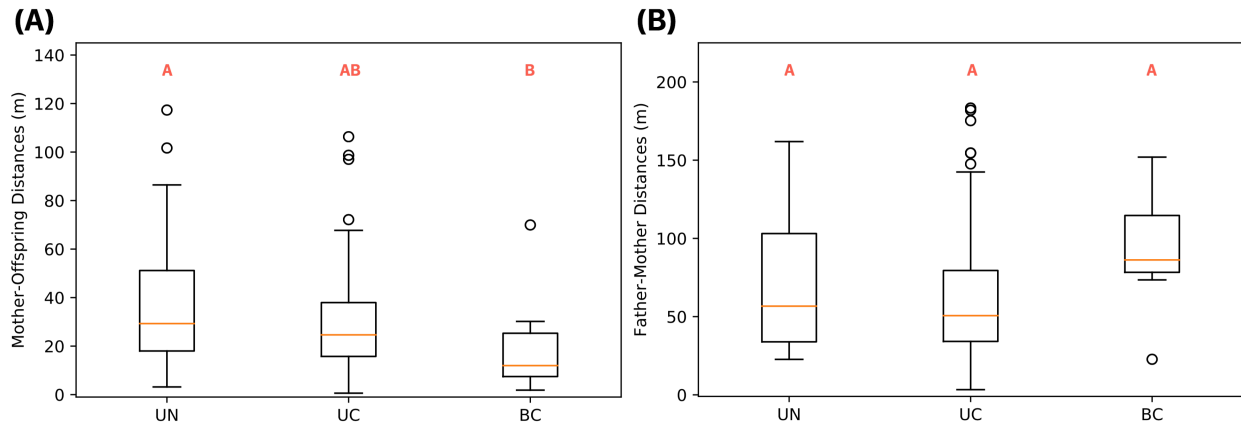


Figure 2.11 Dispersal distances for seed (A) and pollen (B) calculated from the most probable pedigree from parentage analysis, considering only pedigrees with known mothers (A) or known parents (B). Orange letters within each plot show significant differences between medians, as inferred from separate Kruskal-Wallis tests for mother-offspring and father-mother dispersal distances (see main text of Results).

mean seed dispersal = 65m (95% credible interval: 57-75); mean pollen dispersal = 170m (95% CI: 150-190; Figure 2.12).

To take into account uncertainty in parentage assignment, we calculated weighted average dispersal distances for seed and pollen dispersal (see Methods). Assignment to mothers of out-of-plot adults were less common than for assignments to in-plot fathers, as can be seen from the blocks (replicates) within treatment of Figure 2.13. Using this set of fractional parentage,

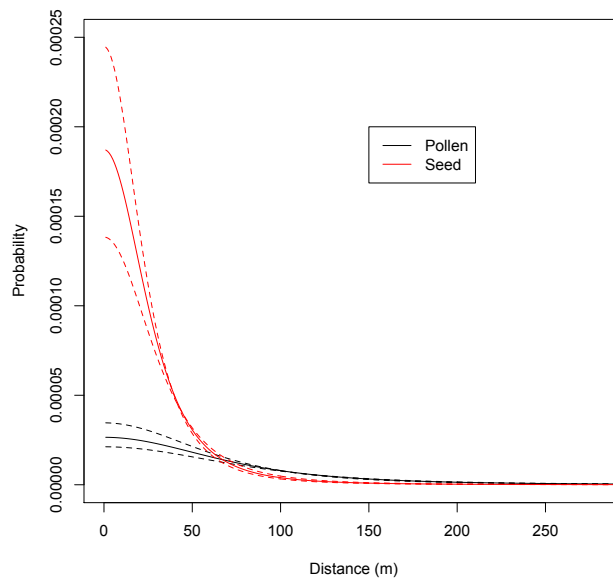


Figure 2.12 Fitted 2D-t dispersal kernel for seed (red) and pollen (black) using shape parameters inferred from parentage analysis (Moran & Clark 2011). Dashed lines show the 95% credible interval. This figure is truncated at the maximum distance within plots ($200\sqrt{2}m$) to focus on differences at short distances.

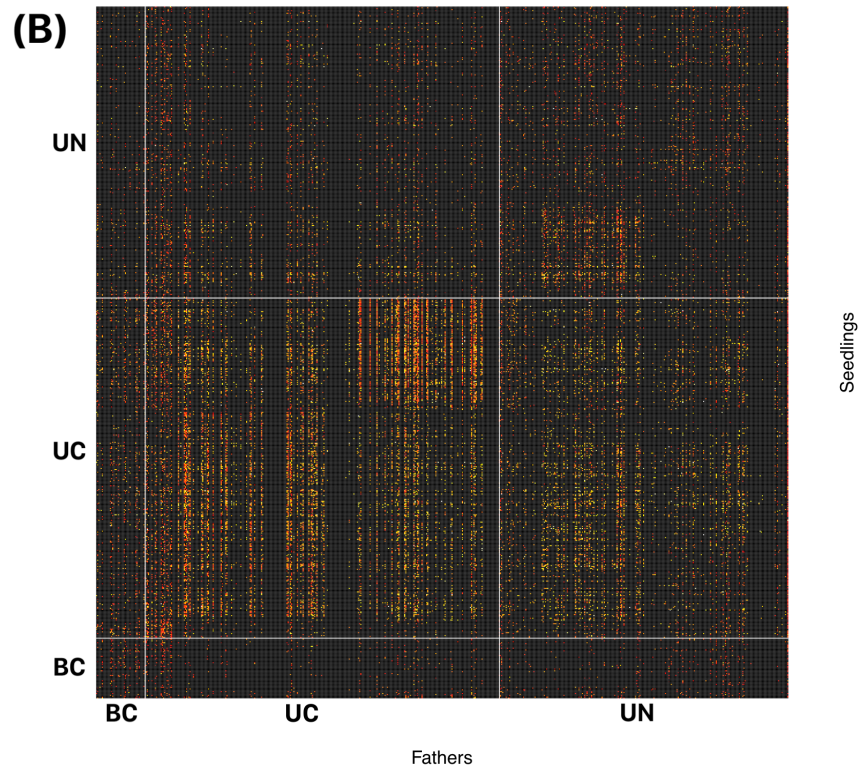
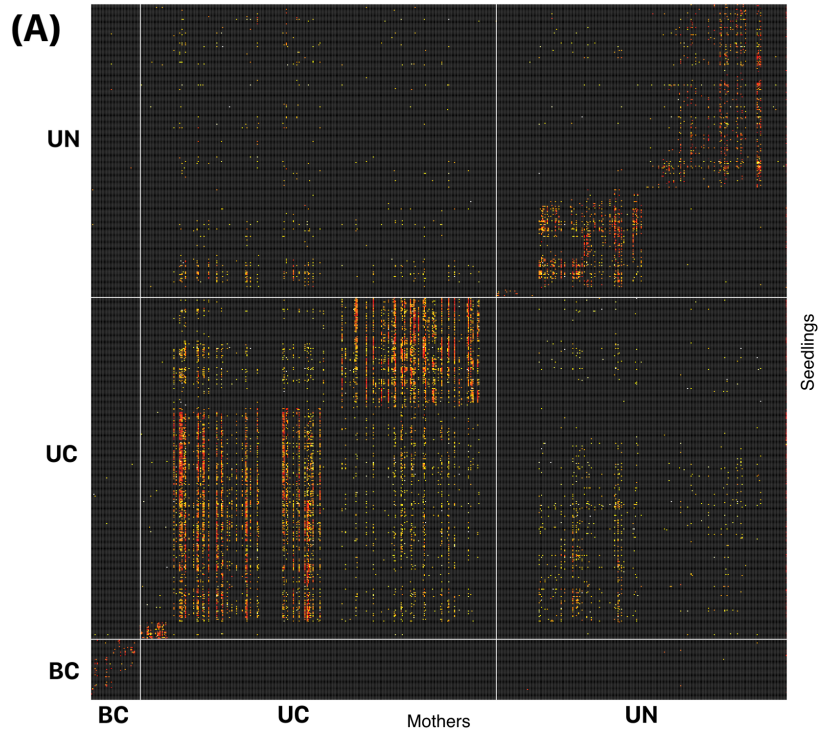


Figure 2.13 Fractional parentage across parentage analysis cycles for (A) maternal assignment and (B) paternal assignment (see Methods) with adult individuals along x-axes and seedling individuals along y-axes. Each cell represents the fraction of the cycles a particular seedling was assigned to a given adult (black ~ 0 to red to orange to yellow to white ~1).

we calculated weighted average distances for each seed and nested these distances within treatments (see Methods). We first considered mother-offspring and father-mother dispersals from fractional parentage where the identified adults could originate in any treatment at TEF. Distances differed significantly by treatment (Figure 2.14A; $H = 7.91$, $p = 0.0191$) where UN and UC were significantly different ($H = 8.11$, $p = 0.0044$) but not between any other comparison (H range = [0.0042,0.6755], $p > 0.4111$). Father-mother distances (Figure 2.14B) also differed by treatment ($H = 41.16$, $p = 1.15E-9$), with median dispersal distance decreasing from BC to UN to UC, where all pairwise considerations were also significant (H range = [5.21,27.18], p range = [1.85E-07, 0.0224]).

Because the proximity of the treatment replicates at TEF may interact with dispersal distance estimates, we also considered dispersal distances within plot tallied within treatments using weighted distances as described above. Median values of mother-offspring in-plot distances decreased with increasing disturbance intensity (Figure 2.14C) and differed by treatment ($H = 47.10$, $p = 5.91E-11$), but only between UN and UC ($H = 4.29$, $p = 0.0382$) and between UN and BC ($H = 5.83$, $p = 0.0253$) and not between UC and BC treatments ($H = 0.95$, $p = 0.3291$). In-plot father-mother distances (Figure 2.14D) were significantly different across treatments ($H = 13.89$, $p = 0.0010$), with BC having greater distances than either UN ($H = 5.83$, $p = 0.0157$) or UC ($H = 5.07$, $p = 0.0242$), and UC exhibiting greater distances than UN ($H = 5.00$, $p = 0.0253$).

Discussion

Frequent fires were commonplace in historical forests of the Sierra Nevada, where forests exhibited relatively lower tree densities and a higher proportion of pine species (Knapp et al. 2017; North et al. 2005). Yet post-settlement fire suppression has led to forest densification that has caused instability in these systems and has increased the chances of uncharacteristic high-severity wildfire. As a result, thinning prescriptions have been put forth as a means by which to increase the resilience of constituent stands (SNFPA 2004; Agee & Skinner 2005; Schwilk et al. 2009; Safford et al. 2009). However, while these prescriptions can mimic the density-reducing

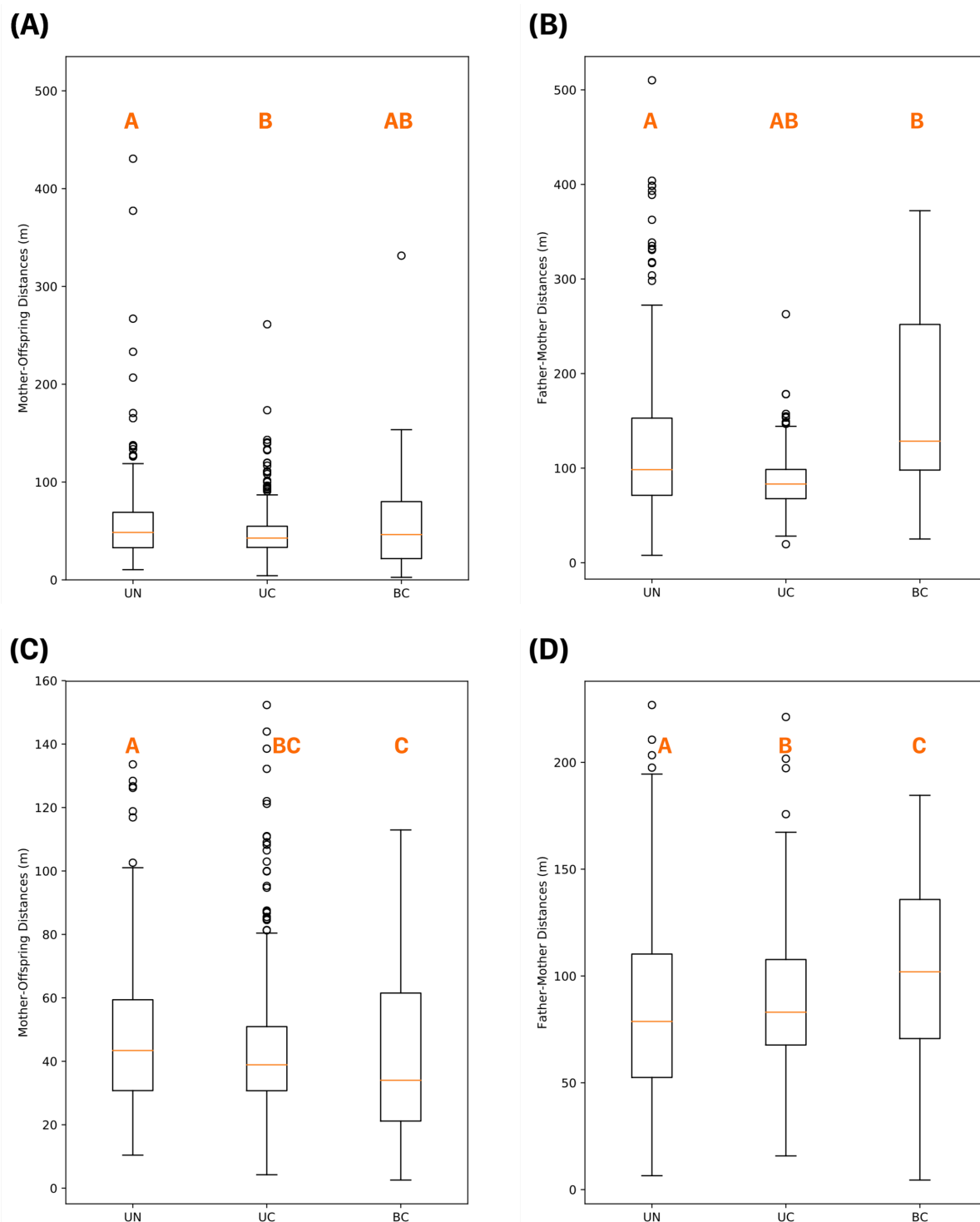


Figure 2.14 Dispersal distances between mothers and offspring (first column) and between fathers and mothers (second column) using assigned adults from any location (A-B) and for only in-plot individuals (C-D). Orange letters within each plot show significant differences between medians, as inferred from separate Kruskal-Wallis tests (see main text of Results).

effects of fire, and reduce subsequent fire severity, it is currently unknown how thinning, in isolation or through its interaction with inevitable wildfire, will alter evolutionary dynamics of ecologically important species such as *P. lambertiana* (SNEP 1996). We characterized spatial and genetic structure of a fire-suppressed forest treated with a common thinning prescription in the Sierra and compared this with its interaction with fire and with a no-thin-no-fire control treatment. Our results suggest that spatial structure of constituent species is a result of the interaction between treatment and ecology where pines are often clustered with other pines, shade tolerant trees are often clustered with other shade tolerant trees, and pine seedlings often are inhibited by both adult pine and shade tolerant individuals. While genetic diversity statistics are informative of stand-level diversity, they are under-informative regarding ongoing evolutionary dynamics as a result of treatment as they do little to predict inbreeding of future generations nor the scale at which mating events occur. From the analysis of spatial genetic structure (Smouse & Peakall 1999), and despite spatial inhibition between adults and seedlings across treatments, our results suggest that unburned thinned stands (UC treatments) result in the increase of fine-scale similarity of adult to seedling genotypes relative to control (UN treatments) or thinned stands which were subsequently burned (BC treatments). Parentage analysis offered additional quantification of fine-scale gene flow and suggested that effective seed (pollen) dispersal within plots generally decreased (increased) with the increasing intensity of disturbance, perhaps due to an increase in microsite suitability for *P. lambertiana* (and the availability of potential mates). Our results were measured from individuals remaining or regenerating 13 years post-treatment, very near the historical fire return interval for this area. Thus, ongoing dynamics should be monitored, and will likely change through time, as either stands continue to develop post-treatment, or in response to subsequent disturbances such as fire.

The genetic effects of forest management

With some exceptions, studies investigating the genetic consequences of forest management have centered around the impact on genetic diversity indices (see Table 1 in Ratnam et al. 2014). This focus is likely due to the fact that highly outcrossing tree species often suffer from elevated inbreeding depression, where survival and reproduction of subsequent generations may be impacted. In such cases, genetic diversity has been used as an index for evolutionary potential, likely attributable to the consequences of the relative contribution of additive genetic variance to phenotypic variance (i.e., narrow-sense heritability) in the breeder's equation (Lynch & Walsh 1998), where the use of heritability itself as a measure of evolvability comes with important caveats (e.g., see Hansen et al. 2011). Further, such diversity indices have been used to assess the relative reduction of alleles due to harvest intensity, where the removal of individuals from stands will likely reduce the diversity of alleles present. Here, management resulting in population bottlenecks is of concern. While these premises are important to investigate, the use of genetic diversity indices as the sole method for inference of management impact are limiting with regard to evolutionary potential. If the focus is to be on management impact on evolutionary processes, processes that influence evolutionary fitness should be investigated instead (e.g., mating systems, effective dispersal, fecundity, spatial genetic structure, pollen pool heterogeneity, juvenile survival), as argued by Lowe et al. (2015). Importantly, many of the traits with fitness consequences in trees are of a polygenic basis (Lind et al. 2018), where any given underlying locus has minimal influence on the trait. In such cases, fixation (as measured by a handful of putatively neutral markers) at some of the underlying loci can be ameliorated by selection for combinations of alleles at other loci. Therefore, while alleles with little to no effect on fitness are informative for demographic processes, these should not be conflated with loci under selection, particularly loci under strong negative selection with important implications for inbreeding depression. Such neutral markers could be better utilized in assessing consequences

within process that directly affect fitness, as argued above. However, in cases where spatial relatedness is increased as a result of management, or individuals become increasingly sparse, wasted reproductive effort (e.g., embryo abortion, or high juvenile mortality) due to increased instances of consanguineous or self mating events may play an important role in ongoing population dynamics (Kärkkäinen et al. 1999; Sorensen 2001), particularly when seed rain of heterospecifics exceeds effective reproductive output of historical or ecologically important species (e.g., as for *P. lambertiana* at TEF, Zald et al. 2008).

Dispersal dynamics of tree species

The analysis of spatial genetic structure and gene flow within and across populations of trees can elucidate ongoing evolutionary dynamics, as this spatial structure is a result of selective and neutral processes acting across temporal and spatial scales (Hardy & Vekemans 1999; Oddou-Muratorio et al. 2004; Robledo-Arnuncio et al. 2004; Oddou-Muratorio et al. 2011). Thus, quantifying dispersal and mating system is an important component in understanding such patterns. There are multiple biological and ecological factors that shape dispersal dynamics and resulting mating systems, such as population density, degree of fragmentation, manner of pollination (e.g., anemophily, entomophily, or zoophily), relative reproductive output, phenotype such as crown shape or height, interannual climatic variation, as well as stochastic variables such as wind direction and strength (Burczyk et al. 1996; Dow & Ashley 1998; Robledo-Arnuncio et al. 2004, Burczyk et al. 2004; O'Connell et al. 2004). Compared with herbaceous and annual plants, trees have more extensive gene flow (Hamrick et al. 1992), though such distances are idiosyncratic to a given population, species, and system. For instance, estimates of pollen dispersal for *Pinus sylvestris* varied from between 17-29m based on paternity assignment (Robledo-Arnuncio et al. 2004) to 136m (Robledo-Arnuncio & Gil 2005) using the TwoGener method (Smouse et al. 2001) where 4.3% of mating events came from pollen dispersed over 30km (Petit & Hampe 2006; Savolainen et al. 2004). Seed dispersal distances can also vary

idiosyncratically, particularly for winged seeds or those that are also dispersed by animals, such as with *P. lambertiana*.

Spatial genetic structure will be a function of these dispersal consequences as well as their ecological interaction with the environment. While much of the quantification of such structure in trees has been carried out at regional or continental scales, examples exist for investigations at fine spatial scales below a few hundred meters. For instance, Marquardt et al. (2007) assessed spatial genetic structure of eastern white pine (*Pinus strobus* L.) as a function of management influence at Menominee Indian Reservation in northeastern Wisconsin. While spatial genetic structure within 100m differed by population, the strongest autocorrelation occurred at the least disturbed site (Marquardt et al. 2007), however while they sampled both adults and natural regeneration they did not distinguish these two groups when inferring spatial genetic structure. Conversely, in Norway spruce (*Picea abies* L. Karst.) populations of northern Italy, Scotti et al. (2008) assessed spatial genetic structure of mitochondrial (maternally inherited) and chloroplast (paternally inherited) across both adults and saplings. While chloroplast haplotypes were uncorrelated across most distance classes up to 90m for both classes, the maternally inherited mitochondrial markers showed strong affinity below 30m, where this affinity was greater for saplings than for adults. This pattern was seen for *P. lambertiana* individuals at TEF as well, where both adults and seedlings were genetically structured at small distance classes in UC treatments, though seedling genotypes were clustered to a higher degree than adults (Figure 2.8). To our knowledge however, few instances in the literature compare both spatial structure of trees with spatial genetic structure of tree genotypes. At TEF, seedlings were clustered across all treatments likely due to microsite suitability, but were only clustered genetically in UC treatments. As such, conclusions without genotypic data may lead to spurious conclusions where it may be assumed that clustering of individuals also indicates clustering of genotypes. Further, ingrowth of *P. lambertiana* in UC treatments will likely be more related to

nearby individuals which may cause inbreeding to a greater degree in subsequent generations than in other treatments at TEF.

The implications of fine scale genetic structure

Our results suggest that management is affecting dispersal through the availability of suitable microsites for seedling establishment, as well as through the reduction of available mates. As disturbance intensity increased at TEF, mean effective seed dispersal generally decreased while effective pollen dispersal generally increased (Figure 2.14A-B), likely due to the proximity of suitable microsites and the availability of potential mates, respectively. Inferences gained from the estimation of the dispersal kernels from parentage analysis, however, allow us to quantify the proportion of dispersal events across small distance classes throughout the TEF watershed (Figure 2.12). Using the inferred dispersal kernels (Figure 2.12), the vast majority of dispersal occurs across small distance classes, with the estimated probability of dispersal of pollen below 150m accounting for more than 90.2% of pollen dispersal events, while dispersal of seed below 50m (150m) accounts for 87.3% (99.2%) of dispersal events across the TEF watershed. Such a dispersal tendency will drive spatial genetic structure and will interact with environment (including management) to ultimately determine the patterns we observe across the landscape. Because UC treatments generally resulted in an increased spatial affinity of alike genotypes between adults and seedlings (Figure 2.8), short-term dynamics (decadal scales) may be dominated by mating events between related individuals. While this may be true, long-term dynamics will likely affect this structure as well. As pointed out by Oddou-Muratorio et al. (2004), the strong levels of spatial genetic structure observed in seedlings have been shown to decrease in adult stages because of self-thinning processes in other tree species (Hamrick et al. 1993; Epperson & Alvarez-Buylla 1997; Chung et al. 2003), where this may occur at TEF as well. Even so, such consequences are dependent upon initial structure which may vary to differing degrees in undisturbed stands, or

across the landscape. Thus, to monitor these effects, long-term dynamics should be observed as these stands continue to develop and respond to contemporaneous ecological pressures.

Conclusion

With the forecast of future fire severity, size and required suppression efforts (Miller et al. 2009), and with fire seasons beginning earlier and lasting longer than historic trends (Westerling 2006), understanding how thinning prescriptions intended to decrease fire severity and restore system resilience influence evolutionary dynamics of historically dominant and ecologically important pine species is of paramount significance, as well as how these prescriptions interact with fire to affect such dynamics. Here we have shown that treatment of fire-suppressed populations of *P. lambertiana* differentially affects fine-scale spatial and genetic structure, and that seed and pollen dispersal respectively increase and decrease with the intensity of disturbance. Such dynamics are likely to remain unequilibrated in the short term, and therefore management would benefit from further monitoring of evolutionary dynamics that affect fitness in these forests. Further monitoring across broader spatial scales would also inform how these management prescriptions affect dynamics across a greater extent of environmental heterogeneity and how these evolutionary dynamics vary by locality. Such information will allow management to prescribe treatments in a regionally specific manner.

Chapter 3.

Local adaptation of *Pinus albicaulis* Engelm. across fine spatial scales of the Lake Tahoe Basin, USA

This work has been published in the following papers:

Lind, BM; CJ Friedline, JL Wegrzyn, PE Maloney, DR Vogler, DB Neale, AJ Eckert (2017) Water availability drives signatures of local adaptation in whitebark pine (*Pinus albicaulis* Engelm.) across fine spatial scales of the Lake Tahoe Basin, USA. *Molecular Ecology* 26 (12): 3168-3185. doi: <http://doi.wiley.com/10.1111/mec.14106>

Abstract

Patterns of local adaptation at fine spatial scales are central to understanding how evolution proceeds, and are essential to the effective management of economically and ecologically important forest tree species. Here, we employ single and multilocus analyses of genetic data ($n = 116,231$ SNPs) to describe signatures of fine-scale adaptation within eight whitebark pine (*Pinus albicaulis* Engelm.) populations across the local extent of the environmentally heterogeneous Lake Tahoe Basin, USA. We show that despite highly shared genetic variation ($F_{ST} = 0.0069$) there is strong evidence for adaptation to the rain shadow experienced across the eastern Sierra Nevada. Specifically, we build upon evidence from a common garden study and find that allele frequencies of loci associated with four phenotypes (mean = 236 SNPs), 18 environmental variables (mean = 99 SNPs), and those detected through genetic differentiation ($n = 110$ SNPs) exhibit significantly higher signals of selection (covariance of allele frequencies) than could be expected to arise, given the data. We also provide evidence that this covariance tracks environmental measures related to soil water availability through subtle allele frequency shifts across populations. Our results replicate empirical support for theoretical expectations of local adaptation for populations exhibiting strong gene flow and high selective pressures, and suggest that ongoing adaptation of many *P. albicaulis* populations within the Lake Tahoe Basin will not be constrained by the lack of genetic variation. Even so, some populations exhibit low levels of heritability for the traits presumed to be related to fitness. These instances could be used to

prioritize management to maintain adaptive potential. Overall, we suggest that established practices regarding whitebark pine conservation be maintained, with the additional context of fine-scale adaptation.

Introduction

The study of local adaptation has been an integral part of evolutionary biology as a whole, as local adaptation influences a wide variety of biological patterns and processes (reviewed in Savolainen *et al.* 2013). Trees in particular have received much attention in this regard because many species are ecologically and economically important, and high outcrossing rates (Neale & Savolainen 2004) result in large effective population sizes (which increase the effectiveness of selection) as well as weak neutral genetic differentiation (which decreases the confounding effects of selection and population structure). Together, these circumstances create ideal conditions in which to detect selective processes in nature (Savolainen & Pyhäjärvi 2007). Investigators seeking to explain the genetic basis of local adaptation in trees, and plants in general, have been motivated by observations of significant differentiation for quantitative genetic variation across populations (e.g., Q_{ST}) where the underlying loci may be differentiated among populations as well (Endler 1977; reviewed in Storz 2005, Haasl & Payseur 2016). In these cases, loci contributing to local adaptation could be identified through genetic indices of differentiation, or by targeting trait- or environmentally-associated loci that stand out above background demography. Yet, theoretical (Latta 2003; Le Corre & Kremer 2003) and empirical (Hall *et al.* 2007; Luquez *et al.* 2007) investigations have shown that discordance between Q_{ST} and F_{ST} of causative loci can occur under adaptive evolution. Moreover, as the number of underlying loci increases, the divergence between these indices increases as well, and the contribution of F_{ST} to any individual underlying locus decreases. In cases that exhibit strong diversifying selection and high gene flow, this adaptive divergence results from selection on segregating genetic variation (Hermisson & Pennings 2005; Barret & Schluter 2008) and is attributable to the among-population component of linkage disequilibrium (Ohta 1982, Latta 1998). In the short term, local adaptation will be

realized through subtle coordinated shifts of allele frequencies across populations causing covariance (i.e., LD) among many underlying loci (Latta 1998; Barton 1999; Latta 2003; McKay & Latta 2002; Kremer & Le Corre 2012; Le Corre & Kremer 2012), such that adaptation need not take place through numerous fixation events or sweeping allele frequency changes (MacKay *et al.* 2009; Pritchard & di Rienzo 2010). Over many thousands of generations, these shifts can lead to concentrated architectures of large-effect loci with a reduction of those with small effect (Yeaman & Whitlock 2011). For studies investigating continuous phenotypes such as those often related to fitness, even among populations with highly differentiated phenotypic traits sampled under a robust design (Lotterhos & Whitlock 2015), it may be difficult to identify many of the loci underlying the quantitative trait in question. Thus, for many species, specifically across fine spatial scales, the signal of local adaptation within much of current genetic data may go largely undetected using only single-locus approaches (Latta 1998; 2003; Le Corre & Kremer 2003; Yeaman & Whitlock 2011; Kemper *et al.* 2014), resulting in calls for theory and empiricism that move beyond single-locus perspectives (Pritchard & di Rienzo 2010; Sork *et al.* 2013; Tiffin & Ross-Ibarra 2014; Stephan 2015).

Populations of forest trees, particularly conifers, have a rich history of common garden, provenance tests, and geneecological studies that demonstrate abundant evidence for local adaptation among populations, even over short geographic distances (e.g., Mitton 1989; 1999; Budde *et al.* 2014; Csilléry *et al.* 2014; Vizcaíno *et al.* 2014; Eckert *et al.* 2015; Holliday *et al.* 2016; Roschanski *et al.* 2016) providing further support that fine spatial scales are relevant to adaptation (Richardson *et al.* 2014). This extensive history has also revealed the highly polygenic nature of adaptive traits (Langlet 1971; Holland 2007). Even so, the majority of these investigations have been limited to single-locus perspectives using either candidate genes (e.g., González-Martínez *et al.* 2008; Eckert *et al.* 2009) or a large set of molecular markers (e.g., Eckert *et al.* 2010) to explain the genetic basis of local adaptation. In most cases, a few loci underlying the adaptive trait in question are identified and generally explain a small to moderate proportion

of the overall heritability of the trait (Neale & Savolainen 2004; Savolainen *et al.* 2007; Ćalić *et al.* 2016). Yet because of the presumed polygenic nature underlying these adaptive phenotypic traits, and because past investigations have generally applied single-locus perspectives, informative biological signals may therefore be overlooked, and it is likely that a majority of the genetic architecture of local adaptation in trees remains undescribed (Savolainen 2007; Sork *et al.* 2013; Ćalić *et al.* 2016).

Spurred in part by the advance of theory and availability of genome-wide marker data, attention has been refocused to describe underlying genetic architectures from a polygenic perspective. This transition began in model organisms (e.g., Turchin *et al.* 2012) and has expanded to other taxa such as stick insects (Comeault *et al.* 2014; 2015), salmon (Bourret *et al.* 2014), and trees (Ma *et al.* 2010; Csilléry *et al.* 2014; Hornoy *et al.* 2015). Indeed, species that occupy landscapes with high degrees of environmental heterogeneity offer exemplary cases with which to investigate local adaptation. Near its southern range limit, whitebark pine (*Pinus albicaulis* Engelm.) populations of the Lake Tahoe Basin (LTB) inhabit a diversity of environmental conditions. As exemplified by the strong west to east precipitation gradient (see Figure 3.1), many of the environmental characteristics of the LTB vary over short physical distances (<1km) and have the potential to shape geographic distributions of *P. albicaulis* at spatial scales below those typically investigated (i.e., range-wide studies) for forest trees. Local spatial scales are of particular interest to resource and conservation agencies as this is the scale at which most management is applied. Here, we build upon past work from a common garden (Maloney *et al.* in review) to investigate the genetic architecture of fine-scale local adaptation across *P. albicaulis* populations of the LTB by exploring the relationships between genotype, 18 environmental variables, and five fitness-related phenotypic traits using both single and multilocus approaches. Specifically, we use the *P. albicaulis* populations of the LTB to address the following three questions: (i) Is there evidence that a long-lived, outcrossing plant species exhibiting high levels of gene flow can be locally adapted across fine spatial scales? (ii) What is the genetic basis and

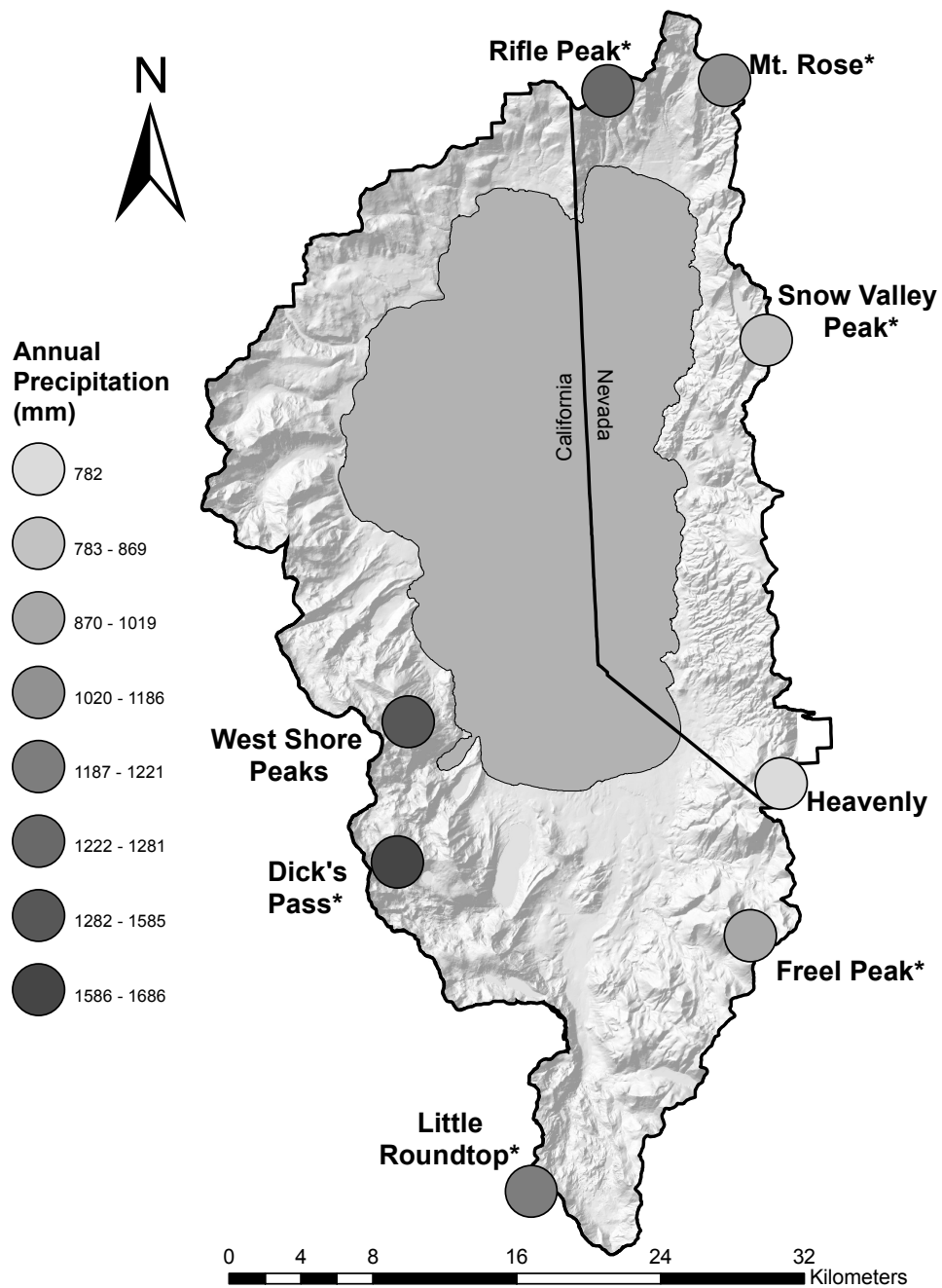


Figure 3.1. Populations used for sampling *P. albicaulis* within the Lake Tahoe Basin (dark outline). Annual precipitation is given for each population to demonstrate the west-east rain shadow experienced across fine spatial scales. Asterisks indicate populations in the common garden study.

relationship among loci underlying adaptation in such a species? (iii) How similar are the genetic bases of fitness-related phenotypes to the loci putatively under selection from the environment? Using this information, we will contextualize how instances of fine-scale adaptation have management implications. This study highlights the advantages of a polygenic perspective and investigates signatures of local adaptation using a large set of null markers to determine the extremity of allele covariance among putatively adaptive loci where others have relied on simulation or null candidate genes. Furthermore, this work provides additional empirical evidence for theoretical predictions of covariance among adaptive loci found by other studies in trees.

Methods

Focal species, study area, and sampling

A foundation species of subalpine, high elevation forests in California and Nevada, *P. albicaulis* plays a vital role in ecosystem function and services including food resources for wildlife, forest cover, watershed protection, protracting snowmelt, and biodiversity (see references in Mahalovich & Strich 2013, Tomback *et al.* 2016). It is threatened by fire-suppression, climate change, the non-native pathogen white pine blister rust, caused by *Cronartium ribicola* J.C. Fisch., and mountain pine beetle, *Dendroctonus ponderosae* Hopkins (Tomback & Achuff 2010; Mahalovich & Strich 2013).

The LTB lies within California and Nevada and experiences a Mediterranean climate. Precipitation falls during the winter months, most often in the form of snow, with a strong west-east gradient (Figure 3.1). Each of the eight study populations (three subplots per population) were located in a distinct watershed and distributed around the Basin to capture variation in the physical environment (e.g., climate, geology, and topography). Needle tissue was sampled in 2008 from 244 *P. albicaulis* trees (Table 3.1). From these eight populations, six populations were chosen to sample cones from 88 of the 244 trees that were sampled for needle tissue. All samples were collected from trees separated by 30 to 1000m, with an average interpopulation distance of

Table 3.1. Population location and associated attributes. Population size – total (maternal trees with seedlings in common garden). Climatic values were ascertained from data spanning 1971-2000. Ann. precipitation – annual precipitation; AWC – available water capacity at 25cm or 50cm soil depth; CEC – cation exchange capacity; GDD – growing degree days above 5°C; Max solar rad input – maximum solar radiation input; WC-15bar – water capacity at -15bar (wilting point); WC- $\frac{1}{3}$ bar – water capacity at $\frac{1}{3}$ bar (field capacity). Asterisks indicates populations from which seeds sampled from cones were planted in a common garden. Environmental variables are averaged across subplots.

	Dick's Pass*	Freel Peak*	Heavenly	Little Round Top*	Mt. Rose Ophir*	Rifle Peak*	Snow Valley Peak*	West Shore Peaks
Population size	25 (15)	48 (19)	25 (0)	25 (14)	49 (11)	24 (15)	24 (14)	24 (0)
Ann. precipitation (mm)	1686	1019	782	1221	1186	1281	869	1585
AWC-25cm (kPa)	1.66	1.57	1.12	1.97	1.95	1.89	2.66	1.20
AWC-50cm (kPa)	2.75	2.38	2.00	2.93	2.75	3.11	4.22	2.02
CEC (cmol _c · kg ⁻¹)	0.00	1.45	0.00	12.50	2.90	0.00	0.00	0.00
Clay (%)	6.50	4.50	6.70	14.60	3.00	6.75	6.80	6.00
Elevation (m)	2806	2865	2851	2875	2717	2819	2740	2780
GDD Aug (days)	295	190	276	211	296	235	289	279.5
GDD May (days)	0	0	6	0	11	0	2	1
Max solar rad input (%)	83.59	79.03	78.40	80.09	90.61	93.28	71.70	76.43
Max Temp – July (°C)	21.1	21.6	23.2	21.5	22.9	22.7	23.4	21.8
Min. Temp – Jan (°C)	-6.5	-8.8	-7.5	-8.0	-7.4	-7.4	-7.7	-6.6
Rock coverage (%)	31.00	18.67	25.00	14.67	7.00	30.00	26.67	42.67
Sand (%)	77.67	87.80	83.50	66.20	90.60	74.00	64.50	85.00
Silt (%)	15.8	7.7	9.7	19.1	6.4	19.2	28.7	9.0
WC-15 bar (kPa)	6.6	4.0	3.3	3.6	5.5	8.7	14.0	2.5
WC- $\frac{1}{3}$ bar (kPa)	9.7	8.0	8.4	7.3	9.8	11.4	14.4	6.2

31km. Universal Transverse Mercator coordinates, elevation, slope, and aspect (USDA FS FHTET) were used with the PRISM climatic model (Daly *et al.* 1994) to determine climatic parameters of sampled areas from 1971-2000, while soil survey data (USDA NRCS 2007) were used to describe the edaphic conditions of the LTB (Table 3.1).

Common gardens and phenotypic measurements

Fitness-related traits related to survival, especially during seedling and juvenile stages, are an important component of total lifetime fitness (e.g., Postma & Ågren 2016), particularly for forest trees, and are likely to be composed of phenotypic traits related to growth, phenology, resource allocation patterns, water-use efficiency, and disease susceptibility. In order to estimate early-lifetime phenotypes of mother trees, seeds sampled from 11 to 19 maternal trees ($n = 88$) located in six of the eight populations were established in a common garden (Table 3.1) using a random block design (for further details see Maloney *et al.* in review). Growth (height), phenology (date of bud flush), water-use efficiency ($\delta^{13}\text{C}$), and resource allocation [root:shoot biomass, $\text{N}(\mu\text{g})$] were measured when seedlings reached ~2 years in age (see Maloney *et al.* in review for details). Height was recorded in April and October 2011, while 2 seedlings per family per block were harvested, clipped above the root collar, dried, and weighed to determine root and shoot biomass. For $\delta^{13}\text{C}$ and $\text{N}(\mu\text{g})$ analysis, needle tissue from 1 seedling per family per block was harvested, coarsely ground, and dried at 60°C for 96 hours. Between 2-3mg of tissue per sample was sent to the Stable Isotope Facility at UC Davis for isotope analyses (<http://stableisotopefacility.ucdavis.edu/>).

DNA extraction, sequencing, and analysis

Total genomic DNA was isolated from needle tissue sampled from 244 trees across all eight populations using the Qiagen DNEasy 96 Plant kit according to protocol (Qiagen, Germantown, MD). Restriction site-associated double digests of total genomic DNA using MseI and EcoRI enzymes (ddRADSeq, Peterson *et al.* 2012) were used to prepare three multiplexed, barcoded

Table 3.2. Signatures of allele frequency shifts associated with environmental distance. Significant Mantel tests (9999 permutations) from comparisons among $pw\widehat{D}_{a(ij)}$ matrices from SNPs associated with environment (first column) against environmental Euclidian distance (second column). Environmental variables as in Table 3.1. † indicates a comparison in which at least one variable is water-related, or was associated with annual precipitation.

$pw\widehat{D}_{a(ij)}$	Environmental Euclidian Distance	Mantel's r	p -value
†Ann. precipitation	Ann. precipitation	0.7135	0.0027
†Longitude	Longitude	0.6522	0.0024
†Longitude	Ann. precipitation	0.7716	0.0016
†Rock coverage	Ann. precipitation	0.5542	0.0221
†Tmin-Jan	Ann. precipitation	0.5765	0.0132
†Longitude	Latitude	-0.4257	0.0347
†Rock coverage	Longitude	0.5566	0.0284
†Tmin-Jan	Longitude	0.4822	0.0273
†WC-15bar	Tmax-July	0.3490	0.0309
†WC-1/3bar	WC-1/3bar	0.4806	0.0361
†WC-1/3bar	Tmax-July	0.4539	0.0037
†WC-1/3bar	AWS0-25	0.4329	0.0384
†WC-1/3bar	AWS0-50	0.4538	0.0464
†WC-1/3bar	WC-15bar	0.5126	0.0335
GDD-May	GDD-May	0.8480	0.0013
Rock coverage	Rock coverage	0.5124	0.0145
Sand	Sand	0.5574	0.0046
Tmin-Jan	Tmin-Jan	0.5791	0.0137
Sand	Clay	0.5345	0.0232
Silt	Sand	0.4408	0.0238

libraries of up to 96 individuals each, as in Parchman *et al.* (2012). Using the QIAquick Gel Extraction Kit (Qiagen), amplified fragments were then isolated near 400bp of pooled PCR product separated in an agarose gel. Single-end sequencing of libraries was carried out on the Illumina HiSeq 2500 platform with a single library per flowcell lane. For added coverage, each library was sequenced twice using 50bp reads and twice for 150bp reads, except Library 3 which was sequenced 4x for 150bp reads to increase optimality of the mapping reference individual. All sequencing was performed at the DNA Sequencing Facility of the University of California at Berkeley (<https://mcb.berkeley.edu/barker/dnaseq/home>). After calling genotypes, SNPs, and further filtering (see Supporting Information), we judged the veracity of our sequence data by mapping

the empirical set of SNPs against the sugar pine (*P. lambertiana* Dougl.) reference genome (v1.0) using 85% similarity and 50% length coverage thresholds (<http://dendrome.ucdavis.edu/>).

Identifying focal sets of loci

To identify genotype-environmental associations, we implemented *bayenv2* (v2.0; Coop *et al.* 2010; Günther & Coop 2013), a Bayesian single-locus approach that accounts for population history and gene flow before performing association analysis (Coop *et al.* 2010). To ensure convergence, we ran five independent chains of *bayenv2* using the empirical SNPs ($n = 116,231$), with 100,000 iterations for each SNP within each chain. MCMC convergence across chains was inspected using the *coda* library in R. For each SNP, we calculated the harmonic mean across chains for the Bayes factor (\overline{BF}) and absolute value of Spearman's ρ (hereafter $\overline{\rho_S}$). When calculating \overline{BF} , if a particular SNP returned Bayes factors greater than one for at least 3/5 chains, we would take the harmonic mean from this subset to avoid underestimation of the Bayes factor. However, if this was not the case ($BF > 1$ in $\leq 2/5$ chains), we would take the harmonic mean from the values that were less than or equal to one. We identified focal SNPs by the intersection between the upper tail (99.5th percentile) of \overline{BF} and the upper tail (99th percentile) of the absolute value of $\overline{\rho_S}$, as recommended in the *bayenv2* manual (v2.0; page 4).

To associate genotype with phenotype, we implemented a Bayesian sparse linear mixed model (BSLMM) from the *GEMMA* software package (Zhou *et al.* 2013). BSLMM is a hybrid of LMM and Bayesian variable selection regression (BVSR) that also offers considerable statistical advantages over single-locus GWAS approaches (Guan & Stephens 2011; Ehret *et al.* 2012; Zhou *et al.* 2013; Moser *et al.* 2015). Specifically, to describe the underlying genetic architecture, BSLMM uses priors (described below) and attributes of the genetic data to estimate the number of underlying SNPs (N_{SNP}), the posterior inclusion probability (γ , hereafter *PIP*) for individual SNPs as well as the proportion of phenotypic variance explained by the polygenic and sparse effects of each SNP (*PVE*).

Before input to *GEMMA*, the empirical set of SNPs was reduced to include only those individuals

with seedlings in the common garden ($n = 88$), and loci which had MAF below 0.01 due to this reduction were eliminated alongside monomorphic SNPs. For each phenotype, we ran four independent chains for the BSLMM, with 1,000,000 warm-up steps and 50,000,000 steps in the MCMC, sampled every 1000th step. Priors for PVE by the model, h , were set as [0.01,0.9], and the \log_{10} inverse number of SNPs, $\log_{10}(1/p)$, [-3.0,0.0], which equates to between 1 and 300 underlying loci (N_{SNP}). Convergence of the MCMC across chains was inspected using the coda library in R. To summarize the GEMMA output, we report means and 95% credible intervals for PVE and N_{SNP} from the posterior distributions. To assess significance of association of a SNP to a phenotype, we used the PIP from all four independent chains to calculate the harmonic mean (\overline{PIP}) and chose SNPs that were greater than or equal to the 99.9th percentile of \overline{PIP} ($n \approx 116$ SNPs/phenotype) for each phenotype. We also explored SNPs with $\overline{PIP} \geq 99.8^{\text{th}}$ percentile ($n \approx 232$ /phenotype).

We implemented the program OutFLANK (Whitlock & Lotterhos 2015) to investigate loci identified as outliers based on population genetic structure (e.g., F_{ST}). Using this approach and excluding loci with expected heterozygosity values below 10% with subsequent trimming of the lower and upper 5% of empirical F_{ST} values, we inferred a null distribution of F'_{ST} and identified outlier loci with a false discovery rate of 5% from the empirical set of SNPs.

Inferring signatures of local adaptation

To determine if individual sets of focal loci (identified from GEMMA, bayenv2, and OutFLANK analyses) collectively exhibited elevated signatures of selection acting across multiple loci, we investigated the level of allele frequency covariance among all SNP pairs within each focal set. For instance, to calculate the covariance of allele frequencies across populations between two SNPs, SNP_i and SNP_j , within a focal set of SNPs associated with a particular phenotype in GEMMA, we used the global minor allele of each SNP, q , according to the interpopulation component of linkage disequilibrium,

$$\widehat{D}_{a(ij)} = \sum_k \frac{n_k}{n} (q_{i,k}q_{j,k} - q_iq_j) \quad \text{Eq. (1)}$$

where n_k is the number of individuals in population k , n is the global population size, $q_{i,k}$ is the allele frequency of the i^{th} SNP in population k , $q_{j,k}$ is the allele frequency of the j^{th} SNP in population k , while q_i and q_j are the respective global allele frequencies of the i^{th} and j^{th} SNP across $k = 6$ populations (Storz & Kelly 2008, their Equation 2; Ma *et al.* 2010, their Equation 3). Because we chose the allele to use in comparisons based on global minor allele frequency, all calculations of $\widehat{D}_{a(ij)}$ are therefore referenced to the global minor allele haplotype for a pair of SNPs. For populations that experience high levels of gene flow and divergent phenotypic optima due to selection, $\widehat{D}_{a(ij)}$ is expected to be positive between allele frequencies of loci conferring a positive effect on the phenotype, negative between those conferring opposite effect, and zero between (conditionally) neutrally loci (eq. [6] in Latta 1998). Because we were not able to discern the direction of effect for alleles within each population (as in e.g., Gompert *et al.* 2015), and to facilitate comparison among analyses, we identified selective signatures by calculating the absolute value of $\widehat{D}_{a(ij)}$ for each locus pair. We also calculated $\widehat{D}_{a(ij)}$ for focal SNPs associated with environmental variables from `bayenv2` and those identified as outliers from `OutFLANK`. In these two cases, we used allele frequencies across all eight populations.

To be able to discern if the level of covariance of allele frequencies among SNPs within a set identified by `GEMMA` (or another method; hereafter focal SNPs) was greater than that from SNPs randomly chosen from our dataset (i.e., than expected given the data), we first separated all SNPs in the dataset by their expected heterozygosity into bins of 0.01 ranging from 0 to 0.50 (e.g., a SNP with H_E of (0.000-0.010] would be binned into the first bin, while an H_E of (0.490-0.500] would be binned into the 50th). We then created a set of SNPs from which to take randomized draws by subtracting the focal SNPs from the full set of SNPs. Next, based on the occupancy of heterozygosity bins for a given focal set, we randomly selected remaining SNPs to create a null set. We chose SNPs randomly in this way, 1000 times, each time calculating the

absolute value of $\widehat{D}_{a(ij)}$ among SNP pairs within each set. From each of these 1000 distributions, we calculated 1000 median absolute $\widehat{D}_{a(ij)}$ values to create a null distribution for use in comparison to the median absolute $\widehat{D}_{a(ij)}$ from the focal set of SNPs. If the median $\widehat{D}_{a(ij)}$ is greater among our focal SNPs than the 95th percentile of the null distribution of 1000 medians, we will conclude that the signature of selection among loci within our focal sets is greater than could have arisen by chance, given the data.

To infer signatures of allele frequency shifts associated to environment, we implemented an approach similar to Equation 1 but instead of estimating $\widehat{D}_{a(ij)}$ across all populations we estimated $\widehat{D}_{a(ij)}$ across populations in a pairwise fashion (hereafter $pw\widehat{D}_{a(ij)}$) using focal SNPs from a given method. In this case, we calculated global allele frequency (q_i or q_j) based on the frequency of allele q across the $k = 2$ populations (pop_l and pop_m) under consideration (where $n_l + n_m = n$). From these estimates, we created a symmetric matrix of $pw\widehat{D}_{a(ij)}$ with columns and rows for populations, and distances within the diagonal set to zero. We then implemented Mantel tests (Mantel 1967) using $pw\widehat{D}_{a(ij)}$ matrices against other population pairwise distance matrices such as geographic distance inferred using great circle distances (km) following Vincenty's method, and Euclidian distance matrices for each of the five phenotypes and 18 environmental variables. Because we chose to take absolute values of $pw\widehat{D}_{a(ij)}$ for each locus pair (as with $\widehat{D}_{a(ij)}$) we note that the sign of the correlation coefficient, r , from Mantel tests may reflect the opposite directionality for any given SNP pair. Mantel tests were run with 9999 iterations using the `skbio` package (v0.4.2) in Python. Each environmental or phenotypic value was centered and standardized across populations before calculating Euclidian distances, but not for $pw\widehat{D}_{a(ij)}$ or geographic distance matrices. For each set of focal SNPs associated with phenotype or environment, we also quantified the mean allele frequency differences across populations and compared this to 1000 sets of random SNPs chosen by H_E .

Results

SNP filtering and characterization

After calling genotypes, SNPs, and filtering (see results section of Supporting Information), we retained 116,231 imputed SNPs for use as the empirical set in downstream analyses (Table S3.1). Of these contigs, 107,354 (92.4%) mapped to the *P. lambertiana* reference genome, thus lending authenticity to our sequence data. However, we avoid further discrimination of loci for (proximity to) genic regions until a future genome update with increased curation and density of annotation.

Overall, populations show little genetic structure with plots accounting for less than 1% of the variance in allele frequencies ($F_{\text{plot,total}} = 0.00687$; 95% credible interval: 0.0067-0.0070). Of this variation, 56.6% was accounted for by populations ($F_{\text{pop,total}} = 0.00389$; 95% CI: 0.0038-0.0040) with the remainder due to plots within populations ($F_{\text{plot,pop}} = 0.00299$; 95% CI: 0.0029-0.0031). We found similar patterns among the locus-specific estimates of F_{ST} (Figure S3.1). Moreover, we found no discernable clustering of populations using PCA, respectively accounting for 5.6% and 1.2% of the variance in allele frequencies (Figure S3.1). To further address applicability of the island model used for calculation of $\hat{D}_{a(ij)}$ and $pw\hat{D}_{a(ij)}$, we analyzed population pairwise F_{ST} according to Weir & Cockerham (1984) using the `hierfstat` package in R. Results show little differentiation among populations (mean = 0.005, max = 0.016) with no evidence of isolation by distance (Mantel's $r = 0.0990$, $p = 0.2310$).

Genotype-environment analysis

To explore the degree of association among environmental variables between populations, we used Mantel tests between Euclidian environmental distance matrices. In most cases, we found significant correlations with many of the edaphic variables measured for this study, as well as between latitude and elevation ($r = 0.3988$, $p = 0.0490$), longitude and annual precipitation ($r = 0.7145$, $p = 0.0030$), and between percent maximum solar radiation and latitude distances ($r = 0.4629$, $p = 0.0370$; Table S3.2). Additionally, geographic distance among populations was only

associated with latitude ($r = 0.9631$, $p = 0.001$), percent maximum solar radiation input ($r = 0.3992$, $p = 0.0468$), and elevation ($r = 0.4062$, $p = 0.0452$), the three of which were correlated environmentally (Table S3.2), but not to any of the remaining environmental variables (Mantel tests $p > 0.3131$, data not shown).

Through the intersection of the top 0.5% of \overline{BF} and top 1% of $\overline{\rho_S}$, bayenv2 analysis revealed between 14 (CEC) and 157 (GDD-Aug) focal SNPs associated with environment (Table S3.3). However, when calculating the \overline{BF} for each SNP, it was never the case that more than two of the five chains produced $BF > 1$, of which chains with large values were driven primarily by seed number (we used additional seed numbers for a small subset of the data during exploration, data not shown). The range of $\overline{\rho_S}$ across all focal SNPs across all environments varied from a minimum of 0.138 to a maximum 0.345 (Table S3.3). Additionally, the focal SNPs identified by bayenv2 displayed a bias towards SNPs with low values of H_E (Figures S3.3-S3.4, see results section of Supporting Information) when compared to the distribution from the full set of SNPs (Figure S3.5). As such, our environmental associations should be interpreted with caution, as we did not have any SNPs with $\overline{BF} > 1$ nor do our harmonic mean nonparametric correlations exceed 0.35. Even so, when we compared absolute estimates of $\widehat{D}_{a(ij)}$ among focal SNPs against the corresponding 1000 null sets of loci, we found that for all focal sets the median $\widehat{D}_{a(ij)}$ was always greater than the 100th percentile of the null distribution (Figure 3.2, Table S3.3). The magnitude of this difference varied across environmental variables, being the smallest for percent clay (1.17x) and largest for annual precipitation (5.10x, Table S3.3). Upon comparison of focal and null sets in both the distribution of single-locus and multilocus F_{ST} , focal sets were representative of single-locus estimates of the null sets (Figures S3.6-S3.7), and generally greater than the distribution of multilocus F_{ST} than the null (Figures S3.8-S3.9). Single-locus results suggest that many focal SNPs are unlikely outliers for F_{ST} , while multilocus comparisons exemplify the elevated frequency covariance of SNPs in focal sets. This data demonstrates that for most environmental variables

the focal SNPs show higher degrees of $\widehat{D}_{a(ij)}$ than could be expected, given the data, despite having low \overline{BF} and $\overline{\rho_S}$.

Through the examination of patterns of allele frequency shifts ($pW\widehat{D}_{a(ij)}$) across loci associated with environment we found no significant associations with geographic distance using Mantel tests ($p > 0.1116$). While this suggests the absence of linear allelic clines, it does not necessarily preclude the presence of environmental gradients or correlated patches as suggested by environmental distance associations (Table S3.2). When we investigated the association between $pW\widehat{D}_{a(ij)}$ matrices against the eponymous environmental distance matrix, we found significant association for annual precipitation, GDD-May, longitude, percent rock coverage, percent sand, minimum January temperature, and field capacity ($WC\text{-}1/3bar$; all $r > 0.4806$; $p = 0.0361$; Table 3.2). Additionally, we examined relationships between a particular $pW\widehat{D}_{a(ij)}$ matrix and the 17 remaining environmental distance matrices and found significant associations in an additional 13 comparisons (Table 3.2), with five of these comparisons having $pW\widehat{D}_{a(ij)}$ associated

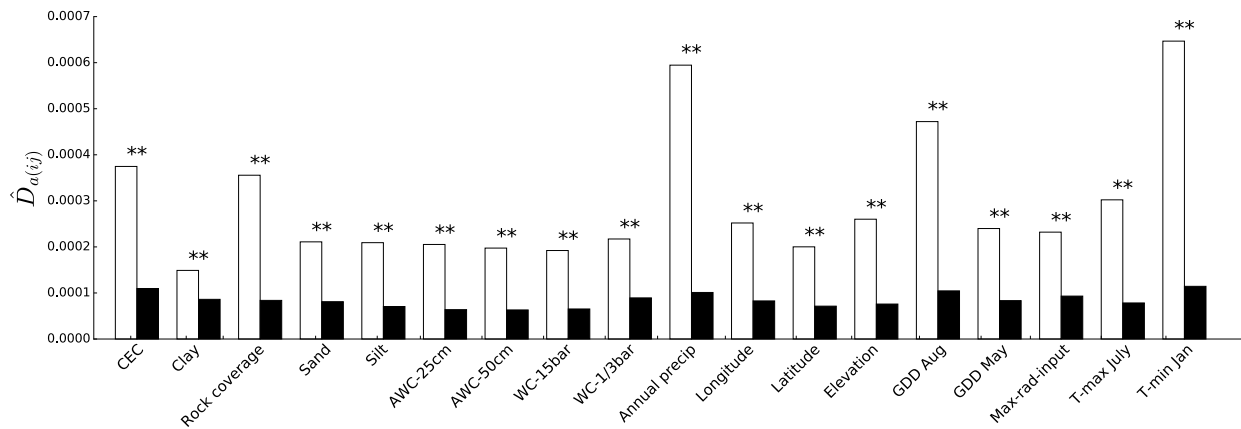


Figure 3.2. Allele frequency covariance ($\widehat{D}_{a(ij)}$) among loci associated to environment by bayenv2. In white are the median values from $\widehat{D}_{a(ij)}$ calculated among focal SNPs associated to environment. Black bars display the 95th percentile of the null distribution of median $\widehat{D}_{a(ij)}$. Environmental variables are grouped by those related to soil (CEC through WC-1/3bar) and those related to climate or geography (Annual precipitation through T-min Jan), with variables related to water availability grouped together in the center of the figure (AWC-25cm through Annual precipitation). All sets of focal loci had median $\widehat{D}_{a(ij)}$ greater than the 100th percentile of the null distribution, as indicated by two stars (**). Environmental variables as in Table 3.1.

with either annual precipitation or longitudinal Euclidian distance. We also observed shifts of alleles associated with longitude or soil water capacity across six of the remaining eight significant associations (Table 3.2), with the remaining two significant associations among edaphic conditions of sand, silt, or clay. The magnitude of the mean allele frequency difference across populations of focal SNPs was subtle as expected (range 0.018-0.029) and were generally slightly larger than that predicted from random SNPs of the same heterozygosity (Figures S3.10-S3.11). Overall, our results indicate that the vast majority of subtle allele frequency shifts among loci associated with environment ($p_w \widehat{D}_{a(ij)}$) have significant associations related to interpopulation distances of water availability (Table 3.2).

Genotype-phenotype analysis

Phenotypic traits were heritable and structured across populations (bud flush $h^2 = 0.3089$, $Q_{ST} = 0.0156$; $\delta^{13}C$ $h^2 = 0.7787$, $Q_{ST} = 0.0427$; height $h^2 = 0.0608$, $Q_{ST} = 0.0418$; $N(\mu g)$ $h^2 = 0.3525$, $Q_{ST} = 0.0191$; root:shoot $h^2 = 0.3240$, $Q_{ST} = 0.0110$; Table S3.4) and were correlated with environmental variables (both climate and soil) in ways unexplainable by neutral evolutionary forces (Maloney *et al.* in review). Additionally, bud flush and $\delta^{13}C$ had significant $Q_{ST} > F_{ST}$ (Maloney *et al.* in review, using the genetic data presented herein). We used a subset of the empirical set of SNPs for use in genotype-phenotype analysis, after filtering we retained 115,632 SNPs. PCA revealed a similar pattern to the empirical set of SNPs (data not shown). Using three significant axes of population structure identified through Tracy-Widom tests, we associated SNPs to phenotypes with BSLMM (Zhou *et al.* 2013) using the top 99.9th and 99.8th percentiles of \overline{PIP} (Figure 3.4, Table S3.5, Figure S3.12). From observations of density and trace plots, we concluded that the posterior distributions across chains were converging (not shown). The H_E of focal loci were generally representative of the empirical set (Figures S3.13-S3.14).

Overall, the genetic variance of SNPs included in the polygenic model explained between 14.4% [$N(\mu g)$] and 37.6% (root:shoot) of the variance in the phenotypes measured in our study

(*PVE*, Figure 3.3, Table S3.4). For many of the measured phenotypes, a considerable proportion of the narrow sense heritability estimated previously was therefore accounted for in the estimates of *PVE* (Table S3.4), as should be the case with sufficient genetic sampling (Gompert *et al.* 2016). Interestingly, in the case of height, *PVE* exceeded the upper confidence interval of the estimated h^2 (Table S3.4).

To acquire estimates of *PVE* from genotype-phenotype associations using single-locus approaches, we used univariate linear mixed models implemented in GEMMA (see Supporting Information, Table S3.6). Across all phenotypes, there were no loci that exceeded the adjusted threshold for inclusion calculated from *q*-values with an FDR of 0.05 (Storey *et al.* 2015; $v2.4.2$), with the minimum *q*-value across SNPs within phenotypes ranging between 0.2046 ($\delta^{13}\text{C}$) and 0.9999 [$\text{N}(\mu\text{g})$] (Table S3.6). Except for root:shoot biomass, the maximum likelihood estimates of *PVE* differed drastically from the estimates from BSLMM, with *PVE* never exceeding $1.08\text{e-}06$ suggesting that a larger proportion of the heritable genetic variation for the traits measured here is explained by multiple SNPs than by individual SNPs alone. Finally, to determine if LMM loci near the threshold were captured by the BSLMM for a particular phenotype, we isolated the loci from univariate LMM above a reduced threshold of $-\ln(p_{\text{wald}}) \geq 10$ (see Figure S3.15, Supporting Information). By this reduced threshold we identified one unique locus for both bud flush and $\text{N}(\mu\text{g})$, four unique loci for both height and root:shoot biomass, and five unique loci for $\delta^{13}\text{C}$ (15 unique loci overall). We examined the focal loci sets identified from the 99.9th percentile of \overline{PIP} in BSLMM for these LMM reduced-threshold loci and found 1 of the 4 LMM loci for both root:shoot biomass and height, and 2 of the 5 loci for $\delta^{13}\text{C}$. When we assessed the set of loci in the 99.8th percentile of BSLMM \overline{PIP} , we recovered all LMM reduced-threshold loci for bud flush and $\text{N}(\mu\text{g})$ ($n = 1$), 1 of 4 loci for root:shoot biomass, 3 of 4 loci for height, and 3 of 5 loci for $\delta^{13}\text{C}$.

To determine if focal loci associated with phenotype by BSLMM exhibited evidence of selection, we estimated allele frequency covariance ($\widehat{D}_{a(ij)}$) among focal SNPs and compared

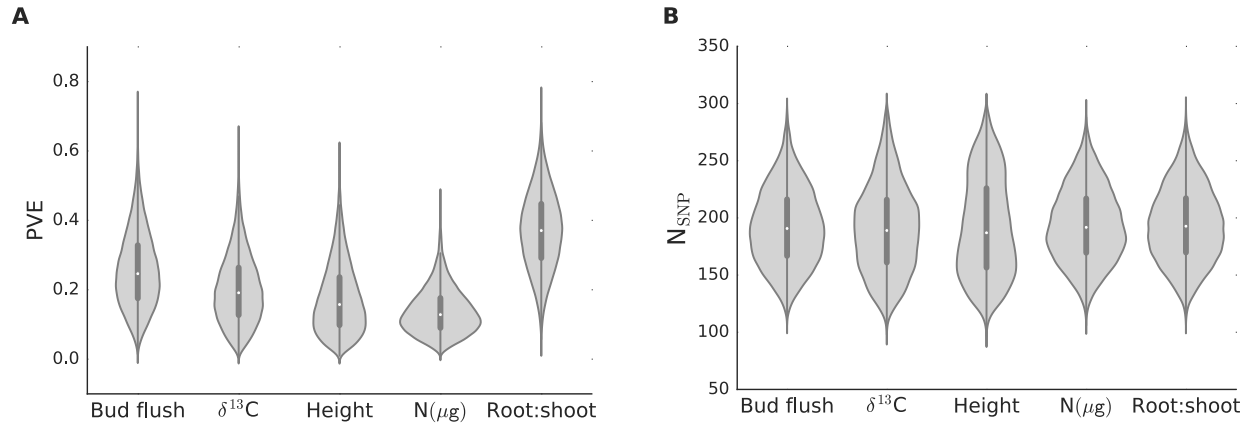


Figure 3.3. Violin plots for the kernel density estimator of the posterior distributions (light grey) taken from Bayesian sparse linear mixed models (BSLMM) executed in GEMMA for (A) the proportion of variance explained by SNPs included in the model (PVE) and (B) the number of SNPs underlying the phenotypic trait (N_{SNP}). Priors for N_{SNP} and PVE were $[1, 300]$ and $[0.01, 0.9]$, respectively. Dark grey vertical bars display the first through third interquartile range, with the median represented by the white dot.

these estimates to 1000 null sets of SNPs. We found evidence for elevated covariance among the 99.9th percentile of \overline{PIP} loci associated with bud flush and root:shoot biomass (Figure 3.4a, Table S3.5), with the latter exceeding the 100th percentile of the null distribution. To consider larger numbers of loci representative of the number of underlying loci estimated by BSLMM, we also isolated SNPs from the top 99.8th percentile of \overline{PIP} . In these sets, we found evidence for elevated signatures of selection acting across multiple loci for all phenotypes except for height, which did not produce a focal median $\widehat{D}_{a(ij)}$ greater than the 95th percentile of null distribution of $\widehat{D}_{a(ij)}$ (Figure 3.4b, Table S3.5). When focal and null sets of SNPs were compared, focal sets were representative of single-locus (Figure S3.16) and multilocus (Figure S3.17) F_{ST} estimates of the null sets.

To identify signatures of allele frequency shifts among focal loci associated with phenotype ($pw\widehat{D}_{a(ij)}$), we ran Mantel tests of $pw\widehat{D}_{a(ij)}$ matrices against geographic distance and environmental Euclidian distance matrices. When considering SNPs identified by the 99.9th percentile of \overline{PIP} , we see substantial evidence for allele frequency shifts of loci associated with

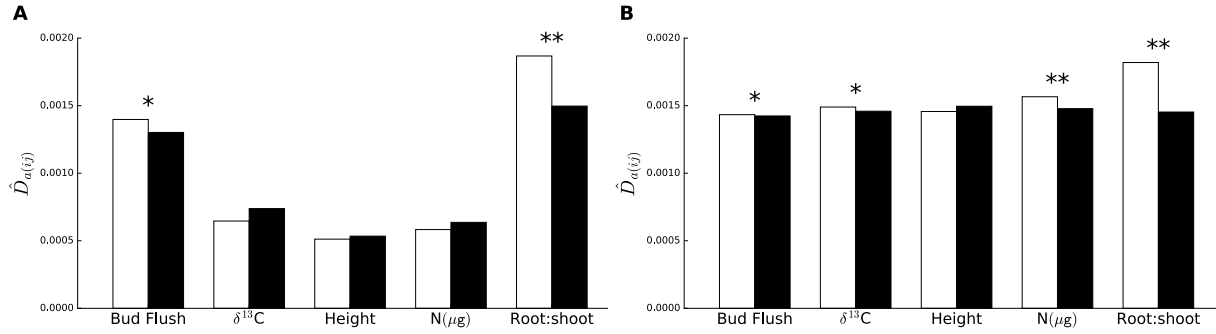


Figure 3.4. Allele frequency covariance ($\hat{D}_{a(ij)}$) among loci associated to phenotype by GEMMA. In white are the median values from $\hat{D}_{a(ij)}$ calculated among focal SNPs associated to phenotype. Black bars display the 95th percentile of the null distribution of median $\hat{D}_{a(ij)}$. (A) SNPs identified in the top 99.9th percentile of \overline{PIP} , (B) SNPs identified in the top 99.8th percentile of \overline{PIP} . One star (*) indicates that the median focal $\hat{D}_{a(ij)}$ was greater than the 95th percentile of the null distribution, whereas two stars (**) indicate that the focal median $\hat{D}_{a(ij)}$ was greater than the 100th percentile of the null distribution.

bud flush to Euclidian distances of GDD-May, GDD-Aug, percent maximum radiation input, and minimum January temperature (Table 3.3). Additionally, when we consider the 99.8th percentile of \overline{PIP} , we show evidence for allele frequency shifts among loci associated with bud flush, height, and $\delta^{13}\text{C}$ with Euclidian distances of annual precipitation, as well as for N(μg) loci with elevation, and bud flush loci with both longitude (a correlate of annual precipitation) and percent maximum radiation input (as in the 99.9th \overline{PIP} set). The strong signal from bud flush and water-related variable in Table 3.3 is intriguing, as bud flush and $\delta^{13}\text{C}$ were the only two phenotypic traits to have significantly larger Q_{ST} than F_{ST} (Maloney et al. in review). The magnitude of the mean focal allele frequency differences across populations were subtle as expected (range: 0.054-0.087) and representative of unassociated SNPs of similar H_E (Figure S3.18).

F_{ST} outlier analysis

OutFLANK analysis revealed 110 focal loci as outliers for F'_{ST} (range: 0.069-0.118). Expected heterozygosity values among the outlier SNPs (Figure S3.19) varied across the distribution from

Table 3.3. Signatures of allele frequency shifts associated with environmental distance. Significant Mantel tests (9999 permutations) from comparisons among allele frequency shifts ($pw\widehat{D}_{a(ij)}$) of SNPs associated with phenotype (second column) against environmental Euclidian distance (third column). Selection Criterion refers to the process used to identify SNPs associated with phenotype. † indicates a comparison in which at least one variable is water-related, or was associated to annual precipitation.

Selection criterion	$pw\widehat{D}_{a(ij)}$	Environmental Euclidian Distance	Mantel's r	p -value
99.9 th \overline{PIP}	Bud flush	GDD August	0.5804	0.0181
	Bud flush	GDD May	-0.5190	0.0458
	Bud flush	Max. radiation input	-0.5486	0.0482
	Bud flush	T_{\min} January	0.6984	0.0191
99.8 th \overline{PIP}	†Bud flush	Annual precipitation	0.4309	0.0140
	†Bud flush	Longitude	0.5532	0.0405
	Bud flush	Max. radiation input	-0.6312	0.0127
	N(μ g)	Elevation	0.5334	0.0246
	†Height	Annual precipitation	0.7210	0.0320
	† $\delta^{13}\text{C}$	Annual precipitation	0.5952	0.0195

the full set of SNPs (Figure S3.5). Upon analysis of patterns of covariance ($\widehat{D}_{a(ij)}$) among the OutFLANK focal SNPs, we found that the median focal $\widehat{D}_{a(ij)}$ (6.08e-03) was 10.6x greater than the 100th percentile of the null distribution of $\widehat{D}_{a(ij)}$ (5.74e-04). However, when we analyzed these outlier SNPs for signatures of allele frequency shifts ($pw\widehat{D}_{a(ij)}$) we found no significant associations with geographic or environmental distances.

Assessing possible artefactual signals of high $\widehat{D}_{a(ij)}$

It was of interest to determine if the elevated $\widehat{D}_{a(ij)}$ inferred for focal sets of loci identified from bayenv2 and OutFLANK were artefactual. For instance, loci identified by either method will likely covary in frequency as a result of the allele frequency differences among populations, perhaps elevating $\widehat{D}_{a(ij)}$ within focal sets. To assess this possible artifact, we randomized individuals across populations 100 times while maintaining original sample sizes for each population, and reran OutFLANK to classify outliers, as described previously. We found that outlier SNPs for all 100 randomizations displayed an elevated $\widehat{D}_{a(ij)}$, as was the case with the original data. To determine if the magnitude of the $\widehat{D}_{a(ij)}$ estimate from our original OutFLANK results was greater than expected, we constructed a distribution by calculating this magnitude from each of the random

runs by dividing the median focal $\widehat{D}_{a(ij)}$ from the outlier set by the 100th percentile of the same run's null distribution. Using this distribution of 100 magnitudinal differences, we found that the elevated $\widehat{D}_{a(ij)}$ calculated from our original set (10.6x greater than 100th percentile of the null) was 4.81 standard deviations from the mean of the distribution, and 1.25x greater than the 100th percentile. This suggests that our original focal set from `OutFLANK` still displayed significantly greater $\widehat{D}_{a(ij)}$ even after methodological artifacts were taken into account. We did not perform a similar analysis with `bayenv2` because this was computationally prohibitive, but we expect similar patterns to have emerged. If we instead use the original results (Figure 3.2) as a surrogate for null expectations, we find that the magnitudinal difference between focal and null $\widehat{D}_{a(ij)}$ (as calculated above) for annual precipitation is 2.11 standard deviations from the mean of the distribution of differences from environmental associations. If we exclude the top three environmental variables with the largest differences (annual precipitation, GDD-Aug, and T_{\min} -Jan; Figure 3.2, Table S3.3), the difference for annual precipitation increases to 3.97 standard deviations from the mean. Because the signal for elevated $\widehat{D}_{a(ij)}$ remains after accounting for artifacts, and given the other biological signals from our dataset, we conclude that the results presented here are consistent with expected signals of local adaptation driven by water availability in this system.

Intersection of SNPs within and across methods

We examined overlap of focal SNPs among the various methods employed in this study (Table S3.7). While there was considerable overlap of loci found between methods, overall there was more overlap of loci associated with multiple phenotypes or with multiple environments than found across methods. For sets of loci associated with environmental variables, the overlap of loci among environments seemed to be driven by the correlations among soil properties, for when ordered by the number of loci within the intersection, 12 of the top 15 comparisons were among edaphic conditions (Table S3.7). Additionally, climate-related variables relating to maximum radiation input, degree growing days, minimum and maximum temperature generally shared loci

among soil variables relating to water availability while annual precipitation shared 18 loci with longitude, among other variables. Very few of the loci identified by `bayenv2` would have been detected through conventional F_{ST} outlier approaches (Figures S3.20-S3.21). Even so, `OutFLANK` captured between 1 to 3 ($n = 18$) of the loci identified across 10 of the 18 environmental associations from `bayenv2`, many of which were water-related (e.g., annual precipitation), and captured 4 of the loci identified in the 99.8th percentile of \overline{PIP} , but not for any of the loci identified from the reduced threshold of LMM. Among loci associated with phenotype (99.8th \overline{PIP}), there were between one and three loci which were found in the intersection among pairwise phenotypic comparisons, yet none of these overlap loci were those identified from LMM. Finally, 15 loci associated with environment overlapped with the 99.8th percentile of \overline{PIP} (including two between $\delta^{13}C$ and longitude, a correlate of annual precipitation) while environmental associations did not capture any of the reduced-threshold loci from univariate LMM (Table S3.7).

Discussion

The spatial extent of local adaptation, particularly in conifers, has generally been investigated at regional scales (Neale & Savolainen 2004; Savolainen *et al.* 2007; Čalić *et al.* 2016). While informative for range-wide inference, management and conservation agencies are often limited to local scales spanning only tens to several hundreds of square kilometers. While there is an expectation that high gene flow (i.e., migration load) exhibited by many conifers can lead to swamping of adaptive alleles, there is mounting empirical evidence that adaptation to the environment can still occur at relatively fine spatial scales (Mitton 1989; 1999; Budde *et al.* 2014; Csilléry *et al.* 2014; Vizcaíno *et al.* 2014; Eckert *et al.* 2015; Holliday *et al.* 2016; Roschanski *et al.* 2016). Thus, studies which investigate adaptation at scales amenable to management may be of relatively greater importance (especially for endangered and threatened species) to reforestation applications such as those carried out through seed sourcing (*sensu* McLane & Aitken 2012) and replanting efforts. Previously, we provided evidence that measured fitness-related pheno-

types are heritable, that population explains a significant proportion of phenotypic variation, and Q_{ST} was significantly greater than F_{ST} for bud flush and $\delta^{13}C$ (Maloney *et al.* in review). Here, our genetic analyses indicate selective pressures of *P. albicaulis* are likely driven by water availability (e.g., precipitation gradients) as well as interactions and correlates of soil properties, which lends replicative support to both theoretical and empirical predictions for the patterns of loci underlying quantitative traits undergoing selection with gene flow. For instance, Ma *et al.* (2010) assessed evidence for diversifying selection within European aspen (*Populus tremula* L.) across 23 candidate genes of the photoperiodic pathway using the covariance of allelic effects among loci, albeit across a geographic region of Sweden spanning 10 latitudinal degrees. From this candidate set, they identified high degrees of covariance among phenotypic effects as predicted from theory (Latta 1998), despite minimal allele frequency differentiation among sampled populations. More recently, Csilléry *et al.* (2014) assessed 53 climate-related candidate genes within European beech (*Fagus sylvatica* L.) providing evidence that covariance among loci is attributable to epistatic selection (*sensu* Ohta 1982) across fine spatial scales of less than 100km². While varying across spatial scales, these studies replicate evidence for a signal of local adaptation in trees through elevated among-population linkage disequilibrium between adaptive loci.

Standing genetic variation for fitness related traits

The populations under study appear to have extensive gene flow, recent divergence, or both. Variation of allele frequencies among populations accounts for less than 1% of the variance observed, which was less than that found for *P. lambertiana* populations within the LTB (Eckert *et al.* 2015), or among isozymes sampled from populations across the Northern *P. albicaulis* range (Krakowski *et al.* 2003). Inspection of PCs showed no distinctive clustering of populations (Figure S3.2) while population pairwise F_{ST} did not exceed 0.016 and a test for isolation by distance was not significant. Consequently, the mean allele frequency differences between populations for focal SNPs were also subtle. Biologically, such a pattern of extensive sharing of alleles across populations has likely resulted from a combination of long-distance pollen movement, and seed

dispersal by Clark's nutcracker (*Nucifraga columbiana* Wilson) which is known to disperse seeds at distances similar to those between our sampled populations (Tomback 1982; Richardson *et al.* 2002 and references therein). Given this pattern of structure, the island model with symmetric migration used to describe the interpopulation component of linkage disequilibrium among loci ($\widehat{D}_{a(ij)}$) and allele frequency shifts ($p_w \widehat{D}_{a(ij)}$) is likely suitable to investigate our dataset for signatures of selection across multiple loci.

While `bayenv2` did not identify any loci strongly associated with environment, as given from small values of Bayes factors (all $\overline{BF} < 1.0$, Table S3.3), there is a strong biological signal for adaptation to soil water availability in our dataset (discussed below), evidence that other white pines within the LTB are also being structured by precipitation differences among populations (Eckert *et al.* 2015), and elevated signals of selection among focal loci associated to annual precipitation which would be unlikely to arise, given the data and methodological artifacts. Thus, it seems unlikely that the focal sets of SNPs associated to annual precipitation, and perhaps GDD-Aug and T_{\min} -Jan, are driven solely by false positives. However, if the majority of loci associated to environment are not an artifact of the method (i.e., are within or in linkage with causative sites), one possible explanation for elevated covariance is that the structure of environmental variables across populations captured variation for unmeasured phenotypic traits which were largely representative of total lifetime fitness (Schoville *et al.* 2012). Structure of unmeasured fitness-related traits is also likely to explain the high covariance of `OutFLANK` loci. Future work could provide validation through functional analyses of loci or from similar patterns found in other systems.

Water availability as a driver of local adaptation

The strongest signal for local adaptation among *P. albicaulis* populations of the LTB came from evidence of adaptation to soil water availability (Figures 3.2 and 3.4; Tables 3.2-3.3). Indeed, water availability is a critical component shaping standing variation across plant taxa (Vicente-Serrano *et al.* 2013), including the distributions of tree species in general (van Mantgem *et al.* 2009; Allen 2010), and southern populations of *P. albicaulis* specifically (Bower & Aitken 2008;

Chang *et al.* 2014). During the Pleistocene-Holocene transition (10,000-12,000yr BP), shifts from mesic to xeric conditions caused proximal *P. albicaulis* populations of the Western Great Basin (~50km distant) to shift from 1380m in elevation to their current position about 1100m down-slope (Nowak *et al.* 1994; *cf.* Table 3.1). Such shifts in climate and local edaphic conditions in the last 10,000yr may in part explain recent (relative to $4N_e$ generations), and ongoing, selective pressures on *P. albicaulis* populations of the LTB. Because of climatic constraints imposed on the southern range of *P. albicaulis*, phenotypic traits affected by precipitation, soil water availability, or soil water capacity likely have fitness-related consequences for this species. Additionally, with climatic models predicting warmer temperatures, reduced snow accumulation, and earlier spring melt across the western USA, it is likely that *P. albicaulis* populations of the Sierra Nevada will continue to face selective pressures of this kind. Even so, many of *P. albicaulis* populations of the LTB exhibit substantial genetic variation for the fitness-related traits measured, suggesting that the majority of ongoing adaptation within the LTB will likely be unconstrained by the lack of genetic variation. Instead, other biotic factors (e.g., white pine-blister rust infection) that can lead to negative population growth rates may be of more immediate concern (see Maloney *et al.* 2012).

Implications to whitebark management

As pointed out by McLane & Aitken (2012), distribution models of many species predict habitat suitability to shift with climate in the upcoming century, leaving great uncertainty that tree species in particular will be able to track suitable environments through natural migration and establishment, as the rate of many of these geographic shifts would far exceed observed post-glacial rates of migration (Davis & Shaw 2001; McLachlan *et al.* 2005). Exacerbating this issue, the presence of *C. ribicola*, and climate-driven outbreaks of mountain pine beetle (*D. ponderosae*) create further challenges to the conservation of *P. albicaulis* (Tomback & Achuff 2010; Mahalovich & Stritch 2013). Without intervention, such cases could lead to population collapse, extirpation, or extinction. As such, assisted gene flow, replanting, or restoration efforts will need to continue to take current and future selective pressures into account (e.g., genetic variation, resistance to

C. ribicola, etc.), as has generally been the standard of practice (Keane *et al.* 2012; Maloney *et al.* 2012; McLane & Aitken 2012).

At the same time, the choice of seed source will also need to take into account local adaptation at fine spatial scales. While a small proportion of the neutral genetic variation (F_{ST}) is found among most tree populations (often less than 5%, Neale & Savolainen 2004), this does not necessarily mean that seeds from within an established seed zone will be optimal for any given constituent environment, particularly if the seed zone exhibits high environmental heterogeneity (such as in montane regions), or if the seed zone is relatively broad compared to these environmental gradients, as is the case in California (see Buck *et al.* 1970). Weak neutral genetic differentiation can be misleading in this way, as polygenic traits influenced by selective processes in the face of gene flow may lead to divergent local adaptation through the covariance of alleles among populations without the buildup of substantial genetic differentiation at any given locus. Particularly in cases where there is evidence of local adaptation, and when ethical (McLachlan *et al.* 2007), appropriate (considering e.g., ecological or demographic factors), or plausible, seed source should come from local sources (i.e., relative to scales of geography and environmental gradients) to maximize adaptive potential (McKay & Latta 2002). This is particularly important when maximal fitness (e.g., reproductive output) is a priority for established trees, as while many genotypes may survive, realized phenotypes related to fitness may be suboptimal for a given environment. In cases where local seed sourcing is not plausible, perhaps due to isolation, or with prohibitively small population sizes, sources likely to perform well are also likely to come from highly correlated environments, particularly if the populations are not highly diverged. In contrast, local seed source may be of lesser importance for populations existing over broad, relatively homogeneous environments. Optimal sources in these cases will also likely come from recently diverged populations (McKay & Latta 2002). With this taken into consideration, management may be able to prioritize populations for restoration through estimates of trait heritability, either through common garden experimentation or estimated from marker data, as estimates of genetic variation

alone may be misleading. For instance, heritability for $\delta^{13}\text{C}$ was near zero for the Rifle Peak population, as was the case for root:shoot ratio in Freer Peak and Little Roundtop, $\text{N}(\mu\text{g})$ in Rifle Peak, and height for four of six populations. While this could mean the presence of recent, strong natural selection, or, conversely, that these traits do not convey an overwhelming adaptive advantage in these populations, candidate traits with substantial evidence for contributing to total lifetime fitness should be monitored nonetheless. In appropriate cases, introducing compatible variation would stand to improve adaptive potential in such populations. While there is no one specific solution to conserve populations of *P. albicaulis* across its range, taking into consideration fine-scale local adaptation in addition to established strategies will likely aid in such endeavors.

Limitations and concluding remarks

While we described associations among genotype, phenotype, and environment that collectively represent strong evidence for adaptive responses of *P. albicaulis* populations to the environment, we acknowledge several limitations. First, our study design was limited in statistical power which could have been improved by increasing the number of individuals sampled, the total number of populations, or both, given an ideal sampling regime (Lotterhos & Whitlock 2015). This would have facilitated use of other methodologies for uncovering evidence of polygenic local adaptation (e.g., Berg & Coop 2014). Second, while we measured fitness-related traits among seedlings of a species whose lifespan differs by several orders of magnitude, establishment success is one of the primary factors influencing dynamics of forest populations, and early life stages of plants have been shown to be a major component of total lifetime fitness (Postma & Ågren 2016). Third, much of the statistical signal for the association of allele frequency shifts to environment would be lost with correction for multiple tests. However, we leverage the fact that, of the few significant $p\widehat{D}_{a(ij)}$ associations, the majority were related to $\delta^{13}\text{C}$, annual precipitation, its correlate of longitude, or measures of soil water availability, which is an outcome highly unlikely by chance alone. Fourth, while we provide evidence for statistical signals predicted by theory, our methodology limited us from making conclusions regarding local adaptation *sensu stricto* as we

utilized just a single common garden without reciprocal transplants and were unable to quantify functional differences of putative loci among populations. Finally, a more fully curated, well-annotated genome assembly and accompanying linkage map would have aided in the detection of physical linkage among SNPs, proximity to genomic regions of estimated effect, and detection of false positives. For instance, the *P. lambertiana* genome used to judge authenticity of sequence data does not yet have the density of annotation needed to draw inferences on the causative sites likely within or linked to the loci described here, as its assembly and curation are still ongoing. We cannot, therefore, conclude that elevated covariances inferred for focal loci are not an artifact due to distant linkage with causative sites of larger effect. For this artifact to be true, however, our results would indicate that many of the loci in focal sets of SNPs would all have to be linked to the same smaller number of larger-effect loci, or that many large-effect loci underlie the measured traits, both unlikely outcomes given the expectations of quantitative traits and the coverage from ddRADseq methods found in other white pines (e.g., Friedline *et al.* 2015). Lastly, while we may not have solely identified causative sites, but instead those linked to causative genomic regions, linked sites can still maintain signals of evolutionary processes (McVean 2007). Future work could address these limitations and lead to the corroboration of our results, particularly in describing patterns exhibited by underlying loci in similar systems.

Our inferences were synthesized from prevailing signals across multiple phenotypic, environmental, and taxonomic (*cf.* Eckert *et al.* 2015) lines of evidence. The results reported here suggest that focal loci collectively show elevated allele frequency covariance (a signal expected between loci undergoing selection with gene flow) across multiple loci than could have been expected to arise from the data by chance or artifact, and that interpopulation levels of allele frequency covariance are often associated with interpopulation distances of soil water availability. Our results further explain a considerable proportion (*PVE*) of the additive genetic variation (h^2) of the quantitative traits under study from a polygenic perspective, as should be the case with sufficient genetic sampling (Gompert *et al.* 2016). Thus, we can posit that the general mode of

adaptation for *P. albicaulis* across the LTB is facilitated by selection on standing levels of genetic variation that is extensively shared throughout the basin and likely improves performance in early life stages. Finally, if soil and climatic variables continue to influence the extant populations within the LTB as evidenced from our analyses, it is likely that these variables will continue to be important to the long-term success of this threatened keystone species.

Acknowledgments

We thank Annette Delfino Mix, Camille Jensen, Tom Burt, and Randi Famula for field, common garden, and lab assistance. Additionally, we thank David Fournier, Joey Keely, Kurt Teuber (USDA Forest Service - LTBMU), Roland Shaw (Nevada Division of Forestry), Bill Champion (Nevada State Parks), Woody Loftis (USDA NRCS) for site information and permission to work on Federal and State lands, the VCU CHiPC for computational resources, Jennifer Ciminelli for GIS tips, and Lindsay Miles who helped improve this manuscript. This work was supported by the Southern Nevada Public Lands Management Act – Rounds 7 and 10, sponsored by the USDA, Forest Service, Pacific Southwest Research Station, Albany, CA. We also thank anonymous reviewers whose suggestions greatly improved this manuscript.

Chapter 4.

Local adaptation and linkage maps I: A first step towards describing genetic architectures of complex traits in natural populations

This work has been published in the following papers:

Friedline CJ, Lind BM, Hobson EM, et al. (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* 11:49. doi: [10.1007/s11295-015-0866-x](https://doi.org/10.1007/s11295-015-0866-x)

Abstract

Explaining the origin and evolutionary dynamics of the genetic architecture of adaptation is a major research goal of evolutionary genetics. Despite controversy surrounding success of the attempts to accomplish this goal, a full understanding of adaptive genetic variation necessitates knowledge about the genomic location and patterns of dispersion for the genetic components affecting fitness-related phenotypic traits. Even with advances in next-generation sequencing technologies, the production of full genome sequences for non-model species is often cost-prohibitive, especially for tree species such as pines where genome size often exceeds 20 to 30 Gbp. We address this need by constructing a dense linkage map for foxtail pine (*Pinus balfouriana* Grev. & Balf.), with the ultimate goal of uncovering and explaining the origin and evolutionary dynamics of adaptive genetic variation in natural populations of this forest tree species. We utilized megagametophyte arrays (n = 76–95 megagametophytes/tree) from four maternal trees in combination with double digest restriction-site associated DNA sequencing (ddRADseq) to produce a consensus linkage map covering 98.58 % of the foxtail pine genome, which was estimated to be 1276 cM in length (95 % CI, 1174 to 1378 cM). A novel bioinformatic approach using iterative rounds of marker ordering and imputation was employed to produce single-tree linkage maps (507–17,066 contigs/map; lengths, 1037.40– 1572.80 cM). These linkage maps were collinear across maternal trees, with highly correlated marker orderings (Spearman's $\rho > 0.95$). A consensus linkage map derived from these single-tree linkage maps contained 12 linkage groups along which 20,655 contigs were non-randomly distributed across 901 unique positions

($n = 23$ contigs/position), with an average spacing of 1.34 cM between adjacent positions. Of the 20,655 contigs positioned on the consensus linkage map, 5627 had enough sequence similarity to contigs contained within the most recent build of the loblolly pine (*Pinus taeda* L.) genome to identify them as putative homologues containing both genic and non-genic loci. Importantly, all 901 unique positions on the consensus linkage map had at least one contig with putative homology to loblolly pine. When combined with the other biological signals that predominate in our data (e.g., correlations of recombination fractions across single trees), we show that dense linkage maps for non-model forest tree species can be efficiently constructed using next-generation sequencing technologies. We subsequently discuss the usefulness of these maps as community-wide resources and as tools with which to test hypotheses about the genetic architecture of local adaptation.

Introduction

Evidence for adaptive evolution among populations of plants is commonly documented at the phenotypic and molecular levels (Kawecki and Ebert 2004; Pannell and Fields 2013), and as such some of the best examples of adaptive evolution within lineages come from the field of plant genetics (e.g., Antonovics and Bradshaw 1970). Despite this evidence, relatively little work has focused explicitly on the genomic organization of loci contributing to these patterns (Hoffmann and Riesberg 2008), which likely stems from a lack of genomic resources for plants relative to animals. Adaptive evolution has been extensively documented for forest trees, especially conifers, with many instances of local adaptation clearly documented over the past century (White et al. 2007; Neale and Kremer 2011). Despite great advances in experimental technology, empirical focus has remained almost fully on the number, effect size, type, and interactions among loci contributing to adaptive evolution (Neale and Kremer 2011; Alberto et al. 2013). A thorough examination of the genetic architecture of fitness-related traits, however, should also include an examination of the genomic organization of the loci contributing to trait variation. We leverage this

idea in the first of a series of papers dissecting the genetic architecture of fitness-related traits in a non-model conifer species, foxtail pine (*Pinus balfouriana* Grev. & Balf.).

The genomic organization of loci contributing to variation in fitness-related traits would follow naturally from the production of a sequenced genome (i.e., a physical map). For many taxa, especially those with small to modest genome sizes, this is monetarily and computationally feasible using next-generation DNA sequencing technologies (Koboldt et al. 2013). For taxa with large or complex genomes, however, even the advent of next-generation DNA sequencing does not solve the complexity and cost hurdles associated with the production of a finished genome sequence. Conifers have large and complex genomes (Murray 1998; Ahuja and Neale 2005), with estimated average genome sizes in *Pinus* in the range of range 20–30Gbp. Several genome projects, each of which involves large consortia, are either underway or have been completed (Mackay et al. 2012). Even these efforts often initially result in limited information, however, as for example, the current assemblies of the Norway spruce (*Picea abies* L.) and loblolly pine (*Pinus taeda* L.) genomes contain millions of unordered contigs with average sizes in the thousands of base pairs (Nystedt et al. 2013; Neale et al. 2014). An alternative, but not mutually exclusive, approach to describing the genome of an organism is that of linkage mapping. In this approach, genetic markers are ordered through observations of recombination events within pedigrees. This approach dates to the beginning of genetics and the logic has remained unchanged since the first linkage maps were created in *Drosophila* (Sturtevant 1913).

Renewed interest in linkage maps has occurred for two reasons. First, linkage maps are often used to order contigs created during genome sequencing projects (Mackay et al. 2012; Martínez-García et al. 2013). In this fashion, linkage maps are used to help create larger contigs from those generated during the assembly. It is these larger contigs that create the utility that most practicing scientists attribute to genome sequences. Second, linkage maps are easy to produce and provide a rich context with which to interpret population and quantitative genetic patterns of variation (e.g., Eckert et al. 2010a, b, 2013, Yeaman 2013). They can also be used to

test explicit hypotheses about the organization of loci contributing to adaptive evolution. For example, Yeaman and Whitlock (2011) developed theoretical predictions about the genomic organization of loci underlying patterns of local adaptation as a function of gene flow, so that loci contributing to local adaptation have differing spatial structure within genomes as a result of differing regimes of gene flow. The relevant scale (*sensu* Houle et al. 2011) in these mathematical formulations is that of recombinational distance among loci, so that when matched with an appropriate study system, linkage maps provide the impetus to test basic evolutionary hypotheses. In this context, future additions of finished genome sequences would add to the interpretation of results.

Construction of linkage maps have a long history within forest genetics, mostly through their use in quantitative trait locus mapping (Ritland et al. 2011). Conifers in particular are highly amenable to linkage mapping, with approximately 25 different species currently having some form of linkage map completed (see Table 5-1 in Ritland et al. 2011). Much of the amenability of conifers to linkage mapping stems from the early establishment of breeding populations in economically important species and from the presence of a multicellular female gametophyte (*i.e.*, the megagametophyte) from which the haploid product of maternal meiosis can be observed (Cairney and Pullma 2007). Indeed, many of the first linkage maps in conifers were generated from collections of megagametophytes made from single trees (Tulsieram et al. 1992; Nelson et al. 1993; Kubisiak et al. 1996). Continued advancements in genetic marker technologies have facilitated rapid development of linkage maps across a diversity of species (*e.g.*, Acheré et al. 2004; Kang et al. 2010, Martínez-García et al. 2013). The development of biologically informative markers for non-economically important conifers, however, is hampered by production costs associated with the creation of characterized genetic markers (*i.e.*, those with a known DNA sequence and/or function). The majority of this cost is in the two-step approach needed to generate biologically meaningful markers: polymorphism discovery via DNA sequencing followed by genotyping of those polymorphisms (*cf.*, Eckert et al. 2013). As a result, the vast majority of

linkage maps outside of economically important species are created with uncharacterized genetic markers (e.g., Travis et al. 1998). Much of the knowledge about the genetic architecture of fitness-related traits, outside of a handful of well-studied conifer species, therefore, encompasses the number and effect size of uncharacterized genetic markers (Ritland et al. 2011). Cost restrictions, however, have largely disappeared. It is now feasible to jointly discover polymorphisms and genotype samples using high-throughput DNA sequencing approaches, such as restriction-site associated DNA sequencing (RADseq; e.g., Peterson et al. 2012).

The generation of linkage maps from RADseq data is a complex endeavor due to the inherent stochasticity and error-prone nature of these data. Recent examples in several crop species highlight the difficulties that must be overcome with respect to missing data and errors in calling polymorphic sites and the resulting genotypes (Pfender et al. 2011; Ward et al. 2013). Despite these difficulties, RADseq has been successively applied to samples taken from natural populations of non-model conifer species (Parchman et al. 2012), but has yet to be applied to linkage mapping in these species. An exploration of these methods to linkage mapping in the large and complex genomes of conifers is thus warranted. Here, we take this approach using megagametophyte arrays from four maternal trees of foxtail pine to generate maternal linkage maps. There are currently no published linkage maps for this species, which is only distantly related to loblolly pine (Eckert and Hall 2006), nor any within the subsection *Balfourianae*. We subsequently discuss the utility of our inferred linkage maps to tests of evolutionary theory addressing local adaptation and its genetic architecture.

Materials and Methods

Focal species

Foxtail pine is a five-needle species of *Pinus* classified into subsection *Balfourianae*, section *Parrya*, and subgenus *Strobus* (Gernandt et al. 2005). It is one of three species within subsection *Balfourianae* (Bailey 1970) and generally is regarded as the sister species to Great Basin bristlecone pine (*P. longaeva* D. K. Bailey; see Eckert and Hall 2006). The natural range of

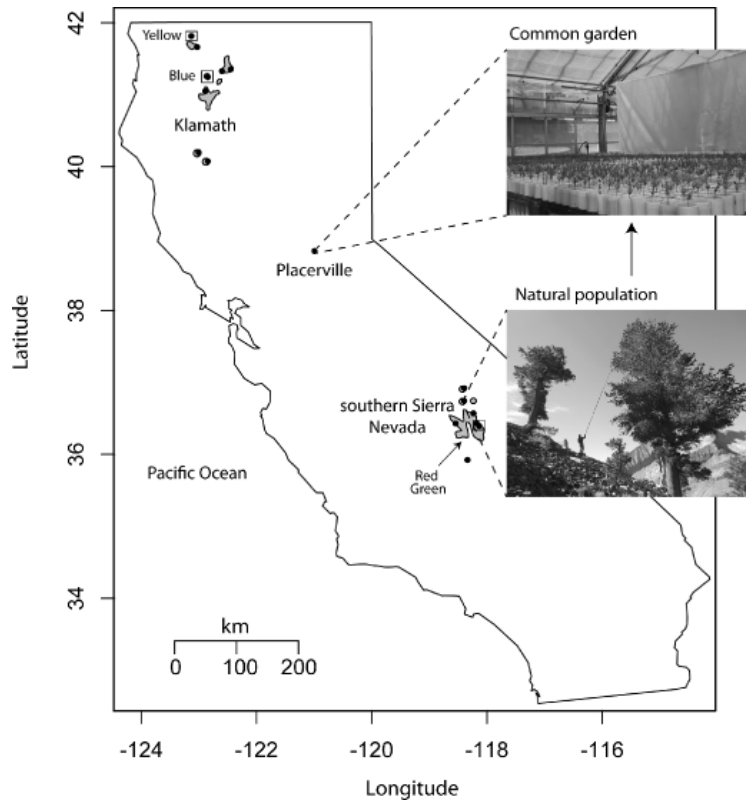


Figure 4.1. Geographical locations of foxtail pine samples used to construct a common garden located in Placerville, CA. Circles denote the 15 unique locations from which 4 to 17 maternal trees were sampled. Circles enclosed in squares denote locations from which maternal trees used in linkage mapping were sampled. Photo credits: lower: T. Burt; upper: A. Delfino Mix.

foxtail pine encompasses two regional populations located within California that are separated by approximately 500 km: the Klamath Mountains of northern California and the Sierra Nevada of southern California (Figure 4.1). These regional populations diverged approximately one million years ago (mya), with current levels of gene flow between regional populations being approximately zero (Eckert et al. 2008). Within each regional population, levels of genetic diversity and the degree of differentiation among local stands differ, with genetic diversity being highest in the southern Sierra Nevada stands and genetic differentiation being the highest among the Klamath stands (Oline et al. 2000; Eckert et al. 2008). These two regional populations have also been recognized as distinct subspecies based on numerous quantitative traits, with *P. balfouriana* subsp. *balfouriana* located in the Klamath region and *P. balfouriana* subsp. *austrina* located in the southern Sierra Nevada mountains (Mastrogiuseppe and Mastrogiuseppe 1980). The two

regional populations of foxtail pine thus represent a powerful natural experiment within which to examine the genomic organization of loci contributing to local adaptation. The first step in using this system to test evolutionary hypotheses is the production of a dense linkage map (cf., Pannell and Fields 2013).

Sampling

Seed collections from 141 maternal trees distributed throughout the natural range of foxtail pine were obtained during 2011 and 2012. Of these 141 maternal trees, 72 were sampled from the Klamath region while 69 were sampled from the southern Sierra Nevada region. These 141 families were divided among 15 local stands ($n = 4$ to 17 trees/stand), with eight stands in the Klamath region and seven stands in the southern Sierra Nevada. Approximately, 50 seeds were germinated from each seed collection and 35 of those 50 seedlings were planted in a common garden located at the USDA Institute of Forest Genetics, Placerville, California (Figure 4.1). The common garden was established using a randomized block design and involved three separate plantings of seeds spanning approximately 1 year (June 6, 2012 until May 20, 2013). Four of the 141 maternal trees were selected at random ($n = 2$ from the Klamath region and $n = 2$ from the southern Sierra Nevada) for linkage analysis. Libraries were color-coded and are referred to as red (southern Sierra Nevada), green (southern Sierra Nevada), blue (Klamath), and yellow (Klamath). For each of these trees, 75 to 100 seeds were germinated and planted in the common garden. Upon germination, haploid megagametophyte tissue was rescued from each seedling, cleaned by removing soil and other extraneous materials with water, and stored for further analysis in 1.5-mL Eppendorf tubes at -20°C .

Library preparation and sequencing

Total genomic DNA was isolated from each rescued megagametophyte using the DNeasy 96 Plant kit following the manufacturer's protocol (Qiagen, Germantown, MD). RADseq (Davey and Blaxter 2010; Parchman et al. 2012; Peterson et al. 2012) was used to generate a genome-wide set of single nucleotide polymorphism (SNP) markers for linkage mapping following the

protocol outlined by Parchman et al. (2012). In brief, this protocol is a double digest RADseq (ddRADSeq) approach based on digestion of total genomic DNA using EcoRI and MseI followed by single-end sequencing on the Illumina HiSeq platform. Single-end sequencing was chosen for reasons related to cost. Paired-end sequencing would have improved the reference assembly, which would have likely improved construction of the linkage map. Since the insert size we selected would have resulted in non-overlapping reads from each end, the improvement to genotype calls is unclear. Following digestion, adapters containing amplification and sequencing primers, as well as barcodes for multiplexing, were ligated to the digested DNA fragments. We chose to multiplex 96 samples using the barcodes available from Parchman et al. (2012). One of these samples, per set of 96, was a pseudo-diploid constructed by pooling five megagametophytes sampled from the same maternal tree, although there is a probability of $0.54 = 0.0625$ that the genotype for any given SNP will be mistakenly called homozygous due to the five megagametophytes all being of the same allele (see Morris and Spieth 1978). These barcodes are a mixture of 8, 9, and 10bp tags that differ by at least four bases. Following ligation, successfully ligated DNA fragments were amplified using PCR and amplified fragments were size selected using gel electrophoresis. We selected fragments in the size range of 400bp (300 to 500bp) by excising and purifying pooled DNA from 2.5% agarose gels using QIAquick Gel Extraction Kits (Qiagen). Further details, including relevant reagents and oligonucleotide sequences, can be found in File S1. All DNA sequencing was performed on the Illumina HiSeq 2000 or 2500 platform at the VCU Nucleic Acids Research Facility (<http://www.narf.vcu.edu/>).

DNA sequence analysis

There are multiple steps involved with the processing of raw DNA sequence reads into a set of SNP genotypes that are useful for linkage mapping: (1) quality control, filtering, and demultiplexing; (2) assembly to generate a reference sequence for mapping reads; (3) mapping of reads to call SNPs and genotypes for each sample; and (4) filtering of SNPs and the resulting genotypes for data quality and biological meaning.

DNA sequence reads were demultiplexed into sample-level fastq files, following quality control and filtering. The filtering pipeline was adapted from Friedline et al. (2012). Briefly, reads containing any N beyond the first base were excluded; however, reads having N as the first base were shifted by one base to the right to exclude it (i.e, a read starting with NTGC would become a read starting with TGC). Additional quality filtering ensured that all reads in the resulting set for downstream processing had a minimum average quality score of 30 over 5-bp sliding windows and that not more than 20% of the bases had quality scores below 30. Reads passing the quality control steps were demultiplexed into sample-specific fastq files by exact pattern matching to known barcodes. Reads that did not match a known barcode were excluded.

The individual with the largest number of reads across all four maternal trees was assembled using Velvet (Zerbino, version 1.2.10), with hash length (k) optimized using parameter sweeps of k through the contributed `velvetOptimiser` (<http://www.vicbioinformatics.com>, version 2.2.5) script (for odd k on $k = [19, 63]$). Assembly robustness was evaluated in each case using the LAP likelihood framework (Ghodsi et al. 2013), version 1.1 (svn commit r186) following mapping of the original reads to the assembly with Bowtie2 (Langmead and Salzberg 2012) (`--local --very-sensitive-local`). The assembly with the maximum likelihood value was chosen as the reference for SNP calling.

SNPs were called for all individuals against the reference using the following methodology. First, reads were mapped to the reference with Bowtie2 (`--local --very-sensitive-local`). The resulting sam files were converted to their binary equivalent (e.g., bam) using `samtools` version 0.1.19 (`view, sort, index`) (Li et al. 2009). SNPs were called using `bcftools` and filtered using `vcfutils` to exclude SNPs with less than 100× coverage. The resulting variant call files (vcf) were further processed using `vcftools` version 0.1.11 (Danecek et al. 2011) to remove indels, exclude genotype calls below a quality threshold of 5, and output as a matrix (`--012`) the haploid genotype of each megagametophyte for each SNP.

We used several thresholds to filter called SNPs for linkage mapping. First, we excluded SNPs using a χ^2 test of homogeneity against an expectation of 1:1 segregation. This segregation pattern was expected because the maternal tree had to be a heterozygote to detect a SNP, and Mendel's first law guarantees that the segregation ratio for this SNP should be 1:1. Significance of each test was assessed using a Bonferroni-corrected significance threshold of $\alpha = 0.05$, where α was corrected using the number of SNPs tested. As reads from each family were mapped against a single reference assembly, we performed the χ^2 test and corrections on a family-wise basis. Second, for each family, we filtered the resulting SNPs based on the genotype of the pseudo-diploid sample in that family, so as to keep only those SNPs where the pseudo-diploid was either (1) called a heterozygote or (2) had a missing genotype call. Lastly, we filtered the resulting SNPs, so as to keep only those that had a minimum of five genotype calls for each of the alternate alleles. These filtering steps were taken to minimize the presence of genotyping errors arising from technical (e.g., read mapping and alignment) and biological (e.g., paralogy) reasons. Previous research within conifer genomes has documented the presence of a large number of paralogues (Keeling et al. 2008; Nystedt et al. 2013; Neale et al. 2014). Although we did not explicitly quantify the degree of paralogy consistent with our data, the filters used during the analysis of DNA sequence reads should flag paralogous loci preferentially. The resulting subset of SNPs was then used as the input to linkage analysis.

Linkage analysis

The production of a linkage map requires three main steps: (1) calculation of pairwise distances between all pairs of loci, (2) clustering (i.e., grouping) of loci based on these pairwise distances, and (3) ordering of loci within each cluster (Cheema and Dicks 2009). A variety of software packages exist to carry out these steps (e.g., Van Ooijen 2011). Traditional software packages for linkage mapping, however, are not amenable to large amounts of missing data and frequent errors in genotype calls. The former causes issues with all aspects of analysis, while the latter primarily affects the genetic distances between markers (Hackett and Broadfoot 2003;

Cartwright et al. 2007). We thus followed the approach of Ward et al. (2013), which was designed specifically for RADseq data.

In brief, this method can be described as follows. Pair-wise distances were estimated and loci were clustered using a custom R script (R Core Team 2013). We used MSTmap (Wu et al. 2008a) to infer marker order and Maskov (Ward et al. 2013) to impute and correct genotypes. The algorithms available in MSTmap can also be used to impute and correct genotype errors (see Wu et al. 2008a), but the amount of missing data and putative genotyping errors in our RADseq data far surpassed those used to develop this software. These two programs were used in an iterative fashion. MSTmap was used initially to order markers, which was followed by the use of Maskov to impute and correct putative genotype errors conditional on this initial marker ordering. A last round of ordering was performed using MSTmap conditional on the imputed and error-corrected genotype data. This general schema was followed for each of the four maternal trees independently.

The relevant pairwise distance for linkage mapping in our haploid case is defined as the probability of observing a recombination event between two haplotypes. This probability can be calculated for a set of biallelic loci using the Hamming distance ($d_{i,j}$). The Hamming distance is the number of differences separating two binary strings (Hamming 1950), which are in this case, the haploid genotypes for a set of two megagametophytes. This distance, scaled by the number of positions (i.e., $d_{i,j}/n$), is the maximum likelihood estimate of the probability of a recombination event with respect to a pair of haplotypes in a double haploid design (Wu et al. 2008a). It is also an estimate of the recombination fraction, so that these distances can be transformed into LOD scores (see Morton 1955). Missing data were dealt with in a pairwise manner, so that each pairwise comparison had missing data removed prior to estimation of $d_{i,j}/n$. When values of $d_{i,j}/n$ exceeded 0.5, which is the theoretical maximum value given the expected 1:1 segregation pattern, they were set to 0.5. The $d_{i,j}/n$ values were used to construct the pairwise distance matrix between all possible pairs of loci passing our quality thresholds.

Loci were clustered hierarchically based on the pair-wise distance matrix using Ward's method as the linkage function (Ward 1963). The values of $d_{i,j}/n$ were squared prior to use of Ward's method in hierarchical clustering. We explored groupings (K) based on clustering on the interval $K = [8, 16]$. This interval was chosen because it brackets the haploid chromosome number of foxtail pine ($1N = 12$). This entailed cutting the resulting dendrogram at a specific height, so that the desired number of groups resulted. Solutions were compared using silhouette widths for each locus (Rousseeuw 1987). The value of K which maximized the fraction of loci for which the silhouette width was maximal across the different values of K was selected as optimal.

Ordering of loci within clusters was carried out using MSTmap (Wu et al. 2008a). This method takes a full, undirected graph where nodes are loci and edges are based on the values of $d_{i,j}/n$ and finds the correct order of markers based on the minimum-weighted traveling salesman path (TSP). Wu et al. (2008a) showed that the minimum-weighted TSP can be found using a minimum spanning tree approach and that it corresponds to the correct order of the loci if the minimum spanning tree on the full, undirected graph is unique. We employed MSTmap using the maximum likelihood objective function, grouping turned off, imputation of missing data turned off, and the Kosambi mapping function (Kosambi 1944). The resulting ordering of loci within each cluster, along with the distances (i.e., cM) in each cluster, were taken as the initial linkage map from which data were error-corrected and imputed.

Data were subsequently imputed and corrected for errors using Maskov (Ward et al. 2013). A full account of the mechanics used in the algorithm of Maskov can be found in Text S1 from Ward et al. (2013). For our purposes, the accuracy of the imputation and error correction depends upon two choices: (1) the threshold for missing data for a given megagametophyte and (2) the number of contiguous loci where genotype errors can occur. We chose a value equal to 90% for the amount of missing data across megagametophytes for the former and a value of 5% of the number of loci in the initial map for each cluster for the latter (cf., Ward et al. 2013).

A final round of ordering was conducted with the imputed and error-corrected data using MSTmap as described previously. Imputation and error correction resulted in many loci where $d_{i,j}/n = 0$. These co-segregating markers were thus mapped to the same bin (Wu et al. 2008a). The collection of resulting ordered clusters was taken as the final linkage map for each of the four maternal trees. The end result of the linkage analysis was thus four independent linkage maps, one per maternal tree.

Consensus map construction and biological interpretation

We took a two-step approach to the inference of the consensus linkage map. First, the four linkage maps, one for each maternal tree, were combined into a framework linkage map using MergeMap (Wu et al. 2008b). We constructed a set of weights with which to rank SNP orderings from each map as more or less probable based on the average amount of missing data, where a higher weight meant that the genotype data used to infer the linkage map had fewer instances of missing data (red: 0.05, green: 0.40, blue: 0.15, yellow: 0.40). Second, the remaining SNPs were added to the framework map by using the weighted average of the observed recombination fractions across libraries and constructing a linkage map as described previously based on these weighted average values. Consistency in the positioning and relative distances among framework markers was assessed using Spearman (1904) and Mantel (1967) correlations. Specifically, pairwise distances (cM) among framework markers were extracted from each linkage group on the framework map built using MergeMap as well as the map resulting from use of the weighted average recombination fractions in MSTmap. A Mantel correlation was used to test the null hypothesis that these distances were not correlated using a Bonferroni-corrected significance threshold of $\alpha = 0.05$. Separate tests were performed for each of the 12 linkage groups. All analysis was conducted in the R ver. 3.0.2 statistical computing environment (R Core Team 2013).

Framework markers on the resulting consensus linkage map were used to estimate the size (Chakravarti et al. 1991) and coverage (Lange and Boehnke 1982) of the foxtail pine genome.

The contigs from the assembly used to discover SNPs that appeared on the consensus linkage map were annotated using BLAST tools (Altschul et al. 1990) and the most recent release of the loblolly pine (*Pinus taeda* L.) genome sequence (v. 1.01, annotation V2). Each contig from the assembly was queried against the set of scaffolds comprising the loblolly pine genome using BLASTN. The hits from each comparison were retained, and these top hits were filtered based on query coverage and the percent identity. As thresholds, we used a minimum of 50 % for the query coverage and 75% for the percent identity. The percent identity for the query coverage was set according to the expected number of substitutions between two sequences ($2\mu t$, see Nei 1987), where the mutation rate (μ) was assumed to be 1×10^{-9} substitutions/site/year and the divergence time (t) was assumed to be 8×10^{-7} years (Willyard et al. 2007). This translated into an average expectation of 16% divergence between any two DNA sequences of loblolly and foxtail pines. We rounded down to 75% to account for a portion of the variance around this expectation. Hits that exceeded these thresholds were transferred as annotations, as obtained from the annotation gff files, to the contig appearing on the consensus linkage map for foxtail pine. The resulting GO annotations were visualized and analyzed with ReviGO (Gene Ontology monthly release 10/2014; UniProt-to-GO mapping 9/30/2014) (Supek et al. 2011) allowing a similarity of 50 % across terms.

Results

DNA sequence analysis

The raw number of reads varied across libraries from a minimum of 71,834,280 (red) to a maximum of 206,365,836 (green), with an average of $153,082,376 \pm 49,855,941$. All raw reads were either 102bp (green, yellow) or 110 (red, blue) in length, depending on sequencing technology (HiSeq 2500 vs 2000, respectively). In general, the libraries run on the Illumina HiSeq 2500 platform had a 1.65-fold greater number of reads than those run on the Illumina 2000 HiSeq platform. Processing of reads for quality reduced these numbers by approximately 1.66-fold, with a range of a 2.56-fold (red) to a 1.33-fold (yellow) reduction. After filtering, the average length of

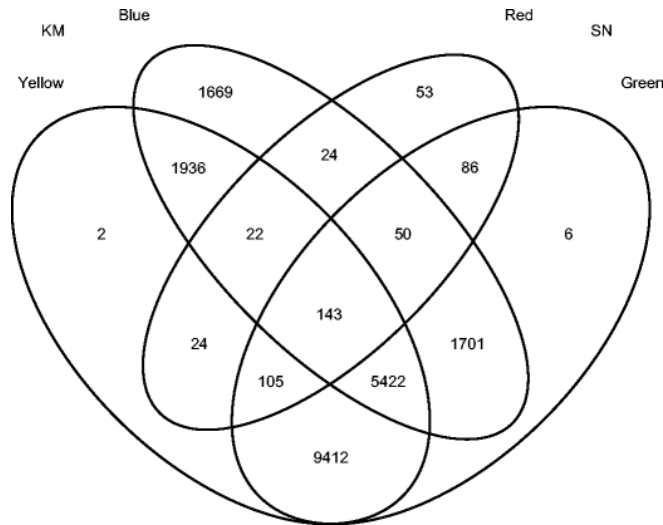


Figure 4.2. Sharing of contigs across maternal tree maps from which the consensus map was constructed. Counts in each cell represent the number of unique contigs appearing on the final consensus map. Unique contigs for the yellow and green maternal trees were largely discarded to make estimation of pairwise recombination fractions computationally feasible (see Materials & Methods section). KM Klamath Mountains, SN Sierra Nevada.

reads was 88 ± 13 bp, with a range of 40 to 102 bp across libraries. The number of quality-filtered reads per megagametophyte also varied 19,741-fold ($\pm 27,069$ -fold) on average across libraries, with average minimums of 753 ± 603 bp to average maximums of $3,421,571 \pm 2,070,990$ bp. After quality filtering, this translated into an average total of $8,137,663,036 \pm 3,658,147,958$ bp generated per library, ranging from 2,436,531,265 bp (red) to 11,643,165,529 bp (yellow).

The largest number of reads ($n = 6, 838, 986$) was obtained for a single megagametophyte in the green library. These reads were used to create an assembly against which all other data were mapped for SNP calling and genotype determination. Optimization of assembly parameters ($k = 31$, $\ln(L) = -110.071$) resulted in an assembly of 231,053 contigs, with an average length of 89 ± 12 bp per contig (range, 61 to 312 bp), and an average per-contig base coverage of 4.5X to 20.0X (range, 1.5X to 5069X). This assembly represented approximately 0.07% of the genome of foxtail pine, which was assumed to be approximately 30 Gbp in size (Murray 1998).

Using this assembly, 349,542 putative SNPs were called (Table 4.1). These 349,542 SNPs were located in 83,051 unique contigs (35.94 % of the total), with a mean of 4 SNPs per

Attribute	Yellow	Blue	Red	Green
Region	Klamath	Klamath	Sierra Nevada	Sierra Nevada
Latitude	44.7483	41.1959	36.4481	36.4481
Longitude	-123.1332	-122.7922	-118.1706	-118.1706
Illumina platform	HiSeq 2500	HiSeq 2000	HiSeq 2000	HiSeq 2500
No. of megagametophytes	95	95	76	73
Reads (total)	174,516,834	159,612,555	71,834,280	206,365,836
Reads (filtered)	131,540,433	80,498,688	28,041,978	129,396,774

Table 4.1. Attributes of the data structure related to maternal tree.

contig (range, 1 to 32). Filtering these SNPs by expected segregation patterns, consistency with heterozygous calls for the pseudo-diploid sample, and minimum sample sizes for genotype calls, resulted in 983, 34261, 21594, and 35304 SNPs for the red, green, blue, and yellow libraries, respectively. The vast majority of SNPs eliminated were for violation of the 1:1 expected pattern of segregation (259,801–268,621), with approximately 95% of these dropped SNPs shared across families. The counts for the yellow and green libraries were also trimmed so as to remove all but a handful ($n = 2$ for the yellow and $n = 6$ for green libraries) of the unique contigs not found as polymorphic in the other libraries. This was done to facilitate the efficiency of the calculation of pairwise recombination fractions. These SNP counts represented 507, 16925, 10967, and 17066 contigs for the red, green, blue, and yellow libraries, respectively. Patterns of shared polymorphic contigs, as well as SNPs, were as expected given the among-region magnitude of genetic differentiation (Figure 4.2, see Eckert et al. 2008), with libraries comprised of megagametophytes sampled from maternal trees located in the same geographical area sharing more polymorphic contigs and SNPs than comparisons of maternal trees from different geographical regions (nonparametric permutation analysis: $P < 0.0001$, see Supplemental Text). On average, megagametophytes in the filtered data set had 79.40% ($\pm 14.7\%$) missing data (i.e., a missing haploid genotype) across SNPs (range, 1.3 to 99.8%), with the green library having the smallest

(74.3 ± 18.9%) and the red library having the largest average amount of missing data per megagametophyte (84.4 ± 15.2 %).

Linkage mapping

Individual linkage maps were constructed for each maternal tree separately using an iterative approach based on imputation. All filtered SNPs for each maternal tree, regardless of being located in the same contig, were assessed for patterns of linkage followed by grouping and ordering of SNPs. Redundant SNPs were filtered post hoc and used to test for biases in our analysis pipeline.

Grouping of pairwise recombination fractions via hierarchical clustering was consistent with 12 linkage groups. This corresponded to a minimum pairwise LOD score of approximately 5.5 for each maternal tree for markers to be placed within the same linkage group. Inspection of the distribution of silhouette values for values of K ranging from 8 to 16 revealed that $K = 12$ was the best clustering solution for each of the four maternal trees (Figure S4.1). This was confirmed by comparison of pairwise LOD scores for SNPs within versus among the 12 linkage groups. Comparisons within linkage groups were on average 3.2-fold larger than among linkage groups, which was significantly greater than expected randomly ($n = 1000$ permutations/maternal tree, $P < 0.015$).

Marker ordering within putative linkage groups using MSTmap resulted in extremely long linkage maps (e.g., >50,000 cM) for each maternal tree. This translated into an average number of recombination events which exceeded 100 per megagametophyte. This pattern is consistent with problems of inference due to missing data and genotyping errors (Ward et al. 2013). To verify this assumption, data for the blue library were split into two sets of 35 megagametophytes—those with the least amount of missing data and those with the largest amount of missing data. As expected, the inferred recombination distances were 3.5-fold smaller for the maps inferred using the megagametophytes with less missing data. Thus, we followed the approach of Ward et al. (2013) to impute and error correct data based on our initial marker orderings.

Attribute	Yellow	Blue	Red	Green	Consensus
Contigs	17,066	10,967	507	16,925	20,655
Positions	728	1101	296	839	901
Contigs/position	23	10	2	20	23
Total bp mapped	1,596,325	1,025,088	47,222	1,583,269	1,931,700
Total length (cM)	1037.40	1263.46	1572.80	1287.48	1192.10
cM/position	1.43	1.15	5.31	1.53	1.32

Table 4.2. Attributes of single-tree and the consensus linkage maps. Values for ratio variables are totals and are not averaged across linkage groups (see Tables S4.1–S4.5).

Imputation and error correction of genotype data for each linkage group for each maternal tree was carried out using Maskov. This process drastically reduced the number of inferred recombination events, including double crossovers, from >100 per megagametophyte to approximately 1 to 2 per megagametophyte. This reduction was controlled by setting a parameter in Maskov so as to produce a number of recombination events per megagametophyte that mirrored those observed previously for linkage mapping within conifers (Eckert et al. 2009; Martínez-García et al. 2013). Changing this parameter had no effect on the downstream ordering of SNPs within linkage groups but only changed the spatial resolution of the resulting linkage map.

The resulting linkage maps for each maternal tree were aligned manually based on the presence of shared contigs. Overall, there was excellent agreement among maps, with only 115 SNPs being mapped to conflicting linkage groups across maternal trees. All 115 SNPs with conflicting group assignments were unique to the red library. These were dropped from further consideration. Within linkage groups, SNPs present in multiple libraries were ordered similarly (pairwise Spearman's $\rho > 0.956$, $P < 0.001$), with conflicting orderings having average differences of 5.91 cM (± 5.64 cM). Inferred linkage maps for each maternal tree also resulted in SNPs from the same contig largely being mapped to the same position, with an average of only 5.8% of SNPs from the same contig being mapped to a different position. Approximately 94 % of the time, these different positions were adjacent on the linkage map. For those SNPs from the same contig that did not map to the same position, the average difference in positioning was 1.64 cM (± 3.01 cM),

with no instances of SNPs from the same contig being located on different linkage groups. We thus pruned multiple SNPs per contig by randomly selecting one SNP per contig from the data set and re-estimated the linkage maps for each maternal tree as described previously. The resulting four linkage maps were taken as the final estimates of linkage relationships among polymorphic contigs in each of the four maternal trees.

The final four linkage maps varied in total length from 1037.40 to 1572.80 cM, with an average of 1290.29 cM (± 219.5 cM; Tables 4.2, S4.1, S4.2, S4.3, and S4.4; Figures 4.3, S4.2, S4.3, S4.4, S4.5, S4.6, and S4.8). In total, 20,655 unique contigs representing 1,931,700 bp of DNA were mapped to a position within at least one linkage map. The number of contigs varied 33.66-fold across linkage maps, with a minimum of 507 (red) to a maximum of 17,066 (yellow). These contigs were organized into an average of 741 (± 335) unique positions, separated on average by 1.77 cM (± 2.36 cM), across linkage maps, with the fewest number of unique positions observed in the linkage map for the red maternal tree ($n = 296$) and the largest number in the linkage map for the blue maternal tree ($n = 1101$). With respect to average distances between adjacent positions, the linkage map for the red maternal tree had the largest (5.53 ± 6.11 cM), while that for the blue maternal tree had the lowest (1.16 ± 0.77 cM). This translated into an average of 15 (± 27) contigs per position on average, with the linkage map for the red maternal tree having the fewest contigs per position on average (2 ± 2) and the linkage map for the yellow maternal tree the most contigs per position on average (27 ± 35). Contigs were also non-randomly distributed across positions for all linkage maps except that for the red maternal tree ($P < 0.0001$, see Supplemental Text), with elevated contig counts typically occurring at the ends of linkage groups (Figure 4.3).

Consensus map construction and biological interpretation

A set of 507 framework SNPs were devised from those contigs shared across at least three of the four linkage maps. These 507 SNPs were used to construct a framework map using MergeMap. The resulting linkage map had an overall length of 1572.80 cM. Comparison of this

map with those for each maternal tree revealed a strong similarity in positioning for each linkage group (Spearman's $\rho > 0.98$, $P < 0.001$). Using an expanded set of SNPs present in at least three families, two of which had to be the green and yellow families, confirmed these patterns, with pairwise correlations among maps on the order of 0.92 or greater. Given this overall similarity, we incorporated the remaining markers into the map by using weighted averages of observed pairwise recombination fractions across maternal trees and inferred a consensus linkage map as outlined previously. Inferred marker positions and distances for the framework markers were highly correlated across linkage groups in this map relative to that inferred using MergeMap and only the framework markers (Mantel's $r > 0.95$, $P < 0.001$). We used this as evidence in support of our approach, and the inferred consensus linkage map was taken as the final consensus estimate of linkage relationships for the 20,655 unique contigs located in the four maternal tree linkage maps.

As with the individual tree maps, $K = 12$ linkage groups was most consistent with the averaged data. This corresponded to a minimum pairwise LOD score of approximately 5.5 for each maternal tree for markers to be placed within the same linkage group. The consensus linkage map was 1192.00 cM in length, with linkage groups varying in length from 88.44 to 108.76 cM (Table S4.5, Figures 4.3, S4.9). There were 901 unique positions across the 12 linkage groups for this map, so that the average number of contigs per position was 23 (± 35). These 901 positions were separated on average by 1.34 cM (± 0.50 cM). As with the individual maternal tree linkage maps, contigs were non-randomly distributed across positions ($P < 0.0001$, see Supplemental Text), with notable enrichment at the ends of inferred linkage groups. Using the 507 framework SNPs and the final consensus linkage map, the estimated genome size of foxtail pine is 1276.04 cM (95% confidence interval, 1174.31–1377.77 cM). As such, the estimated coverage of the genome is 98.58% (LOD threshold = 5.5; maximum distance among adjacent framework markers, 13.4 cM; number of framework markers, 507; $K = 2694$).

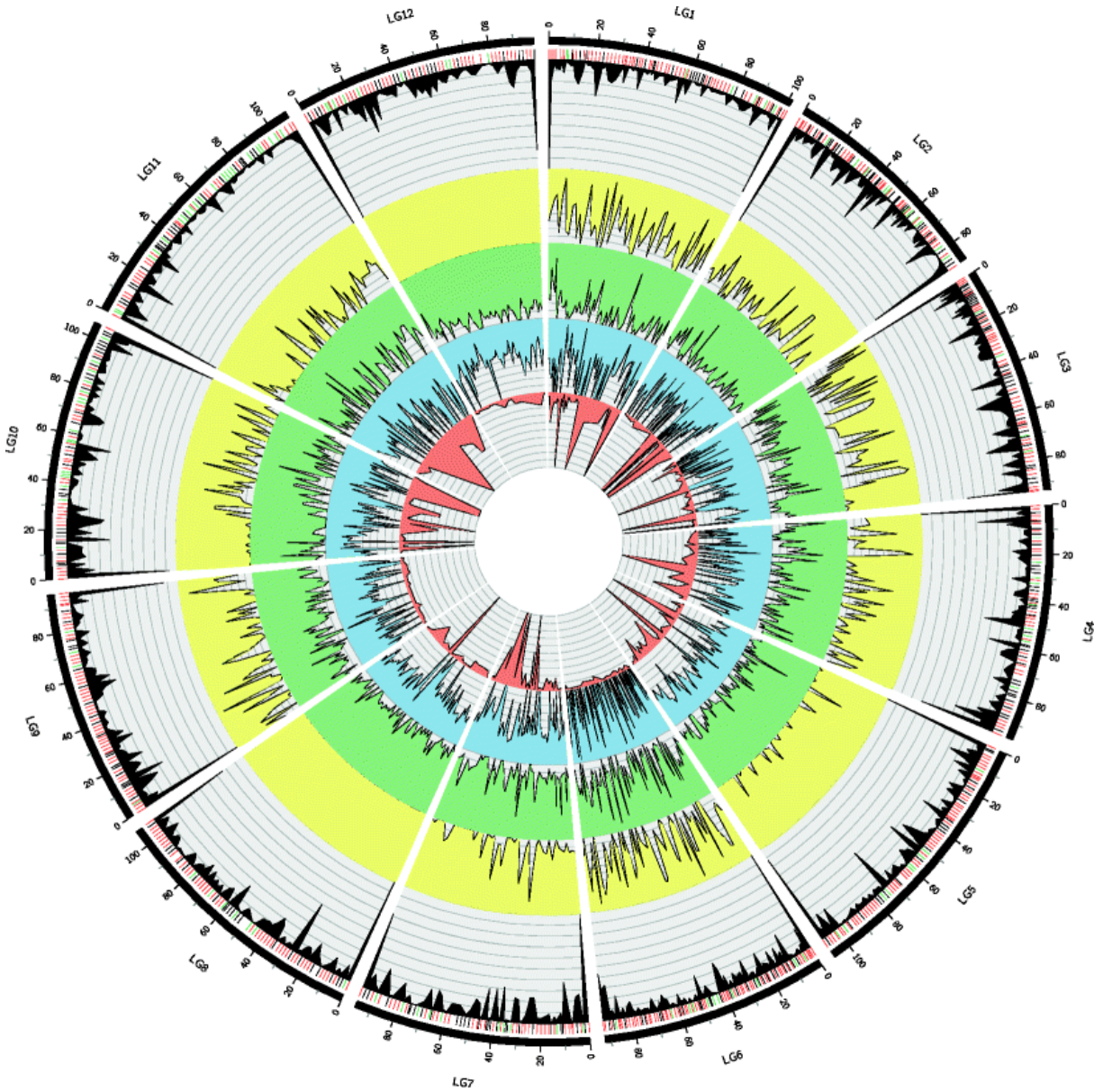


Figure 4.3. Consensus linkage map of 12 linkage groups, derived from SNPs among individuals of four populations. Working inward from the outermost section of the figure, for each linkage group: (1) the solid black bars represent the span of recombination distances (in cM) for markers; (2) the individual tick marks show the locations of the markers, and the colors represent the density of annotation of the SNPs at that position ($\geq 50\%$ = green, $\geq 25\%$ = red, $< 25\%$ = black) to homologous locations in lobolly pine; (3) The black density plot represents the counts of SNPs from all four families mapping to a specific position in the linkage group; (4) the colored density plots show the contribution SNPs from the individual families to the markers on the map at each position and are shown in order by total read count in the library, with yellow having the most and red having the least amount of reads. Linear plots of linkage groups comprising the consensus map are given in Figure S4.9.

Of the 20,655 contigs in the reference assembly which contained SNPs, 5627 (27.2%) contained BLASTN hits ($n = 5853$) which passed the filtering threshold of 50% query length and

75% identity. The averages of query length (bp), query coverage percentage, and identity percentage were 94 ± 7 bp, 86.4 ± 13.0 %, and 90.1 ± 3.9 %, respectively. We found 2802 (48.9%) instances of SNPs mapping to putative genic regions in the loblolly pine genome ($n = 587$ scaffolds) representing 303 unique GO terms. More detailed annotation information (e.g., InterPro IDs and GO terms) can be found in the supplemental file S1.

Discussion

The genetic architecture of fitness-related traits has been a major focus of geneticists for over a century (reviewed by Ellegren and Sheldon 2008). Early efforts to understand the genetic architecture of fitness-related traits focused primarily on the number and effect size of the loci underlying heritable, phenotypic variation (Fisher 1918). Recent work has extended this line of research, with a multitude of studies linking phenotypic with genetic variation through linkage mapping, both within pedigrees (Mauricio 2001; Neale and Kremer 2011; Ritland et al. 2011) and within natural populations (Ingvarsson and Street 2011; Eckert et al. 2013), or through quantitative genetic experimentation (Anderson et al. 2014, 2013; Fournier-Level et al. 2013). Relatively little empirical work outside of model organisms, other than polyploidization or the characterization of genomic islands of divergence (e.g., high F_{ST}), has focused on the genomic organization of loci contributing to fitness differences among individuals (but see Stevison et al. 2011). This is despite clear theoretical predictions relating the evolution of the genetic architecture underlying fitness-related traits to the genomic organization of the loci comprising this architecture (Kirkpatrick and Barton 2006; Yeaman and Whitlock 2011; Yeaman 2013; Akerman and Bürger 2014).

Here, we have provided a high-density linkage map representing over 20,000 unique contigs distributed throughout the 30 Gbp genome of foxtail pine that can be used to aid in the discovery and study of loci contributing to local adaptation. To our knowledge, it represents one of the most dense linkage maps ever produced within forest trees, although the number of unique positions is much less than the number of mapped contigs (i.e., about 1/20th). Approximately 25%

of these contigs had significant similarity to sequences within the draft loblolly pine genome. Importantly, our markers are dispersed in both genic and non-genic regions of the genome. The latter are often ignored in studies of local adaptation utilizing markers based on sequence capture (e.g., Neves et al. 2014) and SNP arrays (e.g., Eckert et al. 2010b), yet it is known that non-genic regions are often involved with adaptation (e.g., Studer et al. 2011). This linkage map, moreover, was created using affordable next-generation sequencing technologies in combination with freely available methods of analysis, which highlights the feasibility of this approach to non-model conifers, where full genome sequencing and assembly are still not quite feasible given realistic research budgets. With regard to map integrity, recombination fractions for pairs of SNPs segregating in multiple trees were highly correlated (Mantel's $r > 0.90$ for all comparisons across linkage groups). This allowed for the creation of a robust consensus linkage map, as well as highlighted the biological signal of linkage apparent even in noisy ddRADseq data. When coupled with the other biological signals in our results (e.g., trees from the same regional population sharing SNPs more often), we can be confident that our inferred linkage maps are based primarily on biological, as opposed to statistical, signals. In further support of this claim, randomly subsampling our data to represent 10,000 contigs and performing linkage mapping as described previously resulted in a consensus linkage map that was indistinguishable from that pictured in Figure 4.3 (Spearman's $\rho = 0.997$, $P < 0.0001$).

The linkage map produced here is valuable in numerous ways. First, it provides a dense resource for quantitative trait locus (QTL) mapping. Our next step using this linkage map is to link fitness-related phenotypic variation with genotypes at mapped markers. We are currently mapping QTLs for $\delta^{13}\text{C}$ to accomplish this goal (cf., Hausmann et al. 2005). Importantly, this will represent one of the first QTL maps in the clade of soft pines outside of section *Quinquefoliae*. Second, the framework provided here is optimal for imputation and phasing of data during population genomic inferences utilizing samples from natural populations (Scheet and Stephens 2006). Third, our linkage map is the foundation upon which theoretical expectations can be tested.

For example, the theory of Yeaman and Whitlock (2011) predicts that the loci contributing to local adaptation should be differentially clustered in genomes as a function of rates of gene flow among populations. Magnitudes of gene flow among stands differ dramatically within the regional populations of foxtail pine, and pairwise plots of synteny across maternal trees revealed several instances of differential marker orderings between regional populations consistent with areas of the genome with structural differences (see Figures S4.7 and S4.8). These areas, however, were the exception, as marker orderings across trees were highly correlated. Fourth, knowledge about the physical ordering of loci allows patterns of linkage disequilibrium (LD) within natural populations to be better characterized. The role of LD in local adaptation has long been recognized (see Akerman and Bürger 2014), yet empirical studies of its role are difficult without some knowledge of physical relationships among loci. This is because LD among physically linked markers is expected to some degree, whereas LD among physically unlinked markers must have originated from some evolutionary process (e.g., genetic drift, natural selection, migration). Patterns of LD across non-genic regions of pine genomes are currently unknown (but see Moritsuka et al. 2012), so additional data in combination with the linkage map provided here allow for rigorous investigations of these patterns. Lastly, continued production of linkage maps across the Pinaceae will aid comparative genomics and evolutionary inference through study of synteny and the evolution of genome structure (Ritland et al. 2011; Pavy et al. 2012).

Despite numerous indicators of biological signals dominating our dataset, caution is still needed when interpreting our results. First, we used novel analysis methods that have not been tested using simulations. For example, the form of hierarchical clustering used here is not employed to our knowledge in any of the available software packages used for linkage mapping. Its utility on data of smaller or larger sizes than that presented here is unknown. Consistency of results across maternal trees, however, indicates that our methods are likely appropriate for our data (see also Tani et al. 2003). Second, error correction and imputation were used, which could have affected marker ordering and distances. Marker order, however, did not change with

increasing stringency of error correction. Only marker distances changed with increased stringency, thus creating clumped distributions of makers. This was also apparent in the total map length, which is at the lower end expected for conifers (cf., Ritland et al. 2011), which is indicative of being conservative with error corrections. The effect of marker clumping on downstream uses of this linkage map, however, is likely to be minimal (e.g., bias in QTL intervals), as this bias would affect QTL size and not necessarily inference of QTL presence or absence. The relative importance of imputation and error correction is to some degree affected by experimental conditions. We did not standardize the total amount of DNA for each megagametophyte prior to construction of libraries (concentration ranges, 10 to > 50 ng/ μ l), which likely affected the average 19,741-fold variation in the number of reads across megagametophytes. Future studies would benefit from considering this prior to library construction. Related to this issue was the poor performance of the red family. In general, the library for this family exhibited signs of low sequence quality, with it having the largest fraction of reads eliminated during quality filtering (Table 4.1), the largest fraction of pseudo-diploid genotypes called as homozygous or missing (99.5%), and the largest fraction of loci deviating from expected segregation patterns (76.8%). This is consistent with lower overall coverage driven by low-quality sequence data, which could have resulted from any of the numerous laboratory steps during creation of the multiplexed libraries (i.e., DNA extractions, restriction digests, ligation, and PCR). Third, we used a form of hierarchical clustering that required the number of groups to be defined subjectively. Post hoc analysis indicated that our clustering solution corresponded to a LOD threshold of approximately 5.5 and that 12 was an optimal number of groups (Figure S4.1). Selection of a larger or smaller number of groups, moreover, did not change marker ordering within groups substantially. Typically, changing K to a larger value broke existing linkage groups into more pieces, whereas changing K to smaller values merged existing linkage groups. Marker orderings within these broken or merged groups, however, did not change. Fourth, we did not explicitly quantify error rates. In theory, error rates can be calculated from summaries derived from the `mpileup` in `samtools`. Given the relatively

low coverage and limited reference assembly for this species (cf., Nystedt et al. 2013; Neale et al. 2014), estimation of error rates would likely be biased. Thus, we preferred to acknowledge the presence of errors, as indicated by the extremely long initial single-tree linkage maps, and use statistical methods to minimize their influence. As next-generation data accumulate for this species, and other conifers in general, precise estimation of error rates will become feasible. Lastly, our sample sizes were not large enough to resolve linkage relationships beyond distances of approximately 1.0 cM (but see Neves et al. 2014). Increased number of sampled megagametophytes would have allowed higher resolution, which could aid in downstream uses of our linkage map. Despite this level of resolution, however, we have produced one of the densest linkage maps to date for a forest tree species (Eckert et al. 2010a; Martínez-García et al. 2013; Neves et al. 2014).

Conifer genomics is emerging as a mature scientific field (Mackay et al. 2012). Draft sequences of genomes and transcriptomes for several species have been released and more are planned. As shown here, production of high-density linkage maps is a fruitful endeavor to accompany this maturation. The results presented here are promising and also provide guidance for future attempts in additional species. Specifically, linkage maps provide ample information about genomic structure that is needed for the study of local adaptation in natural populations (cf., Limborg et al. 2014). Here, we have produced a high-density linkage map for foxtail pine using methods applicable to any non-model conifer species, thus opening the door for further studies of genome structure and the genetic architecture of local adaptation in this rather understudied clade of pines, as well as the Pinaceae as a whole.

Acknowledgments

The authors would like to thank the staff at the USDA Institute of Forest Genetics, the VCU Nucleic Acids Research Facility, and the VCU Center for High Performance Computing. In addition, we would like to thank Tom Blush and Tom Burt for help in obtaining seeds. Funding for this project was made available to AJE via start-up funds from Virginia Commonwealth University. CJF was supported by the National Science Foundation (NSF) National Plant Genome Initiative (NPGI): Postdoctoral Research Fellowship in Biology (PRFB) FY 2013 Award no. NSF-NPGI-PRFB-1306622.

Chapter 5.

Local adaptation and linkage maps II: The QTL landscape of water-use efficiency in *Pinus balfouriana*

This work has been made publicly available in the following preprints:

Eckert AJ, Harwood DE, Lind BM, et al. (2016) The genetic architecture of local adaptation II: The QTL landscape of water-use efficiency for foxtail pine (*Pinus balfouriana* Grev. & Balf.). bioRxiv 1–52. doi: <https://doi.org/10.1101/038240>

Abstract

Water availability is an important driver of the geographic distribution of many plant species, although its importance relative to other climatic variables varies across climate regimes and species. A common indirect measure of water-use efficiency (WUE) is the ratio of carbon isotopes ($\delta^{13}\text{C}$) fixed during photosynthesis, especially when analyzed in conjunction with a measure of leaf-level resource utilization (Nitrogen content, $N_{\mu\text{g}}$). Here, we test two hypotheses about the genetic architecture of WUE for foxtail pine (*Pinus balfouriana* Grev. & Balf.) using a novel mixture of double digest restriction site associated DNA sequencing, species distribution modeling, and quantitative genetics. First, we test the hypothesis that water availability is an important determinant of the geographical range of foxtail pine. Second, we test the hypothesis that variation in $\delta^{13}\text{C}$ and $N_{\mu\text{g}}$ is genetically based, differentiated between regional populations, and has genetic architectures that include loci of large effect. We show that precipitation-related variables structured the geographical range of foxtail pine, climate-based niches differed between regional populations, and $\delta^{13}\text{C}$ and $N_{\mu\text{g}}$ were heritable with moderate signals of differentiation between regional populations. A set of large-effect QTLs ($n = 11$ for $\delta^{13}\text{C}$; $n = 10$ for $N_{\mu\text{g}}$) underlying $\delta^{13}\text{C}$ and $N_{\mu\text{g}}$ variation, with little to no evidence of pleiotropy, was discovered using multiple-marker, half-sibling regression models. Our results represent a first approximation to the genetic architecture of these phenotypic traits, including documentation of several patterns consistent with $\delta^{13}\text{C}$ being a fitness-related trait affected by natural selection.

Introduction

Descriptions of the genetic components underlying fitness-related phenotypic variation have been a focus of quantitative genetics for over a century (Shull 1908; Fisher 1918; Mather 1941; Ford 1975; Mackay et al. 1994; Ritland et al. 2011 and references therein). These descriptions have progressed from identifications of the genetic elements affecting trait variation (e.g. Jermstad et al. 2001) to analysis of interactions among these elements with one another and the environment (e.g. Jermstad et al. 2003). Uniting all these descriptions are foundational questions about the structure, function, and evolution of genotype-phenotype maps in natural populations. For forest trees, these descriptions historically addressed traits of economic importance such as specific gravity of wood (e.g. Groover et al. 1994), microfibril angle (e.g. Sewell et al. 2000), growth (e.g. Wu 1998), and phenology (e.g. Pelgas et al. 2011), with the ultimate goals of marker-assisted breeding (Neale and Savolainen 2004) and trait prediction from genotypic data (Grattapaglia and Resende 2011). These traits, while economically important, often also affect fitness (especially phenology, see Sorensen 1983), so that these efforts can also be leveraged to understand the genetic basis of ecologically relevant trait variation. The linkage between traits measured in common gardens and fitness in natural populations, however, is usually assumed post hoc, which can lead to storytelling (Barrett and Hoekstra 2011) and oversimplification of the ecological ramifications of quantitative genetic results. Here, we address this disconnect through simultaneous use of species distribution modeling and quantitative trait locus (QTL) mapping to dissect the genetic architecture of an ecologically important phenotypic trait for foxtail pine (*Pinus balfouriana* Grev. & Balf).

The spatial and temporal distribution of all viable individuals across the Earth's landscape for a given species is defined as its geographical range (Brown et al. 1996). Evolution of range sizes and structural attributes of these ranges have been studied for a variety of taxa for many decades (e.g. Mayr 1963; Antonovics 1976; Brown et al. 1996; Gaston 2003; Eckert et al. 2008; Sheth and Angert 2014). The common thread underlying these interests is the assumption that

fitness of individuals within species is related to the known geographical range for each species based on the environments defined by this range, other selective pressures (i.e. competition) across this range, and the phylogeographic history that resulted in the current geographical range (Hutchinson 1957; Pulliam 2000; Chuine 2010). For example, relative fitness values within plant populations tend to be highest in their home environments and lower in novel environments at the margin or outside of known geographical ranges (reviewed in Leimu and Fischer 2008). Regardless of the relationship between this pattern and evolutionary concepts such as local adaptation, it is clear that current geographical ranges are to some degree projections of ecological niches (i.e. realized versus fundamental niches), or at least some aspect of these niches, onto geographical space (Pulliam 2000; Ettinger et al. 2011). Knowledge of the environmental and climatic drivers of geographical ranges can therefore be informative about links between traits responsive to these drivers and fitness.

Species distribution models (SDMs) are commonly utilized as predictive tools with which to assess the importance of environmental variables to current geographical ranges of species (Elith et al. 2006). At a minimum, these models are built from known occurrences of a certain species and the environmental and ecological attributes of these locations derived from either field measurements or information stored in geographical information systems (GIS) layers. Numerous approaches are available with which to build models from these data (Segurado and Araujo 2004; Elith et al. 2006; Phillips et al. 2006). Once constructed, SDMs are often used subsequently to study the evolutionary development of ranges (e.g. McCormack et al. 2010), as well as the effects of continued climate change on current geographical ranges (e.g. Pearson and Dawson 2003). However, there are limitations to equating SDMs, even those with good predictive abilities of current geographical ranges, with realized ecological niches and hence measures of fitness limits (Hampe 2004; Soberon and Peterson 2005; Warren and Seifert (2011). For example, individuals used to create SDMs are considered exchangeable, so that fitness variation among individuals is ignored (Hampe 2004). Some of these issues, especially those related to

exchangeability of individuals within species, can be addressed through a careful matching of modeling units (e.g. genetically differentiated populations within species; *sensu* Davis et al. 2005), geographical scale (e.g. the geographical scale relevant to the genetically differentiated populations), and the research questions of interest.

Water is crucial to the survival of all plant species (e.g. Sorenson 1983), although its importance relative to other environmental factors varies depending upon the environmental factors that are most limiting within local environments (Dudley 1996). The intrinsic efficiency by which plants use water (WUE) is defined as the ratio of net assimilation of carbon from CO₂ during photosynthesis to the loss of water during transpiration (Bacon 2004). Carbon isotopic composition ($\delta^{13}\text{C}$) is an indirect measure of intrinsic WUE and is based upon the ratio of two isotopes of carbon (¹³C and ¹²C) within plant tissue standardized to a reference. This ratio is related to WUE because it has been demonstrated that the discrimination by C₃ plants of ¹³CO₂ relative to ¹²CO₂ is correlated to the ratio of carbon assimilation during photosynthesis to stomatal conductance (Farquhar et al. 1982; Farquhar and Richards 1984; e.g. Zhang and Marshall 1994). The physiological and environmental mechanisms, however, driving the linkage between $\delta^{13}\text{C}$ and intrinsic WUE at various levels of biological organization are numerous, so that the expected linear relationship between $\delta^{13}\text{C}$ and WUE may not always hold (Seibt et al. 2008). For example, differences in $\delta^{13}\text{C}$ across individual plants at the leaf level can result from changes in carbon to nitrogen allocation during carboxylation, variation in leaf structure and morphology, and/or variation in available CO₂ (Seibt et al. 2008). Within a common environment, however, it is assumed that variation in available amounts of atmospheric CO₂ is negligible. Variation for $\delta^{13}\text{C}$ across individual plants in these common environments should therefore reflect variation for intrinsic WUE. Indeed, previous research in conifers has established that variation in $\delta^{13}\text{C}$ across individual plants is heritable (Seiler and Johnsen 1988; Cregg 1993; Brendel et al. 2002; Baltunis et al. 2008; Cumbie et al. 2011), is polygenic, yet comprised of a mixture of large and small effect loci (Brendel et al. 2002; Gonzalez-Martinez et al. 2008; Cumbie et al. 2011; Marguerit et al.

2014), and that it often reflects variation for intrinsic WUE through leaf level assimilation (Zhang and Marshall 1994; Brendel et al. 2002; Cumbie et al. 2011; Marguerit et al. 2014).

Water availability is an important driver of tree distributions (Stephenson 1990 and references therein), especially in Mediterranean climates (e.g. Baldocchi and Xu 2007; Lutz et al. 2010). This importance is evident through increased tree mortality as a function of both direct and indirect consequences associated with changing water availability (van Mantgem et al. 2009; Allen et al. 2010). Regional and local water availability will likely be altered, either through changes to annual precipitation totals or the seasonality of precipitation, under most climate change scenarios, especially in ecosystems dependent on residual summer snow-packs (Barnett et al. 2005). The ability of natural populations of forest trees to respond to changing water availability is linked to segregating genetic variation for traits responsive to water availability (Aitken et al. 2008). Knowledge of the genetic architecture of such traits, therefore, provides an important resource for assessing forest health, as well as the genetics of adaptation (Neale and Kremer 2011). Here, we test two hypotheses about the genetic architecture of WUE for foxtail pine – (i) water availability is an important determinant of the geographical range of foxtail pine and hence fitness and (ii) variation in $\delta^{13}\text{C}$ and N_{pg} is genetically based, differentiated between regional populations, and has genetic architectures that include loci of large effect. We subsequently discuss how the integration of results from disparate fields of research (i.e. genomics, ecology, and quantitative genetics) provides information useful to foundational tests about the genetic architecture of local adaptation and its evolution (cf. Friedline et al. 2015).

Materials and methods

Focal species

Foxtail pine is one of three species classified within subsection *Balfourianae* of section *Parrya* within subgenus *Strobus*. It is generally regarded as the sister taxon to Great Basin bristlecone pine (*P. longaeva* D. K. Bailey; see Eckert and Hall 2006). The distribution of this species is relegated to the high elevation mountains of California, with all known occurrences

Table 5.1 Summary of the families ($n=5$) used for QTL mapping. Sibling counts represent the numbers of siblings genotyped and phenotyped for each family. Additional siblings for each family are still growing within the common garden (see Materials & Methods).

	Red	Green	Purple	Blue	Yellow
Latitude	36.448075	36.448075	41.319871	41.195910	41.748267
Longitude	-118.170611	-118.170644	-122.479184	-122.792240	-123.133233
Elevation(m)	3352.80	3352.80	2397.56	2103.12	2103.12
Siblings	35	40	34	40	32
Locality	Cottonwood Pass	Cottonwood Pass	Mt. Eddy	East Lake	Boulder Lake Mountain
Region	SN	SN	KM	KM	KM

being in either the Klamath Mountains of northern California or in the high elevations of the southern Sierra Nevada (Figure S5.1). These two regions are separated by approximately 500 km and differ in climate, soils, and forest composition (Ornduff 1974; Eckert and Sawyer 2002; Barbour et al. 2007).

Common garden

A common garden representing 141 maternal foxtail pine trees was established at the Institute of Forest Genetics (Placerville, CA) during 2011 and 2012 using a randomized block design. Cones were collected from 141 maternal trees sampled range-wide, with 72 sampled from the Klamath Mountains and 69 from the southern Sierra Nevada region. For each maternal tree, 35 – 100 seeds were germinated and grown in standard conditions as outlined in Eckert et al. (2015). More information about the common garden can be obtained from Friedline et al. 2015). Of these 141 maternal trees, offspring, assumed to be half-siblings, from five were selected for analysis of water-use efficiency (see Phenotype determination, Table 5.1). The megagametophyte associated with each germinated seed from these five maternal trees was rescued and used to construct a high-density linkage map based on four of the five maternal trees (Friedline et al. 2015). The seedlings from each maternal tree were allowed to grow for a full year

after which needles were sampled ($n = 32$ to 40 /maternal tree) for determination of phenotypes and genotypes. As done by Friedline et al. (2015), families were named using colors (i.e. these were the colors of family identifier tags in the common garden), with families sampled from the Klamath Mountains being labeled as blue, yellow, and purple and families sampled from the southern Sierra Nevada being labeled as red and green.

Phenotypic determination

Two phenotypic traits were measured from needle tissue sampled from each growing seedling – carbon isotope discrimination ($\delta^{13}\text{C}$) and foliar nitrogen content ($N_{\mu\text{g}}$). These were chosen because ($\delta^{13}\text{C}$) is a proxy for intrinsic WUE (Farquhar et al. 1982; Farquhar and Richards 1984), while $N_{\mu\text{g}}$ is a proxy for plant growth and resource utilization during photosynthesis (Prasolova et al. 2000). Tissue was sampled in year 1 of growth, which was also prior to formation of randomized blocks in the common garden. Given the age of the seedlings, sampling of enough needle tissue for determination of phenotypes and genotypes was destructive. Thus, only a subset of the seedlings per maternal tree was used. For these seedlings, all available needles were sampled, cleaned and separated into those used for genotype determination and those used for phenotype determination. For phenotype determination, needles were placed into a mortar with liquid nitrogen and coarsely ground by hand using a pestle. The resulting needle tissue was then transferred into 20 ml glass vials and oven-dried at 60°C for 96 hrs. Approximately, 2 to 3 mg of ground and dried needle tissue from each seedling was subsequently placed into individual wells comprising a 96 well microtiter plate. Samples were analyzed for $\delta^{13}\text{C}$ and $N_{\mu\text{g}}$ at the Stable Isotope Facility at UC Davis (<http://stableisotopefacility.ucdavis.edu/>). Data are presented as carbon isotope ratios for $\delta^{13}\text{C}$ (‰) and weight for Nitrogen ($N_{\mu\text{g}}$).

Sequence analysis and genotype determination

Total genomic DNA was extracted from the remaining needles from each sampled seedling using Qiagen DNeasy 96 Plant Kits following the manufacturer's protocol. The resulting total genomic DNA for each seedling was quantified using spectrophotometry as implemented

with a Thermo Scientific NanoDrop 8000. Following quantification, samples were prepared for double digest restriction site associated DNA sequencing (ddRADseq) following the protocols of Parchman et al. (2012) as implemented for foxtail pine by Friedline et al. (2015). All samples had concentrations of total genomic DNA in the range of 15 to 60 ng/ μ L. In brief, this protocol proceeds via restriction digests of total genomic DNA for each sample using EcoR1 and Mse1, ligation of adapters that include the Illumina primer, universal M13 primers, and 8 – 10 bp barcodes, PCR amplification, and size selection of the PCR amplified and ligated restriction digests. In our protocol, multiplexing (i.e. pooling) occurred post PCR and size selection was carried out using 1.0% agarose gels run for 1 hour at 110 volts in 1X TAE buffer. All data are based on sequencing fragments in the size range of 300 to 500 bp on the Illumina HiSeq 2500. DNA sequencing was performed at the VCU Nucleic Acid Research Facility (<http://www.narf.vcu.edu/index.html>).

Raw FASTQ sequences were quality-checked and filtered as in Friedline et al. (2015). Briefly, reads must pass a three-stage filtering procedure to be retained for downstream analysis. First, if the average quality for all bases in the read was below 30, the read was discarded. Second, a five-base pair sliding window was evaluated along each raw sequence. Consecutive windows were retained if their mean quality was greater-than or equal-to 30. If the mean score of a window fell below this threshold, the read was trimmed at this point. If the length after trimming was at least 50% of the original read length, the read was kept, otherwise it was discarded. Finally, if 20% of the bases in the original read had quality scores below 30, the entire read was discarded, even if its average quality met the inclusion threshold. The reads that passed quality filtering were demultiplexed and assigned to individual trees in one of five families: Blue, Green, Purple, Red, or Yellow.

Sequences were aligned to the linkage map assembly (Friedline et al. 2015) and read groups were added using Bowtie2 version 2.2.4 (Langmead and Salzberg 2012) using the `–very-sensitive-local` set of options. Each alignment was checked and marked for PCR artifacts using Picard (<http://picard.sourceforge.net, svn 03a1d72>). Variants were called using

species distribution modeling the multiallelic caller from samtools version 1.1 (Li et al. 2009), specifying diploidy for all individuals. The resulting VCF file was processed using VCFtools version 0.1.12.b (Danecek et al. 2011), retaining only biallelic SNPs that mapped to positions on the linkage map defined in Friedline et al. (2015) with quality (`--minQ`) of at least 20. All read processing and variant calling pipeline code, Python 3.4.3 and R version 3.2.0 (R Core Team 2015), can be found as IPython (Pérez and Granger 2007) notebooks and associated files at http://www.github.com/cfriedline/foxtail_wue.

Once genotypes were called for all loci on the linkage map of Friedline et al. (2015), we selected one SNP per position on the linkage map based on minimizing the amount of missing data and being polymorphic in the most families. Missing genotype data were subsequently imputed for each linkage group using the default settings of the program fastPHASE ver. 1.2 (Scheet and Stephens 2006), with families used as populations. To account for uncertainty in genotype imputation, we estimated posterior probabilities of each possible genotype (i.e. 0, 1, or 2) at each locus using 1,000 haplotype reconstructions provided by fastPHASE, which were used subsequently used as weights in a weighted average of the minor allele count. These weighted averages were then rounded to the closest value (0, 1, or 2) following normal rounding rules (i.e. round downward if the tenths position is less than five, otherwise round up).

Species distribution modeling

We used species distribution models (SDMs) to justify water-use efficiency as a fitness-related trait and to quantify niches of each regional population relative to one another. The former provides an a priori justification for the measured traits as ecologically relevant, while the latter provides an estimate of niche differentiation between regional populations comparable to the effect of region on trait differentiation (see Quantitative Genetic Analysis). Species distribution models were used to assess the relative importance of precipitation-related and temperature-related variables to the distribution of foxtail pine. We utilized the approach of maximum entropy (MaxEnt; Phillips et al. 2006) to construct SDMs. Known quantitative genetic analysis locations of

foxtail pine within each regional population ($n = 93$ Klamath Mountains, $n = 207$ southern Sierra Nevada) were gathered from digitized herbarium records available through the Jepson Herbarium located at the University of California, Berkeley (<http://ucjeps.berkeley.edu/>). When the latitude and longitude of locations associated with these herbarium records were missing, visual inspections of maps from Google Earth were used to find the best approximation to the locality described on the herbarium sample. Climate data for each regional population were obtained from WorldClim (<http://www.worldclim.org/>) and are represented as 19 bioclimatic variables, which are functions of temperature and precipitation variables (Table S1), given at a resolution of 30 arc-seconds (~1 km). The generic grid files available from the WorldClim website were trimmed for each climate variable using the raster library in R and the following geographical extent: minimum longitude: -124.0° , maximum longitude: -117.5° , minimum latitude: 35.0° , maximum latitude: 42.5° . Using these trimmed grid files and the location information pruned of duplicate observations ($n_{\text{pruned}} = 65$ Klamath Mountains, $n_{\text{pruned}} = 144$ southern Sierra Nevada), the MaxEnt software version 3.3.3k (<https://www.cs.princeton.edu/~schapire/maxent/>) was used to build a SDM for each regional population. MaxEnt was run using the cross-validation option for model assessment, 10 replicates, a maximum number of background points of 10,000, and jackknife analysis to evaluate variable importance. Measures of variable importance (i.e. variable contribution and permutation importance scores) and the results of the jackknife analyses were used to assess the relative roles of temperature-related and precipitation-related variables to each SDM.

We also used SDMs to quantify niche differentiation between regional populations of foxtail pine (Warren et al. 2008). We tested two null hypotheses. First, we tested the null hypothesis that the two SDMs were based on a single, underlying SDM common to each regional population. Second, we tested the null hypothesis that the two SDMs are no more differentiated than those randomly drawn from a common SDM with non-overlapping geographical distributions for each regional population. Both tests are based on the D and I statistics given by Warren et al.

(2008). The former null hypothesis was tested using the `niche.equivalency.test` function in the `phyloclim` library in R, while the latter null hypothesis was tested using the `bg.similarity.test` function in the same R library. Both tests were based on $n = 100$ permutations to derive null distributions of test statistics.

Quantitative genetic analysis

We performed two sets of analyses to dissect the genetic basis of water-use efficiency for foxtail pine. First, we demonstrated that variation for the measured traits was genetically based using standard methods to decompose trait variance into effects of families, regions, and environment (Lynch and Walsh 1998). Second, we fit single and multiple QTL models to dissect the genetic basis of each trait into their genetic components using the regression methods of Knott et al. (1996).

The genetic basis for each measured trait was assessed using linear models. We fit three different linear models to the observed data for each trait: (1) a fixed effect model containing only a grand mean (i.e. intercept), (2) a linear mixed model with a grand mean as a fixed effect plus a random effect of family, and (3) a linear mixed model of a grand mean as a fixed effect plus a random effect of region plus a random effect of family nested within region. Uncertainty in parameter estimates from each model was assessed using parametric bootstrapping ($n = 1,000$ replicated simulations) as carried out with the `simulate` function in R. Models were compared using the Akaike Information Criterion (AIC), with Akaike weights used to assess the conditional probabilities for each model (Burnham and Anderson 2002). If models containing random effects for families or models containing random effects for regions and families nested within regions fit the data better than a model with only a grand mean, then we concluded that there were non-zero heritabilities for these traits. If we assume that all offspring within each family were half-siblings, we could estimate narrow-sense heritability as $h^2 = 4\sigma_{fam}^2/(\sigma_{fam}^2 + \sigma_{res}^2)$, where σ_{fam}^2 is the variance due to family nested within region and σ_{res}^2 is the residual variance. Given the small number of families, however, we avoided this estimation, as we were interested only in detecting

non-zero heritability and not precise estimation of its magnitude. Linear models with fixed effects were fit using the `lm` function, while linear mixed models were fit using maximum likelihood as employed in the `lmer` function of the `lme4` library of R. Log-likelihood and AIC values were extracted for each fitted model using the `logLik` and `AIC` functions in R, respectively.

The genetic basis of each trait was dissected using the least squares regression approach of Knott et al. (1996) for outbred, half-sibling families, where probabilities of allelic inheritance due to the common parent were used as predictors for each trait. Significance of the regression model was determined using a *F*-test calculated at 1-cM intervals, with the distribution of this statistic under a null model of no QTLs generated via a permutation scheme (Churchill and Doerge 1994). The common parent in our analyses was the maternal tree, we assumed that all offspring per maternal tree were half-siblings, and we used 1,000 permutations to generate null distributions of *F*-statistics. Permutations were used to create null distributions for *F*-statistics at the level of the entire genome (i.e. all linkage groups) and for each chromosome (i.e. linkage group) separately. We initially fit models of one QTL per linkage group using three significance thresholds: (1) $\alpha = 0.05$ at the level of the entire genome (major QTL), (2) $\alpha = 0.01$ at the level of a particular chromosome (minor QTL), and (3) $\alpha = 0.05$ at the level of a particular chromosome (suggestive QTL). For each QTL, we estimated the percent variance explained (PVE) as $PVE = 4[1 - (MSE_{full}/MSE_{reduced})]$, where MSE_{full} and $MSE_{reduced}$ are the mean square errors of the full and reduced models, respectively (cf. Everett and Seeb 2014). Following Knott et al. (1996), estimates of PVE were scaled by $(1 - 2r)^2$, where r is the recombination frequency between the marker and QTL (i.e. $r = 0.01$ for a 1-cM scan of each linkage group). Uncertainty in the position of the QTL was assessed using bootstrapping ($n = 1,000$ replicates). For each linkage group with a statistically significant QTL, we subsequently fit a model of two QTLs using the same approach, with the only differences being the use of asymptotic null distributions to test the statistical significance of the observed *F*-statistics and the lack of adjustments to estimates of the PVE for

Table 5.2 Mean (standard deviation) of read metrics by family.

Family	Number of reads	Length (bp)	Quality	% Aligned
Blue	1,092,446 (319,903)	89.0 (8.28)	38.0 (1.05)	31.00 (4.881)
Green	691,141 (119,272)	87.6 (10.32)	37.5 (1.16)	26.08 (1.614)
Purple	724,998 (126,585)	88.1 (9.98)	37.6 (1.15)	24.81 (1.398)
Red	1,289,156 (304,551)	89.0 (8.10)	38.1 (1.05)	33.14 (3.577)
Yellow	952,597 (377,357)	88.6 (9.17)	37.8 (1.12)	28.89 (4.185)

multiple QTL models. All analyses were conducted with the `HSportlets` module on `GridQTL` ver. 3.3.0 (Seaton et al. 2006; Allen et al. 2012) using the linkage map for foxtail pine reported by Friedline et al. (2015).

Results

Sequence analysis and genotype determination

From two lanes of HiSeq sequencing, we obtained 148,685,598 and 160,770,417 reads from lane 1 (length = 101 bp, %GC = 40) and lane 2 (length = 101 bp, %GC = 41), respectively. Following read filtering, we retained 77,568,370 (length = 49 - 101 bp, %GC = 40) reads from lane 1 and 107,372,313 (length = 49 - 101, %GC = 40) reads from Lane 2. A summary of the sequencing output and quality can be found in Table 5.2. The highest quality and most reads came from the Blue and Red families, while the Green family produced the smallest number of reads. Similarly, the Blue and Red families had the highest percentages of reads mapping to the assembly. The quality of reads across all families was sufficiently high, with average quality of any base of approximately 38. Graphical summaries of missing data and quality metrics are available in Figures S5.2 and S5.3. We filtered SNPs at the same position on the linkage map down to a set of 843 loci with the least amount of missing data and polymorphism in the most families. At these 843 SNPs, missing data averaged 58.0% (0% - 95.6%). Missing data were subsequently imputed using the marker ordering from Friedline et al. (2015) and fastPHASE.

Species distribution modelling

Species distribution models were good predictors of the current geographical ranges for each regional population of foxtail pine (Figures 5.1, S5.1). Estimates of the area under the receiver operating characteristic curves (ROC curves) were near 1.0 for each model for both the training and test set of samples (Figure S5.4). Exceptions to this pattern included low to moderate probabilities of occurrence outside the current geographical distribution for the Klamath

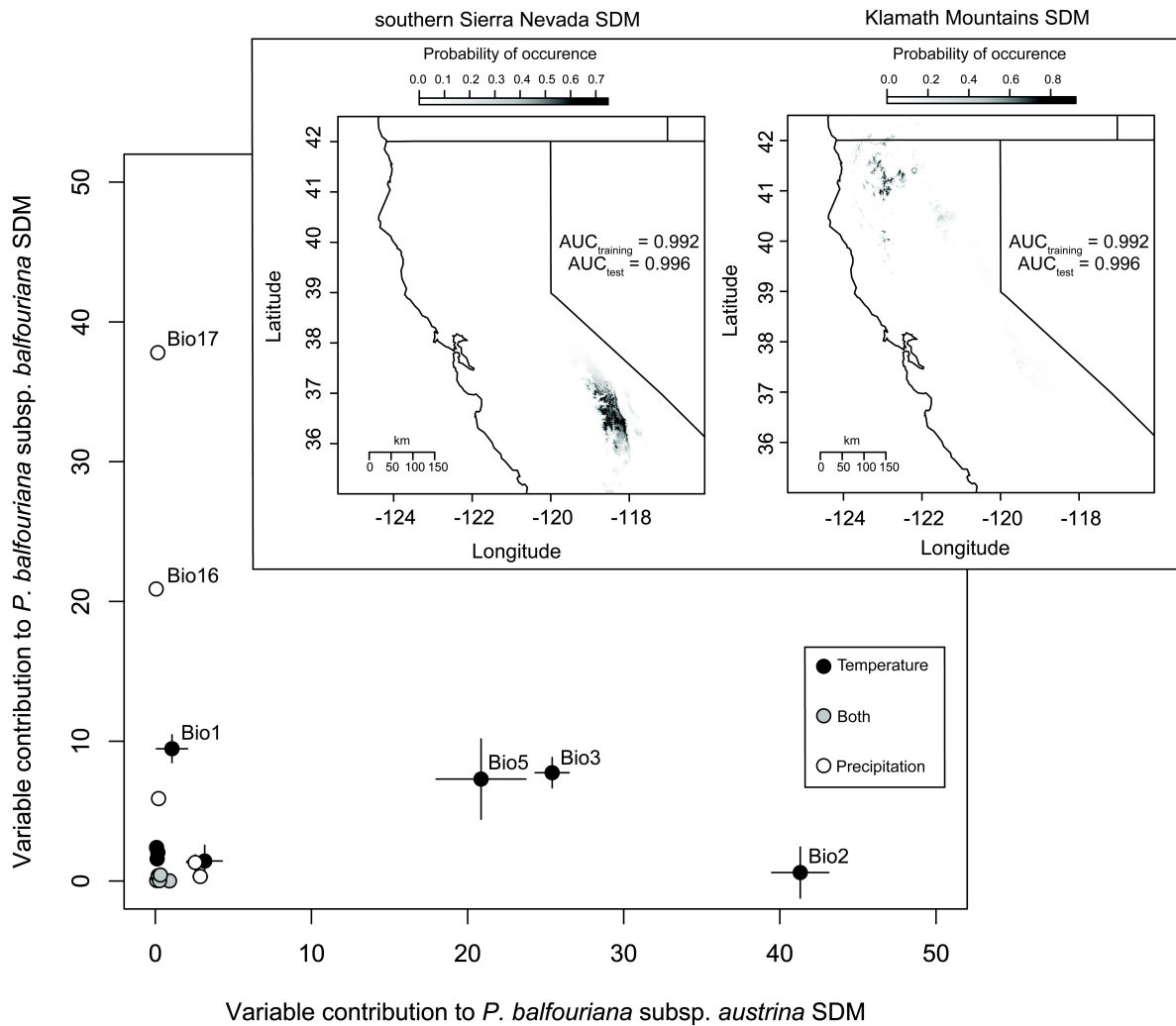


Figure 5.1 Species distribution models (SDMs) created using MaxEnt are good predictors of the current geographical range of foxtail pine (inlaid maps; AUC = area under the receiver operating characteristic curve). Precipitation and temperature-related variables are differentially important, as measured by variable contributions to each model, to the SDM of each regional population of foxtail pine, with precipitation-related variables more important for the Klamath Region and temperature-related variables more important for the southern Sierra Nevada. Variable contribution scores (+/- 1 standard deviation derived from 10 replicated runs of MaxEnt per SDM) are uncorrelated (Spearman's $\rho = -0.065$). For symbols without apparent error bars, the diameter of the circle was greater than the standard deviation.

Mountains, which were centered on the northern Sierra Nevada, and a slightly expanded range north and south of the known range limits in the southern Sierra Nevada. Foxtail pine is known to be absent from these regions. In both cases, the probabilities of occurrence were less, often much less, than 0.40. The SDM based on the Klamath Mountains predicted a near zero probability for cells within the range of the southern Sierra Nevada and vice versa.

Foxtail pine inhabits the cooler portions of each region in which it is currently located (Figures S5.5 – S5.6). For precipitation-related variables, however, foxtail pine in the Klamath Mountains inhabits slightly wetter localities relative to background localities, while in the southern Sierra Nevada foxtail pine inhabits drier localities relative to background localities. The climates inhabited by foxtail pine in each region also differ. In general, differences between the climates inhabited by each regional population were consistent with the Klamath Mountains being warmer, yet less variable in temperature throughout the year, and wetter, yet slightly more variable in precipitation throughout the year, relative to the southern Sierra Nevada. For example, mean annual precipitation was almost twice as high in the Klamath Mountains as in the southern Sierra Nevada (1179 mm versus 650 mm, respectively), yet the distribution of precipitation was slightly more variable throughout the year (e.g. precipitation of the driest month: 11.78 mm versus 12.41 mm, respectively; coefficient of variation across months: 65.86 versus 65.02, respectively).

Bioclimatic variables used to predict occurrences of foxtail pine within each regional population were highly correlated with one another (Figure S5.7). Sets of correlated variables are difficult to evaluate as contributing to SDMs (Warren and Seifert 2011). We, therefore, used several different measures of variable importance. Inspection of variable contribution scores revealed that temperature-related and precipitation-related variables were differentially important across SDMs for each region (Figure 5.1; Table S5.2). Temperature-related variables, specifically mean diurnal range (Bio2), isothermality (Bio3), and maximum temperature of the warmest month (Bio5), were most important for the southern Sierra Nevada population, whereas precipitation-

related variables, specifically precipitation of the driest quarter (Bio17) and precipitation of the wettest quarter (Bio16), were most important for the Klamath Mountains population. This pattern, however, was reversed when using permutation importance scores, despite a moderate correlation between rankings of importance based on variable contribution and permutation importance scores (Figures 5.2, S5.8; Table S5.4). Temperature-related variables became more important for the Klamath Mountains, specifically annual temperature (Bio1), while precipitation-related variables became more important for the southern Sierra Nevada population, specifically precipitation seasonality (Bio15) and mean temperature of the wettest quarter (Bio8). Jackknife

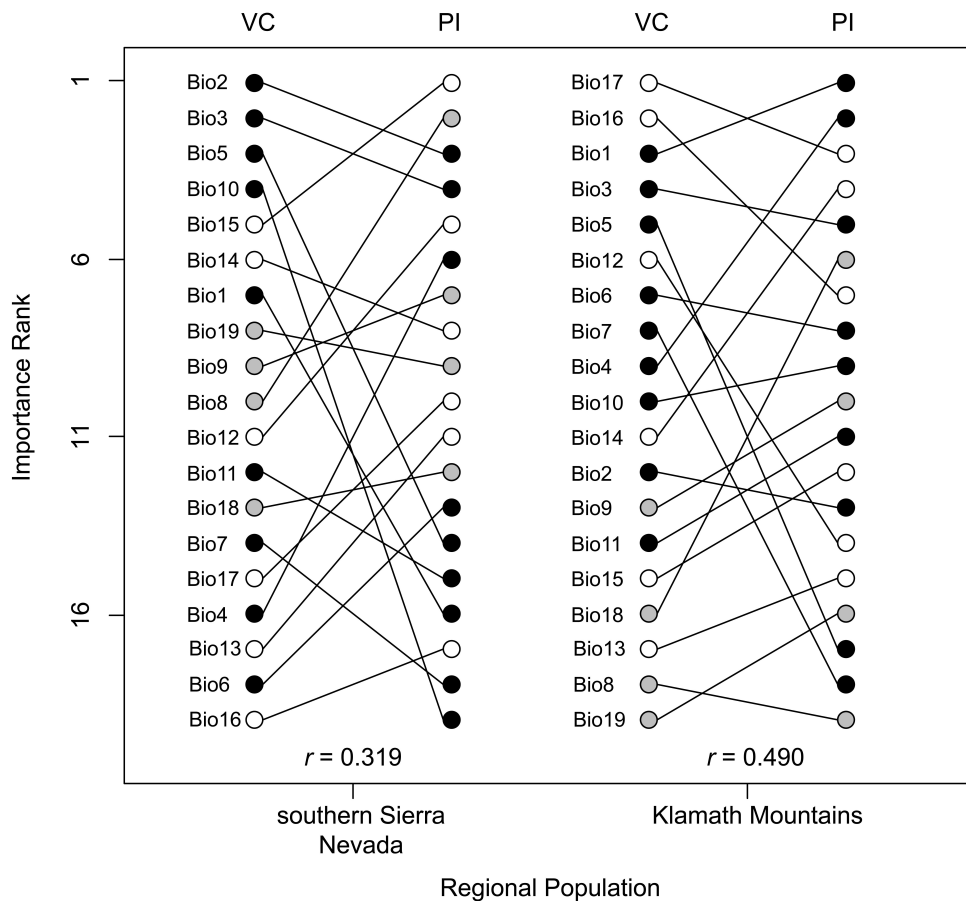


Figure 5.2 Ranks of variable importance (low rank = more important) based on variable contribution (VC) scores and permutation importance (PI) scores to the SDM for each regional population are moderately correlated ($r = \text{Spearman's } \rho$). Variable types are denoted using filled circles, with black used for temperature-related variables, white for precipitation-related variables, and gray for variables related to both temperature and precipitation.

analysis of variable importance based on AUC, test gain, and regularized test gain, however, were consistent with both temperature-related and precipitation-related variables as being important for the Klamath Mountains population (Figures S5.9 – S5.11). For example, mean annual temperature (Bio1), maximum temperature of the warmest quarter (Bio5), mean temperature of the driest quarter (Bio9), mean temperature of the warmest quarter (Bio10), precipitation of the driest quarter (Bio17), and precipitation of the warmest quarter (Bio18) all contributed significantly to the SDM for the Klamath Mountains population (Figure S5.11), although no one variable contained much information that was not present in at least one of the others. In contrast, jackknife analysis of variable importance based on AUC, test gain, and regularized test gain were consistent with primarily temperature-related variables, specifically mean annual temperature (Bio1), mean diurnal range (Bio2), maximum temperature of the warmest month (Bio5), and the mean temperature of the warmest quarter (Bio10), driving the SDM for the southern Sierra Nevada population (Figures S5.12 – S5.14). As with the SDM for the Klamath Mountains population, however, no one variable contained information that was not present in at least one of the others (Figure S5.14).

Predicted niches based on SDMs for each regional population were dissimilar, with estimates of D (0.072) and I (0.258) being much closer to zero (dissimilar) than to 1 (similar) (Figure S5.15). These differences were significant enough to reject a null model of a single shared SDM common to both regional populations ($P < 0.01$ for D and I). Even if differences were accounted for in the background environments of each regional population (Figure S5.5), the predicted niches were statistically different ($P < 0.05$ for both D and I). Replicating the analyses for climate variables related only to temperature or only to precipitation revealed that niche divergence was stronger for precipitation-related variables ($D_{\text{precip}} = 0.074$; $I_{\text{precip}} = 0.271$) relative to temperature-related variables ($D_{\text{temp}} = 0.124$, $I_{\text{temp}} = 0.376$). Therefore, regional populations of foxtail pine have divergent climatic niches, with precipitation-related variables more differentiated than temperature-related variables.

Table 5.3 Attributes of linear mixed models used to estimate familial and regional effects for each phenotypic trait. Values in parentheses are 95% parametric bootstrap confidence intervals (see Materials and Methods).

Model Attribute	$\delta^{13}\text{C}$	$N_{\mu\text{g}}$
logL	-151.705 (-167.029 – -130.520)	-512.587 (-528.588 – -491.444)
Intercept	-30.755 (-31.439 – -30.075)	21.519 (18.615 – 24.596)
Family variance component (σ^2_{fam})	0.159 (0.002 – 0.432)	7.826 (0.000 – 17.521)
Region variance component (σ^2_{reg})	0.167 (0.000 – 0.538)	1.696 (0.000 – 9.933)
Residual variance component (σ^2_{res})	0.316 (0.249 – 0.384)	22.486 (17.927 – 27.912)

Quantitative genetic analysis

Variation across siblings measured within the common garden was genetically based for each trait (Table 5.3). Family identifiers nested within regional populations accounted for sizeable portions of the total variance for $\delta^{13}\text{C}$ ($\sigma^2_{\text{fam}}/[\sigma^2_{\text{reg}} + \sigma^2_{\text{fam}} + \sigma^2_{\text{res}}] = 24.76\%$) and $N_{\mu\text{g}}$ ($\sigma^2_{\text{fam}}/[\sigma^2_{\text{reg}} + \sigma^2_{\text{fam}} + \sigma^2_{\text{res}}] = 24.45\%$). This was consistent with the differences among predicted family means for both traits (Figure 5.2), which were positively correlated (Figure 5.4), but not significantly so (Pearson's $r = 0.415$; $P = 0.487$). Regional identifiers, however, were differentially important across traits, with these identifiers accounting for marginally more variance than family identifiers for $\delta^{13}\text{C}$ (26.01%) but less than 10% of the total variance for $N_{\mu\text{g}}$ (Figure 5.2). The joint effect of family and regional identifiers (i.e. the total genetic effect = $[\sigma^2_{\text{reg}} + \sigma^2_{\text{fam}}]/[\sigma^2_{\text{reg}} + \sigma^2_{\text{fam}} + \sigma^2_{\text{res}}]$), however, was large for each trait ($\delta^{13}\text{C}$: 50.78%; $N_{\mu\text{g}}$: 29.75%). Comparisons of linear models progressing from intercept only to an intercept plus families nested within regions using AIC, revealed that a linear mixed model with an intercept and families was the best fit (AIC = 310.29 for $\delta^{13}\text{C}$; AIC = 1031.26 for $N_{\mu\text{g}}$; Table 5.4). Comparison to other models using AIC weights, however, revealed that the most complex model of an intercept plus region plus families nested within regions had a reasonably high conditional probability (AIC weight = 0.36 $\delta^{13}\text{C}$; AIC weight = 0.28 for $N_{\mu\text{g}}$; Table 4) relative to those for the best model ($\delta^{13}\text{C} = 0.64$; $N_{\mu\text{g}} = 0.72$) for each phenotypic trait.

Table 5.4 Comparisons of linear mixed models using the Akaike Information Criterion (AIC) by trait were used to select the best model (bolded text). In these models, the intercept was a fixed effect, while families nested within regions and regions were random effects. The AIC weight is calculated using the standardized relative likelihoods, where the relative likelihood is given as $e^{(-0.5 \times \Delta AIC)}$. For this calculation, ΔAIC is the difference between the AIC for each model and the AIC for the best model (bolded text), where the best model is the one with the lowest AIC. The weights are then calculated as each of relative likelihoods over the sum of the relative likelihoods, thus making the sum of the weights equal to 1. Akaike weights can also be considered as the conditional probabilities for each model.

Model	$\delta^{13}C$		$N_{\mu g}$	
	AIC	AIC weight	AIC	AIC Weight
Intercept	408.10	3.66×10^{-22}	1071.76	1.16×10^{-9}
Intercept + family	310.29	0.64	1031.26	0.72
Intercept + family + region	311.41	0.36	1033.17	0.28

We dissected the genetic basis of the heritable variation evident for each trait from the linear mixed model analysis using the regression-based approach to QTL mapping of Knott et al. (1996). Application of one-locus models (i.e. a maximum of one-locus per linkage group) resulted in a set of 11 QTLs across all linkage groups and both traits (Table 5.5; Figure 5.6). For $\delta^{13}C$, six QTLs were discovered, with two discovered at the most stringent significance level (genome-wide permutation-based $\alpha = 0.05$) and four at the least stringent significance level (linkage group specific permutation-based $\alpha = 0.05$). Effect sizes for these QTLs were large to moderate, with the percent variation explained (PVE) ranging from 47.807% to 24.066%. For $N_{\mu g}$, five QTLs were discovered, with one QTL at the most stringent significance level, two at the intermediate significance level (linkage group specific permutation-based $\alpha = 0.01$), and two at the least stringent significance level. Effect sizes for these QTLs were also large to moderate, with PVE varying from 39.773% to 25.058%. There was moderate autocorrelation for the F -statistic at a resolution of 6 cM or less for $\delta^{13}C$ and 3 cM or less $N_{\mu g}$ (Figure S5.16), but there was no correlation between F -statistics for each trait (Pearson's r : -0.014, $P = 0.734$; Figure S17). In general, 95% confidence levels of positions for each QTL were large (Table 5.5).

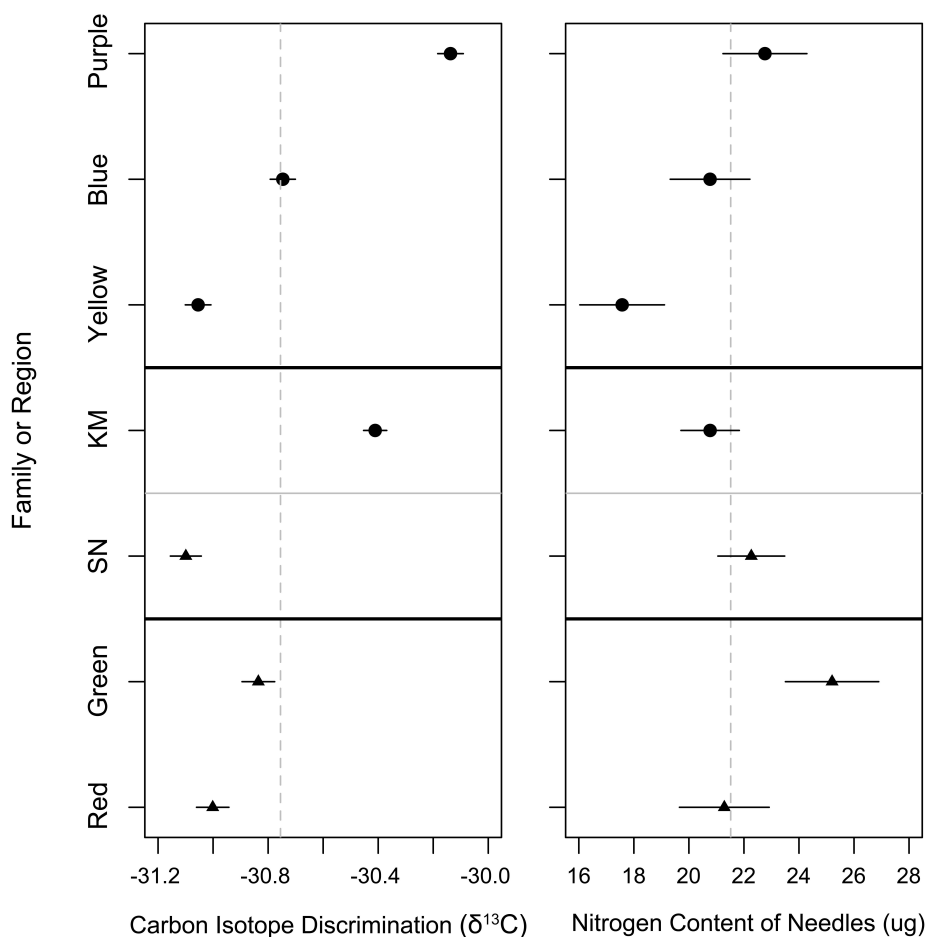


Figure 5.3 Familial and regional level means (± 1 standard error) by trait (left: $\delta^{13}C$, right: $N_{\mu g}$) are differentiated across families and regions relative to the global mean. Dashed gray lines give global means across all families for each trait. Estimates for the Klamath Mountains (KM) are given as filled circles, while estimates for the southern Sierra Nevada (SN) are given as filled triangles. Familial names are given as colors (see Materials and Methods).

For the 11 QTLs detected using one-locus models, 10 were consistent with multiple QTLs using two-locus models (Table 5.6). In general, the QTLs from the one-locus models were one of the pair of QTLs detected in the two-locus models. There were four exceptions to this pattern, with two of these exceptions being a minor modification in position of the original QTL equal to 1.0 cM. The other two exceptions included significant changes to the position of the original QTL, with the QTL on linkage group 3 for $N_{\mu g}$ changing from 93.0 cM to 52.0 cM and 35.0 cM and the

Table 5.5 Summary of QTLs for each trait that survive multiple test corrected significance thresholds at either the level of the whole genome ($\alpha = 0.05$ for $G_{0.05}$) or a chromosome ($\alpha = 0.01$ for $C_{0.01}$, $\alpha = 0.05$ for $C_{0.05}$). ^aLG, Linkage group, ^bPVE, percent variance explained; PVE_c , corrected percent variance explained, ^cThe threshold value for the *F*-statistic under the null model as determined using the listed value of α (0.05 or 0.01) and permutations following Churchill and Doerge (1994) for either individual linkage groups (C) or the entire genome (G). ^d95% CI, 95% confidence interval determined through bootstrap analysis ($n = 1,000$ replicates)

Trait	LG ^a	Position (cM)	F	PVE ^b (PVE ^c)	Threshold ^c	95% CI ^d (cM)
$N_{\mu g}$	1	0.0	4.422	26.540 (25.489)	3.818 (C0.05)	0.0 – 97.0
$\delta^{13}C$	1	98.0	7.506	49.778 (47.807)	5.803 (G0.05)	13.0 – 99.0
$\delta^{13}C$	2	78.0	6.040	39.139 (37.589)	5.803 (G0.05)	3.0 – 78.0
$\delta^{13}C$	3	34.0	4.356	26.092 (25.058)	3.456 (C0.05)	13.0 – 93.0
$N_{\mu g}$	3	93.0	4.475	27.065 (25.993)	3.725 (C0.05)	14.0 – 93.0
$\delta^{13}C$	5	64.0	4.659	28.625 (27.491)	4.008 (C0.05)	17.0 – 103.0
$\delta^{13}C$	6	0.0	4.198	24.825 (23.842)	3.835 (C0.05)	0.0 – 85.0
$N_{\mu g}$	7	62.0	6.351	41.413 (39.773)	6.091 (G0.05)	16.0 – 89.0
$N_{\mu g}$	8	72.0	5.784	37.182 (35.710)	5.559 (C0.01)	1.0 – 100.0
$N_{\mu g}$	9	95.0	5.924	38.237 (36.809)	4.958 (C0.01)	9.0 – 95.0
$\delta^{13}C$	12	23.0	4.105	24.066 (23.113)	4.072 (C0.05)	15.0 – 91.0

QTL on linkage group 6 for $\delta^{13}C$ changing from 0.0 cM to 46.0 cM and 56.0 cM (Tables 5.5 and 5.6). The average spacing between QTLs on the same linkage group was 29.4 cM, with a minimum of 3 cM to a maximum of 85 cM. The multi-QTL PVE for each trait ranged from a minimum of 42.685% to a maximum of 71.315%, with only one instance of positional overlap in QTLs for each trait (linkage group 3 at 34.0 cM for $\delta^{13}C$ and 35.0 cM for $N_{\mu g}$). On average, there was a negative relationship between distance (cM) and the correlation of family effects (Pearson's *r*) between QTLs on the same linkage group (Figure S18), so that strong positive correlations of family effects were observed when QTLs were close together (<15 cM) and strong negative correlations when QTLs were farther apart (>20 cM). QTL effects from the one-locus QTL models were consistent with differentiation between regional populations, with family effects opposite in

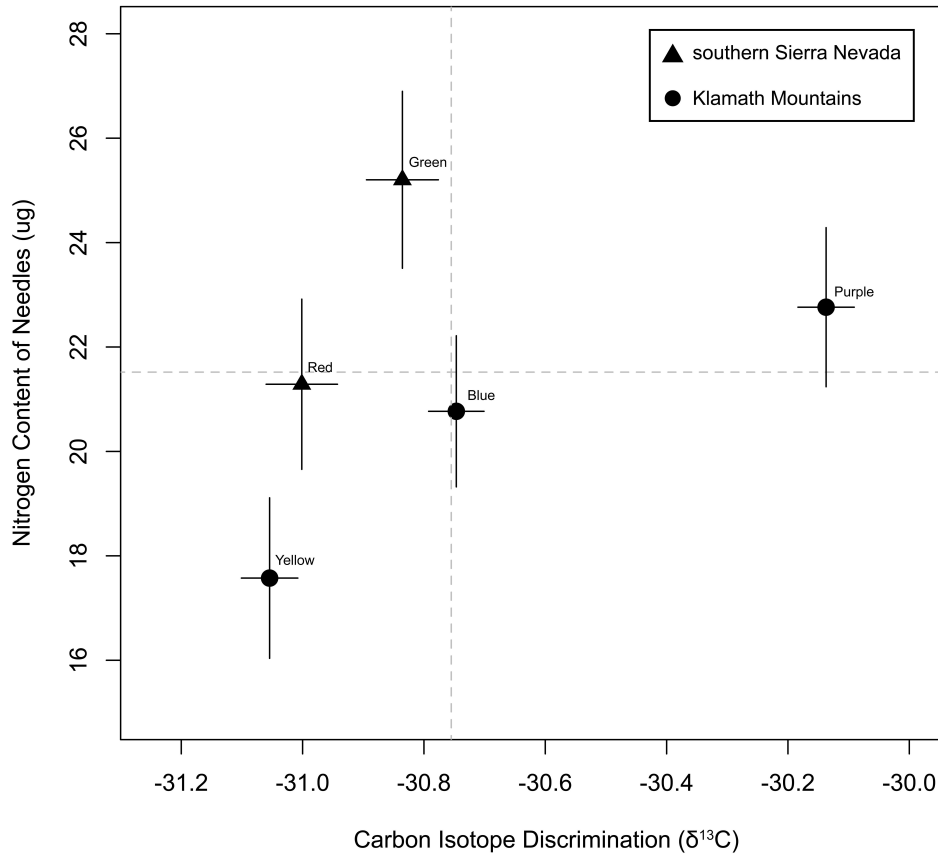


Figure 5.4 The relationship between traits based on family means (± 1 standard error) is positive (Pearson's $r = 0.415$), although statistically non-significant at $\alpha = 0.05$ ($P = 0.487$). Dashed gray lines give global means across all families for each trait.

sign more often than expected by chance for $\delta^{13}\text{C}$ (Fisher's exact test: odds ratio = 0.113, $P = 0.009$), but not for $N_{\mu\text{g}}$ (Fisher's exact test: odds ratio = 1.319, $P = 1.0$). Trait differentiation was similarly structured (Tables 5.3 and 5.4), with the clearest signal of differentiation for $\delta^{13}\text{C}$. The same patterns were observed for family effects in the two-locus models for the original QTL from Table 5.5, but not for the second QTL ($P > 0.05$ for both $\delta^{13}\text{C}$ and $N_{\mu\text{g}}$).

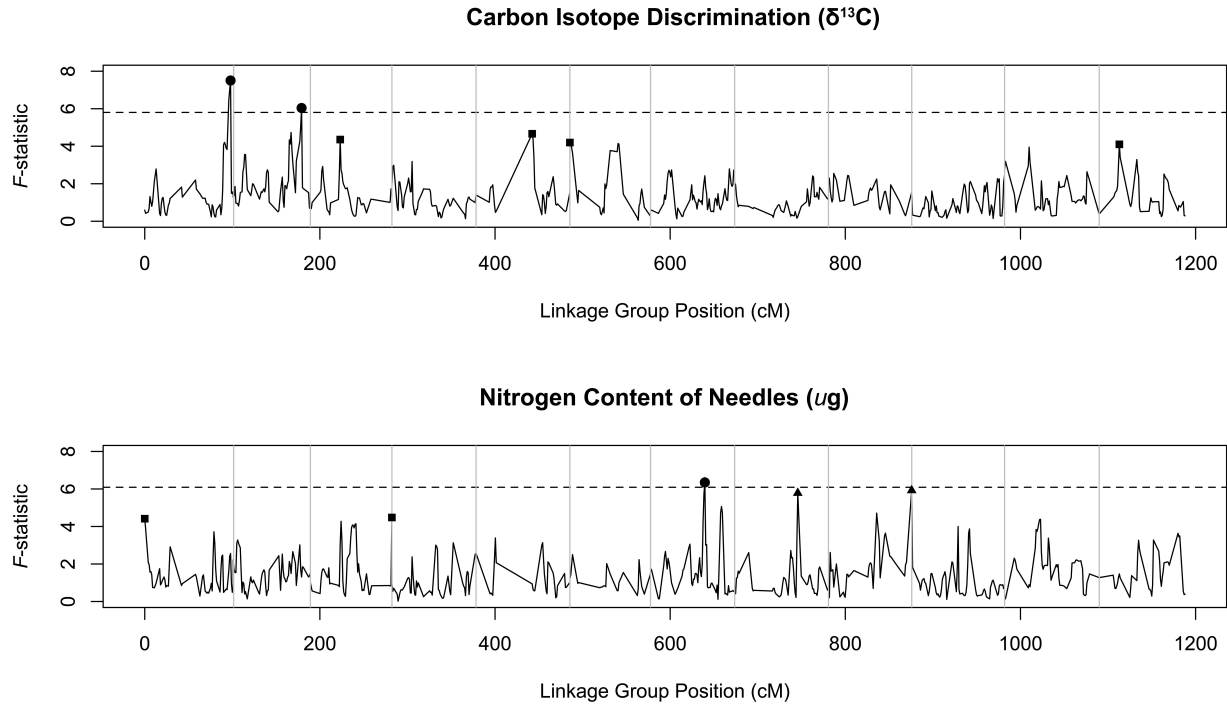


Figure 5.5 The distributions of the F-statistic derived from single QTL models across each linkage group for carbon isotope discrimination and nitrogen content of needles reveals the isolated nature of QTLs. The dashed horizontal line in each panel is the genome-wide significance threshold ($\alpha = 0.05$) for the F-statistic based on the permutation scheme ($n = 1,000$ permutations) suggested by Churchill and Doerge (1994). Significant QTLs are denoted with filled circles ($\alpha = 0.05$, genome-wide), filled triangles ($\alpha = 0.01$, chromosome-wide) or filled squares ($\alpha = 0.05$, chromosome-wide).

Discussion

Climate is one of the main drivers for the distribution and diversification of forest tree species (MacArthur 1972; Royce and Barbour 2000; Ettinger et al. 2011; Alberto et al. 2013). The relative importance of specific climate variables as drivers of natural selection, however, is often assumed. For example, if a phenotypic trait is correlated to water availability in one species, the same trait is often studied in a different focal species without documenting water availability as having a large impact on fitness variation in the latter. The problem lies in the assumption that this correlation is also indicative of similar fitness consequences across species. Here, we address this issue for foxtail pine using a novel combination of species distribution modeling and pine and quantitative genetics. We illustrate the importance of water availability to the distribution of foxtail

Table 5.6 Summary of two QTL models fit to each significant QTL from Table 4. Bolded *P*-values are less than 0.05. ^aLG, Linkage Group; ^bPVE, percent variance explained by both QTLs

Trait	LG ^a	Position (cM)	Position 2 (cM)	<i>F</i>	<i>P</i>	PVE ^b
N _{μg}	1	0.0	79.0	3.89	0.0023	54.518
δ ¹³ C	1	98.0	13.0	2.92	0.0149	64.725
δ ¹³ C	2	78.0	66.0	3.76	0.0030	61.685
δ ¹³ C	3	34.0	14.0	3.18	0.0091	44.459
N _{μg}	3	93.0	35.0	4.24	0.0012	57.594
δ ¹³ C	5	64.0	88.0	1.81	0.1135	37.745
δ ¹³ C	6	0.0	56.0	3.84	0.0026	48.892
N _{μg}	7	62.0	80.0	4.69	0.0005	71.315
N _{μg}	8	72.0	68.0	2.57	0.0287	49.661
N _{μg}	9	95.0	64.0	2.90	0.0155	53.602
δ ¹³ C	12	23.0	43.0	3.20	0.0088	42.685

pine and hence fitness, as well as describe the genetic architecture of WUE, a phenotypic trait responsive to water availability, so that this trait and the markers correlated to it can be used to test hypotheses about local adaptation and its genetic architecture.

Climate drivers of the current geographical distribution and WUE

In many situations, drivers of geographical distributions for tree species are obvious. For example, links between light availability, temperature, precipitation, and phenological traits are commonly noted for forest trees (Howe et al. 2003; Chuine 2010). In other situations, however, climate drivers are less clear, so that quantification of the relative importance for a suite of climate variables is needed. For foxtail pine, the drivers of its current geographical distribution appear to be a mixture of temperature-related and precipitation-related variables, with a clear pattern that precipitation-related variables are necessary to explain the current geographical range. This implies that phenotypic traits correlated to precipitation-related variables likely have fitness

consequences for foxtail pine, as precipitation-related variables appear to structure its current range. Additionally, the importance of these drivers is differentiated between regional populations, with precipitation-related variables more differentiated than temperature-related variables, which mimicked differentiation of phenotypic trait values. Thus, if we leverage the correlations between $\delta^{13}\text{C}$ and water availability, a crucial component of survival and hence fitness, observed in other plant species (Ehleringer et al. 1993) and the conclusion that precipitation-related variables are important for the distribution of foxtail pine, it is likely that $\delta^{13}\text{C}$ variation in foxtail pine is linked with fitness.

In general, increases in $\delta^{13}\text{C}$ reflect higher WUE (Farquhar et al. 1982). Inspection of mean values for $\delta^{13}\text{C}$ for each region (see Figure 5.3), in light of the documented precipitation patterns, however, appears contradictory. On average, maternal trees in the Klamath Mountains had higher $\delta^{13}\text{C}$ values, which suggests higher WUE, yet precipitation is much higher in the Klamath Mountains than in the southern Sierra Nevada. It is well known, however, that soil properties, such as coarseness and depth to bedrock, affect available soil moisture. For example, small differences in soil texture observed across the Southern Sierra Nevada Critical Zone Observatory, a site not far removed from the regional population of foxtail pine in the southern Sierra Nevada, result in large differences in the available soil moisture (Bales et al. 2011). Soil texture also varied by elevation, with soils at the highest elevations being coarser and less developed. As such, water availability in these soils was more limited even though snowfall was typically higher. Soils between regional populations of foxtail pine are fundamentally different, and so is the local distribution of foxtail pine. In the Klamath Mountains, soils are primarily ultramafic, while in the southern Sierra Nevada they are largely granitic. Foxtail pine grows near tops of local peaks in the Klamath Mountains, whereas in the southern Sierra Nevada it is distributed broadly across large swathes of high elevation sites. Thus, one explanation for the apparent contradiction is that soil properties are different, so as to create patterns of soil moisture not reflective of regional mean precipitation patterns. Foxtail pine in the Klamath Mountains often inhabits areas with high

levels of boulder cover (Eckert and Sawyer 2002; Eckert 2006), which are expected to house soils with less capacity to hold water over long periods of time. When coupled with the higher average temperatures in the Klamath Mountains, this suggests that water may be more limited throughout the year (e.g. summer drought) than expected based on annual precipitation totals. Additional work, however, would be needed to quantify trait variation within each regional population and correlate it to both climate and soil characteristics.

Genetic architecture of water-use efficiency

Both $\delta^{13}\text{C}$ and $N_{\mu\text{g}}$ were consistent with non-zero heritabilities. Families and regions accounted for approximately 50% of the total phenotypic variance for $\delta^{13}\text{C}$ and 30% for $N_{\mu\text{g}}$. Models with effects due to families or families nested within regions were also strongly preferred over models without these effects (Table 5.4). The effect of region, however, was highest in magnitude for $\delta^{13}\text{C}$, with the variance component for region larger than that for family. This is consistent with previous estimates of quantitative genetic parameters for these phenotypic traits in other conifers. For example, $\delta^{13}\text{C}$ and $N_{\mu\text{g}}$ are both heritable in a variety of pine species (Brendel et al. 2002; Baltunis et al. 2008; Gonzalez-Martinez et al. 2008; Cumbie et al. 2011; Joao Gaspar et al. 2013; Marguerit et al. 2014; Eckert et al. 2015). Populations within many species are also often differentiated for $\delta^{13}\text{C}$, but not for $N_{\mu\text{g}}$ (e.g. Eckert et al. 2015; Maloney et al. unpublished). Further work, however, would be needed to precisely estimate the level of differentiation for these traits, as well as to test whether this level of differentiation is larger than that expected for neutral loci (i.e. if this pattern is consistent with local adaptation).

Estimates of narrow-sense heritabilities (h^2) resulted in values greater than 1.0 for each phenotypic trait no matter which model with a family effect was used (i.e. families or regions plus families nested within regions). This could be due to tissue sampling occurring prior to formation of randomized blocks in the common garden, as family groups would be confounded with micro-environmental variation. Use of data from Eckert et al. (2015) and Maloney et al. (unpublished data) for sugar pine (*P. lambertiana* Dougl.), western white pine (*P. monticola* Dougl.), and

whitebark pine (*P. albicaulis* Engelm.) grown at the same facility in the same experimental conditions, however, reveals that block effects for $\delta^{13}\text{C}$ were present only for the relatively fast growing western white pine (Type III Wald F-tests with Kenward-Rogers degrees of freedom; sugar pine: $F_{1,416.49} = 3.5166$, $P = 0.06146$; western white pine: $F_{1,630.24}$, $P = 0.00068$; whitebark pine: $F_{1,452.75} = 0.0147$; $P = 0.9037$). In contrast, block had a statistically significant effect on $N_{\mu\text{g}}$ for sugar pine and western white pine ($P < 0.001$), but not whitebark pine ($F_{1,429.22} = 1.6252$, $P = 0.20305$). Thus, our results should be taken with caution, but family effects estimated here were similar in magnitude to those from Eckert et al. (2015) and randomized blocks tended to have no effect on the same phenotypic traits measured in whitebark pine at the same facility, a species with a similar pattern of early slow growth (McCune 1988).

If our results are indicative of true signal, effect sizes could be over-estimated on average due to the small number of sampled families (Beavis 1994). To illustrate this effect, we re-analyzed the data from Eckert et al. (2015) for sugar pine, which was grown in a common garden at the same facility and measured for $\delta^{13}\text{C}$ using the same methodology, by resampling smaller numbers of families ($n = 108$ families resampled in decreasing numbers from 108 to three families) and estimating h^2 . As the number of sampled families decreased, estimates of mean h^2 became larger (Figure S5.19), with a 1.5-fold increase in the mean h^2 as the number of sampled families dropped from 108 to three. This is likely also the case for foxtail pine and for $N_{\mu\text{g}}$. Regardless of the precise value of h^2 , it is clear that at least a moderate amount of segregating genetic variation exists for this trait in natural populations of foxtail pine. There was also a moderate, but statistically insignificant, positive correlation between $\delta^{13}\text{C}$ and $N_{\mu\text{g}}$ (Figure 5.4). This has been noted in other species, such as loblolly pine (Cumbie et al. 2011), although general patterns in the sign of the correlation are lacking. In this context, positive correlations could indicate that WUE is determined primarily through leaf-level assimilation (e.g. Johnson et al. 1999; Prasolova et al. 2005), while a negative correlation could indicate that WUE is determined primarily through stomatal conductance. Despite the observed positive correlation, little evidence of pleiotropy was detected,

with only a single QTL on linkage group 3 shared between traits. The lack of pleiotropy for these traits has been noted in several other conifer species (e.g. Marguerit et al. 2014). Correlations between $\delta^{13}\text{C}$ and $N_{\mu\text{g}}$, or growth traits more generally, can also be driven environmentally and can change depending on water availability. For example, Joao Gaspar et al. (2013) have shown that in water limiting environments $\delta^{13}\text{C}$ correlates with survival, but in less water limited environments $\delta^{13}\text{C}$ correlates with height growth for maritime pine (*P. pinaster* Ait.). A similar case might be occurring for foxtail pine, where in the wetter Klamath Mountains $\delta^{13}\text{C}$ variation is correlated with overall growth and in the more xeric southern Sierra Nevada it is correlated with survival. In this context, WUE would be realized through leaf-level assimilation in the Klamath region (as in Weih et al. 2011 for *Salix*), and through stomatal conductance in the southern Sierra Nevada. Sampling more families, measurement of other traits (e.g. growth), and experimentation in multiple environments, however, would be needed to test these ideas. Importantly, $\delta^{13}\text{C}$ should be measured within natural populations to assess correspondence between inferences from common gardens and natural populations.

Using one-locus QTL models, the observed segregating genetic variance for $\delta^{13}\text{C}$ was dissected into two major QTLs and four suggestive QTLs (Table 5.5, Figure 5.6). Each QTL explained a large fraction of total phenotypic variance (23.113% to 47.807%), which suggests that the genetic architecture of this fitness-related trait includes loci of large effect. Under many models of adaptation, however, is difficult to separate QTLs composed of a single, large-effect locus from those composed of several small-effect loci (Yeaman and Whitlock 2011). The observed large values of PVE may also be over-estimated (Beavis 1994), although there is precedence for large effect QTLs for $\delta^{13}\text{C}$ in other species of *Pinus*, especially those distributed in water-limited regions displaying moderate levels of genetic differentiation among populations. For example, Marguerit et al. (2014) identified a QTL explaining 67% of phenotypic variance for $\delta^{13}\text{C}$ in maritime pine, which is distributed across the Mediterranean regions of Europe and has moderate levels of genetic structure across this range (Eveno et al. 2008). For foxtail pine, water availability

is an important driver of its current geographical distribution and genetic structure is moderate to high between regional populations and among stands within regional populations (Eckert et al. 2008, but see Oline et al. 2000). Furthermore, family effects for these QTLs were consistent with differentiation among regions, so it is plausible that the architecture discovered here for $\delta^{13}\text{C}$ largely represents genomic regions underlying trait divergence between the regional populations. If this is the case, this architecture has evolved since the divergence of the regional populations from their common ancestor on the order of one million years ago (Eckert et al. 2008).

Summaries of the results from two-locus QTL models were largely consistent with those from the one-locus models. For the 11 QTLs reported in Table 5, 10 were consistent with at least two segregating QTLs. This brings the total number of QTLs to four major and seven suggestive QTLs for $\delta^{13}\text{C}$ and two major, four minor, and four suggestive QTLs for $N_{\mu\text{g}}$. Interestingly, the correlation of family-level effects for the two QTLs on the same linkage group was negatively related to the distance between these QTLs, so that QTLs close together tended to have similar patterns of family-level effects, whereas QTLs at larger distances tended to have opposite family-level effects (Figure S5.14). This trend was uncorrelated with the difference in effect sizes between QTLs. When added to the observation that family effects were often consistent within regions and differentiated between regions, a likely explanation for this pattern is some form of natural selection driving clustering of loci dependent on consistency of their effects on a fitness-related trait. The fitness benefit of clustering, however, is related to the level of gene flow (Yeaman and Whitlock 2011), so that clustering of adaptive alleles is expected under high levels of gene flow, reduced recombination, and strong magnitudes of selection. This is especially pronounced when genomic rearrangements are common. Inspection of the family-level linkage maps from Friedline et al. (2015), however, revealed little evidence for clustered QTLs displaying differing marker orders across families more so than random positions on the linkage map. This explanation, however, is complicated given that gene flow is approximately zero between these regions (Eckert et al. 2008) and populations of foxtail pine are unlikely to be at selection –

migration equilibrium due to large effective population sizes and long generation times. For example, patterns of segregating ancestral variation after divergence are similar to those predicted by gene flow (Pamilo and Nei 1988), so that it becomes difficult to separate pattern from process with regard to the effects of gene flow on adaptive genetic architectures. Additional work within natural populations, including fine mapping of trait values in the linkage bins defined by Friedline et al. (2015), would be needed to test these ideas further.

We leveraged the annotations of contigs at or near (± 3 cM) the estimated QTL positions to search for putatively functional genes as the drivers of the genotype-phenotype correlations for each QTL (Table S3). Annotations for foxtail pine contigs were derived through similarity searches against the loblolly pine genome. Annotations were obtained from any locus on a loblolly pine scaffold containing a significant hit to a RADtag from foxtail pine, with significance justified by the estimated substitution rate and divergence time between these species (Friedline et al. 2015). Several statistically significant QTLs had no annotation information available. For example, the QTL on linkage group 1 for $\delta^{13}\text{C}$ had no annotations available within a 6-cM window encapsulating the QTL, despite 24 of 76 RADtags having significant similarity to scaffolds in loblolly pine. This is consistent with reports of gene densities reported for conifers (Nystedt et al. 2013; Neale et al. 2014). For the QTL related to $\delta^{13}\text{C}$ on linkage group 2 (Table 5), however, two of the 18 RADtags for foxtail pine had sequence similarity to loblolly pine scaffolds, with annotated InterPro domains suggestive of loci encoding stress responsive proteins (Table S3; Toka et al. 2010; Karijolich et al. 2015). Another example of potentially biologically informative results included the QTL on linkage group 9 for $N_{\mu\text{g}}$ where putative homologs for proteins with domains such as ribosomal protein L38e, cytochrome P450, and thiolase were present. Proteins containing these domains have been implicated in lipid turnover during leaf senescence (Troncoso-Ponce et al. 2013), as well as plant growth and drought stress response (Tamiru et al. 2015). Care should be taken in interpreting these results, however, as QTL intervals were wide, annotations were based on statements of homology with gene predictions in an early release of the loblolly pine

genome sequence (Wegrzyn et al. 2014), and post hoc explanations linking gene products to phenotypic traits is prone to storytelling (Barrett and Hoekstra 2011; Pavlidis et al. 2012). It is important to note, however, that these concerns are with interpretations of putative functions of genes located within the QTL as sensible in their effect on the measured phenotypic trait, and not with the biological signal of linkage driving the discovery of the QTL.

Conclusions

We have used a mixture of species distribution modeling and quantitative genetics to test two hypotheses about WUE, as measured by $\delta^{13}\text{C}$ and $N_{\mu\text{g}}$, for foxtail pine. We showed that precipitation-related variables structured the geographical range of foxtail pine, that climate-based niches differed between regional populations, and that similar patterns were apparent for $\delta^{13}\text{C}$, which was also demonstrated to be heritable. We subsequently dissected this heritability into a set of large-effect QTLs ($n = 21$ total, with 11 for $\delta^{13}\text{C}$ and 10 for $N_{\mu\text{g}}$), which we interpret in light of population genetic theory about local adaptation. While we cannot definitely say that WUE, as measured by $\delta^{13}\text{C}$, contributes to local adaptation, we have described to a first approximation its genetic architecture, while noting several patterns consistent with $\delta^{13}\text{C}$ being a fitness-related trait affected by natural selection. These are useful results with which to generate further hypotheses about the evolution of genetic architecture contributing to local adaptation in natural populations (e.g. Holliday et al. 2015). Our results also shed light on ecologically relevant phenotypic trait variation useful for management decisions and predictions for range shifts under changing climates.

Chapter 6.

Current state and future directions of forest tree genomics

This work has been published in the following papers:

Lind, BM; M Menon, CE Bolte, TM Faske, AJ Eckert (2018) The genomics of local adaptation in trees: Are we out of the woods yet? *Tree Genetics & Genomes* 14 (2): 29. doi: <https://doi.org/10.1007/s11295-017-1224-y>

Abstract

There is substantial interest in uncovering the genetic basis of the traits underlying adaptive responses in tree species, as this information will ultimately aid conservation and industrial endeavors across populations, generations, and environments. Fundamentally, the characterization of such genetic bases is within the context of a genetic architecture, which describes the multidimensional relationship between genotype and phenotype through the identification of causative variants, their relative location within a genome, expression, pleiotropic effect, environmental influence, and degree of dominance, epistasis, and additivity. Here, we review theory related to polygenic local adaptation and contextualize these expectations with methods often used to uncover the genetic basis of traits important to tree conservation and industry. A broad literature survey suggests that most tree traits generally exhibit considerable heritability, that underlying quantitative genetic variation (Q_{ST}) is structured more so across populations than neutral expectations (F_{ST}) in 69% of comparisons across the literature, and that single-locus associations often exhibit small estimated per-locus effects. Together, these results suggest differential selection across populations often acts on tree phenotypes underlain by polygenic architectures consisting of numerous small to moderate effect loci. Using this synthesis, we highlight the limits of using solely single-locus approaches to describe underlying genetic architectures and close by addressing hurdles and promising alternatives towards such goals, remark upon the current state of tree genomics, and identify future directions for this field. Importantly, we argue, the success of future endeavors should not be predicated on the

shortcomings of past studies and will instead be dependent upon the application of theory to empiricism, standardized reporting, centralized open-access databases, and continual input and review of the community's research.

Introduction

Trees are plants with an arborescent habit, which is loosely defined as a tall-statured growth form usually producing wood (reviewed by Petit & Hampe 2006). Approximately 15% to 25% of plant taxa are classified as trees (Oldfield et al. 1998; Grandtner 2005; Wortley & Scotland 2004), with forested ecosystems accounting for approximately 30% of terrestrial vegetation (Costanza et al. 1997) and providing habitat for terrestrial biodiversity. Indeed, trees play important ecological roles in diverse communities across the globe, such as vertical structural habitat, seeds for wildlife forage, forest cover, the production of oxygen, carbon sequestration, air and water filtration, as well as the reduction of erosion, protracting snowmelt, and desertification. Of these, biological roles are ultimately defined by a set of life history characteristics common to most tree species (Petit & Hampe 2006). These include predominantly outcrossing mating systems with high levels of gene flow and fecundities, as well as long lifespans and generation times (Loehle 1988; Mitton & Williams 2006; Savolainen et al. 2007), although these may differ in, for example, clades of tropical trees. As a result, tree species typically have large effective population sizes, moderate to high levels of genetic diversity, and frequent occurrences of locally adapted ecotypes (Savolainen et al. 2007; Alberto et al. 2013; Sork et al. 2013; Boshier et al. 2015; Prunier et al. 2015; Holliday et al. 2017). Across species, however, rates of morphological and molecular evolution tend to be slow (reviewed in De La Torre et al. 2017). Additionally, genome size varies enormously across species of trees, ranging from 0.4Gbp to 31Gbp (reviewed in Neale et al. 2017). Recent sequencing efforts in gymnosperms, which represent the largest tree genomes, reveal that much of genome size variation is due to transposable element dynamics and gene family evolution (Leitch & Leitch 2012; Morse et al. 2009; Nystedt et al. 2013; Prunier et al. 2015; Neale et al. 2017) where duplication events of select gene families may

contribute to the ability of trees to colonize marginalized habitats (Leitch & Leitch 2012; Prunier et al. 2015; Neale et al. 2017).

In trees, the general presence of large geographical ranges and extensive gene flow also provides an ideal setting to disentangle neutral from selective evolutionary processes (Neale & Kremer 2011). Indeed, their longevity and wide and heterogeneous geographical distributions lend trees suitable for addressing several key evolutionary questions about the importance of historical climatic fluctuations, and local adaptation involving shifts in allele frequencies (Lotterhos & Whitlock 2014; Savolainen et al. 2007, 2013; Platt et al. 2015). As we detail in subsequent sections, evidence consistent with local adaptation in trees is ubiquitous, even across fine spatial scales where it had been hypothesized that gene flow may overcome selection of locally favored alleles (e.g., Mitton et al. 1998; Budde et al. 2014; Csilléry et al. 2014; Vizcaíno-Palomar et al. 2014; Eckert et al. 2015; Holliday et al. 2016; Roschanski et al. 2016; Lind et al. 2017).

Quantitative phenotypes are often used as a proxy for total lifetime fitness, which is composed of two broad components: survival and reproduction. Since most quantitative traits are related to some component of total lifetime fitness, they are often used to assess potential for local adaptation. For many plant taxa, selection pressures are expected to be strongest for variation in survival during the juvenile stages of development (Donohue et al. 2010), particularly for those taxa with high reproductive output, as is the case for many tree species. As such, juvenile stages in plants have been found to contribute substantially to total lifetime fitness (Postma & Agren 2016). Phenotypic traits associated with juvenile survival have thus received the majority of genetic research focus in trees, particularly due to their long-lived nature. Such studies have led to intriguing insights gained through a long history of common garden experimentation (Langlet 1971; Morgenstern 1996). For example, traits such as growth (e.g., height and diameter), form (e.g., specific gravity, straightness), phenology (e.g., bud flush, bud set), juvenile performance (e.g., germination rate, seed traits) and physiology (e.g., cold hardiness, water-use efficiency) have all been shown to be under moderate to high genetic control (reviewed in Corn-

elius 1994, Howe et al. 2003, Alberto et al. 2013; this review). Variation for these traits is also often partitioned among populations (this review), despite the vast majority of neutral variation remaining within populations (Howe et al. 2003; Neale & Savolainen 2004). With few exceptions (e.g., major gene resistance in the white pine-blister rust pathosystem; Kinloch et al. 1970; Liu et al. 2017), variation for these traits forms a continuum across individuals, thus implying that the underlying genetic architecture is composed of a large number of small to moderate effect loci (i.e., a polygenic architecture; concept reviewed in Savolainen et al. 2007, 2013; Gagnaire & Gaggiotti 2016; Hoban et al. 2016; Timpson et al. 2017). There is some uncertainty, however, concerning the properties of the effect size distributions comprising polygenic architectures (sensu Fisher 1930, Kimura 1983, and Orr 1998), the relative importance of various forms of gene actions (e.g., dominance, epistasis) in producing trait variation (Crow 2010, Hansen 2013), how these interact to affect the evolution of polygenic architectures in natural populations (Hansen 2006), and how these factors will ultimately influence evolutionary processes and outcomes in forest trees (Savolainen et al. 2007; Sork et al. 2013; Prunier et al. 2015). Considerable strides, made in the past through genotype-phenotype-environment studies (sensu Sork et al. 2013), have contributed intriguing insight into the genomic basis of local adaptation for tree species. However, given the large genome size of many tree species, such methods have been criticized as lacking in power and sufficient coverage needed to detect small effect loci, which is further exacerbated by rapid decay of linkage disequilibrium (LD) in most forest trees (Mackay 2009; Savolainen et al. 2007). Despite these limitations, association studies have been moderately successful in linking genotypes and phenotypes, including providing information for making inferences about local adaptation.

In this review, we set out to summarize theory related to polygenic local adaptation and, using these expectations, contextualize the progress of describing the genetic architectures underlying traits important to conservation and industry in undomesticated tree species. We first highlight the extensive evidence for local adaptation in trees by reviewing transplant designs often

used in investigations of quantitative genetic differentiation. Using an extensive literature survey across both gymnosperm and angiosperm species, we provide an overview of these transplant methods, give examples of each, and quantify the distribution of narrow sense heritability and Q_{ST} estimates across various trait categories. We further use this survey to establish patterns of comparative quantitative and neutral genetic differentiation (i.e., $Q_{ST}-F_{ST}$ tests) which until this review had not been suitably synthesized in trees. Before we transition into discussing common methods used to uncover loci underlying adaptation, we establish expectations for the genetic architecture of polygenic, fitness-related traits by reviewing the theory available to date. We then provide an extensive review of genotype-phenotype associations in trees and provide the distribution of the percent phenotypic variance explained by empirically associated loci. Using this distribution, we underscore the limitations of using solely single-locus approaches to uncover the loci underlying local adaptation in tree species. Given this synthesis, we highlight exemplary genomic resources available to fill knowledge gaps, identify promising avenues of future research, identify key benchmarks and necessary steps towards truly integrating studies of trees into the genomic era, and address our primary question, “Are we out of the woods yet?”.

Identifying heritable phenotypic variation

Trees have evolved numerous adaptations as a result of their vast ecological breadth. As such, it has long been the goal of forest scientists to understand the traits important to viability and persistence. Among the most frequent designs used, common gardens and reciprocal transplants have aimed at describing genetically based differentiation of measured phenotypes among various source populations of varying sizes and across various geographic scales. Across these designs, investigators seek to better understand the phenotypes relevant to local adaptation and the selective pressures influencing these phenotypes. The exact design chosen, however, is generally based on the questions driving the research endeavor and often by the availability of resources (Morgenstern 1996; Blanquart et al. 2013; de Villemereuil et al. 2015). In this section,

we briefly review these designs, identify relevant questions and inferences, highlight some of the important practical applications of these techniques, and discuss examples of past investigations in various tree species.

There is a rich history of forest scientists using the common garden approach dating back hundreds of years (Langlet 1971; Mátyás 1996). In a broad sense, a common garden design is used to test for differentiation among genetically distinct groups in a homogeneous environment. These groups can be clonal replicates or sibships (families) derived from species or hybrids sampled from various populations, provenances, varieties, cultivars, or agricultural accessions (Cheplick 2015). When individuals from various origins are grown together under the same conditions, the observed phenotypic differentiation is expected to reflect underlying genetic variation, especially when maternal effects are assumed or shown to be absent. Common garden and provenance trial designs can also establish evidence that the phenotypes under study are heritable, a prerequisite for an adaptive response to selective agents (Supplemental Box S1), and that populations exhibit quantitative genetic differentiation (i.e., Q_{ST} ; Spitz 1993). When driven by questions related to differentiation alone, a single common garden approach can be used to describe levels of quantitative genetic variation within and among genetically distinct groups. In these cases, no environmental variables are manipulated, and thus, unequivocal evidence for trait divergence among groups, and the contributing factors influencing this divergence (e.g., neutral or selective processes), is often limited because conclusions must be based on *post hoc* inferences about source environments for the materials established in the common garden. Even so, single common garden approaches can be a powerful tool to demonstrate evidence congruent with local adaptation. For instance, the white carob tree (*Prosopis alba* Griseb., Leguminosae) growing in Argentina is an ideal multipurpose tree that has potential for use in reforestation and afforestation applications in the region. However, this genus is known to invade other regions, encroach on farmland and waterways, and has a thorny growth habit that can cause sepsis in livestock. To better understand how forestry applications can balance the benefits of production

and forest protection, Bessega et al. (2015) used a single common garden representing eight provenances of *P. alba* to compare estimates of neutral genetic patterns to the quantitative genetic variation of life history traits related to economic importance. They found that for most traits there existed considerable underlying genetic variation ($\overline{Q_{ST}} = 0.139$). Additionally, source environments were often correlated with measured trait variation in the common garden, suggesting that the observed differentiation was driven by temperature, precipitation, wind speed, and sunshine fraction, with signals of divergent selection corroborated across Q_{ST} - F_{ST} comparisons and tests for selection (e.g., S test, *sensu* Ovaskainen et al. 2011). Bessega et al. (2015) concluded that the signal of non-neutral differentiation was indicative of divergent phenotypic optima across populations, and that this variation could be used to direct future breeding programs across the region.

When there is evidence that environmental differences among source populations may be driving adaptive divergence, strong environmental candidates can be manipulated (artificially or via site selection) in a multiple common garden design to further investigate hypotheses of differentiation and adaptation. For instance, the sweet chestnut (*Castanea sativa* Mill., Fagaceae), also known for its edible fruit, is distributed across much of Minor Asia and southern Europe and is an ecologically important component of many Mediterranean systems. *Castanea sativa* exhibits ecological, physiological, morphological, and genetic variability as the range overlays a climatic transition from xeric Mediterranean conditions to wetter Euro-Siberian environments (see refs in Lauteri et al. 2004). Previous common garden experiments carried out by Lauteri and colleagues have indicated that populations across this transition are further differentiated by water use efficiency (the ratio of plant carbon gain to water loss) and carbon isotope discrimination, Δ . To further explore variability of drought-related traits, Lauteri et al. (2004) used an *ex situ* multiple common garden design using two water and temperature treatments in individual climatic

chambers to assess differentiation among six populations across Spain, Italy, and Greece. They found *treatment* and *population x treatment* effects were significant, suggesting variation in drought adaptation across populations. Additionally, populations originating from dry sites generally exhibited higher values of Δ , which was also composed of significant additive genetic variation ($h^2 = 0.15-0.52$), and suggests that genetic and physiological mechanisms of drought adaptation confer a capacity to colonize a wide range of environmental conditions, while strong negative relationships between Δ and growth-related traits is suggestive of strong evolutionary constraints at juvenile stages.

While *ex situ* common gardens approaches (e.g., Lauteri et al. 2004) can provide strong evidence of adaptive divergence among populations, and in some cases corroborate putative drivers of observed differentiation, these studies can often exclude key environmental factors, possibly leading to confounding signals of adaptation (Kawecki & Ebert 2004). When *in situ* experimentation is feasible, site selection can be used to test for environmental drivers of local adaptation. For example, Evans et al. (2016) investigated traits related to growth and phenology in juvenile narrowleaf cottonwood (*Populus angustifolia* James, Salicaceae) by planting families from nine populations across the native range into three common gardens, one each at the northern, southern, and interior extent of the range. Using $Q_{ST}-F_{ST}$ comparisons and clinal analyses alongside the quantitative genetic analyses, Evans et al. (2016) concluded that climate cues played a major role in structuring adaptive variation across the range of *P. angustifolia*, and that future industrial and conservation applications should utilize this information to inform source environments for optimal outcomes.

As both *in situ* and *ex situ* common garden trials can include multiple environmental influences in their design, reciprocally transplanting to all source environments is not necessarily a requirement to decompose genetic variation underlying adaptive traits or to provide evidence for, or the drivers of, differentiation among populations. Thus, these designs may preclude

inferences regarding local adaptation *sensu stricto*. To produce such evidence, source populations can be planted in a (full- or incomplete-factorial) reciprocal transplant design and allow for traits related to fitness to be assessed across native and non-native environments. If a population is locally adapted, individuals exposed to their native environments should show increased growth, survival, and reproduction relative to non-native genotypes (Kawecki & Ebert 2004; Leimu & Fischer 2008; Hereford 2009; Savolainen et al. 2013). For example, with the goal of delineating conservation units based on molecular and quantitative trait differentiation, Rodríguez-Quilón et al. (2016) used four reciprocally-transplanted common gardens to assess height and survival of samples from 35 natural populations of maritime pine (*Pinus pinaster* Aiton, Pinaceae). For both traits, Q_{ST} was consistently larger than F_{ST} across the four sites, a pattern suggestive of divergent selection. Six distinct gene pools based on evolutionary history of neutral markers were identified, and because high quantitative differentiation (Q_{ST}) was found within these pools, hierarchical analyses were used to further identify ten adaptive population groups for use in conservation and breeding approaches.

Available evidence suggests that many populations of tree species have substantial heritable genetic variation, and that the quantitative traits under study often show signals of divergent selection across both broad and fine spatial scales. But how broadly can we apply this statement? Are there overall patterns of heritability and quantitative genetic structure across tree species? Because estimates of heritability and Q_{ST} are often only applicable to a specific set of populations, for a specific set of environments, at any specific point in time (e.g., see Figure 6.2D), a large sample of these estimates is therefore necessary to synthesize the current literature with regard to patterns across taxa. To accomplish this aim, we synthesized estimates from 129 published studies with estimates of narrow sense heritability ($n = 114$) from replicated progeny trials and/or estimates of quantitative genetic differentiation (Q_{ST} ; $n = 37$). However, we excluded papers that have been cited for estimates of Q_{ST} or heritability that were calculated *post hoc* from

variance components (i.e., we only recorded estimates that were explicitly reported as h^2 or Q_{ST} in the original publication). For comparison, we further grouped measured traits into 14 broad categories: cold hardiness, disease resistance, drought hardiness, form, growth, herbivore and insect resistance, leaf and needle properties, phenology, plant secondary metabolites, reproduction, resource allocation, seed and early germination properties, survival, and wood properties. Because sample size can influence the precision of both heritability and Q_{ST} , for each trait category we used a weighted average where weights were equal to the number of families used to estimate variance components for each estimate of h^2 and Q_{ST} .

In agreement with Cornelius (1994), our survey found that many of the traits important to conservation and industry exhibit non-zero narrow sense heritability ($\overline{h^2} = 0.367$; File S1; Figures S6.1-S6.4) and are thus amenable to selection. The mean weighted Q_{ST} across traits groups from our survey (Table S6.1; File S6.1) was between 0.10-0.28, except for drought hardiness (0.06) and disease resistance (0.04), with median values from the unweighted distribution generally falling below the weighted average for each trait group (Figure 6.1). This suggests that over various geographic and environmental distances, population histories, and species, there is a general pattern of substantial genetic variation underlying measured traits. Given our synthesis of Q_{ST} estimates in trees, we were curious of the evidence for adaptive divergence among populations ($Q_{ST} > F_{ST}$). Of the 37 articles reporting Q_{ST} estimates in our review, 23 compared Q_{ST} with F_{ST} or G_{ST} estimated from the same populations under study (however, we excluded studies that used F_{ST} measurements taken from the literature, e.g., as in McKay & Latta 2002; Alberto et al. 2013). Indeed, as pointed out by Crnokrak & Merilä (2002), comparisons of Q_{ST} and F_{ST} estimated from different populations and/or at different time points are uninformative. Of these 23 studies, 18 compared Q_{ST} and F_{ST} in a statistical framework while the remaining five studies compared Q_{ST} and F_{ST} numerically. Across numerical and statistical comparisons combined, 67% (254 of 381 traits) exhibited higher Q_{ST} than F_{ST} , with 69% (170 of 246 traits) exhibiting significantly

higher Q_{ST} than F_{ST} . Although we did not tally instances where Q_{ST} was reported to be less than F_{ST} (statistically or otherwise), as this was not the focus of our review, there were some instances in which this was the case. For instance, Lamy et al. (2011) found such patterns when quantifying population genetic differentiation of cavitation resistance across the species range of maritime pine (*Pinus pinaster* Aiton, Pinaceae), while Mahalovich et al. (2011) also found that $Q_{ST} < F_{ST}$ for traits related to white pine-blister rust resistance in inoculated seedlings of whitebark pine (*Pinus albicaulis* Engelm., Pinaceae). While various explanations for such patterns were outlined by Lamy et al. (2011), canalization was argued as the most likely process driving the observed patterns, while Mahalovich et al. (2011) offered similar arguments for selection favoring the same

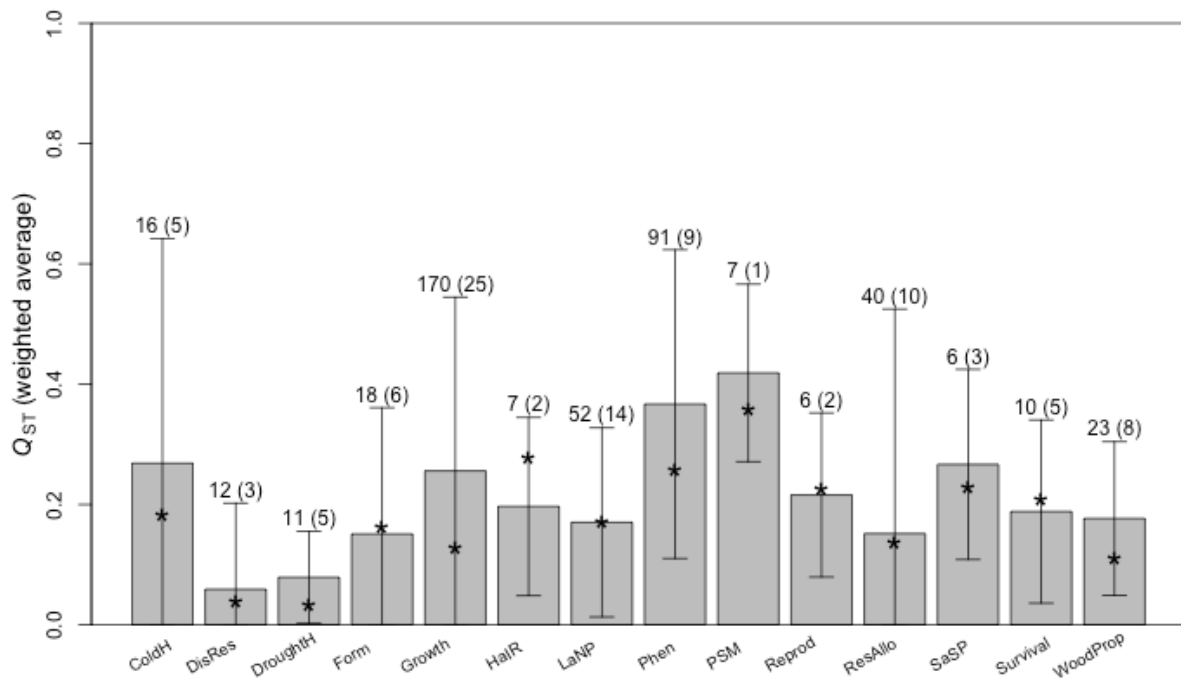


Figure 6.1. Average Q_{ST} for each of 14 trait categories from literature review calculated by weighting each estimate by the number of families used in the estimation. Error bars represent the standard deviation of the weighted averages. Numbers above error bars represent total number of estimates, with total number of unique species in parentheses. Asterisks indicate median values of the unweighted Q_{ST} distribution. ColdH = cold hardiness, DisRes = disease resistance, DroughtH = drought hardiness, HalR = herbivore and insect resistance, LaNP = leaf and needle properties, Phen = phenology, PSM = plant secondary metabolites, Reprod = reproduction, ResAllo = resource allocation, SaSP = seed and seedling properties, WoodProp = wood properties.

genotype in different environments (see Lamy et al. 2012 for more regarding this aspect).

Despite neutral genetic differentiation partitioned primarily within populations, adaptive genetic variation seems to be structured to a greater degree across populations, more often than not, for the various fitness-related traits reviewed here. Such a pattern is indeed consistent with local adaptation, assuming that (among other considerations such as the recency of selection) mutation rates are considerably lower than migration rates in these populations (Whitlock 1999; Hendry 2002; Leinonen et al. 2013). In any case, given an extensive literature supporting the local adaptation hypothesis in trees, our results appear consistent with patterns of selective forces acting on abundant, heritable genetic variation across populations, even in the face of gene flow (discussed further in the next section).

Expectations for the loci underlying quantitative traits

The homogenous environments of the common garden and reciprocal transplant designs are ideally suited to test hypotheses of local adaptation in trees (Sork et al. 2013). However, uncovering the genetic basis and contributory influence of specific loci underlying these adaptive traits is a sizable endeavor on its own, and the success of such pursuits will be determined, in part, by the trait's underlying genetic architecture (i.e., the number, effect size, type, location, expression, pleiotropic effect, environmental influence, and interaction of underlying loci), which is generally not known a priori (Stinchcombe & Hoekstra 2008; Rellstab et al. 2015; Savolainen et al. 2013; Hoban et al. 2016; Burghardt et al. 2017; Wadgyamar et al. 2017). Much of our early understanding of the architectures of complex traits came shortly after Nilsson-Ehle (1909) and East (1910) independently demonstrated evidence for multiple-factor inheritance, where Fisher (1918) laid the groundwork for quantitative genetics by incorporating the additive properties of variance to partition phenotypic variation into components tractable to a model of Mendelian inheritance. It was this work, and that of Fisher's geometric model (1930), which founded the basis for attributing continuous variation of phenotypes to a polygenic model of many underlying heritable components of mainly small effect. From this model, Fisher (1930) concluded that

mutations of small effect were the main drivers of adaptation, suggesting large-effect substitutions to contribute little to adaptation due to negative pleiotropic effects constraining effect size. Therefore, the fate of a given locus would be conditioned on its average, marginal effect on fitness calculated across the species, with non-additive deviations from this linear model of inconsequential influence. This micro-mutationist view, to a large extent, remained the dominant thought for nearly half a century (Orr 2005). It was then that Kimura (1983) established that for an allele to contribute to adaptation, it would need to survive the stochastic nature of drift. Thus, new mutations of low frequency and effect were less likely to contribute substantially to adaptive evolution. Considering the adaptive contribution probability of large and small effect loci, Kimura concluded that mutations of moderate effect would be the most plausible. Years later, Orr (1998) showed that over the entire bout of selection via an adaptive walk, the distribution of fixed substitutions resembles an exponential distribution, with effect size decreasing with the proximity to the phenotypic optimum. In addition, the distribution of fitness effects of beneficial mutations is also expected to be exponential (Orr 2003; for more discussion on this aspect, see also Orr 2006; Eyre-Walker & Keightley 2007; Martin & Lenormand 2008, Kopp & Hermisson 2009b; Keightley & Eyre-Walker 2010, Dittmar et al. 2016). Despite major advances in theory and technology, there still remains substantial uncertainty regarding the exact number of loci underlying many adaptive traits, the effect size distribution of these loci, and how the number of underlying loci and effect distribution may change under various evolutionary regimes (Orr 2001; Slate 2005; Hansen 2006; Mackay et al. 2009). In this section, we describe how various factors can contribute to the (perhaps, effective) number of causative loci, and the distribution of effects underlying continuously distributed adaptive traits, beginning first with aspects of the architecture itself (gene action), and concluding with explanations of how various processes (e.g., selection) play an influential role in the evolution of underlying genetic architectures. Establishing these expectations is essential for assessing common approaches and guiding future directions. In the next section we then compare these expectations with methods used in, and results from, genotype-phenotype

associations in trees. While we discuss these examples in isolation, we highlight the fact that the underlying biological processes are often not independent.

Gene action

The classical genotype-phenotype map is largely one of additive effects, and is represented by a statistical regression of the phenotype on genetic content, as developed by Fisher (1918) and extended by others (e.g., Cockerham 1954; Kempthorne 1954). Indeed, much of the work done in trees has relied on such additive effects to describe heritable and quantitative genetic variation (see previous section). In this model, the phenotypic variance is partitioned into orthogonal (i.e., independent) contributions from the genetic variance (σ_G), environmental variance (σ_E), and the variance due to interaction between genotype and environment ($\sigma_{G \times E}$; Figure 6.2; see Supplemental Box S6.1). Further, σ_G is also the sum of orthogonal variance components, each term representing a different form of gene action. The additive, dominance, and epistatic terms respectfully designate the associated variance contribution of independent alleles, the non-additive contribution to variance of interactions among alleles at the same locus, and the contribution to variance of non-additive interactions among alleles at different loci (the latter of which can take one of many forms such as additive-by-additive, additive-by-dominance, etc.; Lynch & Walsh 1998). As a result, non-additive gene action is minimized as non-linear contributions to the overall phenotype (Moreno 1994; Whitlock et al. 1995) which contributes little to the distinction of the different forms of dominance and epistasis (Cheverud & Routman 1995; Hansen & Wagner 2001; Hermisson et al. 2003; Hansen 2006; Mackay 2014) nor towards the inference of aspects of the underlying genetic architecture in general (Nelson et al. 2013; Huang & Mackay 2016).

These statistical conveniences afforded by Fisher and others led to the notion that such non-additive effects were transient (i.e., are due to LD, which will decay with the relaxation of selection), or that trends of statistical epistasis were representative of functional epistasis in general, and therefore epistasis was unimportant to evolutionary dynamics (e.g., Bulmer 1980;

Crow 2008, 2010; Hill et al. 2008). While minimized in a statistical regression, this does not necessarily mean that epistasis and dominance will not have a profound impact on the genetic architecture, or towards a given population or species' long-term evolutionary trajectory, even if statistical epistatic or dominance variance is minimal (Goodnight 1988; Chevrud & Routman 1995; Hansen & Wagner 2001; Hansen 2013; Nelson et al. 2013; Griswold 2015; Paixão & Barton 2016). Indeed, parameterizing a model in which the type I sums of squares is determined by non-additive parameters, as opposed to additive variance in the conventional regression model, the majority of genetic variation is still captured by the primary effect in the model regardless of the underlying architecture (Huang & Mackay 2016). Given the prevalence of evidence for non-additive contributions (e.g., Phillips 2008; de Visser et al. 2011; see also references in Hansen 2013), it is likely that non-additive effects will play a role in evolutionary outcomes. For instance, Huber et al. (2017) showed that the degree of dominance in *Arabidopsis* is an outcome based upon functional importance and optimal expression level. Further, Carter et al. (2005) show that, relative to a purely additive trait (or with non-directional epistasis) under directional selection, positive and negative epistasis can respectfully increase or decrease the additive genetic variance, and thus increase or decrease the rate of phenotypic response to selection (see also Le Rouzic & Álvarez-Castro 2016). As Jones et al. (2014) show, for a two-trait phenotype controlled by pleiotropic and epistatic effects, epistasis in the presence of selection can also affect the mutational architecture of complex traits, where the average allelic effect evolves to be negatively correlated with the average epistatic coefficient, the strength of which is greater in larger population sizes. Yet, as described by Barton et al. (2016), and further discussed by Barton (2017) and Paixão & Barton (2016), the infinitesimal model can be generalized to include epistatic effects, particularly when the number of underlying loci is large and selection on individual loci is weak. In the case of non-systematic, weak pairwise epistasis, and without mutation or environmental noise, the infinitesimal model holds to a good approximation (Barton et al. 2016). In the case of sparse epistasis with selection and a large number of loci, the change in the trait

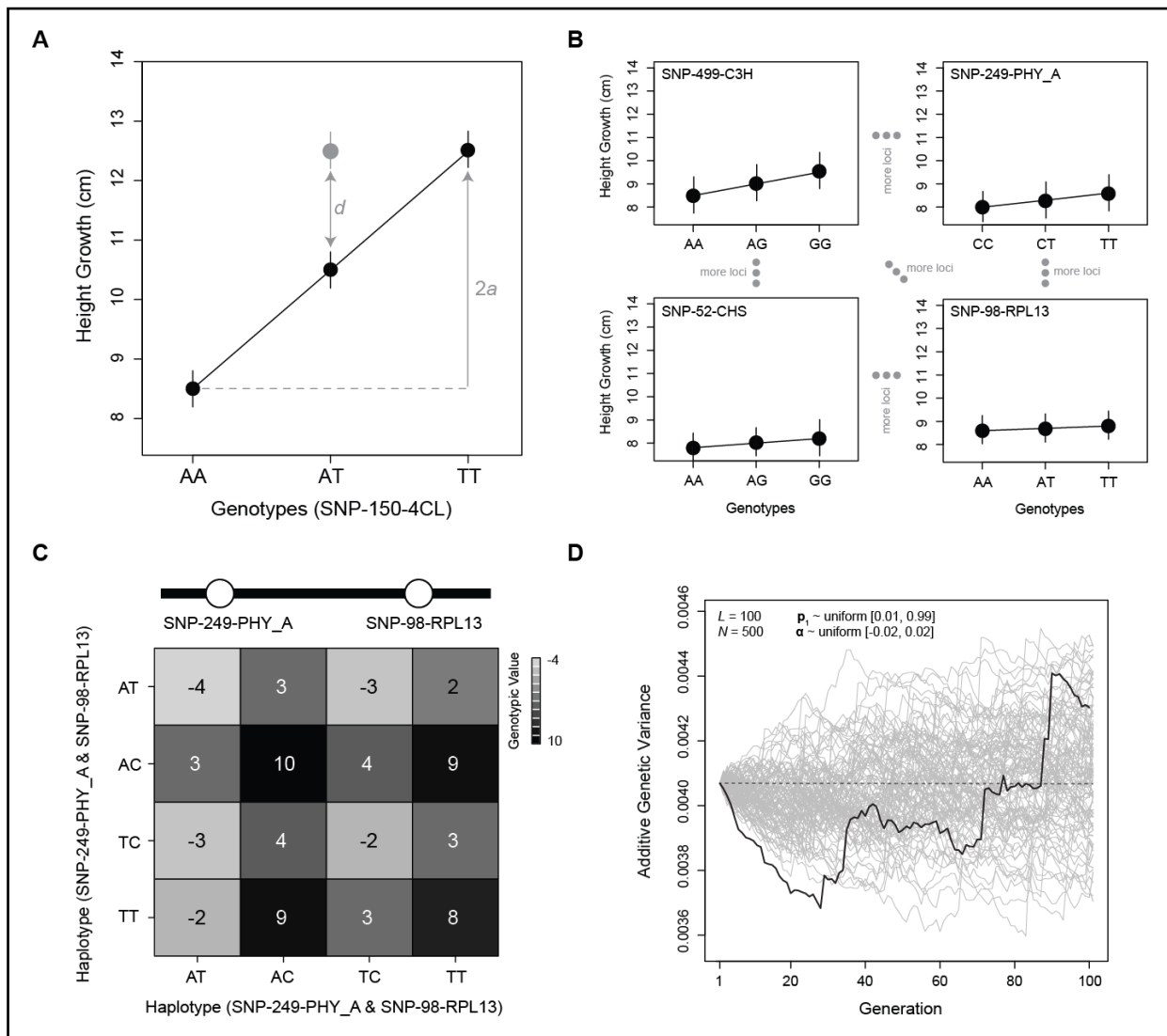


Figure 6.2. Relevant quantitative genetic concepts are needed to understand the evolution of polygenic traits. (A) Additive and non-additive effects at a single locus, where a is defined as the additive effect (also known as the average effect of allelic substitution [α] when there is no dominance) and d is defined as the dominance deviation. With dominance, $\alpha = a[1 + k(p - q)]$, where k is the degree of dominance ($k = 0$: additive, $k = 1$: dominance, $k > 1$: over-dominance, see Lynch & Walsh 1998). (B) Polygenic traits are determined by multiple genes, each with additive (shown) and non-additive (not shown) effects. The total additive effect is the sum of the additive effects at all causative loci. (C) Additive-by-additive epistasis, where the additive effect of an allele at the PHY_A SNP depends on what allele it is paired with at the RPL13 SNP. In this case, the effects can be thought of as dependent in the following manner using the four possible haplotypes at the PHY_A (A/T SNP) and RPL13 (C/T SNP) SNPs – AC: +5, AT: -2, TC: -1, TT: 4. (D) The effect of genetic drift on the additive genetic variance as determined by 100 independent, causative loci. Each line represents a simulation of genetic drift in a constant sized population ($n = 500$ diploids) conditioned on initial allele frequencies across loci (p_i) and effect sizes (α). The expected mean across all 100 simulations is given by the dashed black line. Any given simulation can deviate strongly from this expectation (solid black line). Thus, when the elements of p change over time, in this case due to genetic drift, so does the additive genetic variance. See also Supplemental Box S6.1.

mean over 100 generations is greater than that under a purely additive architecture, and the decrease in additive genetic variance exceeds, to an extent, that of the neutral case after about 30 generations (which is exacerbated with simpler architectures), with a reduction of the frequency of segregating alleles with positive effect on the trait (Barton et al. 2016; Barton 2017). Despite an ongoing debate within the literature (Wright 1932; Whitlock 1995; Crow 2008, 2010; Gibson 2012; Zuk et al. 2012; Hansen 2013; Hemani et al. 2013; Nelson et al. 2013; Mäki-Tanila & Hill 2014; Ávila et al. 2014; Paixão & Barton 2016), and given that there seems to be no general prevalence of either positive or negative epistatic interactions (Mackay 2014), the infinitesimal model is likely to continue to contribute to our understanding of the evolution of complex traits, as exemplified in its application towards breeding applications (Turelli & Barton 1994) and specifically those successfully applied to trees (Savolainen et al. 2007; Thavamanikumar et al. 2013; Isik et al. 2015; Grattapaglia 2017). Ultimately, the success of such models will be conditioned on the context, as well as the distinction between physiological and statistical gene action. Here, (higher order) non-additive contributions to phenotypic variance will likely have minimal deviations from the limit of the infinitesimal model in the short-term, particularly if this is primarily due to independent, low-order interactions, and should thus be applied with this in mind. As such, while short-term evolutionary processes are likely to hold in this limit, identifying the non-additive loci which underlie the trait, and their respective gene action, may still need further inquiry (Grattapaglia 2017). Indeed, it is often argued that non-additive gene action is too often neglected in studies of complex traits (e.g., Carlborg & Haley 2004), possibly due to the large sample sizes required to detect significant interactions, and lack of statistical power incurred due to multiple hypothesis testing (Mackay 2014). Given the recent reduced cost of sequencing technology and availability of novel computational and laboratory tools, future studies incorporating investigations of epistasis and dominance (where appropriate and feasible) would contribute to our understanding of genetic architectures, quantitative trait evolution, and breeding applications in trees (Vitezica et al. 2017). For example, breeding applications assessing hybridization across

divergent backgrounds, as is also prevalent across species in nature, have shown the importance of non-additive effects in phenotypic outcomes (as in *Eucalyptus*, e.g., Tan et al. 2017, and *Pinus*, e.g., Dungey 2001). Even so, the additive model is still a powerful tool to describe the loci underlying adaptive traits.

Pleiotropy is another considerable factor influencing the expectations of the genetic architecture of quantitative traits, its evolution or evolvability, and indeed the genotype-phenotype map (Hansen 2003; Orr 2006; Chevin et al. 2010b; Tenallion 2014). While multiple definitions exist across the literature (see Paaby & Rockman 2013), pleiotropy is generally identified as a single locus influencing multiple phenotypic traits. Other than linkage disequilibrium, pleiotropy is the fundamental cause of genetic covariance among phenotypes (Lande 1980). Given that the number of independent traits under selection is likely limited (Barton 1990), pleiotropy likely plays a substantial role in evolutionary dynamics. It is expected that as the number of traits, n , influenced by a locus increases, the probability of a beneficial mutation will decrease with the effect size of a mutation; where the effect size, r , relative to the distance to the phenotypic optimum, d , must be (much) less than d in order to be beneficial (Fisher 1930; the so-called 'cost of complexity': Orr 2000). Yet, empirical data seem to contradict this hypothetical cost, as the effect size of mutations often do not scale with pleiotropy in this way, and instead increase with the dimensionality of targeted traits (Wagner et al. 2008; Wang et al. 2010). Additionally, universal pleiotropy, where all mutations affect all phenotypes, and where there is no net directionality of mutations (i.e., mutational isotropy; both aspects as in Fisher 1930), has also been challenged by findings which suggest that only a fraction of phenotypic traits are affected by pleiotropic loci (Wagner et al. 2008; Wang et al. 2010). Relaxation of such assumptions from Fisher's geometric model have shown that the total number of traits affected by pleiotropy has a relatively decreased effect on the rate of evolution in more general models (e.g., Martin & Lenormand 2006; see also Simons et al. 2017, and references in Wagner & Zhang 2011 and Tenailon 2014). It seems that if model organisms (e.g., Pickrell et al. 2016, Smith 2016) are taken as a bellwether for

expectations in trees, pleiotropy is likely a contributing factor for many quantitative traits. Thus, the fraction of beneficial mutations is likely limited when the number of traits influenced is large, suggesting that the cost of complexity (or, more precisely, pleiotropy) may be generally robust (Welch & Waxman 2003), particularly when a population is close to its phenotypic optimum where selection acts against dimensionality of pleiotropic effects (Zhang 2012). Thus, the degrees of pleiotropy across underlying loci, distance from phenotypic optima, and covariance among traits under selection can have profound effects on evolutionary outcomes (e.g., as in *Pinus contorta*, Lotterhos et al. 2017). This is particularly true for the evolvability of architectures and distribution of effect sizes, which further depends on the variational autonomy of the traits affected by pleiotropy and the modularity of mutations, the former of which is ultimately determined by the direction and size of effect among a set of pleiotropic loci across a set of characters (see Arnold 1992; Wagner & Altenberg 1996; Hansen 2003, 2006; Wagner et al. 2007; Chevin et al. 2010b; Wagner & Zhang 2011; MacPherson et al. 2015).

In many investigations of local adaptation, the primary interest is in trait evolution and thus the underlying genetic components. As such, environmental effects and interactions are not often pursued, or perhaps even detected (Yoder & Tiffin 2017), particularly in studies of a single common garden or environment, and are instead treated in much the same way as epistatic interactions discussed above. Nonetheless, genotypic effects can evolve through genotype-by-environment interactions with a changing environment just as is the case for the evolution of non-additive interactions with a changing genetic background (Hansen 2006). Indeed, it is likely that consistent fluctuations in the environment would select for environmentally-perceptive responses, which seems to be the case across many tree species (Li et al. 2017). The contribution to the effect size distribution from GxE interactions will be a function of the variation in selection across the environments experienced by the interacting allele(s) as well as the level of gene flow between environments and fitness differences among various genetic backgrounds, but to our knowledge such information (to the extent of that for e.g., selective sweeps) is lacking within the literature.

Negative selection

Negative selection acts against deleterious mutations that arise within populations. It is one, but not the only, mechanism that underlies stabilizing selection, defined at the level of the phenotype where deviations from an optimal value are selected against. Optima in this framework can be thought of either globally (i.e., across all individuals) or locally (i.e., individuals within a population), where the latter can have varying optima across populations. The nature of the optima (i.e., being local or global) affects the detectable trait architecture. For example, trait architecture should be composed of rare alleles with a negative relationship between effect size and allele frequency (cf. Eyre-Walker 2010 and references therein), where this relationship can also be confounded with degree of dominance and gene expression network connectivity (Huber et al. 2017), under models of a single global optimum. From a population genetic perspective, the ubiquity of negative selection is encapsulated in the name background selection, which has extensive reviews about its presence in natural systems (Charlesworth 2013), its importance for the neutral and nearly neutral theories of molecular evolution (Ohta 1992, 1996), and its contribution to observable patterns of hitchhiking (Stephan 2010). Important for the study of polygenic adaptation and its architecture, however, is that loci identified using GWAS may also include segregating deleterious variation (as argued and hinted at in Eckert et al. 2013b; cf. Yang et al. 2017; Gazal et al. 2017) as this creates trait variance, with little known about their prevalence (including differential prevalence across traits), differentiation in frequencies across populations (but see Zhang et al. 2016), and effects on downstream inferences about divergent selection pressures across populations. It is sets of GWAS loci, though, that are currently analyzed for signatures of local adaptation via spatially divergent (i.e., locally positive) natural selection (e.g., Berg & Coop 2014).

Recent exemplary work with expression networks in *Populus tremula* L. (Salicaceae; Mähler et al. 2017) and the herbaceous *Capsella grandiflora* Boiss. (Brassicaceae; Josephs et al. 2015, 2017a) have revealed intriguing insight into the effects of negative selection on the architecture of complex traits in plants, as well as the relationship between network connectivity

and the strength of negative selection. In *P. tremula*, genes with expression levels that were significantly associated with sequence variation were found more often in the periphery of the co-expression network (lower network connectivity) than within network module hubs (higher connectivity), while expression-associated SNPs were negatively correlated with network connectivity and effect size, a pattern also found between connectivity and expression variance, and minor allele frequency and QTL effect size (Mähler et al. 2017). Genes associated with sequence variation had less skewed site-frequency spectra (i.e., the frequency distribution of allelic variants) and lower estimates of nonsynonymous to synonymous divergence (dN/dS) than genes not associated with sequence variation, together suggesting that genes within the periphery of co-expression networks are likely under less selective constraint than those genes with high network connectivity which likely experience greater intensities of purifying selection. These genes thus tend to have more segregating variation and may be those most likely to be detected with current sample sizes utilized in GWAS, which has implications for estimation of trait architecture and its 'degree' of polygenicity. Even so, while there is prevalent evidence of negative selection in trees (e.g., Krutovsky & Neale 2005, Palmé et al. 2009, Eckert et al. 2013a,b; De La Torre et al. 2017), more inquiry is needed.

Positive selection

The temporal and spatial heterogeneity of selection can impact the evolution of genetic architectures underlying adaptation. These impacts are often thought of on a spectrum of trade-offs, with one end being antagonistic pleiotropy where allelic effects vary between positive and negative on fitness across populations, and the other being conditional neutrality where allelic effects on fitness are positive in one or more populations and nearly zero in others (Anderson et al. 2012, Savolainen et al. 2013). For instance, alleles incorporated into a population after a shift in environmental influence can increase from low to high frequency via positive selection. The existence of such a beneficial allele can manifest in several ways: from new mutations,

introgression through gene flow, or molecular reorganization through novel recombination, inversion, transposition, copy number variation, or insertion-deletion events. If there is strong selection acting on this allele ($Nes \gg 1$), it will sweep to high frequency creating a signature of reduced polymorphism at neutral sites physically linked to the allele ('genetic hitchhiking', Maynard Smith & Haigh 1974) resulting in a hard 'selective sweep' (Berry et al. 1991). However, in structured populations with limited gene flow, this process can take significantly longer to reach fixation, resulting in incomplete sweeps (Whitlock 2003). Additionally, Pavlidis et al. (2012) found that, in congruence with Chevin & Hospital (2008), a multilocus genotype often prevents the trajectories of individual alleles from sweeping to fixation, with an increasing number of loci leading to decreasing probability of fixation, and as a result, an altered selective signature at such loci (see also Jain & Stephan 2017). As such, hard selective sweeps in a polygenic architecture are expected to be rare (but not completely absent) under most circumstances, particularly when the shift in environment causes a relatively small deviation from the phenotypic optimum. Thus, hard sweeps most likely apply to loci with relatively large effect above a calculated, context-dependent threshold value (Orr 2005; de Vladar & Barton 2014; Stephan 2015; see specifically Jain & Stephan 2015, 2017).

While early literature (Maynard Smith & Haigh 1974; Kaplan et al. 1989) focused on the rapid sweep of an allele incorporated into a population after an environmental shift, research within the last few decades have focused on 'soft sweeps' resulting from neutral or deleterious mutations that are present in the standing genetic variation prior to the change in the selective environment, wherein the selection coefficient changes with the environmental shift such that the allele(s) become evolutionarily advantageous (reviewed in Hermisson & Pennings 2005, Barret & Schluter 2008, Messer & Petrov 2013, and Hermisson & Pennings 2017; see also Jensen 2014). These allele(s) could manifest via a single low-frequency variant, multiple variants caused by parallel recurrent mutation/reorganization on multiple haplotypes, or multiple unique alleles that arise independently within, perhaps multiple, populations. In such cases where selection acts via

soft sweeps, the rate of evolution at the phenotypic level is expected to exceed those of hard sweeps because the alleles under selection have escaped the stochastic nature of drift to a greater degree and are segregating within multiple individuals and genetic backgrounds within the population. The extent to which soft sweeps alter the effect size distributions underlying the genetic architecture is likely dependent upon both the strength of selection and effect size before and after the environmental change (Messer & Petrov 2013; Matuszewski et al. 2015; Jain & Stephan 2017), while the frequency before selection influences the likelihood of subsequent detection (Innan & Kim 2004). Additionally, if multiple mutations are segregating during the sweep, the probability of fixation for any given locus also decreases (Pennings & Hermisson 2006a, 2006b; Chevin & Hospital 2008; Ralph & Coop 2010). Evidence for hard sweeps in tree species exist within the literature, although they are rare (e.g., disease response genes in *Pinus taeda* Ersoz et al. 2010; see also Table 3 in Siol et al. 2010). However, for many species of trees, which often experience high gene flow and strong diversifying selection across populations, adaptive divergence for polygenic traits is expected to result more often from soft sweeps than hard sweeps, affecting phenotypes by subtle allele frequency changes across populations, such that allele frequency differences of individual loci across populations for neutral and selective sites will often be nearly indistinguishable (Latta 1998, 2003; Barton 1999; Le Corre & Kremer 2012; Stephan 2015; Yeaman 2015; Jain & Stephan 2015, 2017). Indeed, the large effective population sizes found in most tree species would permit large effective mutation rates (or reorganization events) necessary for a soft selective sweep from multiple unique variants, particularly when the phenotype is underlain by a large mutational target. Even so, and as highlighted by Stephan (2015) and Bailey & Bataillon (2016), the extent to which scientists can detect the influence of demographic processes on soft versus hard sweeps, and vice versa, remains challenging (Jensen et al. 2005; Chevin & Hospital 2008; Schrider et al. 2015, 2016; Schrider & Kern 2016; Hermisson & Pennings 2017).

While discrete directional selection events are likely to be a common evolutionary influence across taxa, fluctuating or sustained directional selection (i.e., moving optima) are also likely to be contributory factors influencing the genetic architecture of quantitative traits (reviewed in Kopp & Matuszewski 2013; see also McCandlish & Stoltzfus 2014). For a sustained moving optimum, the effect size distribution of beneficial alleles is expected to be dependent upon the effect distribution of standing or de novo mutations as well as the strength of selection: if the rate of change is dramatic, adaptation from new mutations is expected to occur through intermediate to large-effect loci (Kopp & Hermisson 2009a; Matuszewski et al. 2014) or from small-effect loci when adaptation occurs via standing variation (particularly when epistasis is considered, Matuszewski et al. 2015). Under lesser rates of environmental change, adaptation is expected to proceed through mainly alleles of small-effect (Collins et al. 2007; Kopp & Hermisson 2009a, 2009b) where intermediate effects will dominate the long-term distribution of effect sizes (Kopp & Hermisson 2009b). In the case of fluctuating environments, outcomes often depend directly on the degree of temporal autocorrelation of the changing environment. In such cases of stochastic fluctuation around a linear trend of environmental change, extinction risk increases relative to that of the strictly linear trend (Bürger & Lynch 1995) where local adaptation lags, to some degree, behind any given contemporaneous scenario. In comparison, and similar in some ways, stochastic fluctuations around a constant mean are expected to resemble the dramatic environmental change scenario described above, characterized by strong selection pressures, maladaptation between generations, and a large lag load (Bürger 1999; Chevin 2012; Kopp & Matuszewski 2013). In the case of autocorrelated shifts, the 'predictability' of such fluctuations may decrease the possibility of extinction, increase probability of local adaptation, and lead to similar scenarios as discussed for gradual changes in the environment (Kopp & Matuszewski 2013).

Gene flow

Gene flow, to the extent that would be appreciable to that found in trees (reviewed in Savolainen et al. 2007), is also an important component shaping quantitative expectations. Indeed, since the early 1900s we have known that gene flow can disrupt adaptation if selection is not strong enough to overcome the loss of beneficial alleles (Haldane 1930; Wright 1931; Slatkin 1987; reviewed in Felsenstein 1976, Lenormand 2002, Savolainen et al. 2007, 2013, Feder et al. 2012a, and Tigano & Friesen 2016). Particularly when gene flow is asymmetric between core and peripheral populations, adaptation can be inhibited in marginal habitats (Kirkpatrick & Barton 1997; Kawecki 2008). Even so, there is abundant evidence that gene flow can promote adaptation and maintain polymorphisms within populations, including white sand lizards (Laurent et al. 2016), stick insects (Comeault et al. 2014, 2015), cichlid fishes (Meier et al. 2017), Darwin's finches (Lamichhaney et al. 2015), and lodgepole pine (Yeaman & Jarvis 2006).

The magnitude of gene flow between populations can also impact the distribution of effect sizes, for when gene flow falls below a critical threshold, and over many thousands of generations, there is an increase in the probability of establishment and persistence times of large-effect alleles, thus reducing the proportion of the polymorphism due to small-effect loci (Yeaman and Otto 2011; Yeaman and Whitlock 2011). These dynamics are further influenced by the susceptibility of alleles to 'swamping' (Slatkin 1975; Bürger & Akerman 2011; Lenormand 2002; Yeaman 2015; sensu Haldane 1930). For alleles that are prone to swamping, adaptive phenotypic divergence depends on genetic variation and is driven by allelic covariance among populations particularly when the underlying architecture is highly polygenic, the mutation rate is high, and the number of loci underlying the trait exceeds the number needed to achieve the local optimum phenotype (genetic redundancy; Yeaman 2015). Conversely, when there is little genetic redundancy underlying the trait, limited divergence is observed unless the effect size of a given swamping-prone allele exceeds the critical migration threshold. In these cases where swamping-prone alleles contribute to adaptive divergence, the genetic architecture is transient and any given

locus contributes ephemerally to phenotypic divergence, even for loci of relatively large effect (Yeaman 2015). In the case of swamping-resistant alleles, the evolved architecture is enriched for large-effect loci and adaptive divergence can be maintained with little genetic variation or input from mutation. Yet while the contribution from such loci can last many thousands of generations, the architecture can again become transient as the genetic redundancy or mutation rate increases (Yeaman and Whitlock 2011; Yeaman 2015).

Physical linkage and reduction of recombination between adaptive loci can also play a considerable role in adaptive processes in the face of gene flow (Feder & Nosil 2010; Feder et al. 2012a,b; Yeaman 2013; references therein). In such cases, loci that are tightly linked to other loci already under selection will have an increased probability of contributing to local adaptation, both because of physical linkage as well as by reducing the effective recombination among loci within the sequence block. For instance, Yeaman & Whitlock (2011) showed that under divergent selection with gene flow, the number of contributing loci decreases with increasing recombination while small effect loci tend to cluster in groups that act as a single large effect locus (see also Remington 2015), and strong selection can maintain these clusters of linked loci over greater map distances than can weak selection. More recently, Yeaman (2013) employed individual-based simulations to provide evidence that the clustering of alleles throughout a bout of adaptation is unlikely to be driven mainly by divergence hitchhiking alone, and that instead competition between genetic architectures and chromosomal rearrangements occurring throughout adaptive processes under a range of environmental fluctuation scenarios can lead to the evolution of tightly clustered adaptive loci which persist in the event of gene flow, unlike the clusters identified by Yeaman & Whitlock (2011). Yeaman (2013) found that the level of clustering was a function of the temporal fluctuation period, the rate of rearrangement itself is an important determinant on the evolution of clustered architectures, and clusters can in some cases be evolutionarily disadvantageous. Together, these results suggest that genomic rearrangements (reviewed in Ortiz-Barrientos et al. 2016), including inversions (Kirkpatrick & Barton 2006; reviewed in Hoffman

& Rieseberg 2008), which decrease the effective rates of gene flow among adaptive sequences can be an essential component of local adaptation, and indeed some cases of speciation, in the face of gene flow.

Summary

While we provided an overview of the factors that can influence the genetic architecture of local adaptation, we acknowledge that it is far from exhaustive. Because the phenotypes used in studies of local adaptation (particularly those assumed or corroborated to be a component of total lifetime fitness) often have a continuous distribution, and are thus quantitative in nature, the underlying genetic basis for these traits is likely polygenic and is predicted to be underlain by multiple (often many) segregating loci, many of which may confer small phenotypic effects (and are thus unlikely to be detected using single-locus approaches). Even so, a continuum exists, where the true genetic architecture (the number of contributing loci, as well as their relative locations within the genome, phenotypic effects, and interactions) underlying a given complex trait is itself determined by a combination of evolutionary forces that encompass an interplay between the strength, timing, and direction of (background) selection against the homogenizing effects of gene flow and recombination, disruptive effects of drift, linkage, transposition, inversion, and mutation, interactions between underlying loci as well as between these loci and the environment, structural variation, relationship to gene expression networks, as well as other factors related to life history. Consequently, the contemporary genetic architecture is a result of past evolutionary processes, while the adaptive response to future evolutionary dynamics is influenced in part by the contemporary architecture and genetic variance at hand.

The genomics of local adaptation in trees

Common approaches used to identify adaptive loci

Across taxa, and specifically in trees, the predominant association and outlier methods for uncovering sets of loci underlying local adaptation have relied upon single-locus population

genetic approaches. Putatively adaptive loci are often identified by elevated allele frequency differences among populations relative to patterns genome-wide. Yet, as revealed in the previous section, loci underlying polygenic traits will often be indistinguishable from non-causative sites in this way. Further, outlier tests based on F_{ST} (sensu Lewontin & Krakaur 1973) do not incorporate information regarding putative phenotypic targets of selection nor environmental drivers of differentiation, often do not correct for neutral population structure (but see Lotterhos & Whitlock 2015), and will inevitably isolate a biased set of candidate loci (Hermisson 2009; Cruickshank & Hahn 2014). In the case of single-locus genotype-environment associations (reviewed in Rellstab et al. 2015; see also De Mita et al. 2013), information about possible environmental drivers is incorporated by assessing the association between allele frequencies and environmental heterogeneity, yet without information regarding traits hypothesized to be influenced by selection (Schoville et al. 2012). Single-locus genome wide association studies (see next section; Supplemental Box S3) and quantitative trait loci (QTL) experiments (reviewed in Ritland et al. 2011, Hall et al. 2016) have also been used in trees, quantifying the differential effects of typed alleles on a given phenotype. Despite the shortcomings of these methods, such studies provide candidate loci that can be investigated in further detail (Tiffin & Ross-Ibarra 2014), which is particularly advantageous when resources are limited. Indeed, as discussed below, these approaches dominate the methods used to uncover complex traits (adaptive or otherwise) in trees.

Current progress in trees

In light of the expectations outlined above for the architecture of quantitative traits under various evolutionary regimes, and the methods commonly used to detect these loci, we reviewed the literature of single-locus genotype-phenotype associations (GPAs, which included associations to gene expression levels) from studies in forest trees. In doing so, we identified 52 articles across 10 genera and 24 species with a total of 2113 GPAs (Supplemental Table S6.2, Supplemental File F6.2). Because most studies in trees do not report phenotypic effect sizes of individual loci

(i.e., regression coefficients), we report r^2 values which can be used to quantify the percent phenotypic variance explained by the associated locus. In cases where multiple SNPs from a given locus (e.g., a gene or scaffold) were associated to a trait, we averaged the r^2 values for that locus. As with our review of trait heritability and Q_{ST} , we grouped phenotypic traits used in associations into twelve broad categories (in this case, no phenotypes fell into Survival or Seed and Seedling Properties groups). If traits important to tree conservation and industry are often of a polygenic basis, we would expect small to moderate effects from loci empirically associated to phenotype. Indeed, across the trait groups considered here, the mean r^2 was 0.039, where 80.79% ($n = 1707$) of recorded estimates had r^2 values less than 0.05, 18.78% ($n = 397$) of r^2 values falling between [0.05,0.22], and nine values of r^2 greater than 0.22, which were all related to *Cronartium ribicola* resistance in *Pinus monticola* Douglas ex. D. Don (Figure 6.3a).

Of the twelve trait groups, all but those traits relating to both reproduction and herbivore and insect resistance had r^2 estimates greater than 0.10, with traits relating to disease resistance, growth, leaf and needle properties, phenology, and wood properties each contributing over 10% of these outliers (Figure 6.3b). These small effects tend to also not account for much of the observed heritability, but can explain sizeable fractions in some instances (e.g., primary metabolites in Eckert et al. 2012). Of the loci associated with expression levels, r^2 estimates were between 0.05 and 0.152 in all but one case ($n = 54$). We also assessed the propensity of individual loci to be associated to more than one phenotype or expression level across our literature review. Without correcting for the multiple associations of a locus to yearly phenotypes (e.g., bud flush 2009, bud flush 2010), we found that the average number of loci associated to multiple phenotypes per study was 6.00, while after correcting for multiple years the average number decreased to 5.42. The median number of SNPs utilized for association per study was 206, where 75% (39/52) of studies used less than 1,000 SNPs, eight studies using between 1,000-10,000

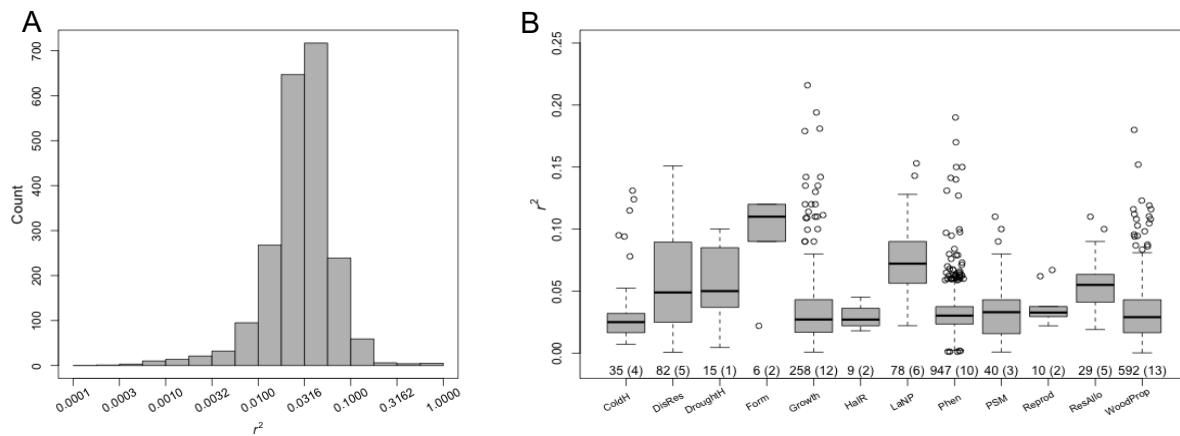


Figure 6.3. Insights from genotype-phenotype literature review. (A) Counts of per-locus percent variance explained (r^2) estimates from single-locus genotype-phenotype associations from literature review. Note logarithmic x-axis. (B) Distribution of per-locus r^2 values for individual trait groups within genotype-phenotype literature review. Values along x-axis are total number of estimates and number of species across estimates. Not shown are nine outliers for disease resistance to *Cronartium ribicola* in *Pinus monticola* (range = [0.402, 1.0]) from Lui et al. 2017. Abbreviations as in Figure 6.1.

SNPs, four studies using between 29,000-35,000 SNPs, and one study utilizing 2,822,609 SNPs for association (all studies with greater than 10,000 SNPs were from either *Pinus* or *Populus* species).

Are we out of the woods yet?

From insight gained from the literature review of genotype-phenotype associations it seems that the vast majority of the genetic architecture of local adaptation and complex traits in trees remains largely unexplained using common GWAS methods (see also Box 1), a consistent pattern across the past decade of research in trees (Neale & Savolainen 2004; Savolainen et al. 2007; Čalić et al. 2015; Hall et al. 2017). Furthermore, it is likely that the estimates for percent variance explained are inflated due to a combination of QTLs that break down into smaller effect loci (Remington 2015), the Beavis effect (Beavis 1994; Xu 2003), and the Winner's Curse (Görning et al. 2001; Zöllner & Pritchard 2007) where locus effects are inflated by using the same data for both gene identification and phenotypic prediction (see Box 1 in Josephs et al. 2017b for a detailed synopsis of these biases). Such a pattern suggests that, indeed, many of the traits important to

evolutionary, breeding, and conservation insight in trees are likely of a polygenic basis and that future studies must take this into account when seeking to identify the underlying loci.

Even within studies of model organisms, missing heritability is nothing new. Across taxa, missing heritability is less frequent within phenotypes of mono- to oligogenic bases (as seen for the Cr2 major-gene resistant locus in *Pinus monticola*, Liu et al. 2017), as would be expected, and is a recurrent, pervasive shortcoming from genotype-phenotype associations of complex traits, particularly those maintaining single-locus perspectives. A number of explanations have been put forth to explain the missing heritability, such as epistasis (Hemani et al. 2013) and its inflationary effect on heritability estimates (Zuk et al. 2012), environmental or epigenetic interactions (Feldman & Lewontin 1975) as well as their inflationary effect on heritability estimates (Zuk et al. 2012), (unmeasured) low-frequency variants of large effect (Dickson et al. 2010), genetic or variance heterogeneity of individual alleles (Leiserson et al. 2013; cf. Box 1 in Nelson et al. 2013), or common variants with effect size below detection thresholds (Yang et al. 2010). As such, here we avoid supporting one causative hypothesis over another, particularly given the ongoing discussion within the literature, for which strengths and weakness for any viewpoint are apparent (e.g., Gibson et al. 2010), and because of the progress yet to be made in trees.

Indeed, the dissection of the genetic architectures underlying complex traits in trees is still in its nascency compared to the progress of model organisms (for which missing heritability is still an issue), and beyond issues of coverage, genomic saturation, and genomic resources (discussed below in The Path Forward), we must approach this issue with all possibilities in mind. Given the unique properties of the life histories, genome size and organization of many tree species, and the limited numbers of studies with large sets of molecular markers, causative sources of the missing heritability should be ruled out, or supported, as with any other hypothesis, particularly as we gain information from contemporary studies of trees that address shortcomings of those in the past. Further, we must keep in mind differences between functional and statistical gene action (Álvarez-Castro et al. 2007; Nelson et al. 2013; Huang & Mackay 2016; Huber et al. 2017). In any

case, it seems that sample sizes of single-locus approaches will need to be increased (Hall et al. 2016), albeit with diminishing returns (Boyle et al. 2017; Simons et al. 2017), to discover a higher proportion of the underlying loci in trees due to small to moderate additive effects. Alongside suggestions outlined in *The Path Forward*, incorporating investigations into such aspects of epistasis, dominance, pleiotropy, expression, GxE effects, and network analyses (when appropriate), may be a worthwhile complement (e.g., Lotterhos et al. 2017, Mähler et al. 2017, Mizrachi et al. 2017; Tan et al. 2017).

While the infinitesimal model will continue to prove to be immensely useful for breeding programs and for short-term evolutionary predictions, and we may find that the missing heritability in trees is truly due to consequences of the infinitesimal regime (as is often cited to be the majority consensus across taxa for missing heritability), it has been argued that the analysis paradigm for such studies is near its limits in describing the functional genetic architecture of quantitative traits, and that it is therefore necessary to move beyond single-locus perspectives and reconsider common practices (Pritchard & Di Rienzo 2010; Nelson et al. 2013; Sork et al. 2013; Tiffin & Ross-Ibarra 2014; Wadgyamar et al. 2017). At this stage, it seems that we investigators seeking to describe the genetic architecture of quantitative traits in trees have some ways yet to go before we are truly out of the woods. In the next section, we describe the path forward to describing genetic architectures from a polygenic and functional perspective, identify resources available to advance our knowledge and fill knowledge gaps, as well as future directions for this research area.

The path forward

As we have outlined, there is still ample room for improvement in our description and understanding of the genetic architecture of quantitative traits in trees (see Table 6.1 and Box 6.1). Importantly, methods used to uncover causative loci should take into consideration the expected degree of polygenicity, the relative contributions of various forms of gene action, as well as how past evolutionary phenomena has likely shaped current adaptive expectations. In this

Box 6.1: A step in the right direction: Synergism between GWAS and Genomic Selection

Early simulations showcased the promise of predicting breeding values from marker data to accelerate domestication and breeding of plants and animals (Meuwissen et al. 2001; Bernardo & Yu 2007; Heffner et al. 2009; Zhong et al. 2009), and particularly under the framework of genomic selection (GS) in trees (Wong & Bernardo 2008; Grattapaglia & Resende 2011; Iwata et al. 2011; defined and reviewed by Grattapaglia 2017). Much of the early exploration into the applicability of GS in trees discounted the utility of marker-assisted selection (MAS) because of the small estimated effects for the few loci significantly associated via single-locus approaches at the time, as well as having concerns related to replication because of the identification of markers across limited parental (genetic) backgrounds (Grattapaglia & Resende 2011; Iwata et al. 2011; Resende et al. 2012a, 2012b). Based on these arguments and results from simulations, genomic selection was identified as a more promising endeavor than MAS, particularly if the breeding cycle can be reduced via efforts such as grafting (Grattapaglia & Resende 2011) or somatic embryogenesis (Resende et al. 2012a).

While GS techniques often can explain a considerable proportion of narrow sense heritability, current implementation of GS in trees is often on par with, or marginally better than, traditional phenotypic selection when evaluating potential within the same generation and environment (see Table 9.1 in Grattapaglia 2017). Further, the predictive accuracy of various models are a function of underlying architecture (Resende et al. 2012c; Grattapaglia 2017). As pointed out by Grattapaglia (2017), current marker densities have produced satisfactory results due to the capture of relatedness between training and validation populations. Here, this success is likely due to the ability of markers to reasonably represent large haplotype blocks (and thus cumulative action of causative effects) due to the high level of relatedness between training and validation populations. Even so, Grattapaglia (2017) recommends higher marker densities so that markers also capture true marker-QTL LD and thus sustain long-term accuracies across generations and environments. We also believe GWAS applications (in the broadest sense) in trees will also see improvements through increased marker densities, the results of which can then be used to further test specific hypotheses regarding underlying architectures and to increase predictive accuracies of GS as well. Incorporating markers that putatively underlie the trait of interest into model prediction may spur opportunities that do not require high degrees of relatedness between training and validation populations, perhaps to the extent of incorporating material from outbred stands using predictive approaches (*sensu* Béréanos et al. 2014; Bontemps 2016) and heritability validation (*sensu* Castellanos et al. 2015) in the field.

In the end, the realized progress of our understanding regarding the genomics of complex traits in trees will therefore be enhanced by the deposition of data from both GS and GWAS (as well as other 'omics') approaches into a centralized open-access database hub such as TreeGenes (treegenesdb.org). Future meta-analyses can then synthesize past inquiry to summarize our current understanding of underlying genetic architectures, ultimately incorporating this knowledge towards future applications in industry and conservation (see The Path Forward; Table 6.1).

Table 6.1. Where to next? The Path Forward identifies meaningful ways in which we can progress our understanding of the architecture underlying complex traits in trees. Here we outline some questions that can be used to guide future inquiry as the number of markers and sequence length increase, and annotation becomes more precise and specific to tree biology.

- 1) Composition and evolution of genetic architectures in trees
 - a. How prevalent are non-additive contributions to underlying genetic architectures in trees? Are there patterns across similar phenotypes or regulatory networks? Is there evidence that such non-additive effects have either constrained or facilitated local adaptation?
 - b. Are adaptive loci most prevalent in areas of low recombination or repetitive sequences (e.g., retrotransposons, clustered gene families)? Do loci of similar effect sizes, expression profiles, or pleiotropic effect (Lotterhos et al. 2017) experience elevated LD within the genome? Should genome size influence our expectations for underlying architectures (Mei et al. 2017)?
 - c. At what frequency does local adaptation result in fitness tradeoffs across environments (Tiffin & Ross-Ibarra 2014; Wadgymar et al. 2017)? And does this interact with demographic history in trees?
 - d. Does pleiotropy play a predictable role in underlying tree genetic architectures (Lotterhos et al. 2017)?
 - e. Which aspects of genetic architectures in trees are likely to exhibit deleterious variation? And how much of this signal are we capturing in genotype-phenotype applications?
- 2) Inter- and intraspecific variation of genetic architectures in trees
 - a. Which aspects of the genetic architecture should we expect to vary across populations or environments?
 - b. Under what conditions in trees are we likely to observe genomic reorganization across species or ecotypes (e.g., physical linkage or dispersion)? Will reference genomes be suitable to assess this question across species or diverged populations, or can long-read sequencing technologies (reviewed in Jiao & Schneeberger 2017) offer appropriate resources?
 - c. What is the degree of convergent and parallel adaptation within polygenic architectures across tree populations and species?
 - d. At what level of the genetic architecture do we see patterns of convergence, parallelism, and divergence? Within core hubs, or perhaps within aspects of the periphery? What does the comparison of the topologies from such architectures tell us about influential evolutionary processes?
 - e. How often are architectures influenced by variation in expression levels rather than structural variation in proteins? Do architectures differ in predictable ways with the prevalence of one or the other? How can we utilize knowledge synthesized across past approaches to spur understanding of underlying genetic architectures in trees (Mizrachi et al. 2017, Lotterhos et al. 2017)?

section, we orient our path forward by first highlighting utilities available to, and underused within, the forest genetics community to describe the genetic architecture of complex traits. We then outline several suggestions to facilitate further progress and advocate for prospective perspectives in future studies such that information and data may continue to be used easily in subsequent syntheses across pathways, environments, species, and towards insight to identify future needed resources as our understanding progresses. While our recommendations are specific to the tree community, we also acknowledge other valuable recommendations from recent reviews (e.g., Savolainen et al. 2013; Tiffin & Ross-Ibarra 2014; Lotterhos & Whitlock 2015; Gagnaire & Gaggiotti 2016; Hoban et al. 2016; Wellenreuther et al. 2016; Burghardt et al. 2017; Wadgymer et al. 2017).

Stepping off the pack – what’s in our pack?

The genetic architecture underlying local adaptation and complex traits likely has a polygenic basis composed of many loci of relatively weak effect yet many of the common association or outlier methods will often fail to detect many of the causative loci of small to moderate influence. Such investigations have so far led to an incomplete description of studied architectures, and, in many cases, have limited our understanding of complex traits in trees to a handful of loci. While we do not advocate that such single-locus methods be avoided in future studies (considered further in the next section), here we outline underused and promising approaches to identify and describe underlying loci that explicitly take into account the polygenic basis of such traits and may help advance our understanding in future studies, including some of the questions we have outlined in Table 6.1. Multivariate, multiple regression, and machine learning techniques are three such examples, and differ from univariate analyses by analyzing patterns among multiple loci simultaneously.

The Bayesian sparse linear mixed model (BSLMM), for instance, such as that deployed in the software package GEMMA (Zhou et al. 2013), is developed for both genomic prediction (see

also Box 1) and mapping of complex traits that offers considerable advantages over single-locus genotype-phenotype approaches (Guan & Stephens 2011; Ehret et al. 2012; Zhou et al. 2013; Moser et al. 2015). This analysis has gained in popularity recently, being used across diverse taxa such as stick insects (Comeault et al. 2015, Riesch et al. 2017), butterflies (Gompert et al. 2015), Darwin's finches (Chaves et al. 2016), and trees (Lind et al. 2017). BSLMM is a hybrid of LMM and Bayesian variable regression that extends the Lande & Arnold (1983) multiple regression approach in an attempt to address the sparsity of common data sets used in genotype associations, where the number of model parameters (loci) is often much greater than the number of observations (sampled individuals; Zhou et al. 2013; Gompert et al. 2016). Specifically, the model takes into account relatedness among individuals and provides a means to summarize estimates of selection across the genome such as the proportion of phenotypic variation explained (PVE) across genotyped markers by estimating the combined influence of markers with either polygenic (infinitesimal) or measurable (moderate to large) effect, the proportion of PVE explained by genetic loci with measurable effects (PGE), and the number of loci with measurable effects that underlie the trait (for more details see Guan & Stephens 2011; Zhou et al. 2014; Gompert et al. 2016). Additionally, GEMMA returns the posterior inclusion probability for each marker providing evidence for association with the phenotype. While the approach remains promising considering its performance in the context of genomic prediction and inference of PVE (e.g., Zhou et al. 2013, Speed & Balding 2014), there has been no attempts, to our knowledge, to assess the approach under various demographic histories, genetic architectures, and sampling designs. A close approximation to this comes from analyses carried out by Gompert et al. (2016), in which GEMMA was evaluated for PVE estimation, estimated effects of causative loci, and the estimated number of underlying SNPs based on various author-specified numbers of causal loci, underlying heritability ranges, and numbers of sampled individuals. In short, the authors convey that GEMMA is promising, but that there are important limitations to consider (Gompert et al. 2016). However, because the authors simulated architectures by randomly assigning effects to

loci from an empirically-derived sequence data set, and while they were thorough in their data exploration, we encourage these results be replicated in silico through full modeling of genomic loci across various demographic, LD, sampling, and architecture scenarios to ensure underlying allele frequencies among populations and LD (within and among populations) reflect realistic patterns which may have an effect on model performance. Such additional analyses will also allow for more specific insight into model performance based on a priori biological insight available to investigators, allowing more informed decisions when choosing an appropriate genotype-phenotype association method such as BSLMM.

Random Forests (Breiman et al. 2001) is a machine learning algorithm used to identify patterns in highly dimensional data sets to further generate predictive models. Alongside uses outside of evolutionary biology, the Random Forests algorithm has gained popularity in association studies across taxa as well as in trees such as that of genotype-phenotype associations in Sitka spruce (*Picea sitchensis*; Holliday et al. 2012) and genotype-environment associations in white spruce (*P. glauca*; Hornoy et al. 2015). Random Forests is based upon classification (for discrete variables, e.g., soil type) and regression (continuous variables; e.g., temperature or phenotypic measurements) trees (so-called CART models). During its implementation, Random Forests creates these decision trees using two layers of stochasticity: the first layer is used to grow each tree by using a bootstrap sample of observations (environmental or phenotypic) while the second uses a random subset of predictors (marker loci) to create a node which is then split based on the best split of the observations across permutations of predictors using the residual mean square error (see Figure 2 in Hornoy et al. 2015). The observations that were not used as training data to create the model are then used to estimate model accuracy, which can be further used to assess variable importance (Holliday et al. 2012; Hornoy et al. 2015; Forester et al. 2017).

While creating a promising alternative to univariate approaches, until recently the Random Forests algorithm has not been fully explored to assess model performance for use in association

studies. Forester et al. (2017) provide a thorough analytical assessment using simulated data to remark on performance for use in genotype-environment association studies (GEA). In their analysis, they used published simulations of multilocus selection (Lotterhos & Whitlock 2014, 2015) of various demographic histories and selection intensities across 100 causative (with 9900 neutral) loci to compare the Random Forests algorithm to the multivariate approaches of constrained ordination (redundancy analysis, RDA, and distance-based RDA, dbRDA - both of which are mechanistically described in Legendre & Legendre 2012, but are multivariate analogs of multiple regression on raw or distance-based data) and to the univariate latent factor mixed model (LFMM). In short, Forester et al. (2017) found that LFMM performed better than Random Forests as a GEA, while constrained ordinations resulted in relatively lower false positive and higher true positive rates across levels of selection than both Random Forests and LFMM. Additionally, the authors found that correction for population structure had little influence on true and false positive rates of ordination methods, but considerably reduced true positive rates of Random Forests. They also note that further testing is needed across various evolutionary scenarios. Even so, constrained ordination provides an effective means by which to detect loci under a range of both strong and weak selection (Forester et al. 2017). While promising under a GEA framework, future analyses may provide evidence that such methods also perform well in genotype-phenotype associations as well. Empirically, it has been used in trees to explore multivariate relationships between phenotypes, genotypes, and environments (e.g., Sork et al. 2016). Additionally, there have been many extensions of the original Random Forests model, such that extensions with purportedly better performance should be assessed alongside other popular association methods in the future.

Once a set of candidate loci have been identified to putatively underlie a phenotype or environment of interest, these loci can be used to further test the hypothesis of polygenic local adaptation. For instance, Berg & Coop (2014) use the significant hits from GWAS data sets to estimate within-population additive genetic values by calculating the frequency-weighted sum of

effects across these loci. These values are then compared to a null model of genetic drift that accounts for population structure to test for an excess of variance among populations, ultimately identifying the populations most strongly contributing to this signal. The excess variance statistic (Q_x) is analogous to Q_{ST} and is composed of two quantities – an F_{ST} -like component describing allele frequency differentiation across populations and a LD-like component describing coordinated and subtle allele frequency shifts across populations. This method thus allows explicit hypothesis tests related to the expected polygenic architecture of local adaptation across populations of trees. It is also noteworthy in that it combines aspects of the genotype-environment-phenotypic spectrum that underlies local adaptation within a single methodological framework (cf. Sork et al. 2013). Prior attempts take a pairwise approach examining each pairwise combination of the genotype-environment-phenotype spectrum (e.g., Eckert et al. 2015). Despite the promising insight from this method, it has not been used widely outside of model organisms. Future applications in trees should consider the number of causal loci identified to be associated with quantitative phenotypes (driven somewhat by the number of loci used in mapping studies), the number of populations needed to increase power, especially in the correlation of genetic values to environmental data, and the ability to reliably estimate genotypic effects.

At the trail junction – where to next?

While we have outlined methods above that have not yet realized their full potential in describing genetic architecture of complex traits in trees, there are several matters that we, as a field, must keep in mind such that we can continue to progress our understanding in the most efficient manner. Here we believe the path forward lies in three critical areas which we discuss in further detail below: 1) needed data, 2) standardized data reporting, and 3) empirical studies in trees designed to test theoretical expectations of genetic architectures.

Needed data

While the common garden approach can facilitate understanding of evolutionary processes without specifically identifying underlying loci (Rausher & Delph 2015), identifying features of the genetic architecture will ultimately inform breeding applications important to management, conservation, and industry, and thus requires knowledge about underlying loci. Consequently, we have not yet had sufficient sampling of both marker densities and studies amenable to replication across systems to truly exhaust the use of single-locus approaches, particularly as the sample size of markers, individuals, and populations increase in the near future. Indeed, Hall et al. (2016) estimated that the number of causative loci underlying quantitative traits in trees is likely in the several hundreds, and to capture 50% of the heritable genetic variation using single-locus approaches, population sizes of about 200 will be needed for mapping disease traits, and about 25,000 for traits such as growth. Even so, we recommend that such single-locus associations should not be used as the sole method of architecture description as we carry out future studies unless justified a priori based on biological principles, knowledge of the expected architecture, and/or for testing specific hypotheses. While the limits of such methods should be considered, these approaches can be used alongside other lines of evidence to either support or spur further testing of underlying loci (*sensu* Sork et al. 2013). For instance, there is little downside to performing both a single-locus association and a multivariate analysis in the same study, even if some or all of the results for a given technique are excluded to the supplement (e.g., Sork et al. 2016). Further, contextualizing genotype-phenotype and genotype-environment relationships with results that describe local adaptation (e.g., phenotype-environment, Q_{ST} - F_{ST} comparisons) can also stimulate further understanding particularly for data that is made publicly available for future synthesis. Specifically, studies which do so within the context of comparisons within and across species (e.g., Yeaman et al. 2016) or environments (Holliday et al. 2016), offer unique circumstances under which to advance our understanding of complex traits in trees (Table 6.1;

Lotterhos & Whitlock 2015; Čalić et al. 2016; Hoban et al. 2016; Ingvarsson et al. 2016; Mahler et al. 2017).

Isozymes (Adams & Joly 1980), restriction fragment length polymorphisms (Devey et al. 1994), randomly amplified DNA (Grattapaglia & Sederoff 1994), and expressed sequence tag polymorphisms (Temesgen et al. 2001) were among the first used to test evolutionary hypotheses in trees related to genome organization and the mapping of complex traits (discussed in Eckert et al. 2009). Marker technology has progressed considerably since this time (dozens of markers) to include markers capable of more densely sampling tree genomes (up to millions of markers). For example, array-based designs (Silva-Junior et al. 2015) and exome capture (Suren et al. 2016) allow for hundreds to tens of thousands of both genic and intergenic markers (which can be dwarfed by the number of subsequently called SNPs) whereas RADseq (reviewed in Parchman et al. in review) is in the range of tens- to hundreds of thousands of markers (e.g., Parchman et al. 2012) and whole genome sequencing in the range of millions (e.g., Stölting et al. 2015). However, while the continual advent of sequencing technology will likely allow for more SNPs and longer sequences, it is ultimately the concordance between polygenic expectations and analytical methods of marker data that will determine the success of such endeavors. With this in mind, future studies aimed at answering outstanding questions (Table 6.1) will benefit from a diverse set of markers that represent both functional proteins (genic regions) as well as those which control aspects of their expression or post-transcriptional regulation. If one lesson is to be gained from the recent discussion of the applicability of reduced representation techniques (Lowry et al. 2016, 2017; Catchen et al. 2017; McKinney et al. 2017), it is that genomic resources are paramount to advancement of knowledge, especially when developed with knowledge of patterns of linkage disequilibrium or, if not with this knowledge, with goal of quantifying it. However, RADseq remains one of the most cost-effective approaches available to trees and should thus be assessed in the specific context of tree species, particularly when strengths and limitations are understood and addressed (as reviewed in Parchman et al. forthcoming). No matter the approach

used for association, some aspect of the architecture is likely to be missed in trees. For example, RADseq-based markers developed within large genomes are not enriched within genic regions where structural changes to proteins are expected to affect phenotypes, although choice of enzymes can affect the relative proportion of genic regions in tree genomes, as evidenced from *in silico* digestions of reference genomes from *Populus*, *Eucalyptus*, *Amborella*, *Pseudotsuga*, and *Pinus* species (Parchman et al. forthcoming). In contrast, exome based approaches are anchored within coding regions thus excluding putative regulatory elements outside of the exomic regions used to develop probes. Recent marker development approaches, such as RAPTURE (Ali et al. 2016), however, have blurred the lines between RADseq and exome based approaches and, in addition to targeted capture approaches, may offer a promising, cost-effective path forward that explicitly avoids biased assumptions about the importance of exomic versus intergenic loci comprising the architecture of local adaptation.

Beyond dense genetic linkage maps (e.g., Friedline et al. 2015) and reference genomes, which undoubtedly should be among our top priorities, other techniques outside of traditional genomics, such as transcriptomics, have the potential to complement genomic studies in many ways without great need for existing species-specific resources (reviewed in Romero et al. 2012, Strickler et al. 2012; Vialette-Guiraud et al. 2016). For instance, comparative transcriptomic techniques in trees can be used to identify putatively orthologous sets of markers (e.g., Wachowiak et al. 2015; Yeaman et al. 2016) that can be used to describe the evolution of architecture (e.g., shared orthologs versus paralogs across species) or for comparative linkage mapping (Ritland et al. 2011) across systems. Additionally, with the appropriate study design, transcriptomics can be implemented in tree species to describe various aspects of differential expression (Cohen et al. 2010; Carrasco et al. 2017; Cronn et al. 2017), selective constraint (Mähler et al. 2017), prevailing selective forces (Hodgins et al. 2016), mapping of disease resistance (Liu et al. 2016; Liu et al. 2017), and regulatory networks (Zinkgraf et al. 2017). The multilocus paradigm of transcriptomics is amenable to identifying and testing hypotheses of the

genetic architecture of complex traits in a network framework (Jansen et al. 2009; Leiserson et al. 2013; Civelek & Lusi 2014; Feltus 2014) and will no doubt provide valuable contributions for tree evolutionary biologists. Other areas amenable to network description such as metabolomics and proteomics would also be a complement (Feltus 2014; Cowen et al. 2017), particularly if genetic studies contextualize results with findings from such approaches and vice versa. Ultimately the goal is to use a priori knowledge synthesized across past studies, techniques, and perspectives to guide further hypotheses about underlying architecture, as exemplified by Mizrachi et al. (2017) and Lotterhos et al. (2017). Finally, high-throughput phenotyping as well as environmental measures at fine spatial scales below square-kilometers will also facilitate and advance our understanding of complex traits in trees (Sork et al. 2013; Rellstab et al. 2015; Leempoel et al. 2017), particularly when measured phenotypes well represent those experiencing selection pressure, and environmental measures well represent the multivariate environment imposing selection (Lotterhos et al. 2017).

Standardized data reporting

As we continue to accrue genotype-phenotype, genotype-environment, and phenotype-environment relationships within and across tree species, authors should consider how their results can most effectively be used in further studies and syntheses, both for the purpose of validation or comparison as well as novel insights yet to be seen. Here we outline a few suggestions that can be broken down into reporting within manuscripts and metadata. For instance, in our survey of common garden studies used to estimate h^2 and Q_{ST} , in many cases the exact design of the study could not be replicated with the information from the manuscript alone. While an abbreviated design may be suitable for the main text, authors can provide much more detail in supplemental materials that can facilitate replication and comparison across studies (e.g., total individuals per garden, family, or block – as opposed to averages or ranges), which will ultimately facilitate syntheses regarding future directions. Further, future studies would benefit

from estimating relatedness using marker data which will ultimately improve the precision of h^2 , Q_{ST} , and missing heritability estimates (de Villemereuil et al. 2016) including those estimates made in the field (Castellanos et al. 2015). For cases in which estimating relatedness from markers is not appropriate or feasible, the field would benefit by authors exploring a range of underlying sibships (e.g., Eckert et al. 2015), which are often assumed to be half-sib relationships. While some studies in our survey assumed a mixed sibship relationship for open-pollinated sources, ultimately such assumptions without data exploration will affect the outcome or conclusions for any given study. A recently released R package by Gilbert and Whitlock (2014) allows for such an exploration of effects of mixed sibships on inference of Q_{ST} and its magnitude relative to F_{ST} . Inclusion of such exploration, even in the supplement, will help contextualize such studies as they are published. For studies estimating causality for genotype to phenotype, it would be worthwhile to include the regression coefficients or other estimates of effect size (e.g., odds ratios) in addition to PVE (r^2). Importantly, the units of the effect size must be explicitly reported (e.g., Julian days versus phenotypic standard deviations), with the standard deviation also reported. For all association studies, supplemental tab- or comma-delimited text files (outside of a word processing document) easily analyzed with programming languages would also facilitate synthesis (even if providing redundant information from the main text), particularly if such files are well described with a README file and contained data regarding marker position, putative orthogroups, hits to reference genomes, effect size, PVE, genotypes by individual identifiers, individual population assignments, and if the sequence or marker was significantly associated to phenotype or environment. Such an operating procedure may work well in the short term, however in the long term such information will need to be easily accessible from one or a central hub of repositories.

Data standardization, the inclusion of meta-information, and compilation of these data specific to trees into a database with common terminology will be crucial to future inquiries with the purpose of synthesizing evidence for underlying architectures across species and

environmental systems (e.g., as for human GWAS data: <https://www.ebi.ac.uk/gwas/>). If the data generated by tree biologists is disparate and housed across databases and journal supplements this impedes synthesis first by forcing scientists to collate information across sources, which may be further impeded by data redundancies or inconsistencies in data format and utilized nomenclature (Wegrzyn et al. 2012). While many journals have required submission of sequence data to repositories such as NCBI, such databases are lacking with regard to information pertaining to phenotypic, environmental, and geographic information upon which much of the foundation of our field is built. Submissions to Dryad somewhat overcome this, but there is no standardization within the community for content for such submissions and important information may be lacking. Currently, this information is often appended in supplemental files that cannot be readily accessed, compared, or queried in an efficient manner. Hierarchical ontologies can be used to ease this burden. Gene Ontology is likely the most recognizable to evolutionary biologists, but there also exist Plant Ontologies for organismal structure and developmental stages, Environmental Ontologies for habitat categorization, and Phenotypic, Attribute, and Trait Ontologies for the annotation of phenotypes. Such ontologies not only standardize nomenclature, but also assist in database queries. The utilization of such databases will no doubt encourage comparative studies and syntheses, as infrastructure and data accessibility are essential to the comparative approach (Neale et al. 2013; Ingvarsson et al. 2016; Plomion et al. 2016). Luckily, such a database exists for the broader tree genetics community. The open-source genomics and phenomics database, called TreeGenes (treegenesdb.org), is part of a central hub of repositories, including the Hardwood Genomics Project (hardwoodgenomics.org) and the Genome Database for Roseaceae (rosaceae.org), that communicate and integrate data from each other. Unlike many other repositories for tree genomic data, TreeGenes is not project or institution specific. The data and metadata for roughly 1700 species housed on TreeGenes can be accessed, queried, and visualized through DiversiTree, a web-based, desktop-style interface (Wegrzyn et al. 2008). DiversiTree connects to the geographical interface CartograTree (Vasquez-Gross et al. 2013) to

encourage comparative synthesis by providing technology to filter and visualize geo-referenced biotic and abiotic data housed on TreeGenes. As promising as such database hubs are, they are only as useful as the data that is deposited to them. While TreeGenes will regularly import and enhance data from public repositories (through e.g., sequence alignment to published genomes, or data from Genbank, Phytozome, PLAZA, etc), often pertinent metadata necessary for comparative synthesis is lacking (Wegrzn et al. 2008, 2012). Indeed, from our survey of published GPA since the release of the database in 2008, less than 13% (6/48) of the studies submitted their data directly to TreeGenes. To better prepare for future synthesis, we advocate that authors submit their data to the TreeGenes database and that reviewers and editors enforce this habit, as currently implemented for linkage maps published in *Tree Genetics & Genomes*. Consolidated, open-source resources will be crucial to the advancement of this field (Neale et al. 2013), and will no doubt spur knowledge that would not have been recognized otherwise. Prime examples of advancement to knowledge because of these types of resources and community-wide efforts come from the human GWAS literature where such resources provide crucial information necessary to study polygenic adaptation (e.g., Berg & Coop 2014).

Empirical tests of theory

In combination with the development of truly genome-wide public resources, there is need to use these resources to validate and better characterize foundational ideas and assumptions in the theory of polygenic adaptation relative to the life history strategies of tree species. For example, Gagnaire & Gaggiotti (2016) highlight that the degree of polygenicity can be tested as a function of the number of GWAS hits relative to the length of contigs or chromosomes containing these markers. Simple models of polygenicity predict that there should be a positive correlation between these quantities. Thus, rather than assuming some functional form of a polygenic architecture (i.e., an approximate infinitesimal model) during analysis, researchers can strive to characterize, or at least exclude some forms of, the underlying genetic architecture prior to interpretation. In a

related fashion, publicly available data sets would spur comparisons across species and study systems to test hypotheses about polygenic architectures (e.g., the modularity of genetic architectures as in Lotterhos et al. 2017, or perhaps genomic organization or effect size distribution) due to the relative timing of selection, degree of environmental contrast (e.g., diversifying selection and changes to the strength of negative selection), selection strength, and level of gene flow across diverging lineages. As an example, much of the theory of polygenic adaptation requires assumptions about simplistic demographics (where violations have consequences for standing levels of non-neutral diversity, e.g., Wang et al. 2017) and the equilibration among co-acting evolutionary forces over a large number of generations (Brandvain & Wright 2016). Indeed, differing architectures are expected as a function of the timing for the onset of selection (Le Corre & Kremer 2003; Kremer & Le Corre 2012), with subtle allele frequency shifts across populations dominating architectures near the onset of selection and larger allele frequency shifts much later in time. While there is need for empirical validation of this theory, there is also a need to characterize the prevalence of its predicted patterns across differing clades of tree species. In other words, researchers could imagine testing the theory itself in natural populations (e.g., as begun by Le Corre & Kremer 2012) or assuming its validity and characterizing the circumstances under which to expect large shifts in allele frequencies across tree species with differing life history strategies. Little of any of this (Table 6.1), however, will be possible without development of needed data and its deposition into publicly available, standardized databases.

Concluding remarks

The path forward provides a means by which we can most efficiently describe the underlying genetic architectures of traits important to management, conservation, and industry, which can ultimately be used to expedite breeding projects (Box 1). The past evolutionary history will have a profound effect on the underlying genetic architecture of such traits, and thus strengths and weakness of the data and methods used to uncover such architecture should be specifically

addressed in the future, particularly in how utilized methods perform across various demographic and architecture scenarios. Insights gained from empirically testing theory will also contribute to the advancement of this field and will ultimately quantify the variation in architecture across environments and species and inform effective management. Importantly, the success of future genotype-phenotype efforts should not be predicated on past studies using single-locus approaches and small numbers of markers, and instead on overcoming such shortcomings by applying theoretical expectations to empirical inquiry. Even so, until sequencing technologies allow for cost-effective whole genome sequencing of individual trees, most genotype-phenotype studies (GS included) will be carried out via reduced representation techniques (i.e., a subset of all sites within the genome). Therefore, it is essential that processed data be uploaded to a repository that, in addition to raw sequences, includes genotypic, environmental, and spatial data, facilitates user-friendly queries, and allows for future meta-analysis. The future is bright, but we are not yet out of the woods. As such, efficient advancement in this field relies on community efforts, standardized reporting, centralized open-access databases, and continual input and review within the community's research.

Acknowledgements

The authors would like to thank S. González-Martínez for inviting this review, Chris Friedline for stimulating conversations when developing content, and Justin Bagley, Jill Wegrzyn, and two anonymous reviewers for providing helpful comments to earlier versions of this manuscript. Brandon Lind is supported through a Dissertation Fellowship provided by the Graduate School of Virginia Commonwealth University. Andrew Eckert is supported through the National Science Foundation (EF-1442486) and the United States Department of Agriculture (USDA 2016-67013-24469).

Chapter 7. Conclusions

Throughout my dissertation I have utilized natural populations of *Pinus* species to better understand the natural and anthropogenic drivers of tree evolutionary dynamics. Throughout these systems, I have provided evidence that the environment is driving dynamics within and between populations of trees. Such dynamics are an important consideration when making decisions regarding conservation and industrial endeavors. Below I outline the main conclusions from each of my chapters.

Chapter 2. Effects of forest management on fine-scale gene flow in *Pinus lambertiana*

1. Univariate statistics of spatial aggregation suggest that *P. lambertiana* adults often cluster with other *P. lambertiana* adults, that shade-tolerant *Abies* spp. often cluster with other *Abies* spp., and *P. lambertiana* seedlings cluster with other *P. lambertiana* seedlings
2. Bivariate statistics of spatial aggregation suggest that *P. lambertiana* adults and seedlings are both spatially inhibited by the presence of shade-tolerant *Abies* spp.
3. Genetic diversity statistics reveal that alleles are structured across plots, but not with respect to any pattern of increasing disturbance intensity
4. Analysis of spatial genetic structure reveals that while *P. lambertiana* seedlings are clustered spatially, alike genotypes of seedlings are only clustered in UC treatments; adult *P. lambertiana* individuals are clustered only in UN and UC plots, but alike genotypes of adults and seedlings are only clustered in UC treatments
5. Effective dispersal among known samples is influenced by treatment, where weighted dispersal distances from fractional parentage suggests that as disturbance intensity increases, seed (pollen) dispersal decreased (increases)

6. Estimates of the seed dispersal kernel reveal that the vast majority of pollen and seed dispersal events occur below 150m
7. The genetic consequences of forest management result from biological and ecological interaction from standing structure and treatment, where ongoing dynamics are likely to be influenced by treatment and further ecological dynamics. Thus, while genetic structure is elevated in UC treatments, this may change over time.

Chapter 3. Water availability drives signatures of local adaptation across fine spatial scales

1. Using the common garden, we provided evidence that measured fitness-related phenotypes are heritable, that population explains a significant proportion of phenotypic variation, and Q_{ST} was significantly greater than F_{ST} for bud flush and $\delta^{13}C$
2. Using both single- and multilocus approaches to identify focal loci of interest (i.e., those associated to environment and/or phenotype) our analyses suggest that selective pressures of *P. albicaulis* are likely driven by water availability
3. Despite multiple lines of evidence suggesting fine-scale adaptation, we found little evidence for classical genetic signatures of large allele frequency differences for loci putatively underlying adaptation.
4. Our analyses suggest that adaptation within the Lake Tahoe Basin is facilitated by subtle allele frequency shifts across populations, which is suggestive of soft sweeps within a polygenic architecture
5. Our results replicate evidence for soft sweeps in adaptive traits in trees which are enabled through linkage disequilibrium among populations
6. Conservation and replanting efforts of the Lake Tahoe Basin Management Unit should take into consideration local adaptations at fine spatial scales. Seed should be sourced

not by zone, but instead from environments highly correlated with the (future) environment in which they'll be planted.

Chapter 4. A first step towards describing genetic architectures of complex traits in natural populations

1. Precipitation-related variables have structured the geographical range of foxtail pine
2. Climate-based niches differed between regional populations
3. $\delta^{13}\text{C}$ and $N_{\mu\text{g}}$ were heritable with moderate signals of differentiation between regional populations
4. A set of QTLs associated with these phenotypes explained a large proportion of the heritability of these traits, but showed no evidence of pleiotropy with this data set

Chapter 5. The QTL landscape of water-use efficiency for foxtail pine

1. Precipitation-related variables structured the geographical range of foxtail pine where climate-based niches differed between regional populations
2. $\delta^{13}\text{C}$ and $\delta^{15}\text{N}$ were heritable with moderate signals of differentiation between regional populations
3. A set of large-effect QTLs ($n = 11$ for $\delta^{13}\text{C}$; $n = 10$ for $\delta^{15}\text{N}$) underlying $\delta^{13}\text{C}$ and $\delta^{15}\text{N}$ variation, with little to no evidence of pleiotropy, was discovered using multiple-marker, half-sibling regression models

Chapter 6. The genomics of local adaptation in trees: Are we out of the woods yet?

1. There is substantial evidence to suggest that trees across both angiosperms and gymnosperms are locally adapted to many different environments.

2. For the circumstances most likely experienced by many tree species, fitness-related traits are most likely to experience soft sweeps than hard sweeps where genetic signatures of adaptive loci will likely be indistinguishable from other non-causative sites in the genome.
3. The vast majority of genotype-phenotype associations in trees use single-locus approaches that look for signals that are likely infrequent occurrences in natural populations.
4. To most efficiently advance knowledge in this field authors should explore various association methods (particularly those that take into account the polygenic nature of the traits), document how well utilized methods perform under the expected demographic history and genetic architectures of the studied species, and ensure that their results are deposited to open source databases that can be easily synthesized in future studies.

References

- Acheré V, Faivre-Rampant P, Jeandroz S, Besnard G, Markussen T, Aragonés A, Fladung M, Ritter E, Favre JM (2004) A full saturated linkage map of *Picea abies* including AFLP, SSR, ESTP, 5S rDNA and morphological markers. *Theoretical and Applied Genetics* 108:1602–1613.
- Adamack AT, Gruber B (2014) PopGenReport: simplifying basic population genetic analyses in R. *Methods in Ecology & Evolution* 5:384–387.
- Adams WT, Joly RJ (1980) Linkage relationships among twelve allozyme loci in loblolly pine. *Journal of Heredity* 71:199–202.
- Adams WT, Zuo J, Shimizu JY, Tappeiner JC (1998) Impact of alternative regeneration methods on genetic diversity in coastal Douglas-fir. *Forest Science* 44:390–396.
- Agee JK, Skinner CN (2005) Basic principles of forest fuel reduction treatments. *Forest Ecology and Management* 211:83–96.
- Ahuja MR, Neale DB (2005) Evolution of genome size in conifers. *Silvae Genetica* 54:126–137.
- Aitken SN, Yeaman S, Holliday JA, Wang T, Curtis-McLane S (2008) Adaptation, migration, or extirpation: climate change outcomes for tree populations. *Evolutionary Applications* 1: 95–111.
- Akerman A, Bürger R (2014) The consequences of gene flow for local adaptation and differentiation: a two-locus two-deme model. *Journal of Mathematical Biology* 68:1135–1198.
- Alberto FJ, Aitken SN, Alía R (2013) Potential for evolutionary responses to climate change—evidence from tree populations. *Global Change Biology* 19:1645–1661.
- Ali OA, O'Rourke SM, Amish SJ, Meek MH, Luikart G, Jeffres C, Miller MR (2016) RAD capture (Rapture): flexible and efficient sequence-based genotyping. *Genetics* 202:389–400.
- Allen CD, Macalady AK, Chenchouni H, Bachelet D, McDowell N, Vennetier M, Kitzberger T, Rigling A, Breshears DD, Hogg EHT, Gonzalez P, Fensham R, Zhang Z, Castro J, Demidova N, Lim J-H, Allard G, Running SW, Semerci A, Cobb N (2010) A global overview of drought and heat-induced tree mortality reveals emerging climate change risks for forests. *Forest Ecology and Management* 259:660–684.
- Allen J, Scott D, Illingworth M, Dobrzelecki B, Virdee D, Thorn S, Knott S (2012) CloudQTL: Evolving a Bioinformatics Application to the Cloud. *Digital Research 2012*, September 10–12, 2012. Oxford, UK.
- Altschul S, Gish W, Miller W, Myers E, Lipman D (1990) Basic local alignment search tool. *Journal of Molecular Biology* 215:403–410.

- Álvarez-Castro JM, Carlborg O (2007) A unified model for functional and statistical epistasis and its application in quantitative trait loci analysis. *Genetics* 176:1151–1167.
- Anderson JT, Lee C-R, Rushworth CA, Colautti RI, Mitchell-Olds T (2012) Genetic trade-offs and conditional neutrality contribute to local adaptation. *Molecular Ecology* 22:699–708.
- Anderson JT, Wagner MR, Rushworth CA, Prasad KVSK, Mitchell-Olds T (2014) The evolution of quantitative traits in complex environments. *Heredity* 195:1353–1372.
- Antonovics J, Bradshaw AD (1970) Evolution in closely adjacent plant populations VIII. Clinal patterns at a mine boundary. *Heredity* 25:349–362.
- Arnold SJ (1992) Constraints on phenotypic evolution. *The American Naturalist* 140:S85–S107..
- Ashander J, Chevin L-M, Baskett ML (2016) Predicting evolutionary rescue via evolving plasticity in stochastic environments. *Proceedings of the Royal Society B: Biological Sciences* 283:20161690.
- Ávila V, Pérez-Figueroa A, Caballero A, et al. (2014) The action of stabilizing selection, mutation, and drift on epistatic quantitative traits. *Evolution* 68:1974–1987.
- Bacon MA (2004) Water use efficiency in plant biology. In: Bacon MA (ed) *Water Use Efficiency in Plant Biology*, Blackwell Publishing Ltd, Oxford, UK, pp 1-26.
- Baddeley A, Rubak E, Turner R (2015) *Spatial point patterns: Methodology and applications with R*. London: Chapman and Hall/CRC press, 2015. URL: <http://www.crcpress.com/Spatial-Point-Patterns-Methodology-and-Applications-with-R/Baddeley-Rubak-Turner/9781482210200/>
- Bailey DK (1970) Phytogeography and taxonomy of *Pinus* subsection *Balfourianae*. *Annals of the Missouri Botanical Garden* 57:210–249.
- Bailey SF, Bataillon T (2016) Can the experimental evolution programme help us elucidate the genetic basis of adaptation in nature? *Molecular Ecology* 25:203–218.
- Baldocchi DD, Xu L (2007) What limits evaporation from Mediterranean oak woodlands – the supply of moisture in the soil, physiological control by plants or the demand by the atmosphere? *Advanced Water Resources* 30:2113-2122.
- Bales RC, Hopmans J, O'Green AT, Meadows M, Hartsough PC, Kirchner P, Hunsaker CT, Beaudette D (2011) Soil moisture response to snowmelt and rainfall in a Sierra Nevada mixed-conifer forest. *Vadose Zone Journal* 10:786-799.
- Baltunis BS, Martin TA, Huber DA, Davis JM (2008) Inheritance of foliar stable carbon isotope discrimination and third-year height in *Pinus taeda* clones on contrasting sites in Florida and Georgia. *Tree Genetics & Genomes* 4:797-807.

- Barbour M, Keeler-Wolf T, Schoenherr AA (2007) *Terrestrial Vegetation of California*. Third edition. University of California Press, Berkeley, California, USA.
- Barker MS, Kane NC, Matvienko M, Kozik A, Michelmore RW, et al. 2008. Multiple paleopolyploidizations during the evolution of the Compositae reveal parallel patterns of duplicate gene retention after millions of years. *Molecular Biology and Evolution* 25:2445–55
- Barnett TP, Adam JC, Lettenmaier (2005) Potential impacts of a warming climate on water availability in snow-dominated regions. *Nature* 438:303-309.
- Barrett RD, Hoekstra HE (2011) Molecular spandrels: tests of adaptation at the genetic level. *Nature Reviews Genetics* 12:767-780.
- Barrett R, Schluter D (2008) Adaptation from standing genetic variation. *Trends in Ecology & Evolution* 23:38–44.
- Barton NH (2017) How does epistasis influence the response to selection? *Heredity* 118:96–109.
- Barton NH (1990) Pleiotropic models of quantitative variation. *Genetics* 124:773–782.
- Barton NH (1999) Clines in polygenic traits. *Genetical Research* 74:223–236.
- Barton NH, Etheridge AM, Véber A (2016) The infinitesimal model. *bioRxiv* 1–54.
- Beavis W (1994) The power and deceit of QTL experiments: lessons from comparative QTL studies. Proceedings of the forty-ninth annual corn and sorghum industry research conference. Chicago, IL, USA: American Seed Trade Association, 250–266.
- Béréanos C, Ellis PA, Pilkington JB, Pemberton JM (2014) Estimating quantitative genetic parameters in wild populations: a comparison of pedigree and genomic approaches. *Molecular Ecology* 23:3434–3451.
- Berg JJ, Coop G (2014) A Population Genetic Signal of Polygenic Adaptation. *PLoS Genetics* 10:e1004412.
- Bernardo R, Yu JM. 2007. Prospects for genomewide selection for quantitative traits in maize. *Crop Science* 47:1082–1090.
- Berry AJ, Ajioka JW, Kreitman M (1991) Lack of polymorphism on the *Drosophila* fourth chromosome resulting from selection. *Genetics* 129:1111–1117.
- Bessega C, Pometti C, Ewens M et al. (2015) Evidences of local adaptation in quantitative traits in *Prosopis alba* (Leguminosae). *Genetica* 143: 31.
- Blanc GG, Wolfe KHK. 2004. Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. *Plant Cell* 16:1667–78

- Blanquart F, Kaltz O, Nuismer SL, Gandon S (2013) A practical guide to measuring local adaptation. *Ecology Letters* 16:1195–1205.
- Bonello P, Gordon TR, Herms DA, Wood DL, Erbilgin N (2006) Nature and ecological implications of pathogen-induced systemic resistance in conifers: A novel hypothesis. *Physiological and Molecular Plant Pathology* 68: 95–104.
- Bontemps et al. (2016) In situ marker-based assessment of leaf trait evolutionary potential in a marginal European beech population. *Journal of Evolutionary Biology* 29: 514-527.
- Boshier D, Broadhurst L, Cornelius J, et al (2015) Is local best? Examining the evidence for local adaptation in trees and its scale. *Environmental Evidence* 4:1–10.
- Bourret V, Dionne M, Bernatchez L (2014) Detecting genotypic changes associated with selective mortality at sea in Atlantic salmon: polygenic multilocus analysis surpasses genome scan. *Molecular Ecology* 23: 4444–4457.
- Bower AD, Aitken SN (2008) Ecological genetics and seed transfer guidelines for *Pinus albicaulis* (Pinaceae). *American Journal of Botany* 95:66–76.
- Boyle EA, Li YI, Pritchard JK (2017) An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell* 169:1177–1186.
- Branca AA, Paape TDT, Zhou PP, Briskine RR, Farmer ADA, et al. 2011. Whole-genome nucleotide diversity, recombination, and linkage disequilibrium in the model legume *Medicago truncatula*. *Proceedings of the National Academy of Sciences. USA* 108:E864–70
- Brandvain Y, Wright SI (2016) The Limits of Natural Selection in a Non-equilibrium World. *Trends in Genetics* 32:201–210.
- Breiman L (2001) Random Forests. *Mach Learn* 45:5–32.
- Brendel O, Pot D, Plomion C, Rozenberg P, Guehl J-M (2002) Genetic parameters and QTL analysis of $\delta^{13}\text{C}$ and ring width in maritime pine. *Plant Cell Environ* 25:945-953.
- Brown JH, Stevens GC, Kaufman DM (1996) The geographic range: size, shape, boundaries, and internal structure. *Annual Review of Ecology and Systematics* 27:597-623.
- Browning BL, Browning SR (2016) Genotype imputation with millions of reference samples. *The American Journal of Human Genetics* 98:116–126.
- Buchert GP, Rajora OP, Hood JV, Dancik BP (1997) Effects of Harvesting on Genetic Diversity in Old-Growth Eastern White Pine in Ontario, Canada. *Conservation Biology* 11:747–758.
- Buck JM, Adams RS, Cone J, Conkle MT, Libby WJ, Eden CJ, Knight MJ (2007) California seed zones. San Francisco: California Region, Forest Service, United States Department of Agriculture, 5p.

- Budde KB, Heuertz M, Hernandez-Serrano A et al. (2014) In situ genetic association for serotiny, a fire-related trait, in Mediterranean maritime pine (*Pinus pinaster*). *New Phytologist* 201:230–241.
- Bulmer MG (1980) The mathematical theory of quantitative genetics. *Genetical research* 19: 17-25.
- Burczyk J, Adams WT, Shimizu JY (1996) Mating patterns and pollen dispersal in a natural knobcone pine (*Pinus attenuata* Lemmon.) stand. *Heredity* 77:251–260.
- Bürger R (1999) Evolution of Genetic Variability and the Advantage of Sex and Recombination in Changing Environments. *Genetics* 153:1055–1069.
- Bürger R, Akerman A (2011) The effects of linkage and gene flow on local adaptation: A two-locus continent-island model. *Theoretical Population Biology* 80:272–288.
- Bürger R, Lynch M (1995) Evolution and extinction in a changing environment: a quantitative-genetic analysis. *Evolution* 49:151–163.
- Burghardt LT, Young ND, Tiffin P (2017) A Guide to Genome-Wide Association Mapping in Plants. *Current Protocols in Plant Biology*.
- Burnham KP, Anderson DR (2002) Model Selection and Multimodal Inference – A Practical Information-Theoretic Approach. Springer, New York, New York, USA.
- Caballero A, Tenesa A, Keightley PD (2015) The nature of genetic variation for complex traits revealed by GWAS and regional heritability mapping analyses. *Genetics* 201:1601-1613.
- Cairney J, Pullman GS (2007) The cellular and molecular biology of conifer embryogenesis. *New Phytologist* 176:511–536.
- Ćalić I, Bussotti F, Martínez-García PJ, Neale DB (2015) Recent landscape genomics studies in forest trees. *Tree Genetics & Genomes* 1–7.
- Carlborg Ö, Haley CS (2004) Epistasis: too often neglected in complex trait studies? *Nature Reviews Genetics* 5:618–625.
- Carrasco A, Wegrzyn JL, Durán R, Fernández M, Donoso A, Rodriguez V, Neale DB, Valenzuela S (2017) Expression profiling in *Pinus radiata* infected with *Fusarium circinatum*. *Tree Genetics & Genomes* 13:1665.
- Carter AJR, Hermisson J, Hansen TF (2005) The role of epistatic gene interactions in the response to selection and the evolution of evolvability. *Theoretical Population Biology* 68:179–196.
- Cartwright DA, Troggio M, Velasco R, Gutin A (2007) Genetic mapping in the presence of genotyping errors. *Genetics* 176:2521– 2527.

- Catchen JM, Hohenlohe PA, Bernatchez L, et al (2017) Unbroken: RADseq remains a powerful tool for understanding the genetics of adaptation in natural populations. *Molecular Ecology Resources* 17:362–365.
- Chakravarti A, Lasher LK, Reefer JE (1991) A maximum-likelihood method for estimating genome length using genetic linkage data. *Genetics* 128:175–182
- Chang T, Hansen AJ, Piekielek N (2014) Patterns and Variability of Projected Bioclimatic Habitat for *Pinus albicaulis* in the Greater Yellowstone Area (B Bond-Lamberty, Ed.). *PLoS ONE*, 9, e111669.
- Charlesworth, B. 2013. Background selection 20 years on: The Wilhelmine E. Key 2012 Invitational Lecture. *Journal of Heredity* 104: 161-171.
- Charlesworth B (2009) Effective population size and patterns of molecular evolution and variation. *Nature Reviews Genetics* 10:195–205.
- Charlesworth B, Charlesworth D (2010) *Elements of Evolutionary Genetics*. Greenwood Village, CO: Roberts and Company Publishers. 734 pp.
- Chaves JA, Cooper EA, Hendry AP, Podos J, De León LF, Raeymaekers JAM, MacMillan WO, Uy JAC (2016) Genomic variation at the tips of the adaptive radiation of Darwin's finches. *Molecular Ecology* 25:5282–5295.
- Cheema J, Dicks J (2009) Computational approaches and software tools for genetic linkage map estimation in plants. *Briefings in Bioinformatics* 10:595–608.
- Cheliak WM, Murray G, Pitel JA (1988) Genetic effects of phenotypic selection in white spruce. *Forest Ecology and Management* 24:139–149.
- Cheverud JM, Routman EJ (1995) Epistasis and its contribution to genetic variance components. *Genetics* 139:1455–1461.
- Chevin L-M (2012) Genetic constraints on adaptation to a changing environment. *Evolution* 67:708–721.
- Chevin L-M, Hoffmann AA (2017) Evolution of phenotypic plasticity in extreme environments. *Philosophical Transactions of the Royal Society B: Biological Sciences* 372:20160138–12.
- Chevin L-M, Lande R, Mace GM (2010b) Adaptation, plasticity, and extinction in a changing environment: towards a predictive theory. *PLoS Biology* 8:e1000357.
- Chevin L-M, Martin G, Lenormand T (2010a) Fisher's model and the genomics of adaptation: restricted pleiotropy, heterogeneous mutation, and parallel evolution. *Evolution* 64:3213–3231.
- Chevin L-M, Hospital F (2008) Selective sweep at a quantitative trait locus in the presence of background genetic variation. *Genetics* 180:1645–1660.

- Chevin L-M, Lande R (2011) Adaptation to marginal habitats by evolution of increased phenotypic plasticity. *Journal of Evolutionary Biology* 24:1462–1476.
- Cheplick GP (2015) *Approaches to Plant Evolutionary Ecology*. Oxford University Press, USA.
- Chaine I (2010) Why does phenology drive species distribution? *Philosophical Transactions of the Royal Society B: Biological Sciences* 365:3149-3160
- Churchill GA, Doerge RW (1994) Empirical threshold values for quantitative trait mapping. *Genetics* 138:963-971
- Civelek M, Lusk AJ (2014) Systems genetics approaches to understand complex traits. *Nature Reviews Genetics* 15:34–48.
- Cloutier D, Kanashiro M, Ciampi AY, Schoen DJ (2006) Impact of selective logging on inbreeding and gene dispersal in an Amazonian tree population of *Carapa guianensis* Aubl. *Molecular Ecology* 16:797–809.
- Cockerham CC (1954) An Extension of the Concept of Partitioning Hereditary Variance for Analysis of Covariances among Relatives When Epistasis Is Present. *Genetics* 39:859–882.
- Cohen D, Bogaert-Triboulot M-B, Tisserant E, Balzergue S, Martin-Magniette M-L, Lelandais G, Ningre N, R, J-P, Tamby J-P, Le Thiec D, Hummel I (2010) Comparative transcriptomics of drought responses in *Populus*: a meta-analysis of genome-wide expression profiling in mature leaves and root apices across two genotypes. *BMC Genomics* 11:630.
- Collins BM, Everett RG, Stephens SL (2011) Impacts of fire exclusion and recent managed fire on forest structure in old growth Sierra Nevada mixed-conifer forests. *Ecosphere* 2:1–14.
- Collins S, de Meaux J, Acquisti C (2007) Adaptive walks toward a moving optimum. *Genetics* 176:1089–1099.
- Comeault AA, Flaxman SM, Riesch R, et al. (2015) Selection on a genetic polymorphism counteracts ecological speciation in a stick insect. *Current Biology* 25:1975–1981.
- Comeault AA, Soria-Carrasco V, Gompert Z, et al. (2014) Genome-wide association mapping of phenotypic traits subject to a range of intensities of natural selection in *Timema cristinae*. *The American Naturalist* 183:711–727.
- Coop G, Witonsky D, Di Rienzo A, Pritchard JK (2010) Using environmental correlations to identify loci underlying local adaptation. *Genetics*, 185, 1411–1423.
- Cornelius J (1994) Heritabilities and additive genetic coefficients of variation in forest trees. *Canadian Journal of Forest Research* 24:372–379.
- Costanza R, d'Arge R, De Groot R, et al. (1997) The value of the world's ecosystem services and natural capital. *Nature* 387:253-260.

- Covington WW, Everett RL, Steele R et al. (1994) Historical and anticipated changes in forest ecosystems of the inland west of the United States. *Journal of Sustainable Forestry* 2:13–63.
- Cowen L, Ideker T, Raphael BJ, Sharan R (2017) Network propagation: a universal amplifier of genetic associations. *Nature Ecology and Evolution* 18:551–562.
- Cregg, BM 1993. Seed source variation in water relations, gas exchange, and needle morphology of mature ponderosa pine trees. *Canadian Journal of Forrest Research* 23:749-755.
- Critchfield WB, Little EL (1966) Geographic distribution of the pines of the world (No. 991). US Department of Agriculture, Forest Service.
- Cronn R, Dolan PC, Jogdeo S, Wegrzyn JL, Neale DB, St. Clair JB, Denver DR (2017) Transcription through the eye of a needle: daily and annual cycles of gene expression variation in Douglas Fir needles. *bioRxiv*.
- Crow JF (2010) On epistasis: why it is unimportant in polygenic directional selection. *Philosophical Transactions of the Royal Society B: Biological Sciences* 365:1241–1244.
- Crow JF (2008) Maintaining evolvability. *Journal of Genetics* 87:349–353.
- Cruickshank TE, Hahn MW (2014) Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology* 23:3133–3157.
- Csilléry K, Lalagüe H, Vendramin GG, González-Martínez SC, Fady B, Oddou-Muratorio S (2014) Detecting short spatial scale local adaptation and epistatic selection in climate-related candidate genes in European beech (*Fagus sylvatica*) populations. *Molecular Ecology* 23:4696–4708.
- Cumbie WP, Eckert A, Wegrzyn J, Whetten R, Neale D, Goldfarb B (2011) Association genetics of carbon isotope discrimination, height and foliar nitrogen in a natural population of *Pinus taeda* L. *Heredity* 107:105-114.
- De La Torre AR, Li Z, Van de Peer Y, Ingvarsson PK (2017) Contrasting Rates of Molecular Evolution and Patterns of Selection among Gymnosperms and Flowering Plants. *Molecular Biology and Evolution* 34:1363–1377.
- Daly C, Neilson RP, Phillips DL (1994) A statistical-topographic model for mapping climatological precipitation over mountainous terrain. *Journal of Applied Meteorology* 33:140–158.
- Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. (2011) The variant call format and VCFtools. *Bioinformatics*, 27, 2156–2158.
- Davey JW, Blaxter ML (2010) RADSeq: next-generation population genetics. *Briefings in Functional Genomics* 9:416–423.

- Davis MB, Shaw RG (2001) Range shifts and adaptive responses to Quaternary climate change. *Science*, 292, 673-679.
- Davis MB, Shaw RG, Etterson JR (2005) Evolutionary responses to changing climate. *Ecology* 86:1704-1714.
- De Mita S, Thuillet A-C, Gay L, Ahmadi N, Manel S, Ronfort J, Vigouroux Y (2013) Detecting selection along environmental gradients: analysis of eight methods and their effectiveness for outbreeding and selfing populations. *Molecular Ecology* 22:1383–1399.
- Denti D, Schoen DJ (1988) Self-fertilization rates in white spruce: effect of pollen and seed production. *Journal of Heredity* 79:284—288.
- Devey ME, Fiddler TA, Liu BH, Knapp SJ, Neale DB (1994) An RFLP linkage map for loblolly pine based on a three-generation outbred pedigree. *Theoretical and Applied Genetics* 88:273–278
- de Villemereuil P, Gaggiotti OE, Mouterde M, Till-Bottraud I (2015) Common garden experiments in the genomic era: new perspectives and opportunities. *Heredity* 116:249–254. doi:
- de Visser JAGM, Cooper TF, Elena SF (2011) The causes of epistasis. *Proceedings of the Royal Society B: Biological Sciences* 278:3617–3624.
- de Vladar HP, Barton N (2014) Stability and response of polygenic traits to stabilizing selection and mutation. *Genetics* 197:749-767.
- Devlin B, Roeder K (1999) Genomic control for association studies. *Biometrics* 55:997-1004.
- Dickson SP, Wang K, Krantz I, Hakonarson H, Goldstein DB (2010) Rare variants create synthetic genome-wide associations. *PLoS Biology* 8:e1000294.
- Dittmar EL, Oakley CG, Conner JK, Gould BA, Schemske DW (2016) Factors influencing the effect size distribution of adaptive substitutions. *Proceedings of the Royal Society B: Biological Sciences* 283:3065–3068.
- Donohue K, Rubio de Casas R, Burghardt L, Kovach K, Willis CG (2010) Germination, post-germination adaptation, and species ecological ranges. *Annual Review of Ecology Evolution and Systematics* 41:293–319.
- Dow BD, Ashley MV (1998) Factors influencing male mating success in bur oak, *Quercus macrocarpa*. *New Forest* 15:161–180.
- Du J, Groover A (2010) Transcriptional regulation of secondary growth and wood formation. *J Integrative Plant Biology* 52:17-27.
- Dudley SA (1996) Differing selection on plant physiological traits in response to environmental water availability: a test of adaptive hypotheses. *Evolution* 50:92-102.

- Dungey HS (2001) Pine hybrids—a review of their use performance and genetics. *Forest Ecology & Management* 148:243-258.
- Dyer (2016) gstudio: Tools related to the spatial analysis of genetic marker data. R package version 1.5.0. <https://dyerlab.ces.vcu.edu/2015/01/16/gstudio/>
- East EM (1910) A Mendelian interpretation of variation that is apparently continuous. *The American Naturalist* 44:65–82.
- Echt CS, MayMarquardt P, Hseih M, Zahorchak R (1996) Characterization of microsatellite markers in eastern white pine. *Genome* 39:1102–1108.
- Eckert AJ (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* 93:1615-1624.
- Eckert AJ, Bower AD, González-Martínez SC, Wegrzyn JL, Coop G, Neale DB (2010a) Back to nature: ecological genomics of loblolly pine (*Pinus taeda*, Pinaceae). *Molecular Ecology* 19:3789–3805.
- Eckert AJ, Bower AD, Jermstad KD, Wegrzyn JL, Knaus BJ, Syring JV, Neale DB (2013b) Multilocus analyses reveal little evidence for lineage-wide adaptive evolution within major clades of soft pines (*Pinus* subgenus *Strobus*). *Molecular Ecology* 22: 5635–5650.
- Eckert AJ, Hall BD (2006) Phylogeny, historical biogeography, and patterns of diversification for *Pinus* (Pinaceae): phylogenetic tests of fossil-based hypotheses. *Molecular Phylogenetics and Evolution* 40:166–182
- Eckert AJ, Maloney PE, Vogler DR et al (2015) Local adaptation at fine spatial scales: an example from sugar pine (*Pinus lambertiana*, Pinaceae). *Tree Genetics & Genomes* 11:1–17.
- Eckert AJ, Pande B, Ersoz ES, Wright MH, Rashbrook VK, Nicolet CM, Neale DB (2009) High-throughput genotyping and mapping of single nucleotide polymorphisms in loblolly pine (*Pinus taeda* L.). *Tree Genetics & Genomes* 5:225-234.
- Eckert AJ, Sawyer JO (2002) Foxtail pine importance and conifer diversity in the Klamath Mountains and southern Sierra Nevada, California. *Madroño* 49:33-45.
- Eckert AJ, Tearnse BR, Hall BD (2008) A phylogeographical analysis of the range disjunction for foxtail pine (*Pinus balfouriana*, Pinaceae): the role of Pleistocene glaciation. *Molecular Ecology* 17: 1983–1997.
- Eckert AJ, van Heerwaarden J, Wegrzyn JL, Nelson CD, Ross-Ibarra J, González-Martínez SC, Neale DB (2010b) Patterns of population structure and environmental associations to aridity across the range of loblolly pine (*Pinus taeda* L., Pinaceae). *Genetics* 185:969–982

- Eckert AJ, Wegrzyn JL, Cumbie WP, et al. (2012) Association genetics of the loblolly pine (*Pinus taeda*, Pinaceae) metabolome. *New Phytologist* 193:890–902.
- Eckert AJ, Wegrzyn JL, Liechty JD, Lee JM, Cumbie WP, Davis JM, Goldfarb B, Loopstra CA, Palle SR, Quesada T, Langley CH, Neale DB (2013a) The evolutionary genetics of the genes underlying phenotypic associations for loblolly pine (*Pinus taeda*, Pinaceae). *Genetics* 195:1353–1372.
- Eckert AJ, Wegrzyn JL, Pande B, Jermstad KD, Lee JM, Liechty JD, Tearse BR, Krutovsky KV, Neale DB (2009) Multilocus patterns of nucleotide diversity and divergence reveal positive selection at candidate genes related to cold hardiness in coastal douglas fir (*Pseudotsuga menziesii* var. *menziesii*). *Genetics*, 183: 289–298.
- Eckert AJ, van Heerwaarden J, Wegrzyn JL, Nelson CD, Ross-Ibarra J, González-Martínez SC, Neale DB (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 CG, Samis KE, Loughheed SC (2008) Genetic variation across species' geographical ranges: the central-marginal hypothesis and beyond. *Molecular Ecology* 17:1170-1188.
- Ehleringer JR, Hall AE, Farquhar GD (1993) Stable isotopes and plant carbon/water relations. Academic Press, San Diego, California, USA.
- Ehret GB, Lamparter D, Hoggart CJ, Whittaker JC, Beckmann JS, Kutalik Z, Genetic Investigation of Anthropometric Traits Consortium (2012) A multi-SNP locus-association method reveals a substantial fraction of the missing heritability. *The American Journal of Human Genetics*, 91, 863–871.
- El-Kassaby YA, Dunsworth BG, Krakowski J (2003) Genetic evaluation of alternative silvicultural systems in coastal montane forests: western hemlock and amabilis fir. *Theoretical & Applied Genetics* 107:598–610.
- Elith J, Graham CH, Anderson RP, Dudik M, Ferrier S, et al. (1996) Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29:129-151.
- Ellegren H, Sheldon BC (2008) Genetic basis of fitness differences in natural populations. *Nature* 452:169–175.
- Endler JA (1977) Geographic variation, speciation, and clines. No. 10, Princeton University Press.
- Ennos RA. (1994) Estimating relative rates of pollen and seed migration among plant populations. *Heredity* 72:250–59.
- Ersoz ES, Wright MH, González-Martínez SC, Langley CH, Neale DB (2010) Evolution of disease response genes in loblolly pine: insights from candidate genes. *PLoS ONE* 5:e14234.

- Ettinger AK, Ford KR, HilleRisLambers J (2011) Climate determines upper, but not lower, altitudinal range limits of Pacific Northwest conifers. *Ecology* 92:1323-1331.
- Eveno E, Collada, Angeles Guevara M, Leger V, Soto 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.
- Everett MV, Seeb JE (2014) Detection and mapping of QTL for temperature tolerance and body size in Chinook salmon (*Oncorhynchus tshawytscha*) using genotyping by sequencing. *Evolutionary Applications* 7:480-492.
- Eyre-Walker A, Keightley PD (2007) The distribution of fitness effects of new mutations. *Nature Reviews Genetics* 8:610–618.
- Eyre-Walker A, Keightley PD. 2009. Estimating the rate of adaptive molecular evolution in the presence of slightly deleterious mutations and population size change. *Molecular Biology and Evolution* 26:2097–108
- Evans, L. M., Kaluthota, S., Pearce, D. W., Allan, G. J., Floate, K., Rood, S. B., & Whitham, T. G. (2016). Bud phenology and growth are subject to divergent selection across a latitudinal gradient in *Populus angustifolia* and impact adaptation across the distributional range and associated arthropods. *Ecology and Evolution* 6:4565-4581.
- Fageria MS, Rajora OP (2013a) Effects of harvesting of increasing intensities on genetic diversity and population structure of white spruce. *Evolutionary Applications* 6:778–794.
- Fageria MS, Rajora OP (2013b) Effects of silvicultural practices on genetic diversity and population structure of white spruce in Saskatchewan. *Tree Genetics & Genomes* 10:287–296.
- Falconer DS (1989) *Introduction to quantitative genetics*, 3d ed. Longman, New York.
- Faltus FA (2014) Systems genetics: A paradigm to improve discovery of candidate genes and mechanisms underlying complex traits. *Plant Science* 223: 45-48.
- FAO (2010) *Global Forest Resources Assessment 2010*. Rome, Italy.
- Farris MA, Mitton JB (1984) Population density, outcrossing rate, and heterozygote superiority in ponderosa pine. *Evolution* 8:1151–1154.
- Farquhar GD, Richards RA (1984) Isotopic composition of plant carbon correlates with water-use efficiency of wheat genotypes. *Australian Journal of Plant Physiology* 11:539-552.
- Farquhar GD, O’Leary MH, Berry JA (1982) On the relationship between carbon isotope discrimination and the intercellular carbon dioxide concentrations in leaves. *Australian Journal of Plant Physiology* 9:121-137.

- Feder JL, Egan SP, Nosil P (2012a) The genomics of speciation-with-gene-flow. *Trends in Genetics* 28:342–350.
- Feder JL, Gejji R, Yeaman S, Nosil P (2012b) Establishment of new mutations under divergence and genome hitchhiking. *Philosophical Transactions of the Royal Society B: Biological Sciences* 367:461–474.
- Feder JL, Nosil P (2010) The efficacy of divergence hitchhiking in generating genomic islands during ecological speciation. *Evolution* 64:1729–1747.
- Feldman M, Lewontin R (1975) The heritability hang-up. *Science* 190:1163–1168.
- Felsenstein J (1976) The theoretical population genetics of variable selection and migration. *Annual Review of Genetics* 10:253–280.
- Finkeldey R, Ziehe M (2004) Genetic implications of silvicultural regimes. *Forest Ecology and Management* 197:231–244.
- Fisher RA (1918) The correlation between relatives on the supposition of Mendelian inheritance. *Transactions of the Royal Society of Edinburgh* 52:399–433.
- Fisher RA (1930) *The genetical theory of natural selection: a complete variorum edition*. Oxford University Press
- Ford EB (1975) *Ecological Genetics*, 4th edition. Chapman and Hall, London, UK.
- Forester BR, Lasky JR, Wagner HH, Urban DL (2017) Using genotype-environment associations to identify multilocus local adaptation. *bioRxiv* 1–24.
- Fournier-Level A, Wilczek AM, Cooper MD, Roe JL, Anderson J, Eaton D, Moyers BT, Petipas RH, Schaeffer RN, Pieper B, Raymond M, Koornneef M, Welch SM, Remington DL, Schmitt J (2013) Paths to selection on life history loci in different natural environments across the native range of *Arabidopsis thaliana*. *Molecular Ecology* 22:3552–3566
- Fowells HA, Schubert GH (1956) *Seed crops of forest trees in the pine region of California*. USDA Forest Service, Technical Bulletin 150, Washington, D.C.
- Frankham R (1995) Effective population size/adult population size ratios in wildlife: a review. *Genetics Research* 66:95-107.
- Franks SJ, Weber JJ, Aitken SN (2013) Evolutionary and plastic responses to climate change in terrestrial plant populations. *Evolutionary Applications* 7:123–139.
- Friedline CJ, Franklin RB, McCallister SL, Rivera MC (2012) Bacterial assemblages of the eastern Atlantic Ocean reveal both vertical and latitudinal biogeographic signatures. *Biogeosciences* 9: 2177–2193

- Friedline CJ, Lind BM, Hobson EM, Harwood DE, Mix AD, Maloney PE, Eckert AJ (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 Genet Genomes* 11:49.
- Furnier OR, Adams WT (1986) Mating system in natural populations of Jeffrey pine. *American Journal of Botany* 73:1002–1008.
- Gagnaire P-A, Gaggiotti OE (2016) Detecting polygenic selection in marine populations by combining population genomics and quantitative genetics approaches. *Current Zoology* 62:603–616.
- García-Ramos G, Kirkpatrick M (1997) Genetic models of adaptation and gene flow in peripheral populations. *Evolution* 51:21–28.
- Gaston KJ (2003) *The Structure and Dynamics of Geographic Ranges*. Oxford University Press, Oxford, UK.
- Gazal S, Finucane HK, Furlotte NA, Loh P, Palamara PF, Liu X, Schoech A, Bulik-Sullivan B, Neale BM, Gusev A, Price A (2017) Linkage disequilibrium–dependent architecture of human complex traits shows action of negative selection. *Nature Genetics* 49:1421-1427.
- Geremew A, Woldemariam MG, Kefalew A, Stiers I, Triest L (2018) Isotropic and anisotropic processes influence fine-scale spatial genetic structure of a keystone tropical plant. *AoB Plants*. Early online.
- Gernandt DS, Lopez GG, Garcia SO, Liston A (2005) Phylogeny and classification of *Pinus*. *Taxon* 54:29–42
- Ghodsi M, Hill CM, Astrovskaya I, Lin H, Sommer DD, Koren S, Pop M (2013) De novo likelihood-based measures for comparing genome assemblies. *BMC Research Notes* 6:334
- Gibson G (2012) Rare and common variants: twenty arguments. *Nature Reviews Genetics* 13:135–145.
- Gilbert, K. J., & Whitlock, M. C. (2015). Q_{ST} – F_{ST} comparisons with unbalanced half-sib designs. *Molecular Ecology Resources* 15:262-267.
- Gladman S, Seeman T (2016) Velvet Optimiser v2.5.5, available online <http://bioinformatics.net.au/software.velvetoptimiser.shtml>.
- Goddard ME, Wray NR, Verbyla K, Visscher PM (2009) Estimating effects and making predictions from genome-wide marker data. *Statistical Science* 24:517-529.
- Gömöry D (1992) Effect of stand origin on the genetic diversity of Norway spruce (*Picea abies* Karst.) populations. *Forest Ecology and Management* 54:215–223.

- Gompert, Z., Egan, S. P., Barrett, R. D. H., Feder, J. L., & Nosil, P. (2016). Multilocus approaches for the measurement of selection on correlated genetic loci. *Molecular Ecology*, 1–18.
- Gompert Z, Jahner JP, Scholl CF, Wilson JS, Lucas LK, Soria-Carrasco V, Fordyce JA, Nice CC, Buerkle CA, Forister ML (2015) The evolution of novel host use is unlikely to be constrained by trade-offs or a lack of genetic variation. *Molecular Ecology* 24:2777–2793.
- González-Martínez SC, Huber D, Ersoz E, Davis JM, Neale DB (2008) Association genetics in *Pinus taeda* L. II. Carbon isotope discrimination. *Heredity*, 101, 19–26.
- Goodnight CJ (1988) Epistasis and the effect of founder events on the additive genetic variance. *Evolution* 42:441-454.
- Göring HHH, Terwilliger JD, Blangero J (2001) Large Upward Bias in Estimation of Locus-Specific Effects from Genomewide Scans. *The American Journal of Human Genetics* 69:1357–1369.
- Gossmann TI, Song B-H, Windsor AJ, Mitchell-Olds T, Dixon CJ, et al. 2010. Genome wide analyses reveal little evidence for adaptive evolution in many plant species. *Molecular Biology and Evolution* 27:1822–32
- Gossmann TI, Keightley PD, Eyre-Walker A. 2012. The effect of variation in the effective population size on the rate of adaptive molecular evolution in eukaryotes. *Genome Biology and Evolution* 4(5):658–67
- Grandtner, M. M. (2005). Elsevier's Dictionary of Trees: Volume 1: North America. Elsevier.
- Grattapaglia D (2017) Status and perspectives of genomic selection in forest tree breeding. In ME Sorrells, *Genomic Selection for Crop Improvement* (pp. 199-257). Cham, Switzerland: Springer.
- Grattapaglia D, Resende MDV (2011) Genomic selection in forest tree breeding. *Tree Genetics & Genomes* 7: 241-255.
- Grattapaglia D, Sederoff R (1994) Genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudo-test cross: mapping strategy and RAPD markers. *Genetics* 137:1121–1137
- Griswold CK (2015) Additive genetic variation and evolvability of a multivariate trait can be increased by epistatic gene action. *Journal of Theoretical Biology* 387:241–257.
- Groover A (2015) Genomic science provides new insights into the biology of forest trees. *New Phytologist*, 208, 302–305.
- Groover A, Devey M, Fiddler T, Lee J, Megraw R, Mitchell-Olds T, Sherman B, Vujcic S, Williams C, Neale DB (1994) Identification of quantitative trait loci influencing wood specific gravity in an outbred pedigree of loblolly pine. *Genetics* 138:1293-1300.

- Guan Y, Stephens M (2011) Bayesian variable selection regression for genome-wide association studies and other large-scale problems. *Annals of Applied Statistics* 5:1780–1815.
- Günther T, Coop G (2013) Robust identification of local adaptation from allele frequencies. *Genetics*, 195, 205–220.
- Haasl RJ, Payseur BA (2016) Fifteen years of genomewide scans for selection: trends, lessons and unaddressed genetic sources of complication. *Molecular Ecology*, 25, 5–23.
- Hackett CA, Broadfoot LB (2003) Effects of genotyping errors, missing values and segregation distortion in molecular marker data on the construction of linkage maps. *Heredity* 90:33–38
- Haldane JBS (1930) A mathematical theory of natural and artificial selection.(Part VI, Isolation.). 26:220–230.
- Hall D, Hallingbäck HR, Wu HX (2016) Estimation of number and size of QTL effects in forest tree traits. *Tree Genetics & Genomes* 1–17.
- Hall D, Luquez V, Garcia VM, St Onge KR, Jansson S, Ingvarsson PK (2007) Adaptive population differentiation in phenology across a latitudinal gradient in European aspen (*Populus tremula*, L.): A comparison of neutral markers, candidate genes, and phenotypic traits. *Evolution*, 61, 2849–2860.
- Hamming RW (1950) Error detecting and error correcting codes. *Bell System Technical Journal* 29:147–160
- Hampe A (2004) Bioclimate envelope models: what they detect and what they hide. *Global Ecology and Biogeography* 13:469-471.
- Hamrick JL, Godt MJW, Sherman-Broyles SL (1992) Factors influencing levels of genetic diversity in woody plant species. *New Forests*. 6:95–124.
- Hansen TF (2006) The Evolution of Genetic Architecture. *Annual Review of Ecology, Evolution, and Systematics* 37:123–157.
- Hansen TF, Pelabon C, Houle D (2011) Heritability is not evolvability. *Evolutionary Biology* 38:258-277
- Hansen TF (2013) Why epistasis is important for selection and adaptation. *Evolution* 67:3501–3511.
- Hansen TF (2003) Is modularity necessary for evolvability? *Biosystems* 69:83–94.
- Hansen TF, Wagner GP (2001) Modeling genetic architecture: a multilinear theory of gene interaction. *Theoretical Population Biology* 59:61–86.

- Hardy OJ, González-Martínez SC, Colas B, Fréville H, Mignot A, Olivieri I (2004). Fine-scale genetic structure and gene dispersal in *Centaurea corymbosa* (Asteraceae). II. Correlated paternity within and among sibships. *Genetics* 168:1601–1614.
- Hardy OJ, Vekemans X (1998) Isolation by distance in a continuous population: reconciliation between spatial autocorrelation analysis and population genetics models. *Heredity* 83:145–154.
- Hausmann NJ, Juenger TE, Sen S, Stowe KA, Dawson TE, Simms EL (2005) Quantitative trait loci affecting $\delta^{13}\text{C}$ and response to differential water availability in *Arabidopsis thaliana*. *Evolution* 59:81–96
- Heffner EL, Sorrells ME, Jannink JL (2009) Genomic selection for crop improvement. *Crop Science* 49: 1–12.
- Hemani G, Knott S, Haley C (2013) An Evolutionary Perspective on Epistasis and the Missing Heritability. *PLoS Genetics* 9:e1003295.
- Henderson CR (1975) Best linear unbiased estimation and prediction under a selection model. *Biometrics* 31:423–447.
- Hendry AP (2002) $Q_{\text{ST}} \geq \neq < F_{\text{ST}}$? *Trends in Ecology and Evolution* 17:502–502.
- Hendry AP (2016) Key Questions on the Role of Phenotypic Plasticity in Eco-Evolutionary Dynamics. *Journal of Heredity* 107:25–41.
- Hereford J (2009) A Quantitative Survey of Local Adaptation and Fitness Trade-Offs. *The American Naturalist* 173:579–588.
- Hermisson J (2005) Soft sweeps: molecular population genetics of adaptation from standing genetic variation. *Genetics* 169: 2335–2352.
- Hermisson J (2009) Who believes in whole-genome scans for selection? *Heredity* 103:283–284.
- Hermisson J, Hansen TF, Wagner GP (2003) Epistasis in Polygenic Traits and the Evolution of Genetic Architecture under Stabilizing Selection. *American Naturalist* 161:708–734.
- Hermisson J, Pennings PS (2005) Soft sweeps: molecular population genetics of adaptation from standing genetic variation. *Genetics* 169:2335–2352.
- Hermisson J, Pennings PS (2017) Soft sweeps and beyond: understanding the patterns and probabilities of selection footprints under rapid adaptation. *Methods in Ecology and Evolution* 8:700–716.
- Hewitt GM (1999) Post-glacial recolonization of European biota. *Biological Journal of the Linnean Society* 68:87–112.

- Hill WG (2010) Understanding and using quantitative genetic variation. *Philosophical Transactions of the Royal Society B: Biological Sciences* 365:73–85.
- Hill WG, Goddard ME, Visscher PM (2008) Data and theory point to mainly additive genetic variance for complex traits. *PLoS Genetics* 4:e1000008.
- Hirschhorn JN, Daly MJ (2005) Genome-wide association studies for common diseases and complex traits. *Nature Reviews Genetics* 6:95–108.
- Hoban S, Kelley JL, Lotterhos KE et al (2016) Finding the genomic basis of local adaptation: Pitfalls, practical solutions, and future directions. *Am Nat* 188:379–397.
- Hodgins KA, Yeaman S, Nurkowski KA, et al. (2016) Expression Divergence Is Correlated with Sequence Evolution but Not Positive Selection in Conifers. *Molecular Biology and Evolution* 33:1502–1516.
- Hoffmann AA, Rieseberg LH (2008) Revisiting the Impact of Inversions in Evolution: From Population Genetic Markers to Drivers of Adaptive Shifts and Speciation? *Annual Review of Ecology, Evolution, and Systematics* 39:21–42.
- Holland J (2007) Genetic architecture of complex traits in plants. *Current Opinion in Plant Biology*, 10, 156–161.
- Holliday JA, Aitken SN, Cooke JEK, Fady B, González-Martínez SC, Heuertz M, Jaramillo-Correa JP, Lexer C, Staton M, Whetten RW, Plomion C (2017) Advances in ecological genomics in forest trees and applications to genetic resources conservation and breeding. *Molecular Ecology* 26:706–717.
- Holliday, JA, Wang T, Aitken SN (2012). Predicting adaptive phenotypes from multilocus genotypes in Sitka spruce (*Picea sitchensis*) using random forest. *G3-Genes Genomes and Genetics* 2:1085–1903.
- Holliday JA, Zhou L, Bawa R, Zhang M, Oubida RW (2016) Evidence for extensive parallelism but divergent genomic architecture of adaptation along altitudinal and latitudinal gradients in *Populus trichocarpa*. *New Phytologist* 209:1240–1251.
- Holsinger KE (2000) Reproductive systems and evolution in vascular plants. *Proceedings of the National Academy of Sciences* 97:7037–7042.
- Holsinger KE, Weir BS (2009) Genetics in geographically structured populations: defining, estimating and interpreting *F*. *Nature Reviews Genetics* 10:639–650.
- Hornoy B, Pavy N, Gérardi S, Beaulieu J, Bousquet J (2015) Genetic adaptation to climate in white spruce involves small to moderate allele frequency shifts in functionally diverse genes. *Genome Biology and Evolution* 7:3269–3285.
- Hough J, Williamson RJ, Wright SI (2013). Patterns of selection in plant genomes. *Annual Review of Ecology, Evolution, and Systematics* 44:31–49.

- Houle D, Pelabon C, Wagner GP, Hansen TF (2011) Measurement and meaning in biology. *The Quarterly Review of Biology* 86:3–34
- Howe GT, Aitken SN, Neale DB, Jermstad KD, Wheeler NC, Chen THH (2003) From genotype to phenotype: unraveling the complexities of cold adaptation in forest trees. *Canadian Journal of Botany* 81:1247–1266.
- Huang W, Mackay TFC (2016) The genetic architecture of quantitative traits cannot be inferred from variance component analysis. *PLoS Genetics* 12:e1006421.
- Huber CD, Durvasula A, Hancock AM, Lohmueller KE (2017). Gene expression drives the evolution of dominance. *bioRxiv*.
- Hutchinson GE (1957) Concluding remarks. *Cold Spring Harb Symp Quant Biol* 22:415–427.
- Ingvarsson PK, Hvidsten TR, Street NR (2016) Towards integration of population and comparative genomics in forest trees. *New Phytologist* 212:338–344.
- Ingvarsson PK, Street NR (2011) Association genetics of complex traits in plants. *New Phytologist* 189:909–922
- Innan H, Kim Y (2004) Pattern of polymorphism after strong artificial selection in a domestication event. *Proceedings of the National Academy of Sciences* 101:10667–10672.
- Isik F, Kumar S, Martínez-García PJ, Iwata H, Yamamoto T (2015) Acceleration of forest and fruit tree domestication by genomic selection. *Adv Bot Res* 74:93–124.
- Iwata et al (2011) Prospects for genomic selection in conifer breeding: a simulation study of *Cryptomeria japonica*. *Tree Genet Genomes* 7:747–758.
- Jain K, Stephan W (2015) Response of polygenic traits under stabilizing selection and mutation when loci have unequal effects. *G3-Genes Genom Genet* 5:1065–1074.
- Jain K, Stephan W (2017) Rapid Adaptation of a Polygenic Trait After a Sudden Environmental Shift. *Genetics* 206:389–406.
- Jansen RC, Tesson BM, Fu J, Yan Y, McIntyre LM (2009) Defining gene and QTL networks. *Current Opinion in Plant Biology* 12:241–246.
- Jensen JD (2014) On the unfounded enthusiasm for soft selective sweeps. *Nature Communications* 5:5281.
- Jensen JD, Kim Y, DuMont VB, Aquadro CF, Bustamante CD (2005) Distinguishing between selective sweeps and demography using DNA polymorphism data. *Genetics* 170:1401–1410.

- Jermstad KD, Bassoni DL, Wheeler NC, Anekonda TS, Aitken SN, Adams WT, Neale DB (2001) Mapping of quantitative trait loci controlling adaptive traits in coastal Douglas-fir. II. Spring and fall cold-hardiness. *Theor Appl Genet* 102:1152-1158.
- Jermstad KD, Bassoni DL, Jech KS, Ritchie GA, Wheeler NC, Neale DB (2003) Mapping of quantitative trait loci controlling adaptive traits in coastal Douglas-fir. III. Quantitative trait loci-by-environment interactions. *Genetics* 165:1489-1506.
- Jiao WB, Schneeberger K (2017) The impact of third generation genomic technologies on plant genome assembly. *Curr Opin Plant Biol* 36:64-70.
- Joao Gaspar M, Velasco T, Feito I, Alia R, Majada J (2013) Genetic variation of drought tolerance in *Pinus pinaster* at three hierarchical levels: a comparison of induced osmotic stress and field testing. *PLoS ONE* 8:e79094.
- Johnson KH, Flanagan LB, Huber DA, Major JE (1999) Genetic variation in growth, carbon isotope discrimination, and foliar N concentration in *Picea mariana*: analyses from a half-diallel mating design using field-grown trees. *Canadian Journal of Forest Research* 29:1727-1735.
- Johnson RC, Nelson GW, Troyer JL, Lautenberger JA, Kessing BD, Winkler CA, O'Brien SJ (2010) Accounting for multiple comparisons in a genome-wide association study (GWAS). *BMC Genomics* 11:724.
- Jones AG, Bürger R, Arnold SJ (2014) Epistasis and natural selection shape the mutational architecture of complex traits. *Nature Communications* 5:3709.
- Joo JWJ, Hormozdiari F, Han B, Eskin E (2016) Multiple testing correction in linear mixed models. *Genome Biology* 17:62.
- Josephs EB, Wright SI, Stinchcombe JR, Schoen DJ (2017a) The relationship between selection, network connectivity, and regulatory variation within a population of *Capsella grandiflora*. *Genome Biology and Evolution* 9:1099–1109.
- Josephs EB, Lee YW, Stinchcombe JR, Wright SI (2015) Association mapping reveals the role of purifying selection in the maintenance of genomic variation in gene expression. *Proceedings of the National Academy of Sciences* 112:15390–15395.
- Josephs EB, Stinchcombe JR, Wright SI (2017b) What can genome-wide association studies tell us about the evolutionary forces maintaining genetic variation for quantitative traits? *New Phytologist* 214:21–33.
- Kang BY, Mann IK, Major JE, Rajora OP (2010) Near-saturated and complete genetic linkage map of black spruce (*Picea mariana*). *BMC Genomics* 11:515
- Kaplan NL, Hudson RR, Langley CH (1989) The “hitchhiking effect” revisited. *Genetics* 123:887–899.

- Karijolic J, Yi C, Yu Y-T (2015) Transcriptome-wide dynamics of RNA pseudouridylation. *Nature Reviews Molecular and Cell Biology* 16:581-585.
- Kärkkäinen K, Savolainen O, Koski V (1999) Why do plants abort so many developing seeds: bad offspring or bad maternal genotypes? *Evolutionary Ecology* 13:305-317.
- Kawecki TJ (2008) Adaptation to Marginal Habitats. *Annual Review of Ecology, Evolution, and Systematics* 39:321–342.
- Kawecki TJ, Ebert D (2004) Conceptual issues in local adaptation. *Ecology Letters* 7:1225–1241.
- Keeling CI, Weisshaar S, Lin RPC, Bohlmann J (2008) Functional plasticity of paralogous diterpene synthases involved in conifer defense. *Proceedings of the National Academy of Sciences* 105:1085–1090.
- Keightley PD, Eyre-Walker A (2010) What can we learn about the distribution of fitness effects of new mutations from DNA sequence data? *Philosophical Transactions of the Royal Society B: Biological Sciences* 365:1187–1193.
- Kemper KE, Saxton SJ, Bolormaa S, Hayes BJ, Goddard ME (2014) Selection for complex traits leaves little or no classic signatures of selection. *BMC Genomics*, 15, 246–260.
- Kempthorne O (1954) The correlation between relatives in a random mating population. *Proceedings of the Royal Society of London B: Biological Sciences* 143:103–113.
- Kimura M (1983) *The neutral theory of molecular evolution*. Cambridge University Press
- Kinloch Jr, B. B., Parks, G. K., & Fowler, C. W. (1970). White pine blister rust: simply inherited resistance in sugar pine. *Science*, 193-195.
- Kirkpatrick M, Barton NH (1997) Evolution of a species' range. *The American Naturalist* 150:1–23.
- Kirkpatrick M, Barton NH (2006) Chromosome Inversions, Local Adaptation and Speciation. *Genetics* 173:419–434.
- Knott SA, Elsen, JM, Haley CS (1996) Methods for multiple-marker mapping of quantitative trait loci in half-sib populations. *Theoretical and Applied Genetics* 93:71–80.
- Knowles P, Furnier GR, Aleksasuk MA, Perry DJ (1987) Significant levels of self-fertilization in natural populations of tamarack. *Canadian Journal of Botany* 65:1087—1091.
- Koboldt DC, Steinberg KM, Larson DE, Wilson RK, Mardis ER (2013) The next-generation sequencing revolution and its impact on genomics. *Cell* 155:27–38.
- Kopp M, Hermisson J (2009a) The Genetic Basis of Phenotypic Adaptation I: Fixation of Beneficial Mutations in the Moving Optimum Model. *Genetics* 182:233–249.

- Kopp M, Hermisson J (2009b) The genetic basis of phenotypic adaptation II: the distribution of adaptive substitutions in the moving optimum model. *Genetics* 183:1453–1476.
- Kopp M, Matuszewski S (2013) Rapid evolution of quantitative traits: theoretical perspectives. *Evolutionary Applications* 7:169–191.
- Kosambi DD (1944) The estimation of map distance from recombination values. *Ann Eugen* 12:172–175
- Krakowski J, Aitken SN, El-Kassaby YA (2003) Inbreeding and conservation genetics in whitebark pine. *Conservation Genetics* 4, 581–593.
- Kramer AT, Ison JL, Ashley MV, Howe HF (2008) The Paradox of Forest Fragmentation Genetics. *Conservation Biology* 22:878–885.
- Kremer A, Le Corre V (2012) Decoupling of differentiation between traits and their underlying genes in response to divergent selection. *Heredity* 108:375–385.
- Kremer A, Ronce O, Robledo-Arnuncio JJ, Guillaume F, Bohrer G, Nathan R, Bridle JR, Gomulkiewicz, Klein EK, Ritlan K, Kuparinen A, Gerber S, Schueler S (2012) Long-distance gene flow and adaptation of forest trees to rapid climate change. *Ecology Letters* 15:378–392.
- Krutovsky KV, Neale DB (2005) Nucleotide diversity and linkage disequilibrium in cold-hardiness- and wood quality-related candidate genes in Douglas fir. *Genetics* 171:2029–2041.
- Kubisiak TL, Nelson CD, Name WL, Stine M (1996) Comparison of rapid linkage maps constructed for a single longleaf pine from both haploid and diploid mapping populations. *Forest Genetics* 3:203–211
- Lamichhaney S, Berglund J, Almén MS, Maqbool K, Grabherr M, Martinez-Barrio A, Promerová M, Rubin C-J, Wang C, Zamani N, Grant BR, Grant PR, Webster MT, Andersson L (2015) Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature* 518:371–375.
- Lamy JB, Bouffier L, Burlett R, Plomion C, Cochard H, Delzon S (2011) Uniform selection as a primary force reducing population genetic differentiation of cavitation resistance across a species range. *PLoS ONE* 6:e23476.
- Lamy JB, Plomion C, Kremer A, Delzon S (2012) $Q_{ST} < F_{ST}$ as a signature of canalization. *Molecular Ecology* 21:5646–5655.
- Lande R (1980) The genetic covariance between characters maintained by pleiotropic mutations. *Genetics* 94:203–215.
- Lande R (2009) Adaptation to an extraordinary environment by evolution of phenotypic plasticity and genetic assimilation. *Journal of Evolutionary Biology* 22:1435–1446.

- Lande R, Arnold S (1983) The measurement of selection on correlated characters. *Evolution*, 37, 1210–1226.
- Lange K, Boehnke M (1982) How many polymorphic marker genes will it take to span the human genome. *American Journal of Human Genetics* 128:842–845
- Langlet O (1971) Two hundred years genecology. *Taxon* 20:653–721.
- Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9, 357–359.
- Laporte V, Charlesworth B (2002) Effective population size and population subdivision in demographically structured populations. *Genetics* 162: 501-519.
- Larson AJ, Churchill D (2012) Tree spatial patterns in fire-frequent forests of western North America, including mechanisms of pattern formation and implications for designing fuel reduction and restoration treatments. *Forest Ecology and Management* 267:74–92.
- Latta RG (1998) Differentiation of allelic frequencies at quantitative trait loci affecting locally adaptive traits. *The American Naturalist* 151:283–292.
- Latta RG (2003) Gene flow, adaptive population divergence and comparative population structure across loci. *New Phytologist* 161:51–58.
- Laurent S, Pfeifer SP, Settles ML, Hunter SS, Hardwick KM, Ormond L, Sousa VC, Jensen JD, Rosenblum EB (2016) The population genomics of rapid adaptation: disentangling signatures of selection and demography in white sands lizards. *Molecular Ecology* 25:306–323.
- Lauteri M, Pliura A, Monteverdi MC, Brugnoli E, Villani F, Eriksson G (2004) Genetic variation in carbon isotope discrimination in six European populations of *Castanea sativa* Mill. originating from contrasting localities. *J Evol Biol* 17:1286–1296.
- Le Corre V, Kremer A (2012) The genetic differentiation at quantitative trait loci under local adaptation. *Molecular Ecology* 21:1548–1566.
- Le Rouzic A, Álvarez-Castro JM (2016) Epistasis-induced evolutionary plateaus in selection responses. *American Naturalist* 188:E134–E150.
- Ledig FT (1988) The conservation of diversity in forest trees: why and how should genes be conserved? *BioScience* 38:471–479.
- Ledig FT (1992) Human impacts on genetic diversity in forest ecosystems. *Oikos* 63:87–108.
- Leempoel K, Duruz S, Rochat E, Widmer I, Orozco Wengel P, Joost S (2017) Simple rules for an efficient use of Geographic Information Systems in molecular ecology. *Frontiers in Ecology and Evolution* 5:33.

- Legendre, P., & Legendre, L. F. (2012). Numerical ecology (Vol. 24). Elsevier.
- Leimu R, Fischer M (2008) A meta-analysis of local adaptation in plants. PLoS ONE 3:e4010.
- Leinonen T, McCairns RJS, O'Hara RB, Merilä J (2013) Q_{ST} – F_{ST} comparisons: evolutionary and ecological insights from genomic heterogeneity. Nature Reviews Genetics 14:179–190.
- Leinonen PH, Sandring S, Quilot B et al (2009) Local adaptation in European populations of *Arabidopsis lyrata* (Brassicaceae). American Journal of Botany 96:1129–1137.
- Leiserson M, Eldridge JV, Ramachandran S (2013) Network analysis of GWAS data. Current opinion in Genetics and Development 23:602-610.
- Leitch AR, Leitch IJ (2012) Ecological and genetic factors linked to contrasting genome dynamics in seed plants. New Phytologist 194:629-646.
- Lenormand T (2002) Gene flow and the limits to natural selection. Trends in Ecology & Evolution 17:183–189.
- Lewontin RC, Krakauer J (1973) Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. Genetics 74: 175-195
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup (2009) The sequence alignment/map format and SAMtools. Bioinformatics, 25, 2078–2079.
- Li Y, Suontoma M, Burdon RD, Dungey HS (2017) Genotype by environment interactions in forest tree breeding: review of methodology and perspectives on research and application. Tree Genetics & Genomes 13:60.
- Limborg MT, Waples RK, Seeb JE, Seeb LW (2014) Temporally isolated lineages of pink salmon reveal unique signatures of selection on distinct pools of standing genetic variation. Journal of Heredity 105:741– 751.
- Lind BM, Friedline CJ, Wegrzyn JL, Maloney PE, Vogler DR, Neale DB, Eckert AJ (2017) Water availability drives signatures of local adaptation in whitebark pine (*Pinus albicaulis* Engelm.) across fine spatial scales of the Lake Tahoe Basin, USA. Molecular Ecology 26:3168–3185.
- Loehle C (1988) Tree life history strategies: the role of defenses. Canadian Journal of Forest Research 18: 209-222.
- Long AD, Langley CH (1999) The power of association studies to detect the contribution of candidate genetic loci to variation in complex traits. Genome Research 9:720-731.
- Lopez GA, Potts BM, Vaillancourt RE, Apiolaza LA (2003). Maternal and carryover effects on early growth of *Eucalyptus globulus*. Canadian Journal of Forest Research 33:2108–2115.

- Liu J-J, Williams H, Li XR, Schoettle AW, Sniezko RA, Murray M, Zamany A, Roke, G, Chen H (2017) Profiling methyl jasmonate-responsive transcriptome for understanding induced systemic resistance in whitebark pine (*Pinus albicaulis*). *Plant Molecular Biology* 95:359-374.
- Liu J-J, Schoettle AW, Sniezko RA et al (2016) Genetic mapping of *Pinus flexilis* major gene (Cr4) for resistance to white pine blister rust using transcriptome-based SNP genotyping. *BMC Genomics* 17:753.
- Lotterhos KE, Hodges K, Yeaman S, Degner J, Aitken S (2017) Modular environmental pleiotropy of genes involved in local adaptation to climate despite physical linkage. *bioRxiv*
- Lotterhos KE, Whitlock MC (2014) Evaluation of demographic history and neutral parameterization on the performance of F_{ST} outlier tests. *Molecular Ecology* 23:2178–2192.
- Lotterhos KE, Whitlock MC (2015) The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. *Molecular Ecology* 24:1031–1046.
- Lowe AJ, Cavers S, Boshier D, Breed MF, Hollingsworth PM (2015) The resilience of forest fragmentation genetics - no longer a paradox - we were just looking in the wrong place. *Heredity* 115:97–99.
- Lowry DB, Hoban S, Kelley JL, Lotterhos KE, Reed LK, Antolin MF, Storfer A (2016) Breaking RAD: an evaluation of the utility of restriction site-associated DNA sequencing for genome scans of adaptation. *Molecular Ecology Resour* 17:142–152.
- Lowry DB, Hoban S, Kelley JL, Lotterhos KE, Reed LK, Antolin MF, Storfer A (2017) Responsible RAD: Striving for best practices in population genomic studies of adaptation. *Molecular Ecology Resources* 17:366–369.
- Lutz JA, van Wagtendonk JW, Franklin JF (2010) Climatic water deficit, tree species ranges, and climate change in Yosemite National Park. *Journal of Biogeography* 37:936-950.
- Luquez V, Hall D, Albrechtsen BR, Karlsson J, Ingvarsson P, Jansson S (2007) Natural phenological variation in aspen (*Populus tremula*): the SwAsp collection. *Tree Genetics & Genomes*, 4, 279–292.
- Lynch M, Walsh B (1998) *Genetics and analysis of quantitative traits*. Sinauer Sunderland, MA
- Ma XF, Hall D, Onge KRS, Jansson S, Ingvarsson PK (2010) Genetic differentiation, clinal variation and phenotypic associations with growth cessation across the *Populus tremula* photoperiodic pathway. *Genetics*, 186, 1033–1044.
- MacArthur RH (1984) *Geographical Ecology: Patterns in the Distribution of Species*. Princeton University Press, Princeton, New Jersey, USA.
- Macdonald SE, Thomas BR, Cherniawsky DM, Purdy BG (2001) Managing genetic resources of lodgepole pine in west-central Alberta: patterns of isozyme variation in natural populations and effects of forest management. *Forest Ecology and Management* 152:45–58.

- Mackay TFC (2001) The genetic architecture of quantitative traits. *Annual Review of Genetics* 35:303–339.
- Mackay TFC (2014) Epistasis and quantitative traits: using model organisms to study gene-gene interactions. *Nature Reviews Genetics* 15:22–33. doi:
- Mackay J, Dean JFD, Plomion C, Peterson DG, Canovas FM, Pavy N, Ingvarsson PK, Savolainen O, Guevara MA, Fluch S, Vinceti B, Abarca D, Diaz-Sala C, Cervera MT (2012) Towards decoding the conifer giga-genome. *Plant Molecular Biology* 80:555–569
- Mackay TFC, Stone EA, Ayroles JF (2009) The genetics of quantitative traits: challenges and prospects. *Nature Reviews Genetics* 10:565–577.
- MacPherson A, Hohenlohe PA, Nuismer SL (2015) Trait dimensionality explains widespread variation in local adaptation. *Proceedings of the Royal Society of London B: Biological Sciences* 282: 20141570.
- Mahalovich MF, Hipkins VD (2011) Molecular genetic variation in whitebark pine (*Pinus albicaulis* Engelm.) in the Inland West. In: Keane, Robert E.; Tomback, Diana F.; Murray, Michael P.; and Smith, Cyndi M., eds. 2011. The future of high-elevation, five-needle white pines in Western North America: Proceedings of the High Five Symposium. 28-30 June 2010; Missoula, MT. Proceedings RMRS-P-63. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 376 p.
- Mahalovich MF, Stritch L (2013) *Pinus albicaulis*. The IUCN Red List of Threatened Species. e.T39049A2885918.
- Mähler N, Wang J, Terebieniec BK, Ingvarsson PK, Street NR, Hvidsten TR (2017) Gene co-expression network connectivity is an important determinant of selective constraint. *PLoS Genet* 13:e1006402.
- Mahler DL, Weber MG, Wagner CE, Ingram T (2017) Pattern and Process in the Comparative Study of Convergent Evolution. *American Naturalist* 190:S13–S38.
- Maloney PE, Vogler DR, Eckert AJ, Jensen CE, Delfino-Mix A (in review) Ecological genetics of three white pine species from the Lake Tahoe Basin, USA: Implications for conservation and evolutionary potential. 1–67.
- Maloney PE, Vogler DR, Eckert AJ, Jensen CE, Delfino-Mix A (2012) Ecology of whitebark pine in relation to white pine-blister rust infection in subalpine forests of the Lake Tahoe Basin, USA: Implications for restoration. *Forest Ecology and Management*, 280, 166-175.
- Mäki-Tanila A, Hill WG (2014) Influence of gene interaction on complex trait variation with multilocus models. *Genetics* 198:355–367.
- Mantel N (1967) The detection of disease clustering and a generalized regression approach. *Cancer Research*, 27, 209–220.

- Marguerit E, Bouffier L, Chancerel E, Costa P, Lagane F, Guehl J-M, Plomion C, Brendel O. The genetics of water-use efficiency and its relation to growth in maritime pine. *Journal of Experimental Botany* 65:4757-4768.
- Marquardt PE, Echt CS, Epperson BK, Pubanz DM (2007) Genetic structure, diversity, and inbreeding of eastern white pine under different management conditions. *Canadian Journal of Forest Research* 37:2652–2662.
- Martin G, Lenormand T (2006) A general multivariate extension of Fisher's geometrical model and the distribution of mutation fitness effects across species. *Evolution* 60:893–16.
- Martin G, Lenormand T (2008) The distribution of beneficial and fixed mutation fitness effects close to an optimum. *Genetics* 179:907–916.
- Martínez-García PJ, Stevens KA, Wegrzyn JL, Liechty J, Crepeau M, Langley CH, Neale DB (2013) Combination of multipoint maximum likelihood (MML) and regression mapping algorithms to construct a high-density genetic linkage map for loblolly pine (*Pinus taeda* L.) *Tree Genet Genomes* 9:1529–1535
- Mastrogiuseppe RJ, Mastrogiuseppe JD (1980) A study of *Pinus balfouriana* Grev. & Balf. (Pinaceae). *Systematic Botany* 5:86–104
- Mather K (1941) Variation and selection of polygenic characters. *J Genetics* 41:159-193.
- Matuszewski S, Hermisson J, Kopp M (2015) Catch me if you can: Adaptation from standing genetic variation to a moving phenotypic optimum. *Genetics* 200:1255–1274.
- Matuszewski S, Hermisson J, Kopp M (2014) Fisher's Geometric Model with a moving optimum. *Evolution* 68:2571–2588.
- Mátyás C (1996) Climatic adaptation of trees: rediscovering provenance tests. *Euphytica* 92:45–54.
- Mauricio R (2001) Mapping quantitative trait loci in plants: uses and caveats for evolutionary biology. *Nature Reviews Genetics* 2:370–381
- Maynard Smith JH, Haigh J (1974) The hitch-hiking effect of a favourable gene. *Genetical Research* 23:23–35.
- Mayr E (1963) *Animal Species and Evolution*. Belknap Press, Cambridge, Massachusetts, USA.
- McCandlish DM, Stoltzfus A (2014) Modeling evolution using the probability of fixation: history and implications. *The Quarterly Review of Biology* 89:225–252.
- McCormack JE, Zellmer AJ, Knowles LL (2010) Does niche divergence accompany allopatric divergence in *Aphelocoma* jays as predicted under ecological speciation?: Insights from tests with niche models. *Evolution* 64:1231-1244.

- McCune B (1988) Ecological diversity in North American pines. *American Journal of Botany* 75:353-368
- McKay JK, Latta RG (2002) Adaptive population divergence: markers, QTL and traits. *Trends Ecol Evol* 17:285–291.
- McKinney GJ, Larson WA, Seeb LW, Seeb JE (2017) RADseq provides unprecedented insights into molecular ecology and evolutionary genetics: comment on Breaking RAD by Lowry et al. (2016). *Molecular Ecology Resources* 17:356–361.
- McLachlan JS, Clark JS, Manos PS (2005) Molecular indicators of tree migration capacity under rapid climate change. *Ecology*, 86, 2088-2098.
- McLachlan JS, Hellman JJ, Schwartz MW (2007) A framework for debate of assisted migration in an era of climate change. *Conservation Biology*, 21, 297-302.
- McLane SC, Aitken SN (2012) Whitebark pine (*Pinus albicaulis*) assisted migration potential: testing establishment north of the species range. *Ecological Applications*, 22, 142–153.
- McVean, G (2007) The structure of linkage disequilibrium around a selective sweep. *Genetics*, 175, 1395-1406.
- Mei W, Stetter MG, Gates DJ, Stitzer MC, Ross-Ibarra J (2017) Adaptation in plant genomes: bigger is different. *bioRxiv*
- Meier JI, Sousa VC, Marques DA, Selz OM, Wagner CE, Excoffier L, Seehausen O (2017) Demographic modelling with whole-genome data reveals parallel origin of similar *Pundamilia* cichlid species after hybridization. *Molecular Ecology* 26:123–141.
- Messer PW, Petrov DA (2013) Population genomics of rapid adaptation by soft selective sweeps. *Trends in Ecology & Evolution* 28:659–669.
- Meuwissen TH, Hayes BJ, Goddard ME. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157: 1819–1829.
- Millar C, Graumlich LJ, Delany DL, Westfall RD, King JC (2002) Response of high-elevation conifers. In: *Proceedings of the Eighteenth Annual Pacific Climate Workshop* (p. 57).
- Miller JD, Safford HD, Crimmins M, Thode AE (2009) Quantitative Evidence for Increasing Forest Fire Severity in the Sierra Nevada and Southern Cascade Mountains, California and Nevada, USA. *Ecosystems* 12:16–32.
- Mitton JB, Grant MC, Yoshino AM (1998) Variation in allozymes and stomatal size in pinyon (*Pinus edulis*, Pinaceae), associated with soil moisture. *American Journal of Botany* 85:1262–1265.
- Mitton JB, Stutz HP, Schuster WS (1989) Genotypic differentiation at PGM in Engelmann spruce from wet and dry sites. *Silvae Genetica*, 38, 217–221.

- Mitton JB, Williams CG (2006) Gene Flow in Conifers. In: Landscapes, Genomics, and Transgenic Conifers. (ed Williams CG), pp. 147–168. Springer Netherlands, Dordrecht.
- Mizrachi E, Verbeke L, Christie N, Fierro AC, Mansfield SD, Davis MF, Gjersing E, Tuskan GA, Montagu MV, de Peer YV, Marchal K, Myburg AA (2017) Network-based integration of systems genetics data reveals pathways associated with lignocellulosic biomass accumulation and processing. *Proceedings of the National Academy of Sciences* 114:1195–1200.
- Moran EV, Clark JS (2011) Estimating seed and pollen movement in a monoecious plant: a hierarchical Bayesian approach integrating genetic and ecological data. *Molecular Ecology* 20:1248–1262.
- Moreno G (1994) Genetic architecture, genetic behavior, and character evolution. *Annual Review of Ecology and Systematics* 25:31–44.
- Morgenstern EK (1996) Geographic variation in forest trees: genetic basis and application of knowledge in silviculture. UBC press.
- Moritsuka E, Histaka Y, Tamura M, Uchiyama K, A W Tsmura Y, Tachida H (2012) Extended linkage disequilibrium in non- coding regions in a conifer, *Cryptomeria japonica*. *Genetics* 190:1145–1148
- Morris RW, Spieth PT (1978) Sampling strategies for using female gametophytes to estimate heterozygosity in conifers. *Theoretical and Applied Genetics* 51:217–222
- Morse AM, Peterson DG, Islam-Faridi MN, et al. (2009) Evolution of genome size and complexity in *Pinus*. *PLoS ONE* 4:e4332.
- Morton NE (1955) Sequential tests for the detection of linkage. *American Journal of Human Genetics* 7:277–318
- Moser G, Lee SH, Hayes BJ, Goddard ME, Wray NR, Visscher PM (2015) Simultaneous discovery, estimation and prediction analysis of complex traits using a Bayesian mixture model. *PLoS Genetics*, 11, e1004969.
- Murray BG (1998) Nuclear DNA amounts in gymnosperms. *Annals of Botany* 82:3–15
- Mouillot D, Graham NAJ, Villéger S, Mason NWH, Bellwood DR (2013) A functional approach reveals community responses to disturbances. *Trends in Ecology & Evolution* 28:167–177.
- Namkoong G (1979) Introduction to Quantitative Genetics in Forestry. Technical Bulletin No. 1588. Washington, D. C. USDA Forest Service. 342 pp.
- Namroud M-C, Bousquet J, Doerksen T, Beaulieu J (2012) Scanning SNPs from a large set of expressed genes to assess the impact of artificial selection on the undomesticated genetic diversity of white spruce. *Evolutionary Applications* 5:641–656.

- Neale DB (1983) Population genetic structure in Douglas-fir shelterwood regeneration system in southwest Oregon. PhD dissertation. Oregon State University.
- Neale DB, Kremer A. (2011). Forest tree genomics: growing resources and applications. *Nature Reviews Genetics* 12:111–122.
- Neale DB, Langley CH, Salzberg SL, Wegrzyn JL (2013) Open access to tree genomes: the path to a better forest. *Genome Biology* 14:6 14:120.
- Neale DB, Martínez-García PJ, La Torre De AR, Montanari S, Wei X-X (2017) Novel in-sights into tree biology and genome evolution as revealed through genomics. *Annual Review of Plant Biology* 68:457–483.
- Neale DB, Savolainen O (2004) Association genetics of complex traits in conifers. *Trends in Plant Science* 9:325–330.
- Neale DB, Wegrzyn JL, Stevens KA, Zimin AV, Puiu D et al (2014) Decoding the massive genome of loblolly pine using haploid DNA and novel assembly strategies. *Genome Biology* 15:R59
- Nei M (1987) *Molecular evolutionary genetics*. Columbia University Press
- Nelson CD, Nance WL, Doudrick RL (1993) A partial genetic linkage map of slash pine (*Pinus elliottii* Engelm. var. *elliottii*) based on random amplified polymorphic DNA. *Theoretical and Applied Genetics* 87:145– 151
- Nelson RM, Pettersson ME, Carlborg Ö (2013) A century after Fisher: time for a new paradigm in quantitative genetics. *Trends in Genetics* 29:669–676.
- Neves LG, Davis JM, Barbazuk WB, M K (2014) A high-density gene map of loblolly pine (*Pinus taeda* L.) based on exome sequence capture genotyping. *G3-Genes Genomes Genomics* 4:29–37
- NIFC (2014) National Interagency Coordination Center Incident Management Situation Report - October 2014. National Interagency Fire Center, 1–7.
- NIFC (2015) National Interagency Coordination Center Incident Management Situation Report - September 2015. National Interagency Fire Center, 1–9.
- Nilsson-Ehle H (1909) Kreuzungsuntersuchungen an Hafer und Weizen. *Lunds Universitets Arsskrift* 5:1–122.
- North MP (2002) The Teakettle Experiment. *Proceedings of a Symposium on the Kings River Sustainable Forest Ecosystems Project: progress and current status: January 26, 1998, Clovis, California*, 47.
- North MP, Hurteau MD, Fiegenger R, Barbour M (2005) Influence of fire and El Nino on tree recruitment varies by species in Sierran mixed conifer. *Forest Science* 51:187–197.

- North MP, Stephens SL, Collins BM et al. (2015) Reform forest fire management. *Science* 349:1280–1281.
- Nowak CL, Nowak RS, Tausch RJ, Wigand PE (1994) A 30000 year record of vegetation dynamics at a semi-arid locale in the Great Basin. *Journal of Vegetation Science*, 5, 579-590.
- Nystedt B, Street NR, Wetterbom A, Zuccolo A, Lin Y-C, Scofield DG, Vezzi F, Delhomme N, Giacomello S, Alexeyenko A, Vicedomini R, Sahlin K, Sherwood E, Elfstrand M, Gramzow L, Holmberg K, Hällman J, Keech O, Klasson L, Koriabine M, Kucukoglu M, Källner, Luthman J, Lysholm F, Nittylä T, Olson Å, Rilakovic N, Ritland C, Rosselló, Sena J, Svensson T, Talavera-López C, Theißen G, Tuominen H, Vanneste K, Wu Z-Q, Zhang B, Zerbe P, Arvestad L, Bhalerao R, Bohlmann J, Bousquet J, Gil RG, Hvidsten TR, de Jong P, MacKay J, Morgante M, Ritland K, Sundberg B, Thompson SL, de Peer YV, Andersson B, Nilsson O, Ingvarsson PK, Lundeberg J, Jansson S (2013) The Norway spruce genome sequence and conifer genome evolution. *Nature* 497:570– 584.
- O’Connell LM, Russell J, Ritland K (2004) Fine-scale estimation of outcrossing in western redcedar with microsatellite assay of bulked DNA. *Heredity* 93:443–449.
- Oddou-Muratorio S, Demesure-Musch B, Pelissier R (2004) Impacts of gene flow and logging history on the local genetic structure of a scattered tree species, *Sorbus torminalis* L. Crantz. *Molecular Ecology* 13(2), 3689–3702.
- Oddou-Muratorio S, Klein EK, Vendramin GG, Fady B (2011) Spatial vs. temporal effects on demographic and genetic structures: the roles of dispersal, masting and differential mortality on patterns of recruitment in *Fagus sylvatica*. *Molecular Ecology* 20:1997–2010.
- Ohta T (1982) Linkage disequilibrium with the island model. *Genetics* 101:139–155.
- Ohta, T. 1992. The nearly neutral theory of molecular evolution. *Annual Review of Ecology and Systematics* 23:263-286.
- Ohta, T. 1996. The current significance and standing of neutral and nearly neutral theories. *BioEssays* 18: 673-684.
- Oldfield, S., Lusty, C., & MacKinven, A. (1998). *The world list of threatened trees*. World Conservation Press.
- Oline DK, Mitton JB, Grant MC (2000) Population and subspecific genetic differentiation in the foxtail pine (*Pinus balfouriana*). *Evolution* 54:1813–1819
- Omduff R (1974) *Introduction to California Plant Life*. University of California Press, Berkeley, California, USA.
- Orr HA (1998) The population genetics of adaptation: the distribution of factors fixed during adaptive evolution. *Evolution* 52:935.

- Orr HA (2005) The genetic theory of adaptation: a brief history. *Nature Reviews Genetics* 6:119–127.
- Orr HA (2003) The Distribution of Fitness Effects Among Beneficial Mutations. *Genetics* 163:1519–1526.
- Orr HA (2006) The distribution of fitness effects among beneficial mutations in Fisher's geometric model of adaptation. *Journal of Theoretical Biology* 238:279–285.
- Orr HA (2001) The “sizes” of mutations fixed in phenotypic evolution: a response to Clarke and Arthur. *Evolution & Development* 3:121–123.
- Orr HA (2000) Adaptation and the cost of complexity. *Evolution* 54:13–20.
- Ortiz-Barrientos D, Engelstädter J, Rieseberg LH (2016) Recombination Rate Evolution and the Origin of Species. *Trends in Ecology & Evolution* 31:226–236.
- Ovaskainen O, Karhunen M, Zheng CH, Cano Arias JM, Merilä J (2011) A new method to uncover signatures of divergent and stabilizing selection in quantitative traits. *Genetics* 189:621–632.
- Paaby AB, Rockman MV (2013) The many faces of pleiotropy. *Trends in Genetics* 29:66–73.
- Paape T, Bataillon T, Zhou P, Kono T, Briskine R, et al. 2013. Selection, genome-wide fitness effects and evolutionary rates in the model legume *Medicago truncatula*. *Mol. Ecol.* 22:3525–38
- Paixão T, Barton NH (2016) The effect of gene interactions on the long-term response to selection. *Proceedings of the National Academy of Sciences* 113:4422–4427.
- Palmé AE, Pyhajarvi T, Wachowiak W, Savolainen O (2009) Selection on nuclear genes in a *Pinus* phylogeny. *Molecular Biology and Evolution* 26:893–905.
- Pamilo P, Nei M (1988) Relationships between gene trees and species trees. *Molecular Biology and Evolution* 5:568-583.
- Pannell JR, Fields PD (2013) Evolution in subdivided plant populations: concepts, recent advances and future directions. *New Phytologist* 201:417–432
- Parchman TL, Gompert Z, Mudge J, Schilkey FD, Benkman CW, Buerkle CA (2012) Genome-wide association genetics of an adaptive trait in lodgepole pine. *Molecular Ecology* 21: 2991–3005.
- Parchman TL, Jahner JP, Uckele K, Galland LM (forthcoming) RADseq approaches and applications for forest tree genetics.
- Patterson N, Price AL, Reich D (2006) Population structure and eigenanalysis. *PLoS Genetics*, 2, e190.

- Patterson HD, Thompson R (1971) Recovery of interblock information when block sizes are unequal. *Biometrika* 58:545-554.
- Pavlidis P, Metzler D, Stephan W (2012) Selective sweeps in multilocus models of quantitative traits. *Genetics* 192:225–239.
- Pavy N, Pelgas B, Laroche J, Rigault P, Isabel N, Bousquet J (2012) A spruce gene map infers ancient plant genome reshuffling and subsequent slow evolution in the gymnosperm lineage leading to extant conifers. *BMC Biology* 10:84
- Pearson RG, Dawson TP (2003) Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? *Global Ecol Biogeogr* 12:361-371.
- Pelgas B, Bousquet J, Merimans PG, Ritland K, Isabel N (2011) QTL mapping in white spruce: gene maps and genomic regions underlying adaptive traits across pedigrees, years and environments. *BMC Genomics* 12:145.
- Pennings PS, Hermisson J (2006a) Soft sweeps III: the signature of positive selection from recurrent mutation. *PLoS Genetics* 2:e186.
- Pennings PS, Hermisson J (2006b) Soft Sweeps II--Molecular Population Genetics of Adaptation from Recurrent Mutation or Migration. *Molecular Biology and Evolution* 23:1076–1084.
- Pérez F, Granger BE (2007) IPython: A system for interactive scientific computing. *Computing in Science and Engineering* 9: 21-29.
- Perry DJ, Bousquet J (2001) Genetic diversity and mating system of post-fire and post-harvest black spruce: an investigation using codominant sequence-tagged-site (STS) markers. *Canadian Journal of Forest Research* 31:32–40.
- Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS ONE* 7: e37
- Petit RJ, Hampe A (2006) Some evolutionary consequences of being a tree. *Annual Review of Ecology* 37:187–214.
- Pfender WF, Saha MC, Johnson EA, Slabaugh MB (2011) Mapping with RAD (restriction-site associated DNA) markers to rapidly identify QTL for stem rust resistance in *Lolium perenne*. *Theoretical and Applied Genetics* 122:1467–1480
- Phillips PC (2008) Epistasis—the essential role of gene interactions in the structure and evolution of genetic systems. *Nature Reviews Genetics* 9:855–867.
- Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. *Ecol Model* 190:231-259.

- Piepho HP, Möhring J, Melchinger AE, Büchse A (2008) BLUP for phenotypic selection in plant breeding and variety testing. *Euphytica* 161:209-228.
- Pickrell JK, Berisa T, Liu JZ, Ségurel L (2016) Detection and interpretation of shared genetic influences on 42 human traits. *Nature Genetics* 48:709–717.
- Platt A, Vilhjalmsón BJ, Nordborg M (2010) Conditions under which genome-wide association studies will be positively misleading. *Genetics* 186:1045-1052.
- Plomion C, Bastien C, Bogeat-Triboulot M-B, Bouffier L, Déjardin A, Duplessis S, Fady B, Geuert M, Le Gac A-L, Le Provost G, Legué V, Lelu-Walter M-A, Leplé J-C, Maury S, Morel A, Oddou-Muratorio S, Pilate G, Sanchez L, Scotti I, Scotti-Saintagne C, Segura V, T J-F, Vacher C (2016) Forest tree genomics: 10 achievements from the past 10 years and future prospects. *Annals of Forest Science* 73:77–103.
- Postma FM, Ågren J (2016) Early life stages contribute strongly to local adaptation in *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences* 113:7590–7595.
- Prasolova NV, Xu ZH, Farquhar GD, Saffigna PG, Dieters MJ (2000) Variation in canopy $\delta^{13}\text{C}$ of 8-year-old hoop pine families (*Araucaria cunninghamii*) in relation to canopy nitrogen concentration and tree growth in subtropical Australia. *Tree Physiology* 20:1049-1055.
- Prasolova NV, Lundkvist K, Xu ZH (2005) Genetic variation in foliar nutrient concentration in relation to foliar carbon isotope composition and tree growth with clones of the F1 hybrid between slash pine and Caribbean pine. *Forest Ecology and Management* 210:173–191.
- Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics* 38:904-909.
- Pritchard JK, Di Rienzo A (2010) Adaptation – not by sweeps alone. *Nature Reviews Genetics*, 11, 665–667.
- Prout T, Barker JSF (1993) *F* statistics in *Drosophila buzzatii*: selection, population size and inbreeding. *Genetics* **134**: 369–375.
- Prunier J, Laroche J, Beaulieu J, Bousquet J (2011) Scanning the genome for gene SNPs related to climate adaptation and estimating selection at the molecular level in boreal black spruce. *Molecular Ecology* 20:1702–1716.
- Prunier J, Verta J-P, MacKay JJ (2015) Conifer genomics and adaptation: at the crossroads of genetic diversity and genome function. *New Phytologist* 209:44–62.
- Pulliam HR (2000) On the relationship between niche and distribution. *Ecol Lett* 3:349-361.
- Quesada T, Li Z, Dervinis C, Bock P, Tuskan GA, Casella G, Davis JM, Kirst M (2008) Comparative analysis of the transcriptomes of *Populus trichocarpa* and *Arabidopsis thaliana*

- suggests extensive evolution of gene expression regulation in angiosperms. *New Phytologist* 180:408–420.
- R Core Team (2015) R: A language and environment for statistical computing R Foundation for Statistical Computing, Vienna, Austria URL <http://www.R-project.org/>
- R Core Team (2017) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Rajora OP, Pluhar SA (2003) Genetic diversity impacts of forest fires, forest harvesting, and alternative reforestation practices in black spruce (*Picea mariana*). *Theoretical and Applied Genetics* 106:1203–1212.
- Rajora OP, Rahman MH, Buchert GP, Dancik BP (2000) Microsatellite DNA analysis of genetic effects of harvesting in old-growth eastern white pine (*Pinus strobus*) in Ontario, Canada. *Molecular Ecology* 9:339–348.
- Ralph P, Coop G (2010) Parallel adaptation: one or many waves of advance of an advantageous allele? *Genetics* 186:647–668.
- Rellstab C, Gugerli F, Eckert AJ, Hancock AM, Holdregger R (2015) A practical guide to environmental association analysis in landscape genomics. *Molecular Ecology* 24:4348–4370.
- Remington DL (2015) Alleles versus mutations: Understanding the evolution of genetic architecture requires a molecular perspective on allelic origins. *Evolution* 69:3025–3038.
- Resende MFR Jr, Muñoz P, Acosta JJ, Peter GF, Davis JM, Grattapaglia D, Resende MDV, Kirst M (2012a) Accelerating the domestication of trees using genomic selection: accuracy of prediction models across ages and environments. *New Phytologist* 193:617–624.
- Resende MDV, Resende MFR Jr, Sansaloni CP, Petrolí CD, Missiaggia AA, Aguiar AM, Abad JM, Takahashi EK, Rosado AM, Faria DA, Pappas GJ Jr, Kilian A, Grattapaglia D (2012b) Genomic selection for growth and wood quality in Eucalyptus: capturing the missing heritability and accelerating breeding for complex traits in forest trees. *New Phytologist* 194:116–128.
- Resende MFR, Muñoz P, Resende MDV, Garrick DJ, Fernando RL, Davis JM, Jokela EJ, Martin TA, Peter GF, Kirst M (2012c) Accuracy of genomic selection methods in a standard data set of loblolly pine (*Pinus taeda* L.) *Genetics* 190:1503–1510.
- Richardson BA, Brunfeldt SJ, Klopfenstein NB (2002) DNA from bird-dispersed seed and wind-disseminated pollen provides insights into postglacial colonization and population genetic structure of whitebark pine (*Pinus albicaulis*). *Molecular Ecology* 11:215–227.
- Richardson JL, Urban MC, Bolnick DI, Skelly DK (2014) Microgeographic adaptation and the spatial scale of evolution. *Trends in Ecology & Evolution*, 29, 165–176.

- Riesch R, Muschick M, Lindtke D, Villoutreix R, Comeault AA, Farkas TE, Lucek K, Hellen E, Soria-Carrasco V, Dennis SR, de Carvalho CF, Safran RJ, Sandoval CP, Feder J, Gries R, Crespi BJ, Gries G, Gompert Z, Nosil P (2017) Transitions between phases of genomic differentiation during stick-insect speciation. *Nat Ecol Evol* 1:0082.
- Ritland K, Krutovsky KV, Tsumura Y, Pelgas B, Isabel N, and Bousquet, J (2011) Genetic mapping in conifers. In: *Genetics, genomics and breeding of conifers*, pp.196-238.
- Ritland K, Ritland C (1996) Inferences about quantitative inheritance based on natural population structure in the yellow monkeyflower, *Mimulus guttatus*. *Evolution* 50:1074–1082.
- Robledo-Amuncio JJ, Alía R, Gil L (2004) Increased selfing and correlated paternity in a small population of a predominantly outcrossing conifer, *Pinus sylvestris*. *Molecular Ecology* 13:2567–2577.
- Rockman, M. V. (2012, January). The QTN program and the alleles that matter for evolution: All that's gold does not glitter. *Evolution* 66:1-17.
- Rodríguez-Quilón I, Santos-del-Blanco L, Serra-Varela MJ, Koskela J, González-Martínez SC, Alía R (2016) Capturing neutral and adaptive genetic diversity for conservation in a highly structured tree species. *Ecological Applications* 26:2254–2266.
- Romero IG, Ruvinsky I, Gilad Y (2012) Comparative studies of gene expression and the evolution of gene regulation. *Nature Reviews Genetics* 13:505–516.
- Roschanski AM, Csilléry K, Liepelt S, Oddou-Muratorio S, Ziegenhagen B, Huard F, Ullrich KK, Postolache D, Vendramin GG, Fady B (2016) Evidence of divergent selection at landscape and local scales in *Abies alba* Mill. in the French Mediterranean Alps. *Molecular Ecology* 25:776–794.
- Rousseeuw PJ (1987) Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics* 20:53–65
- Royce EB, Barbour MG (2000) Mediterranean climate effects. II. Conifer growth phenology across a Sierra Nevada ecotone. *American Journal of Botany* 88:919-932.
- Safford HD, Schmidt DA, Carlson CH (2009) Effects of fuel treatments on fire severity in an area of wildland–urban interface, Angora Fire, Lake Tahoe Basin, California. *Forest Ecology and Management* 258:773–787.
- Savolainen O, Lascoux M, Merilä J (2013) Ecological genomics of local adaptation. *Nature Reviews Genetics* 14:807–820.
- Savolainen O, Pyhäjärvi T, Knürr T (2007) Gene Flow and Local Adaptation in Trees. *Annual Review of Ecology, Evolution, and Systematics* 38:595–619.

- Savolainen O, Pyhäjärvi T (2009) Genomic Diversity in Forest Trees. *Current Opinion in Plant Biology* 10: 162-167.
- Scheet P, Stephens M (2006) A fast and flexible statistical model for large-scale population genotype data: applications to infer- ring missing genotypes and haplotypic phase. *American Journal of Human Genetics* 78:629–644
- Schoen DJ, Brown AHD (1991) Intraspecific variation in population gene diversity and effective population size correlates with the mating system in plants. *Proceedings of the National Academy of Sciences* 88:4494–4497.
- Schoville SD, Bonin A, Francois O, et al. (2012) Adaptive Genetic Variation on the Landscape: Methods and Cases. *Annual Review of Ecology, Evolution, and Systematics* 43:23–43.
- Schrider DR, Kern AD (2016) S/HIC: Robust Identification of Soft and Hard Sweeps Using Machine Learning. *PLoS Genetics* 12:e1005928–31.
- Schrider DR, Mendes FK, Hahn MW, Kern AD (2015) Soft shoulders ahead: spurious signatures of soft and partial selective sweeps result from linked hard sweeps. *Genetics* 200:267–284.
- Schrider DR, Shanku AG, Kern AD (2016) Effects of linked selective sweeps on demographic inference and model selection. *Genetics* 204:1207–1223.
- Schwilk DW, Keeley JE, Knapp EE et al. (2009) The National Fire and Fire Surrogate study: effects of fuel reduction methods on forest vegetation structure and fuels. *Ecological applications* 19:285–304.
- Scotti I, Gugerli F, Pastorelli R, Sebastiani F, Vendramin GG (2008) Maternally and paternally inherited molecular markers elucidate population patterns and inferred dispersal processes on a small scale within a subalpine stand of Norway spruce (*Picea abies* [L.] Karst.). *Forest ecology and management* 255:3806-3812.
- Searle SR, Casella G, McCulloch CE (1992) *Variance Components*. New York, NY: John Wiley & Sons, Inc. 528 pp.
- Seaton G, Hernandez J, Grunchev JA, White I., Allen J, De Koning DJ, Wei W, Berry D, Haley C, Knott S (2006) GridQTL: A Grid Portal for QTL Mapping of Compute Intensive Datasets. *Proceedings of the 8th World Congress on Genetics Applied to Livestock Production*, August 13-18, 2006. Belo Horizonte, Brazil.
- Segurado P, Araujo MB (2004) An evaluation of methods for modelling species distributions. *Journal of Biogeography* 31:1555-1568.
- Seibt U, Rajabi A, Griffiths H, Berry JA (2008) Carbon isotopes and water use efficiency: sense and sensibility. *Oecologia* 155:441-454.

- Seiler JR, Johnsen JD (1988) Physiological and morphological response of three half-sib families of loblolly pine to water-stress conditioning. *Forest Science* 34:487-495.
- Sewell MM, Bassoni DL, Megraw RA, Wheeler NC, Neale DB (2000) Identification of QTLs influencing wood property traits in loblolly pine (*Pinus taeda* L.). I. Physical wood properties. *Theoretical and Applied Genetetics* 101:1273-1281.
- Sheth SN, Angert AL (2014) The evolution of environmental tolerance and range size: a comparison of geographically restricted and widespread *Mimulus*. *Evolution* 68:2917- 2931.
- Shull GH (1908) The composition of a field of maize. *American Breeders Association Report* 5:51-59.
- Sierra Nevada Ecosystem Project - SNEP (1996) Final Report to Congress: vols.1–3. University of California, Centers for Water and Wildland Resources, Davis, CA.
- Sierra Nevada Forest Plan Ammendment – SNFPA (2004) Final Environmental Impact Statement: vols.1–6. USDA Forest Service, Pacific Southwest Region, Vallejo, CA.
- Silva-Junior O, Faria DA, Grattapaglia D (2015) A flexible multi-species genome-wide 60K SNP chip developed from pooled resequencing of 240 Eucalyptus tree genomes across 12 species. *New Phytologist* 206:1527-1540.
- Simons YB, Bullaughey K, Hudson RR, Sella G (2017) A model for the genetic architecture of quantitative traits under stabilizing selection. *arXiv* 1–76.
- Siol M, Wright S, Barrett S (2010) The population genomics of plant adaptation. *New Phytologist* 188: 313-332.
- Slate J (2005) Quantitative trait locus mapping in natural populations: progress, caveats and future directions. *Molecular Ecology* 14:363–379.
- Slatkin M (1987) Gene flow and the geographic structure of natural populations. *Science* 236:787–793.
- Slatkin M (1975) Gene flow and selection in a two-locus system. *Genetics* 81:787–802.
- Slavov GT, DiFazio SP, Strauss SH (2004) Gene flow in forest trees: gene migration patterns and landscape modeling of transgene dispersal in hybrid poplar. In: *Introgression from Genetically Modified Plants into Wild Relatives*, ed. HCM den Nijs, D Bartsch, J Sweet, pp. 89–106. Wallingford, UK: CAB Int.
- Slotte T, Foxe JP, Hazzouri KM, Wright SI. 2010. Genome-wide evidence for efficient positive and purifying selection in *Capsella grandiflora*, a plant species with a large effective population size. *Molecular Biology and Evolution* 27:1813–21.
- Smith SD (2016) Pleiotropy and the evolution of floral integration. *New Phytologist* 209:80–85.

- Smouse PE, Dyer RJ, Westfall RD, Sork VL (2001) Two-generation analysis of pollen flow across a landscape. I. Male gamete heterogeneity among females. *Evolution* 55:260–271.
- Smouse PE, Peakall R (1999) Spatial autocorrelation analysis of individual multiallele and multilocus genetic structure. *Heredity* 82:561–573.
- Soberon J, Peterson AT (2005) Interpretation of models of fundamental ecological niches and species' distributional areas. *Biodiversity Informatics* 2:1-10.
- Sorenson FC (1983) Geographic variation in seedling Douglas-fir (*Pseudotsuga menziesii*) from the western Siskiyou Mountains of Oregon. *Ecology* 64:696-702.
- Sork VL, Aitken SN, Dyer RJ, Eckert AJ, Legendre P, Neale DB (2013) Putting the landscape into the genomics of trees: approaches for understanding local adaptation and population responses to changing climate. *Tree Genet Genomes* 9:901–911.
- Sork VL, Squire K, Gugger PF, Steele SE, Levy ED, Eckert AJ (2016) Landscape genomic analysis of candidate genes for climate adaptation in a California endemic oak, *Quercus lobata*. *American Journal of Botany*. 103:33-46.
- Spearman C (1904) The proof and measurement of association between two things. *American Journal of Psychology* 15:72–101
- Speed D, Balding DJ (2014) MultiBLUP: improved SNP-based prediction for complex traits. *Genome Research* 24:1550–1557.
- Spitze K (1993) Population structure in *Daphnia obtusa*: quantitative genetic and allozyme variation. *Genetics* 135:367-374.
- Spencer CCA, Su Z, Donnelly P, Marchini J (2009) Designing genome-wide association studies: sample size, power, imputation, and the choice of genotyping chip. *PLoS Genetics* 5:e1000477.
- St Clair JB, Mandel NL, Vance-Borland KW (2005) Genecology of Douglas fir in western Oregon and Washington. *Annals of Botany* 96:1199–1214.
- Stephan W (2010) Genetic hitchhiking versus background selection: the controversy and its implications. *Philosophical Transactions of the Royal Society B: Biological Sciences* 365:1245–1253.
- Stephan W (2015) Signatures of positive selection: from selective sweeps at individual loci to subtle allele frequency changes in polygenic adaptation. *Molecular Ecology* 25:79–88.
- Stephens M, Balding DJ (2009) Bayesian statistical methods for genetic association studies. *Nature Reviews Genetics* 10:681-690.

- Stephenson NL (1990) Climatic control of vegetation distribution - the role of the water balance. *The American Naturalist* 135:649-670.
- Sterck L, Rombauts S, Jansson S, Sterky F, Rouzé P, Van de Peer Y. 2005. EST data suggest that poplar is an ancient polyploid. *New Phytologist* 167(1):165–70
- Stevison LS, Hoehn KB, Noor MAF (2011) Effects of inversions on within- and between-species recombination and divergence. *Genome Biology and Evolution* 3:830–841
- Stinchcombe JR, Hoekstra HE (2008) Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. *Heredity* 100:158–170.
- Stölting KN, Paris M, Meier C, Heinze B, Castiglione S, Bartha D, Lexer C (2015) Genome-wide patterns of differentiation and spatially varying selection between postglacial recolonization lineage of *Populus alba* (Salicaceae), a widespread forest tree. *New Phytologist* 207:723-734.
- Storey, JD, Bass AJ, Dabney A, Robinson D (2015) qvalue: Q-value estimation for false discovery rate control R package version 2.4.2.
- Storey JD, Tibshirani R (2003) Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences* 100:9440–9445.
- Storz JF (2005) Using genome scans of DNA polymorphism to infer adaptive population divergence. *Molecular Ecology* 14:671–688.
- Storz JF, Kelly JK (2008) Effects of spatially varying selection on nucleotide diversity and linkage disequilibrium: Insights from deer mouse globin genes. *Genetics* 180:367–379.
- Strasburg JL, Kane NC, Raduski AR, Bonin A, Michelmore R, Rieseberg LH. 2011. Effective population size is positively correlated with levels of adaptive divergence among annual sunflowers. *Molecular Biology and Evolution* 28:1569–80
- Strickler SR, Bombarely A, Mueller LA (2012) Designing a transcriptome next-generation sequencing project for a nonmodel plant species1. *American Journal of Botany* 99:257–266.
- Studer A, Zhao Q, Ross-Ibarra J, Doebley J (2011) Identification of a functional transposon insertion in the maize domestication gene *tb1*. *Nature Genetics* 43:1160–1163
- Sturtevant AH (1913) The linear arrangement of six sex-linked factors in *Drosophila*, as shown by their mode of association. *Journal of Experimental Zoology* 14:43–59
- Supek F, Bosnjak M, Skunca N, Smuc T (2011) REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS ONE* 6: e21

- Suren H, Hodgins KA, Yeaman S, Nurkowski KA, Smets P, Rieseberg LH, Aitken SN, Holliday JA (2016) Exome capture from the spruce and pine giga-genomes. *Molecular Ecology* 16:1136-1146.
- Tan B, Grattapaglia D, Wu HX, Ingvarsson PK (2017) Genomic prediction reveals significant non-additive effects for growth in hybrid *Eucalyptus*. *bioRxiv*, 1–35.
- Tani N, Takahashi T, Iwata H, Mukai Y, Ujino-Ihara T et al (2003) A consensus linkage map for sugi (*Cryptomeria japonica*) from two pedigrees, based on microsatellites and expressed sequence tags. *Genetics* 165:1551–1568
- Tamiru M, Undan JR, Takagi H, Abe A, Yoshida K, et al. (2015) A cytochrome P450, OsDSS1, is involved in growth and drought stress responses in rice (*Oryza sativa* L.). *Plant Mol Biol* 88:85-99.
- Temesgen B, Brown GR, Harry DE, Kinlaw CS, Sewell MM, Neale DB (2001) Genetic mapping of expressed sequence tag polymorphism (ESTP) markers in loblolly pine (*Pinus taeda* L.) *Theoretical and Applied Genetics* 102:664–675.
- Tenaillon O (2014) The Utility of Fisher's Geometric Model in Evolutionary Genetics. *Annual Review of Ecology, Evolution, and Systematics* 45:179–201.
- Thavamanikumar S, Southerton SG, Bossinger G, Thumma BR (2013) Dissection of complex traits in forest trees — opportunities for marker-assisted selection. *Tree Genetics & Genomes* 9:627–639.
- Thomas BR, Macdonald SE, Hicks M, Adams DL, Hodgetts RB (1999) Effects of reforestation methods on genetic diversity of lodgepole pine: an assessment using microsatellite and randomly amplified polymorphic DNA markers. *Theoretical and Applied Genetics* 98:793–801.
- Tierney L, Rossini AJ, Li N, Sevcikova H (2016) snow: Simple Network of Workstations. R package version 0.4-2. <https://CRAN.R-project.org/package=snow>
- Tiffin P, Ross-Ibarra J (2014) Advances and limits of using population genetics to understand local adaptation. *Trends in Ecology & Evolution* 29:673–680.
- Tigano A, Friesen VL (2016) Genomics of local adaptation with gene flow. *Molecular Ecology* 25:2144–2164.
- Timpson NJ, Greenwood CMT, Soranzo N, Soranzo Lawson DJ, Richards JB (2018) Genetic architecture: the shape of the genetic contribution to human traits and disease. *Nature Reviews Genetics* 19:110-124.

- Toka I, Planchais S, Cabassa C, Justin AM, De Vos D, Richard L, Savouré A, Carol P (2010) Mutations in the hyperosmotic stress-responsive mitochondrial BASIC AMINO ACID CARRIER2 enhance proline accumulation in *Arabidopsis*. *Plant Physiology* 152:1851-1862
- Tomback DF (1982) Dispersal of whitebark pine seeds by Clark's nutcracker: a mutualism hypothesis. *The Journal of Animal Ecology*, 51, 451–467
- Tomback DF, Achuff P (2010) Blister rust and western forest biodiversity: ecology, values and outlook for white pines. *Forest Pathology*, 40, 186-225
- Tomback DF, Resler LM, Keane RE, Pansing ER, Andrade AJ, Wagner AC (2016) Community structure, biodiversity, and ecosystem services in treeline whitebark pine communities: potential impacts from a non-native pathogen. *Forests* 7:1-8
- Travis SE, Ritland K, Whitham TG, Keim P (1998) A genetic linkage map of pinyon pine (*Pinus edulis*) based on amplified fragment length polymorphisms. *Theoretical and Applied Genetics* 97:871–880
- Troncoso-Ponce MA, Cao X, Yang Z, Ohlrogge JB (2013) Lipid turnover during senescence. *Plant Science* 205-206:13-19.
- Tulsieram LK, Glaubitz JC, Kiss G, Carlson JE (1992) Singletree genetic linkage mapping in conifers using haploid DNA from megagametophytes. *Biotechnology* 10:3–34
- Turchin MC, Chiang CW, Palmer CD, Sankararaman S, Reich D, Hirschhorn JN (2012) Evidence of widespread selection on standing variation in Europe at height-associated SNPs *Nature Genetics* 44: 1015–1019.
- Turelli M, Barton NH (1994) Genetic and statistical analyses of strong selection on polygenic traits: what, me normal? *Genetics* 138:913–941.
- USDA Forest Service, Forest Health Technology Enterprise Team Ft Collins, CO.
- USDA Forest Service (2004) Sierra Nevada Forest Plan Amendment. Final Supplemental Environmental Impact Statement, Record of Decision. Pacific Southwest Region, R5-MB-046, 1–72.
- USDA Forest Service (2016a) Fiscal Year 2016 Budget Overview. 1–57.
- USDA Forest Service (2016b) Fiscal Year 2016 Budget Justification. 1–553.
- USDA NRCS, 2007 Soil survey of the Tahoe Basin area, California and Nevada.
- Vander Wall SB (2002) Masting in animal-dispersed pines facilitates seed dispersal. *Ecology* 83:3508-3516.
- van Mantgem PJ, Stephenson NL, Byrne JC, Daniels LD, Franklin JF, Fulé PZ, Harmon ME, Larson AJ, Smith JM, Taylor AH, Veblen TT (2009) Widespread increase of tree mortality in the western United States. *Science*, 323, 521-524.

- Van Ooijen JW (2011) Multipoint maximum likelihood mapping in a full-sib family of an outbreeding species. *Genetical Research* 93:343–349
- Vasquez-Gross, H. A., Yu, J. J., Figueroa, B., Gessler, D. D., Neale, D. B., & Wegrzyn, J. L. (2013). CartograTree: connecting tree genomes, phenotypes and environment. *Molecular Ecology Resources* 13: 528-537.
- Verner J, McKelvey KS, Noon BR, Gutierrez RJ, Gould Jr., GI, Beck TW (1992) The California spotted owl: a technical assessment of its current status. General Technical Report PSW-133. USDA Forest Service, Pacific Southwest Research Station, Albany, CA, 285 pp.
- Via S, Lande R (1985) Genotype-environment interaction and the evolution of phenotypic plasticity. *Evolution* 39:505–522.
- Vialette-Guiraud ACM, Andres-Robin A, Chambrier P, Tavares R, Scutt CP (2016) The analysis of Gene Regulatory Networks in plant evo-devo. *Journal of Experimental Botany* 67:2549–2563.
- Vicente-Serrano SM, Gouveia C, Camarero JJ, Beguería S, Trigo R, López-Moreno JI, Azorín-Molina C, Pasho E, Lorenzo-Lacruz J, Revuelto J, Morán-Tejeda E, Sanchez-Lorenzo A (2013) Response of vegetation to drought time-scales across global land biomes. *Proceedings of the National Academy of Sciences* 110: 52–57.
- Vitezica ZG, Legarra A, Toro MA, Varona L (2017) Orthogonal Estimates of Variances for Additive, Dominance, and Epistatic Effects in Populations. *Genetics* 206:1297–1307.
- Vizcaíno-Palomar N, Revuelta-Eugercios B, Zavala MA, Alia R, González-Martínez SC (2014) The role of population origin and microenvironment in seedling emergence and early survival in Mediterranean maritime pine (*Pinus pinaster* Aiton). *PLoS ONE* 9: e109132.
- Wachowiak W, Trivedi U, Perry A, Cavers S (2015) Comparative transcriptomics of a complex of four European pine species. *BMC Genomics* 16:234.
- Wadgyamar SM, Lowry DB, Gould BA, Byron CN, Mactavish RM, Anderson JT (2017) Identifying targets and agents of selection: innovative methods to evaluate the processes that contribute to local adaptation. *Methods in Ecology and Evolution* 8:738–749.
- Wagner GP, Altenberg L (1996) Perspective: complex adaptations and the evolution of evolvability. *Evolution* 50:967.
- Wagner GP, Kenney-Hunt JP, Pavlicev M, Peck JR, Waxman D, Cheverud JM (2008) Pleiotropic scaling of gene effects and the 'cost of complexity'. *Nature* 452:470–472.
- Wagner GP, Pavlicev M, Cheverud JM (2007) The road to modularity. *Nature Reviews Genetics* 8:921–931.

- Wagner GP, Zhang J (2011) The pleiotropic structure of the genotype–phenotype map: the evolvability of complex organisms. *Nature Reviews Genetics* 12:204–213.
- Wang L, Beissinger TM, Lorant A, Ross-Ibarra C, Ross-Ibarra J, Hufford M (2017) The interplay of demography and selection during maize domestication and expansion. *bioRxiv*
- Wang Z, Liao BY, Zhang J (2010) Genomic patterns of pleiotropy and the evolution of complexity. *Proceedings of the National Academy of Science* 107:18034–18039.
- Ward JHJ (1963) Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association* 58:236–244
- Ward J, Bhangoo J, Fernandez-Fernandez F, Moore P, Swanson J, Viola R, Velasco R, Bassil N, Weber C, Sargent D (2013) Saturated linkage map construction in *Rubus idaeus* using genotyping by sequencing and genome-independent imputation. *BMC Genomics* 14:2
- Warren DL, Glor RE, Turelli M (2008) Environmental niche equivalency versus conservatism: quantitative approaches to niche evolution. *Evolution* 62:2868–2883.
- Warren DL, Seifert SN (2011) Ecological niche modeling in Maxent: the importance of model complexity and the performance of model selection criteria. *Ecological Applications* 21:335–342.
- Wegrzyn JL, Lee JM, Tearse BR, Neale DB (2008) TreeGenes: A Forest Tree Genome Database. *International Journal of Plant Genomics* 2008:1–7.
- Wegrzyn JL, Liechty JD, Stevens KA, Wu L-S, Loopstra CA, et al. (2014) Unique features of the loblolly pine (*Pinus taeda* L.) megagenome revealed through sequence annotation. *Genetics* 196:891–909.
- Wegrzyn JL, Main D, Figueroa B, Choi M, Yu J, Neale DB, Jung S, Lee T, Stanton M, Zheng P, Ficklin S, Cho I, Peace C, Evans K, Volk G (2012) Uniform standards for genome databases in forest and fruit trees. *Tree Genet Genomes* 8:549–557.
- Weih M, Bonosi L, Ghelardini L, Ronnberg-Wastljung AC (2011) Optimizing nitrogen economy under drought: increased leaf nitrogen is an acclimation to water stress in willow (*Salix* spp.). *Annals of Botany* 108:1347–1353.
- Weir BC, Cockerham CC (1984) Estimating *F*-statistics for the analysis of population structure. *Evolution*, 1358–1370.
- Welch JJ, Waxman D (2003) Modularity and the cost of complexity. *Evolution* 57:1723–13.
- Wellenreuther M, Hansson B (2016) Detecting Polygenic Evolution: Problems, Pitfalls, and Promises. *Trends in Genetics* 32:155–164.

- Westerling AL (2006) Warming and Earlier Spring Increase Western U.S. Forest Wildfire Activity. *Science* 313:940–943.
- Whitaker (1975) *Communities and Ecosystems*. Macmillian, London.
- White TL, Adams WT, Neale DB (2007) *Forest genetics*. CABI Publishing, Cambridge.
- Whitlock MC (1999) Neutral additive genetic variance in a metapopulation. *Genetical Research* 74:215–221.
- Whitlock MC (2003) Fixation probability and time in subdivided populations. *Genetics* 164:767–779.
- Whitlock MC, Gilbert KJ (2012) Q_{ST} in a hierarchically structured population. *Molecular Ecology Resources* 12:481–483.
- Whitlock, M. C. and F. Guillaume. 2009. Testing for spatially divergent selection: Comparing Q_{ST} to F_{ST} . *Genetics* 183: 1055-1063.
- Whitlock MC, Lotterhos KE (2015) Reliable detection of loci responsible for local adaptation: Inference of a null model through trimming the distribution of F_{ST} . *The American Naturalist* 186:S34–S36.
- Whitlock MC, Phillips PC, Moore FB (1995) Multiple fitness peaks and epistasis. *Annual Review of Ecology and Systematics* 26:601–629.
- Willis KJ, van Andel TH (2004) Trees or no trees? The environments of central and eastern Europe during the last glaciation. *Quaternary Science Reviews* 23:2369–87
- Willyard A, Syring J, Gernandt D, Liston A, Cronn R (2007) Fossil calibration of molecular divergence in *Pinus*: inferences for ages and mutation rates. *Molecular Biology and Evolution* 24:90–101
- Wofford AM, Finch K, Bigott A, Willyard A (2014) A Set of Plastid Loci for Use in Multiplex Fragment Length Genotyping for Intraspecific Variation in *Pinus* (Pinaceae). *Applications in Plant Sciences* 2:1400002.
- Wong CK, Bernardo R (2008) Genomewide selection in oil palm: increasing selection gain per unit time and cost with small populations. *Theoretical and Applied Genetics* 116:815–824
- Wortley, A. H., & Scotland, R. W. (2004). Synonymy, sampling and seed plant numbers. *Taxon* 53:478-480
- Wright J (1976) *Introduction to forest genetics*. Academic Press, New York.
- Wright S (1969) *Evolution and the Genetics of Populations Vol. 2*. Univ. Chicago Press, Chicago, Illinois

- Wright S (1933) Inbreeding and homozygosis. *Proceedings of the National Academy of Science USA* 19:411–420
- Wright S (1938) Size of population and breeding structure in relation to evolution. *Science* 87:430–431
- Wright S (1932) The roles of mutation, inbreeding, crossbreeding, and selection in evolution. *Proceedings of the Sixth International Congress on Genetics* 1:356–366
- Wright S (1931) Evolution in Mendelian populations. *Genetics* 16:97–159
- Wu RL (1998) Genetic mapping of QTLs affecting tree growth and architecture in *Populus*: implication for ideotype breeding. *Theoretical and Applied Genetics* 96:447–457.
- Wu Y, Bhat PR, Close TJ, Lonardi S (2008a) Efficient and accurate construction of genetic linkage maps from the minimum spanning tree of a graph. *PLoS Genetics* 4:e1000
- Wu Y, Close TJ, Lonardi S (2008b) On the accurate construction of consensus genetic maps. *Computational Systems and Bioinformatics Conference* 7:285–296
- Xu S (2003) Theoretical Basis of the Beavis Effect. *Genetics* 165:2259–2268.
- Yang J, Benyamin B, McEvoy BP, et al. (2010) Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics* 42:565–569.
- Yang, J., S. Mezouk, A. Baumgarten, E. S. Buckler, K. E. Guill, M. D. McMullen, and J. Ross-Ibarra. 2017. Incomplete dominance of deleterious alleles contribute substantially to trait variation and heterosis in maize. *bioRxiv*.
- Yeaman S (2013) Genomic rearrangements and the evolution of clusters of locally adaptive loci. *Proceedings of the National Academy of Sciences* 110:E1743–51.
- Yeaman S (2015) Local Adaptation by Alleles of Small Effect. *The American Naturalist* 186:S74–S89.
- Yeaman S, Hodgins KA, Lotterhos KE, et al. (2016) Convergent local adaptation to climate in distantly related conifers. *Science* 353:1431–1433.
- Yeaman S, Jarvis A (2006) Regional heterogeneity and gene flow maintain variance in a quantitative trait within populations of lodgepole pine. *Proceedings of the Royal Society B: Biological Sciences* 273:1587–1593.
- Yeaman S, Otto SP (2011) Establishment and maintenance of adaptive genetic divergence under migration, selection, and drift. *Evolution* 65:2123–2129.
- Yeaman S, Whitlock MC (2011) The genetic architecture of adaptation under migration-selection balance. *Evolution* 65:1897–1911.

- Yoder JB, Tiffin P (2017) Effects of gene action, marker density, and timing of selection on the performance of landscape genomic scans of local adaptation. *Journal of Heredity* 109:16-28.
- Young ND, Debelle F, Oldroyd GE, Geurts R, Cannon SB, et al. (2011) The *Medicago* genome provides insight into the evolution of rhizobial symbioses. *Nature* 480:520–24
- Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB, Kresovich S, Buckler ES (2005) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genetics* 38:203–208.
- Zald HSJ, Gray AN, North MP, Kern RA (2008) Initial tree regeneration responses to fire and thinning treatments in a Sierra Nevada mixed-conifer forest, USA. *Forest Ecology and Management* 256:168–179.
- Zerbino DR, Birney E (2008) Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research*, 18, 821–829.
- Zhang J, Marshall JD (1994) Population differences in water-use efficiency of well-watered and water-stressed western larch seedlings. *Canadian Journal of Forest Research* 24:92-99.
- Zhang X-S (2012) Fisher's geometric model of fitness landscape and variance in fitness within a changing environment. *Evolution* 66:2350–2368.
- Zhang M, Zhou L, Bawa R, Suren H, Holliday JA (2016) Recombination rate variation, hitchhiking, and demographic history shape deleterious load in poplar. *Molecular Biology and Evolution*. 33:2899-2910.
- Zhong S, Dekkers JC, Fernando RL, Jannink JL (2009) Factors affecting accuracy from genomic selection in populations derived from multiple inbred lines: a barley case study. *Genetics* 182: 355–364.
- Zhou, X., Carbonetto, P., & Stephens, M. (2013). Polygenic modeling with Bayesian sparse linear mixed models. *PLoS Genetics*, 9(2), e1003264.
- Zinkgraf M, Liu L, Groover A, Filkov V (2017) Identifying gene coexpression networks underlying the dynamic regulation of wood-forming tissues in *Populus* under diverse environmental conditions. *New Phytologist* 214:1464–1478.
- Zöllner S, Pritchard JK (2007) Overcoming the Winner's Curse: Estimating Penetrance Parameters from Case-Control Data. *The American Journal of Human Genetics* 80:605–615. doi:
- Zuk O, Hechter E, Sunyaev SR (2012) The mystery of missing heritability: genetic interactions create phantom heritability. *Proceedings of the National Academy of Sciences* 109:1193-1198.

Vita

Brandon Michael Lind was born on September 14, 1984 in Dallas, Texas. He graduated from Mason City High School in Mason City, Iowa and received bachelor's degrees in chemistry and biology from the University of Northern Iowa in Cedar Falls, Iowa. Brandon continued his education at Iowa State University as a post-baccalaureate following graduation.