bioRxiv preprint doi: https://doi.org/10.1101/056317; this version posted May 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

When local means local: Polygenic signatures of local adaptation within whitebark pine (*Pinus albicaulis* Engelm.) across the Lake Tahoe Basin, USA

Brandon M. Lind[*1], Christopher J. Friedline[†], Jill L. Wegrzyn[‡], Patricia E. Maloney[§], Detlev R. Vogler[**], David B. Neale[††], and Andrew J. Eckert[†2]

[*]Integrative Life Sciences program, Virginia Commonwealth University, Richmond, Virginia 23284 USA

[†]Department of Biology, Virginia Commonwealth University, Richmond, Virginia 23284 USA

[‡]Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT 06269 USA

[§]Department of Plant Pathology and Tahoe Environmental Research Center, University of California, Davis, California 95616 USA

[**]USDA, Forest Service, Pacific Southwest Research Station, Institute of Forest Genetics, 2480 Carson Road, Placerville, California 95667 USA

[††]Department of Plant Sciences, University of California, Davis California 95616 USA

Data reference numbers (XXXXXXXXXXX)

[1] Corresponding Author. Current Address: Integrative Life Sciences PhD. program, Virginia Commonwealth University, Richmond, VA 23284. E-mail: lindb@vcu.edu @iowensis

[2] Corresponding Author. Current Address: Department of Biology, Virginia Commonwealth University, Richmond, VA 23284. E-mail: aeckert2@vcu.edu

bioRxiv preprint doi: https://doi.org/10.1101/056317; this version posted May 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Running Title: Local adaptation of *P. albicaulis*

Key words:

Local adaptation, *Pinus albicaulis*, divergent selection, gene flow, linkage disequilibrium

Corresponding Authors:

Brandon M. Lind

1000 West Cary Street

Life Science Bldg. - Department of Biology Rm 126

Virginia Commonwealth University

Richmond, Virginia 23284 USA

e-mail: lindb@vcu.edu, fax: 804-828-0820

Andrew J. Eckert

1000 West Cary Street

Life Science Bldg. - Department of Biology Rm 126

Virginia Commonwealth University

Richmond, Virginia 23284 USA

e-mail: aeckert2@vcu.edu, fax: 804-828-0800

bioRxiv preprint doi: https://doi.org/10.1101/056317; this version posted May 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

ABSTRACT

For populations exhibiting high levels of gene flow, the genetic architecture of fitness-related traits is expected to be polygenic and underlain by many small-effect loci that covary across a network of linked genomic regions. For most coniferous taxa, studies describing this architecture have been limited to single-locus approaches, possibly leaving the vast majority of the underlying genetic architecture undescribed. Even so, molecular investigations rarely search for patterns indicative of an underlying polygenic basis, despite prior expectations for this signal. Here, using a polygenic perspective, we employ single and multilocus analyses of genome-wide data (n = 116,231 SNPs) to describe the genetic architecture of adaptation within whitebark pine (*Pinus albicaulis* Engelm.) 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 polygenic adaptation to the rain shadow experienced across the eastern Sierra Nevada. Specifically, we find little evidence for large-effect loci and that the frequencies of loci associated with 4/5 phenotypes (mean = 236 SNPs), 18 environmental variables (mean = 99 SNPs), and those detected through genetic differentiation (n = 110 SNPs) exhibit significantly higher covariance than random SNPs. We also provide evidence that this covariance tracks environmental measures related to soil water availability through subtle allele frequency shifts across populations. Our results provide replicative support for theoretical expectations and highlight advantages of a polygenic perspective, as unremarkable loci when viewed from a single-locus perspective are noteworthy when viewed through a polygenic lens, particularly when considering protective measures such as conservation guidelines and restoration strategies.

bioRxiv preprint doi: https://doi.org/10.1101/056317; this version posted May 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

INTRODUCTION

Shortly after Nilsson-Ehle (1909) and East (1910) independently demonstrated evidence for multiple-factor inheritance, 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 infinitesimal model (1930), which founded the basis for attributing continuous variation of phenotypes to a polygenic model of many underlying heritable components of small effect. There, Fisher (1930) characterized adaptation as the non-random advance of a population's mean phenotype towards an optimum that best fits its environment. Correspondingly, when selective forces are spatially heterogeneous, populations can become locally adapted. Indeed, over the subsequent century since Fisher's insight, local adaptation has been demonstrated to occur across numerous taxa. As such, 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; Tigano and Friesen 2016). Plants in particular have received much attention in this regard due in part to ideal characteristics within native populations and environments that lend themselves to such analyses. Through these investigations, local adaptation in plants appears to be common, yet the genetic architecture (i.e., the number, effect size, type, and interaction of loci) of local adaptation in natural populations remains largely undescribed (Leimu and Fischer 2008; Hereford 2009).

Investigators seeking to explain the genetic basis of local adaptation in plants have been motivated by observations of significant phenotypic differentiation among populations (e.g., Q_{ST}). If such phenotypes have a genetic basis, the underlying QTL may be differentiated among populations as well (Endler 1977; reviewed in Storz 2005, Haasl and Payseur 2016). In such 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 and Kremer 2003) and

bioRxiv preprint doi: https://doi.org/10.1101/056317; this version posted May 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

empirical (Hall *et al.* 2007; Luquez *et al.* 2007) investigations exploring the relationship between phenotypic differentiation (e.g., Q_{ST}) and that of the underlying loci (e.g, G_{ST} or F_{ST}) have shown that discordance between these two structural indices 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 and Pennings 2005; Barret and Schluter 2008) and is attributable to the between-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 loci (Latta 1998; Barton 1999; Latta 2003; McKay and Latta 2002; Kremer and Le Corre 2012; Le Corre and Kremer 2012), such that adaptation need not take place through numerous fixation events or sweeping allele frequency changes (MacKay *et al.* 2009; Pritchard and di Rienzo 2010). Over many generations these shifts can lead to concentrated architectures of large-effect loci with a reduction of those with small effect (Yeaman and 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 and Whitlock 2015), it may be difficult to discern if the focal loci offer a representative picture of the underlying genetic architecture. Thus for many species, specifically across fine spatial scales, the signal of polygenic local adaptation within much of current genetic data may go largely undetected using only single-locus approaches (Latta 1998; 2003; Le Corre and Kremer 2003; Yeaman and Whitlock 2011; Kemper *et al.* 2014), resulting in calls for theory and empiricism that move beyond single-locus perspectives (Pritchard and di Rienzo 2010; Sork *et al.* 2013; Tiffin and Ross-Ibarra 2014; Stephan 2015).

Populations of forest trees, particularly conifers, have a rich history of common garden, provenance tests, and genecological studies that demonstrate abundant evidence for local

bioRxiv preprint doi: https://doi.org/10.1101/056317; this version posted May 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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) 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 moderate- to large-effect loci underlying the adaptive trait in question explain were identified (Neale and 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, it is likely that a majority of the genetic architecture of local adaptation in trees remains undescribed as well (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 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

bioRxiv preprint doi: https://doi.org/10.1101/056317; this version posted May 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 address the following four questions: (i) What is the number, effect size distribution, and relationship among SNPs associated with environment or phenotype? (ii) To what degree is there overlap of the loci identified through environmental association with those identified through phenotypic association? (iii) Do focal loci show higher degrees of evidence for natural selection acting on a polygenic architecture (covariance of allele frequencies) than random loci within the genome? (iv) Is the covariance of allele frequencies across population pairs associated with environmental heterogeneity? This study highlights the advantages of a polygenic perspective and investigates signatures of local adaptation using a large set of null markers to judge the extremity of allele covariation among putatively adaptive loci where others have relied on simulation or null candidate genes. Furthermore, this work provides additional replication for the support of theoretical predictions for the covariation among adaptive loci found by other studies in trees.

MATERIALS and METHODS

Focal species, study area, and sampling

A principal component of high elevation forests in California and Nevada, *P. albicaulis* is widespread throughout subalpine and treeline environments and plays a vital role in ecosystem function and services including food resources for wildlife, forest cover, watershed protection, protracting snowmelt, and biodiversity (Hutchins and Lanner 1982; Farnes 1990; Tomback et al., 2001; McKinney et al., 2009; Tomback and Achuff 2010; Tomback et al. 2016). Most of the species' distribution is outside of California, extending northward into Oregon, Washington, British Columbia, and Alberta and eastward into northern Nevada, Idaho, Montana, and Wyoming (Critchfield and Little 1966; Tomback and Achuff 2010). Whitebark pine is a foundation species

bioRxiv preprint doi: https://doi.org/10.1101/056317; this version posted May 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

in subalpine ecosystems throughout most of its range in western North America (Ellison et al. 2005) and 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, *Dendroctonous ponderosae* Hopkins (Tomback and Achuff 2010; Mahalovich and Stritch 2013).

The LTB lies within California and Nevada in the north-central Sierra Nevada range, varies in elevation from 1900 to 3300m, and is flanked to the west by the Sierra Nevada crest and to the east by the Carson Range. The LTB experiences a Mediterranean climate with warm, dry summers and cool, wet winters. Precipitation falls during the winter months, most often in the form of snow, with a strong west-east gradient. The geology of the region is dominated by igneous intrusive rocks, typically granodiorite, and igneous extrusive rocks, typically andesitic lahar, with small amounts of metamorphic rock (USDA NRCS 2007).

Each of the eight study populations (three subplots per population) was located in a distinct watershed and were distributed around the Basin to capture variation in the physical environment (e.g., climate, geology, and topography; Figure 1). Needle tissue was sampled in the summer of 2008 from 244 *P. albicaulis* trees (Table 1). From these eight populations, six populations were chosen to sample cones from 88 of the trees that were sampled for needle tissue. All samples were collected from trees separated by 30 to 1000m, with an average interpopulation distance of 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 1).

Common gardens and phenotypic measurements

For populations of forest trees, fitness-related traits associated with survival, especially during seedling and juvenile stages, are an important component of total lifetime fitness 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

bioRxiv preprint doi: https://doi.org/10.1101/056317; this version posted May 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 1) using a random block design (for further details see Maloney *et al. in review*). Growth (height, root:shoot biomass), phenology (budset), water-use efficiency (δ^{13}C), and resource allocation (δ^{15}N) 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 δ^{13}C and δ^{15}N 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/).

Values for each phenotype were estimated for maternal trees (i.e. families) using linear mixed models of the form:

$$Y_{ijklm} = \mu + pop_i + fam(pop)_{j(i)} + block_k + date_l + \varepsilon_{ijklm},$$

where Y_{ijklm} is the phenotype of the mth seedling sowed in the lth year within the kth block originating from the jth family nested within the ith population, μ is the grand mean, pop_i is the random effect of the ith population, $fam(pop)_{j(i)}$ is the random effect of the jth family nested within the ith population, $block_k$ is the fixed effect of the kth block, $date_l$ is the effect of the lth sowing year, and ε_{ijklm} is the residual error of the mth seedling sowed in the lth year within the kth block originating from the jth family nested within the ith population. Separate models were fit to the data for each phenotype using restricted maximum likelihood estimation (REML) as employed in the `lme4` library (v1.1-12) in `R` (v3.2.2, R Core Team 2015). Effects of families were estimated as the sum of the grand mean, the population effect, and the effect of family. Estimated values were reported on the original scale of measurements of each phenotype, which were then used in downstream analyses. In a previous study (Maloney *et al. in review*) we also estimated narrow sense heritability and Q_{ST} (see Appendix).

bioRxiv preprint doi: https://doi.org/10.1101/056317; this version posted May 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

DNA extraction, sequencing, and analysis

Total genomic DNA was isolated from finely ground needle tissue sampled from 244 trees across all 8 populations using the Qiagen DNEasy 96 Plant kit according to protocol (Qiagen, Germantown, MD), except that each sample was eluted twice with 50µL of the elution buffer to increase DNA yield, as recommended by Qiagen. 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 libraries of up to 96 individuals each, as in Parchman *et al.* (2012). For each library, one individual was duplicated in a separate well to increase coverage for the downstream reference assembly. Total genomic DNA was digested and ligated with adapters containing amplification and sequencing primers as well as barcodes for multiplexing. These 96 barcodes (Parchman *et al.* 2012) are of 8-10bp sequences that differ by at least four bases. Ligated DNA was then amplified using high-fidelity PCR according to manufacturer's specifications. Using the QIAquick Gel Extraction Kit (Qiagen), amplified fragments were then isolated near 400bp by excising the 300-500bp window of pooled PCR product separated in a 1% agarose gel at 100V for one hour. 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).

Reads were assigned to individual sample IDs based on 100% match to the barcode sequence or were otherwise discarded. The reads from the individual with the greatest number of reads was assembled using `Velvet` (v1.2.1, Zerbino 2008) optimized for hash length *k* for odd *k* on *k* ([37,49]) using `VelvetOptimiser` (v2.2.5, Gladman and Seemann 2016) where the `-short` and `-short2` flags were used to distinguish the 150bp and 50bp reads, respectively. To call SNPs for all individuals, reads were mapped to the reference assembly using `Bowtie2`

bioRxiv preprint doi: https://doi.org/10.1101/056317; this version posted May 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(v2.2.6, `--local -D 20 -R 3 -N 1 -L 20 -i S, 1, 0.50`; Langmead and Salzberg 2012). `Samtools` (v1.3, Li *et al.* 2009) was used to convert the resulting `.sam` files into their binary (`.bam`) equivalent (`view`, `sort`, `index`). Genotypes were then called using `bcftools` (v1.3) based on likelihoods estimated by `samtools` in `mpileup`. Genotypes were filtered with `vcftools` (v0.1.13, Danecek *et al.* 2011) to enforce diallelic loci, remove indels and minor allele frequency (MAF) of ≤ 0.01, and exclude sites with >50% missing data.

Two datasets were created each starting with the same set of SNPs remaining after filtering. The first dataset was left as-is to leave missing data (missing dataset, hereafter MDS). The second dataset was created in which the missing data were imputed (imputed dataset, hereafter IDS) using `Beagle` (v4.0, Browning and Browning 2016). Next, each of these two datasets was analyzed using custom `Python` (v2.7.11) scripts to filter SNPs by removing those with minor allele frequency <0.01 and global outliers for Wright's fixation index ($F > 0.5$ and $F < -0.5$). To remove low coverage and artifacts of amplification, SNPs were removed if read depth was < 100 or ≥ 1500. Each dataset was then reduced further to include only the intersection of loci remaining between the two sets. To reduce physical linkage within our data, one SNP per contig from the MDS was chosen by the least missing data. These SNPs were used to define a subset from the IDS as the final empirical set of SNPs used for downstream analyses. To judge veracity of sequence data we mapped 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/ftp/Genome_Data/genome/pinerefseq/Pila/v1.0/).

Identifying loci under selection

Linear mixed models (LMM) and sparse regression models, such as Bayesian variable selection regression (BVSR, e.g., Guan and Stephens 2011) have been used to uncover adaptive traits under a multilocus perspective. Yet the underlying assumptions of these models differ in meaningful ways, particularly for polygenic modeling. In the case of LMM, the number of underlying loci which affect the phenotype is assumed to be large, effectively every variant, with

bioRxiv preprint doi: https://doi.org/10.1101/056317; this version posted May 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

a normal distribution of effect sizes (Zhou *et al.* 2013). Conversely, BVSR assumes that the number of variants affecting the phenotype is small and represented by a point-normal distribution of effect sizes (i.e., the effects come from a mixture of a normal distribution and a point mass at 0; Guan and Stephens 2011). In this way, the LMM effectively assumes a large number of small effects while the BVSR assumes a small number of larger effects (Zhou *et al.* 2013). Model selection between such methods should therefore depend upon the underlying genetic architecture, which is generally not known beforehand. Because the genetic architecture of the traits investigated here is unknown, 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 BVSR that also offers considerable statistical advantages over single-locus GWAS approaches (Guan and Stephens 2011; Ehret *et al.* 2012; Zhou *et al.* 2013; Moser *et al.* 2015). BSLMM accounts for population structure and relatedness then subsequently identifies which relevant genetic variants to include in a multiple regression of the phenotype. Specifically, to describe the underlying genetic architecture, BSLMM uses priors (described below) and attributes of the genetic data to estimate the number of SNPs underlying a given trait (N_{SNP}), the posterior inclusion probability (γ, hereafter *PIP*) for individual SNPs, the random (α; i.e., polygenic) effect of SNPs included in the model (in standard deviations), the fixed (β, i.e., sparse) effects of SNPs included in the model (in standard deviations), the total effect ($\alpha + \beta\gamma$; i.e., the combined effects of small- and large-effect SNPs), as well as the proportion of phenotypic variance explained by the fixed effects and random effects combined (*PVE*, i.e., by the total effect estimated from SNPs included in the model).

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. To account for population structure, principal component analyses (PCAs) were calculated using the centered and standardized mean genotypes of each SNP (based on the global MAF, following Patterson *et al.*

bioRxiv preprint doi: https://doi.org/10.1101/056317; this version posted May 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

2006) and the `prcomp()` function in `R`. Significant PCs (α = 0.05) were identified using a Tracy-Widom test and were then used in linear models predicting each previously-estimated phenotypic value from the common garden (Maloney *et al. in review*) using the `lm()` function in `R` (phenotype ~ PCs). From these linear models, the trait-specific residuals were quantile-transformed and were used to create the phenotypic in-files to `GEMMA` (as recommended by Guan and Stephens 2011; Zhou *et al.* 2013). 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 which were sampled every 1000th step. Priors for the proportion of variance explained 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 visually 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 posterior inclusion probability 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 ≈ 116 SNPs) for each phenotype. We also explored SNPs with $\overline{PIP}$ ≥99.8th percentile (n ≈ 232). Finally, we calculated harmonic mean across chains for fixed ($\bar{\beta}$), random ($\bar{\alpha}$), and total effects ($\hat{b} = \bar{\alpha} + \bar{\beta} \cdot \overline{PIP}$).

To identify genotype-environmental associations, we implemented `bayenv2` (v2.0; Coop *et al.* 2010; Günther and 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 IDS SNPs (n = 116,231), with 100,000 iterations for each SNP within each chain. Convergence of the MCMC across chains was visually inspected using the `coda` library in `R`. For each SNP, we took the harmonic mean across chains for the Bayes factor ($\overline{BF}$) and absolute value of Spearman's ρ (hereafter $\overline{\rho_S}$). When calculating $\overline{BF}$, we further checked for convergence by flagging SNPs

bioRxiv preprint doi: https://doi.org/10.1101/056317; this version posted May 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

which had large differences between values across chains. 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 ≤ 2/5 chains) we took the harmonic mean from the values that were less than or equal to one. We identified SNPs as those most likely to be associated with environmental variables 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).

We implemented the program `OutFLANK` (Whitlock and Lotterhos 2015) to investigate the opportunity of detecting the environmentally- and phenotypically-associated loci using outlier approaches based on population genetic structure alone (e.g., F_{ST}). While F_{ST} outlier approaches do not take on a polygenic perspective *per se*, they do have the advantage of not requiring the investigator to identify, *a priori*, the phenotypic and environmental variables most important to local adaptation. `OutFLANK` is an approach which uses empirical data to infer the null distribution of F_{ST} (F'_{ST}) for loci unlikely to be under spatially heterogeneous selection (upper tail of F'_{ST}) or homogenous balancing selection (lower tail of F'_{ST}). From this null distribution, focal loci can be identified from the empirical set which show signatures of additional evolutionary processes, such as spatially heterogeneous selection, with a lower false discovery rate and comparable power relative to other outlier methods (Whitlock and Lotterhos 2015). 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

Because of the polygenic basis expected from fitness-related traits, we investigated the level of covariance of allele frequencies among focal SNPs identified from `GEMMA`, `bayenv2`, and `OutFLANK` analyses with estimates of covariance among random SNPs chosen from bins

bioRxiv preprint doi: https://doi.org/10.1101/056317; this version posted May 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

based on expected heterozygosity (H_E). 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,

$$\widehat{D}_{a(ij)} = \sum_k \frac{n_k}{n}\left(q_{i,k}q_{j,k} - q_i q_j\right) \qquad \text{Eq. (1)}$$

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 ith SNP in population k, $q_{j,k}$ is the allele frequency of the jth SNP in population *k*, while q_i and q_j are the respective global allele frequencies of the ith and j^{th} SNP across k = 6 populations (Storz and Kelly 2008, their Equation 2; Ma *et al.* 2010, their Equation 3). In some populations q_k was the major allele because we chose the allele to use in comparisons based on global minor allele frequency. Therefore, all calculations of $\widehat{D}_{a(ij)}$ are referenced to the global minor allele haplotype for a pair of SNPs.

To be able to discern if the level of covariance of allele frequencies among SNPs identified by GEMMA (or another method; hereafter focal SNPs) was greater than that from SNPs randomly chosen from our dataset, we first divided all SNPs in the dataset by their expected heterozygosity into bins of 0.01 ranging from 0 to 0.50. For instance, 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 number of focal SNPs, a random set of SNPs equal in total size was selected, as well as in the same number of SNPs from a given heterozygosity bin. 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 $\widehat{D}_{ij}$ from the focal set of SNPs. If the median $\widehat{D}_{a(ij)}$ is

bioRxiv preprint doi: https://doi.org/10.1101/056317; this version posted May 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

greater among our focal SNPs than the 95th percentile of the null distribution of 1000 medians, we will conclude that the focal SNPs identified by a given method are in higher degrees of covariance than expected by chance. 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 alleles of loci conferring a positive effect on the phenotype, negative between those alleles among loci 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), we chose to identify extreme values by taking the absolute value of $\widehat{D}_{a(ij)}$ for each locus pair. We also calculated covariance 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 infer signatures of allele frequency shifts, 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. To discern signals of allele frequency shifts associated with environment, we 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, Euclidian distance matrices for each of the five phenotypes and for each of the 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`

bioRxiv preprint doi: https://doi.org/10.1101/056317; this version posted May 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 quantified the mean allele frequency differences across populations and compared this to 1000 sets of random SNPs chosen by H_E. To investigate evidence for associations between sampled locations among environmental variables or among phenotypes, we ran additional Mantel tests for each comparison.

Data availability

Sequence data is deposited in the short read archive of the National Center for Biotechnology Information (project number: TBD). Scripts used in analyses can be found in IPython notebook format (Pérez and Granger 2007) at https://github.com/brandonlind/whitebark_pine.

RESULTS

SNP filtering and characterization

The sample chosen to make the reference assembly was that with the greatest number of reads (N = 23,363,768), which was the individual duplicated in the third library. Optimization with `VelvetOptimiser` (k_{opt} = 45) resulted in an assembly with 391,957 contigs (maximum 330bp per contig, 48,906,035 total bases across contigs). Using the reference assembly, 2,892,582 SNPs were called using `samtools`. Initial filtering with `vcftools` left 1,300,961 SNPs. Next, we created the missing data set (MDS) by leaving our SNPs as-is and an imputed set (IDS) by imputing the MDS with `Beagle`. After filtering each set for monomorphic SNPs, minor allele frequency, and Wright's *F* outliers, the MDS retained 778,406 SNPs while the IDS retained 1,029,063 SNPs. We reduced each set further to include only the intersection of loci between the data sets, resulting in two data sets (MDS, IDS) each with 713,745 SNPs. Finally, we reduced these sets further by choosing from each contig in the MDS the SNP with the least amount of missing data. In cases where multiple SNPs across the same contig were equal in missing data, we chose randomly from this subset of SNPs. We then removed the remaining

bioRxiv preprint doi: https://doi.org/10.1101/056317; this version posted May 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

SNPs on the contig from both IDS and MDS. After these filtering steps, we retained 116,231 SNPs from the IDS for use as the empirical set in downstream analyses (Table S1). Of these contigs, 107,354 (92.4%) mapped to the *P. lambertiana* reference genome with 85% similarity and 50% query length coverage thresholds, 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.

Phenotypic traits were heritable and structured across populations – bud flush h^2 = 0.3089 Q_{ST} = 0.0156; δ^{13}C h^2 = 0.7787, Q_{ST} = 0.0427; height h^2 = 0.0608, Q_{ST} = 0.0418; δ^{15}N h^2 = 0.3525, Q_{ST} = 0.0191; root:shoot h^2 = 0.3240, Q_{ST} = 0.0110; Table S4). Overall, populations show little genetic structure with plots accounting for less than 1% of the variance in allele frequencies ($F_{\mathrm{plot,total}}$ = 0.00687; 95% credible interval: 0.0067-0.0070). Of this variation, 56.6% was accounted for by populations ($F_{\mathrm{pop,total}}$ = 0.00389; 95% CI: 0.0038-0.0040) with the remainder due to plots within populations ($F_{\mathrm{plot,pop}}$= 0.00299; 95% CI: 0.0029-0.0031). We found similar patterns among the locus-specific estimates of F_{ST} (Figure S3). 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 S11). To further address applicability of the island model used for calculation of allelic covariance ($\widehat{D}_{a(ij)}$) and allele frequency shifts ($pw\widehat{D}_{a(ij)}$) we analyzed population pairwise F_{ST} according to Weir and 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 analyses

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

bioRxiv preprint doi: https://doi.org/10.1101/056317; this version posted May 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

precipitation (r = 0.7145, p = 0.0030), and between percent maximum solar radiation and latitude distances (r = 0.4629, p = 0.0370; Table 3). 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 3), 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 2). 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. 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 2). Additionally, the focal SNPs identified by `bayenv2` displayed a bias towards SNPs with low values of H_E (Figures S1-S2) when compared to the distribution from the full set of SNPs (Figure S13). As such, our environmental associations should be interpreted with caution, as we did not have any SNPs with $\overline{BF}$ > 1 nor do our nonparametric correlations exceed 0.35. Even so, when we compared absolute estimates of allele frequency covariance ($\widehat{D}_{a(ij)}$) among focal SNPs against the corresponding 1000 sets of random SNPs (the null sets) equal in set size as well as within H_E bins, we found that for all focal sets the median $\widehat{D}_{a(ij)}$ was always greater than the 100th percentile of the null distribution (Table 2). 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 2). The percentile of the focal $\widehat{D}_{a(ij)}$ distribution corresponding to the 100th percentile of the associated null set varied across environmental variables as well, reaching just the 3rd percentile for minimum January temperature and the 43rd percentile for percent clay (Table 2), suggesting that for most environmental variables the focal SNPs show higher degrees of covariance than expected by chance, despite having low $\overline{BF}$ and $\overline{\rho_S}$.

bioRxiv preprint doi: https://doi.org/10.1101/056317; this version posted May 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Through the examination of patterns of allele frequency shifts ($pw\hat{D}_{a(ij)}$) across loci associated with environment we found no significant associations with geographic distance using Mantel tests (p > 0.1116, data not shown). 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 3). When we investigated the association between allele frequency shift ($pw\hat{D}_{a(ij)}$) matrices against the eponymous environmental distance matrix, we found significant association for annual precipitation (r = 0.7134, p = 0.0027), GDD-May (r = 0.8480, p = 0.0013), longitude (r = 0.6522, p = 0.0024), percent rock coverage (r = 0.5124, p = 0.0145), percent sand (r = 0.5574, p = 0.0046), minimum January temperature (r = 0.5791, p = 0.0137), and WC-⅓bar (field capacity, r = 0.4806, p = 0.0361; Table 5). Additionally, we examined relationships between a particular $pw\hat{D}_{a(ij)}$ matrix and the 17 remaining environmental distance matrices and found significant associations in an additional 13 comparisons (Table 5), with five of these comparisons having $pw\hat{D}_{a(ij)}$ associated 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 5), 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 (range 0.018-0.029) were generally larger than that predicted from random SNPs of the same heterozygosity (Figures S15-S16). Overall, our results indicate that the vast majority of subtle allele frequency shifts among loci associated with environment also have significant associations related to annual precipitation, longitude, or available soil water capacity.

Genotype-phenotype analyses

We used a subset of the empirical set of SNPs for use in genotype-phenotype analysis to account for only those populations with individuals contributing to the common garden (n = 88

bioRxiv preprint doi: https://doi.org/10.1101/056317; this version posted May 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

trees, see Table 1). We filtered SNPs with MAF < 0.01 across these 6 populations to retain 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 $\overline{PIP}$ (Table 4, Figure S4). 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 S18-19). While the phenotypic main effect ($\bar{\beta}$) of loci across $\overline{PIP}$ sets ranged from 0-0.2 (Figure S5), the random effects ($\bar{\alpha}$) and total effect ($\bar{\alpha} + \bar{\beta} * \overline{PIP}$) were generally well below 6e-04 (Figures S6-7) suggestive of small effects of similar magnitude across focal loci.

Overall, the genetic variance of SNPs included in the polygenic model explained between 14.4% (δ^{15}N) and 37.6% (root:shoot) of the variance in the phenotypes measured in our study (*PVE*, Figure 2, Table S4). For many of the measured phenotypes, a considerable proportion of the narrow sense heritability estimated previously (Table S4) was accounted for in the estimates of *PVE*. Interestingly, in the case of height, *PVE* exceeded the upper confidence interval of the estimated h^2 (Table S4). Even so, *PVE* estimates were subject to uncertainty, particularly root:shoot biomass (Figure 2), and thus, *PVE* could be larger or smaller than estimated here. Similarly, the estimates for the number of SNPs underlying the phenotype also showed uncertainty, and we acknowledge that these estimates could be larger or smaller than that estimated by the mean (Figure 2, Table S4).

To acquire estimates of *PVE* from the identification of loci with large effects on phenotype, we conducted single-locus association using univariate linear mixed models implemented in GEMMA (see Appendix, Table S2). 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

bioRxiv preprint doi: https://doi.org/10.1101/056317; this version posted May 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

0.2046 (δ^{13}C) and 0.9999 (δ^{15}N) (Table S2). Except for root:shoot biomass, the maximum likelihood estimates of *PVE* differed drastically from the estimates from BSLMM, with *PVE* never exceeding 1.08e-06 (Table S2) suggesting that a larger proportion of the heritable genetic variation for the traits measured here is explained by multiple SNPs than by individual SNPs of large effect. Finally, to determine if any relatively large-effect loci from the univariate LMMs that were 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 S10, Appendix). By this reduced threshold we identified one unique locus for both bud flush and δ^{15}N, four unique loci for both height and root:shoot biomass, and five unique loci for δ^{13}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 1/4 for both root:shoot biomass and height, and 2/5 for δ^{13}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 δ^{15}N (n = 1), 1/4 loci for root:shoot biomass, 3/4 loci for height, and 3/5 loci for δ^{13}C.

To determine if focal loci associated with phenotype by BSLMM showed signatures of an underlying polygenic architecture under selection, we estimated allele frequency covariance ($\widehat{D}_{a(ij)}$) among focal SNPs and compared these estimates to 1000 sets of SNPs each randomly sampled from H_{E} bins represented in the focal set. We found evidence for elevated covariance among the 99.9th percentile of $\overline{PIP}$ loci associated with bud flush and root:shoot biomass (Table 4), with the latter exceeding the 100th percentile of the random distribution. To consider larger numbers of loci representative of the number of underlying loci estimated by BSLMM (Figure 2b, Table S4), we also isolated SNPs from the top 99.8th percentile of $\overline{PIP}$. In this set we found evidence for elevated covariance 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 $\widehat{D}_{a(ij)}$ (Table 4).

To identify signatures of allele frequency shifts among focal loci associated with

bioRxiv preprint doi: https://doi.org/10.1101/056317; this version posted May 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 bud flush to Euclidian distances of GDD-May, GDD-Aug, percent maximum radiation input, and minimum January temperature (Table 6). 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 δ^{13}C to Euclidian distances of annual precipitation, as well as for δ^{15}N loci to elevation, and bud flush loci to both longitude (a correlate of annual precipitation) and to percent maximum radiation input (as in the 99.9th $\overline{PIP}$ set). The magnitude of the allele frequency differences across populations were subtle (range: 0.054-0.087) and representative of unassociated SNPs of similar H_E (Figure S17). Taken together, the lack of any significant large-effect loci, the low *PVE* for univariate models compared with BSLMM, the covariance of allele frequencies among associated SNPs, and the evidence for coordinated shifts in allele frequencies relative to environmental distances (particularly to bud flush, soil water availability and correlated environmental variables), our data suggests an overall polygenic basis of local adaptation.

F_{ST} outlier analysis

From the empirical set of SNPs `OutFLANK` analysis revealed 110 focal loci as outliers for F'_{ST} (range: 0.069-0.118). Expected heterozygosity values among the outlier SNPs (Figure S14) varied across nearly the entire distribution from the full set of SNPs (Figure S13). Upon analysis of patterns of covariance ($\widehat{D}_{a(ij)}$) among the `OutFLANK` focal SNPs against 1000 sets of random SNPs equal in total set size as well as within expected heterozygosity bins (the null sets), 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). Moreover, the maximum median $\widehat{D}_{a(ij)}$ from null sets corresponded to just the 12th percentile (743/5995, where $\binom{110}{2}$ = 5995) of the focal $\widehat{D}_{a(ij)}$ values, suggesting that the majority of SNPs within the focal set showed higher levels of

bioRxiv preprint doi: https://doi.org/10.1101/056317; this version posted May 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

covariance among other outliers than expected by chance. 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.

Intersection of SNPs within and across methods

We examined overlap of focal SNPs among the various methods employed in this study (Table S3). 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, there was a considerable number of loci that were found to overlap among comparisons. This seemed to be driven by the correlations among environments, for when ordered by the number of loci within the intersection, 12 of the top 15 comparisons were among edaphic conditions. Overall, `OutFLANK` captured 4 of the loci identified in the 99.8th percentile of $\overline{PIP}$, between 1-3 of the loci identified across 10/18 environmental associations from `bayenv2` (including annual precipitation), but not for any of the moderate-effect 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 loci were those identified from LMM. Very few of the loci identified by `bayenv2` would have been detected through conventional F_{ST} outlier approaches (Figures S8-9). Finally, there was little overlap among loci associated with environment with the 99.8th percentile of $\overline{PIP}$ (including two between δ^{13}C and longitude) and environmental associations did not capture any of the reduced-threshold loci from univariate LMM (Table S3).

DISCUSSION

The spatial extent of local adaptation, particularly in conifers, has generally been investigated at regional scales using single-locus perspectives (Neale and Savolainen 2004; Savolainen *et al.* 2007; Ćalić *et al.* 2016). While informative for range-wide inference, management and conservation agencies are often limited to local scales spanning only several

bioRxiv preprint doi: https://doi.org/10.1101/056317; this version posted May 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

to tens 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), particularly when the environment is highly heterogeneous and selective forces are strong. Thus, studies which investigate adaptation at scales amenable to management may be of relatively greater importance, especially for endangered and threatened species, for reforestation applications such as those carried out through seed sourcing (*sensu* McLane and Aitken 2012) and replanting efforts. Here, our results also highlight the advantage of a polygenic perspective. For instance, through this approach we found that the degrees of covariance among loci which may have been otherwise overlooked using a single locus perspective (`bayenv2` loci with $\overline{BF} < 1$) tracked environmental distance often related to water availability, a pattern corroborated by similar inferences gained from a multilocus approach to phenotypic association. Together, this evidence lends replicative support to both recent and long-standing theory for the patterns of loci underlying quantitative traits undergoing selection with gene flow.

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). Additionally, inspection of PCs showed no distinctive clustering of populations (Figure S11) while population pairwise F_{ST} did not exceed 0.016 and a test for isolation by distance was not significant. For focal SNPs identified by `bayenv2`, average allele frequency differences between populations were generally slightly larger than for the majority of SNPs in the null set (Figures S15-S16). Contrastingly, SNPs identified from `GEMMA` were roughly representative of the average allele frequency differences from the null set (Figure

bioRxiv preprint doi: https://doi.org/10.1101/056317; this version posted May 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

S17). The magnitude of allele frequency differences across populations of loci associated by `bayenv2` may have been less than those associated by `GEMMA` due to reducing the sample size to only trees contributing seedlings to the common garden. Biologically, this 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 ($\widehat{D}_{\mathrm{a(ij)}}$) and allele frequency shifts ($pw\widehat{D}_{\mathrm{a(ij)}}$) is likely suitable to investigate our dataset for signatures of polygenic adaptation.

P. albicaulis of the Lake Tahoe Basin exhibit substantial genetic variation for the fitness-related traits measured (Figure 2), suggesting that ongoing adaptation within the LTB will likely be unconstrained by the lack of genetic variation and instead by other factors such as population growth (i.e., declination) rates. A much larger proportion of the heritable variation (estimated in Maloney *et al*. *in review*) can be explained by multiple SNPs than by individual SNPs of large effect. Together our analyses suggest that selection among the LTB populations is acting through many loci of small to moderate effect (Figure 2, Table S4). As prolonged bouts of selection with gene flow are expected to result in reduced variation through the concentration of architectures with fewer, larger, and more strongly linked adaptive loci (Yeaman and Whitlock 2011), our results may also suggest that selection has been recent (relative to $4N_{\mathrm{e}}$ generations). 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 1). Such shifts in climate and local edaphic conditions in the last 10,000yr may in part explain a recent selective episode on *P. albicaulis* populations of the LTB.

bioRxiv preprint doi: https://doi.org/10.1101/056317; this version posted May 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Polygenic signals of local adaptation

Studies of tree species have described adaptive loci showing evidence of moderate to large effect on fitness-related traits (Neale and Savolainen 2004; Savolainen *et al.* 2007; Ćalić *et al.* 2016). Much of this work in this regard occurred across relatively large geographic scales and generally without further description of signals predicted from an underlying polygenic basis. In contrast, 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 short spatial scales. While varying across spatial scales, they replicate evidence for a polygenic signal of local adaptation through linkage disequilibrium among adaptive loci.

Although we expected to uncover some alleles with relatively large effect, a polygenic architecture for the measured fitness-related traits was expected, given the multifaceted and quantitative nature of these traits. With levels of high covariance of allele frequencies among loci associated with phenotypes, environment, and genetic differentiation, our results corroborate this expectation. For example, loci associated with phenotype (≥ 99.8th percentile $\overline{PIP}$) showed high covariance for all traits except for height (Table 4), while all cases with `OutFLANK` and `bayenv2`, the median levels of covariance among focal SNPs were between 1.17-10.6x greater than the 100th percentile of the null distribution (Table 2, Results). This suggests that the covariance among these identified loci was greater than expected by chance. However, the low levels of expected heterozygosity among SNPs associated with environment were unexpected (Figures S1-S2). To explore this bias, we examined uncorrected correlations among allelic

bioRxiv preprint doi: https://doi.org/10.1101/056317; this version posted May 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

frequencies and environmental measures. From this, we found no apparent pattern between expected heterozygosity and environmental correlation across the whole dataset, within the top SNPs correlated to environment (same *N* as identified by `bayenv2`), the top 500 SNPs, or through the inclusion of an over-representation of loci with large H_E among these three sets of SNPs (data not shown). This suggests that loci of greater H_E weren't excluded from association because they contained more information regarding demographic structure or that the `bayenv2` analyses were driven by statistical outliers. Even so, `bayenv2` did not identify any loci strongly associated with environment, as given from small values of Bayes factors (all $\overline{BF}$ < 1.0, Table 2), which is consistent with an underlying architecture with many loci of small effect. Yet, given the strong biological signal for adaptation to soil water availability in our dataset (discussed below), as well as evidence that other white pines within the LTB are also being structured by precipitation differences among populations (Eckert *et al*. 2015), it seems unlikely that the focal sets of SNPs are driven by false positives. However, if the majority of loci across `bayenv2` are true positives and not an artifact of the method, one possible explanation for elevated covariance among focal loci 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 could also 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.

The strongest signal for local adaptation among *P. albicaulis* populations of the LTB came from evidence of adaptation to soil water availability, as well as other environmental variables correlated with annual precipitation (e.g., longitude; see Figure 1). For example, water-use efficiency as measured by $\delta^{13}C$ was one of the phenotypes with loci that exhibited high levels of covariance (Table 4). Our results provide evidence that the covariance of allele frequencies among adaptive loci is also tracking soil moisture conditions among the studied populations. Of the six significant associations between population pairwise environmental

bioRxiv preprint doi: https://doi.org/10.1101/056317; this version posted May 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

distance and allele frequency shifts ($pw\widehat{D}_{a(ij)}$) of loci associated with phenotype, four were related to annual precipitation or longitude where δ^{13}C $pw\widehat{D}_{a(ij)}$ itself was associated with interpopulation differences in annual precipitation (Table 6), including height, which did not exhibit elevated covariance among associated loci (Table 4). Of the loci associated with environment, both annual precipitation and longitudinal $pw\widehat{D}_{a(ij)}$ were among those (n = 7) associated with the eponymous environmental distance while, of the 13 non-eponymous associations, 11 were related with annual precipitation, longitude, or soil water capacity (Table 5). Together, this evidence suggests that *P. albicaulis* populations are undergoing selection for soil water availability limits across the LTB despite high levels of gene flow.

Water availability as a driver of local adaptation

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 and Aitken 2008; Chang *et al*. 2014). Because of climatic constraints imposed on the southern range of *P. albicaulis*, phenotypic traits which are correlated to precipitation, soil water availability, or soil water capacity likely have fitness-related consequences for this species. 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 continuing selective pressures of this kind.

Past research regarding variation in δ^{13}C among conifers has shown that this trait displays substantial levels of heritability across species (Seiler and Johnson 1988; Cregg 1993; Brendel *et al.* 2002; Baltunis *et al.* 2008; Cumbie *et al.* 2011; Eckert *et al.* 2015), and consists of a polygenic architecture with constituent loci being comprised of both large and small effect (Brendel *et al.* 2002; González-Martínez *et al.* 2008; Cumbie *et al.* 2011; Marguerit *et al.* 2014). Indeed, we have found significant heritability for the measured phenotypes with the observed

bioRxiv preprint doi: https://doi.org/10.1101/056317; this version posted May 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

phenotypic variation (Q_{ST}) structured across populations (Table S4; Maloney *et al*. *in review*). Here, we have found segregating genetic variation of a polygenic architecture that explains a considerable portion of the heritability for this trait (Table S4). Additionally, there was a significant association between phenotypic differences of δ^{13}C and δ^{15}N (Table 3), which have been noted in other species such as *P. taeda* L. (Cumbie *et al*. 2011) and *P. lambertiana* (Eckert *et al*. 2015). This suggests that studied populations of *P. albicaulis* are capable of maximizing nitrogen-use efficiency under low availability of water (e.g., Livingston *et al*. 1999) and perhaps that water-use efficiency is determined through leaf assimilation to a larger extent than stomatal conductance (e.g., Prasolova *et al*. 2005). While we did not measure pleiotropy directly (as in Gompert *et al*. 2015), and despite the association, δ^{13}C and δ^{15}N shared just one focal locus (Table S3). Because this locus was not one of those identified from the relaxed thresholds for the univariate LMM, it is possible that this locus has a minor effect on either trait.

Concluding remarks, limitations, and future work

The results reported here suggest that the genetic architecture for variation in fitness-related phenotypic traits of *P. albicaulis* consists of numerous loci of small to moderate effects, that these loci show higher covariance than expected by chance, and that this covariance is often associated with interpopulation levels of soil water availability. Our results further explain a considerable portion (*PVE*) of the additive genetic variation (h^2) of the quantitative traits under study from a polygenic perspective. 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.

While we described associations among genotype, phenotype, and environment that reflect

bioRxiv preprint doi: https://doi.org/10.1101/056317; this version posted May 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

strong evidence for adaptive responses of *P. albicaulis* populations to the environment, we acknowledge several shortcomings. 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 and Whitlock 2015). 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 is most likely a major component of total lifetime fitness. Third, much of the statistical signal for the association of allele frequency shifts to environment would be lost with correction for multiple tests yet we leverage the fact that, of the few significant associations, the majority were related to $\delta^{13}C$, annual precipitation, its correlate of longitude, or soil water capacity, 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. Reciprocal transplants would have allowed us to differentiate pleiotropic effects and facilitate direct measures of fitness through survival and growth across environments. 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, (non)synonymous mutations, 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. Because of this, we cannot confidently estimate the proportion of the polymorphism due to coding and non-coding sites nor conclude that the relatively small effects inferred for focal loci are not an artifact due to distant linkage with causative sites of larger effect. Future work could address these shortcomings and lead to the corroboration of our results, particularly

bioRxiv preprint doi: https://doi.org/10.1101/056317; this version posted May 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

in describing patterns exhibited by underlying loci in other systems. However, while considered an inconsistency by some (Ćalić *et al.* 2016), we have doubts as to whether the loci of small effect uncovered here would show consistent discovery across conifer species, for when adaptation occurs in the face of gene flow the architecture itself is often transient (Yeaman 2015), as no locus makes any considerable contribution for an extended period of time. Even for such cases of loci of large effect, the genetic architecture of local adaptation can be transient if genetic redundancies and effective mutation rates are high.

Lastly, we highlight calls from others (e.g., Pritchard and di Rienzo 2010; Sork *et al.* 2013; Csilléry et al. 2014; Tiffin and Ross-Ibarra 2014; Hornoy *et al*. 2015; Stephan 2015; Tigano and Friesen 2016) for future investigations into the genetic architectures of local adaptation of fitness-related traits to continue to address a multilocus perspective and we relate this to a recent meeting conferring forest tree genomic scientists (Groover 2015). There, delegates identified challenges facing the advancement of insight into forest biology. While improvements in computational, genomic, and sequencing technologies will continue to aid the capacity of research, delegates specified that it would be the untested hypotheses that will bring about the most fruitful insight. However, for this to take place, empiricism will need to continue to test hypotheses currently at hand, as well as to develop and improve our overall evolutionary understanding. With this, advances in the expectations of polygenic adaptation through theory (of e.g., the effect of recombination, LD, genomic networks, as well as anisotropic gene flow or selection pressures on the prediction of transient dynamics of small-effect alleles and genomic clustering via rearrangement) will inform study and sampling design across spatial scales and provide new and interesting models with which to contrast to evolving populations in nature.

bioRxiv preprint doi: https://doi.org/10.1101/056317; this version posted May 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Acknowledgements

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.

bioRxiv preprint doi: https://doi.org/10.1101/056317; this version posted May 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

REFERENCES

Allen C. D., Macalady A. K., Chenchouni H., Bachelet D., McDowell N., Vennetier M., Kitzberger T., Rigling A., Breshears D. D., Hogg E. H. T., Gonzalez P., Fensham R., Zhang Z., Castro J., Demidova N., Lim J.-H., Allard G., Running S. W., 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.

Baltunis B. S., Martin T. A., Huber D. A., Davis J. M., 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.

Barrett R., Schluter D., 2008 Adaptation from standing genetic variation. Trends in Ecology & Evolution **23**: 38–44.

Barton N. H., 1999 Clines in polygenic traits. Genetical Research **74**: 223–236.

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 A. D., Aitken S. N., 2008 Ecological genetics and seed transfer guidelines for *Pinus albicaulis* (Pinaceae). American Journal of Botany **95**: 66–76.

Brendel O., Pot D., Plomion C., Rozenberg P., Guehl J. M., 2002 Genetic parameters and QTL analysis of δ13C and ring width in maritime pine. Plant, Cell & Environment **25**: 945–953.

Browning B. L., Browning S. R., 2016 Genotype imputation with millions of reference samples. The American Journal of Human Genetics **98**: 116–126.

Budde K. B., Heuertz M., Hernández-Serrano A., Pausas J. G., Vendramin G. G., Verdú M.,

bioRxiv preprint doi: https://doi.org/10.1101/056317; this version posted May 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

González-Martínez S. C., 2014 In situgenetic association for serotiny, a fire-related trait, in Mediterranean maritime pine (*Pinus pinaster*). New Phytologist **201**: 230–241.

Ćalić I., Bussotti F., Martínez-García P. J., Neale D. B., 2015 Recent landscape genomics studies in forest trees. Tree Genetics & Genomes **12**: 1–7.

Chang T., Hansen A. J., 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.

Comeault A. A., Flaxman S. M., Riesch R., Curran E., Soria-Carrasco V., Gompert Z., Farkas T. E., Muschick M., Parchman T. L., Schwander T., Slate J., Nosil P., 2015 Selection on a genetic polymorphism counteracts ecological speciation in a stick insect. Current Biology **25**: 1975–1981.

Comeault A. A., Soria-Carrasco V., Gompert Z., Farkas T. E., Buerkle C. A., Parchman T. L., Nosil P., 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 J. K., 2010 Using environmental correlations to identify loci underlying local adaptation. Genetics **185**: 1411–1423.

Cregg B. M., 1993 Seed-source variation in water relations, gas exchange, and needle morphology of mature ponderosa pine trees. Canadian Journal of Forest Research **23**: 749–755.

Critchfield W. B., Little E. L., 1966 Geographic distribution of the pines of the world (No. 991). US Department of Agriculture, Forest Service.

Csilléry K., Lalagüe H., Vendramin G. G., González-Martínez S. C., Fady B., Oddou-Muratorio

bioRxiv preprint doi: https://doi.org/10.1101/056317; this version posted May 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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 W. P., Eckert A. J., 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.

Daly C., Neilson R. P., Phillips D. L., 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 C. A., Banks E., DePristo M. A., Handsaker R. E., Lunter G., Marth G. T., Sherry S. T., others, 2011 The variant call format and VCFtools. Bioinformatics **27**: 2156–2158.

East E. M., 1910 A Mendelian interpretation of variation that is apparently continuous. The American Naturalist **44**: 65–82.

Eckert A. J., Maloney P. E., Vogler D. R., Jensen C. E., Mix A. D., Neale D. B., 2015 Local adaptation at fine spatial scales: an example from sugar pine (*Pinus lambertiana, Pinaceae*). Tree Genetics & Genomes **11**: 42.

Eckert A. J., van Heerwaarden J., Wegrzyn J. L., Nelson C. D., Ross-Ibarra J., González-Martínez S. C., Neale D. B., 2010 Patterns of population structure and environmental associations to aridity across the range of loblolly pine (*Pinus taeda* L., *Pinaceae*). Genetics **185**: 969–982.

Eckert A. J., Wegrzyn J. L., Pande B., Jermstad K. D., Lee J. M., Liechty J. D., Tearse B. R.,

bioRxiv preprint doi: https://doi.org/10.1101/056317; this version posted May 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Krutovsky K. V., Neale D. B., 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.

Ehret G. B., Lamparter D., Hoggart C. J., Whittaker J. C., Beckmann J. S., Kutalik Z., Consortium G. I. O. A. T., 2012 A multi-SNP locus-association method reveals a substantial fraction of the missing heritability. The American Journal of Human Genetics **91**: 863–871.

Ellison A. M., Bank M. S., Barton D. C., Colburn E. A., Elliot K., Ford C. R., Foster D. R., Kloeppel B. D., Knoepp J. D., Lovett G. M., Mohan J., 2005 Loss of foundation species: consequences for the structure and dynamics of forested ecosystems. Frontiers in Ecology and the Environment **3**: 479-486.

Endler J. A., 1977 Geographic variation, speciation, and clines. No. 10, Princeton University Press.

Farnes P. E., 1990 SNOTEL and snow course data: describing the hydrology of whitebark pine ecosystems. General Technical Report INT (USA).

Fisher R. A., 1918 The correlation between relatives on the supposition of Mendelian inheritance. Transactions of the Royal Society of Edinburgh **52**: 399–433.

Fisher R. A., 1930 *The genetical theory of natural selection*. Oxford University Press.

Gladman S., Seeman T., 2016 Velvet Optimiser v2.5.5, available online http://bioinformatics.net.au/software.velvetoptimiser.shtml.

Gompert Z., Jahner J. P., Scholl C. F., Wilson J. S., Lucas L. K., Soria-Carrasco V., Fordyce J. A., Nice C. C., Buerkle C. A., Forister M. L., 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–

bioRxiv preprint doi: https://doi.org/10.1101/056317; this version posted May 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

2793.

González-Martínez S. C., Huber D., Ersoz E., Davis J. M., Neale D. B., 2008 Association genetics in *Pinus taeda* L. II. Carbon isotope discrimination. Heredity **101**: 19–26.

Groover A., 2015 Genomic science provides new insights into the biology of forest trees. New Phytologist **208**: 302–305.

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 R. J., Payseur B. A., 2016 Fifteen years of genomewide scans for selection: trends, lessons and unaddressed genetic sources of complication. Molecular Ecology **25**: 5–23.

Hall D., Luquez V., Garcia V. M., St Onge K. R., Jansson S., Ingvarsson P. K., 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.

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.

Holland J., 2007 Genetic architecture of complex traits in plants. Current Opinion in Plant

bioRxiv preprint doi: https://doi.org/10.1101/056317; this version posted May 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Biology **10**: 156–161.

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.

Hutchins H. E., Lanner R. M., 1982 The central role of Clark's nutcracker in the dispersal and establishment of whitebark pine. Oecologia **55**: 797-807.

Kemper K. E., Saxton S. J., Bolormaa S., Hayes B. J., Goddard M. E., 2014 Selection for complex traits leaves little or no classic signatures of selection. BMC Genomics **15**: 246–260.

Krakowski J., Aitken S. N., El-Kassaby Y. A., 2003 Inbreeding and conservation genetics in whitebark pine. Conservation Genetics **4**: 581–593.

Kremer A., Le Corre V., 2012 Decoupling of differentiation between traits and their underlying genes in response to divergent selection. Heredity **108**: 375–385.

Langlet O., 1971 Two hundred years of genecology. Taxon: 653–721.

Langmead B., Salzberg S. L., 2012 Fast gapped-read alignment with Bowtie 2. Nature Methods **9**: 357–359.

Latta R. G., 1998 Differentiation of allelic frequencies at quantitative trait loci affecting locally adaptive traits. The American Naturalist **151**: 283–292.

Latta R. G., 2003 Gene flow, adaptive population divergence and comparative population structure across loci. New Phytologist **161**: 51–58.

Le Corre V., Kremer A., 2003 Genetic variability at neutral markers, quantitative trait loci and

bioRxiv preprint doi: https://doi.org/10.1101/056317; this version posted May 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

trait in a subdivided population under selection. Genetics **164**: 1205–1219.

Le Corre V., Kremer A., 2012 The genetic differentiation at quantitative trait loci under local adaptation. Molecular Ecology **21**: 1548–1566.

Leimu R., Fischer M., 2008 A meta-analysis of local adaptation in plants. PLoS ONE **3**: e4010.

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.

Livingston N. J., Guy R. D., Sun Z. J., 1999 The effects of nitrogen stress on the stable carbon isotope composition, productivity and water use efficiency of white spruce (*Picea glauca* (Moench) Voss). Plant **22**: 281–289.

Lotterhos K. E., Whitlock M. C., 2015 The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. Molecular Ecology **24**: 1031–1046.

Luquez V., Hall D., Albrectsen B. R., Karlsson J., Ingvarsson P., Jansson S., 2007 Natural phenological variation in aspen (*Populus tremula*): the SwAsp collection. Tree Genetics & Genomes **4**: 279–292.

Ma X. F., Hall D., Onge K. R. S., Jansson S., Ingvarsson P. K., 2010 Genetic differentiation, clinal variation and phenotypic associations with growth cessation across the *Populus tremula* photoperiodic pathway. Genetics **186**: 1033–1044.

Mackay T. F. C., Stone E. A., Ayroles J. F., 2009 The genetics of quantitative traits: challenges and prospects. Nature Reviews Genetics **10**: 565–577.

bioRxiv preprint doi: https://doi.org/10.1101/056317; this version posted May 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Mahalovich M. F., Stritch L., 2013 *Pinus albicaulis*. The IUCN Red List of Threatened Species. e.T39049A2885918. http://dx.doi.org/10.2305/IUCN.UK.2013-1.RLTS.T39049A2885918.en.

Maloney P. E., Vogler D. R., Eckert A. J., Jensen C. E., 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.

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., 2014 The genetics of water-use efficiency and its relation to growth in maritime pine. Journal of Experimental Botany **65**: 4757–4768.

McKay J. K., Latta R. G., 2002 Adaptive population divergence: markers, QTL and traits. Trends in Ecology & Evolution **17**: 285–291.

McKinney S. T., Fiedler C. E., Tomback D. F., 2009 Invasive pathogen threatens bird-pine mutalism: implications for sustaining a high-elevation ecosystem. Ecological Applications **19**: 597-607.

McLane S. C., Aitken S. N., 2012 Whitebark pine (*Pinus albicaulis*) assisted migration potential: testing establishment north of the species range. Ecological Applications **22**: 142–153.

Mitton J. B., Grant M. C., Yoshino A. M., 1998 Variation in allozymes and stomatal size in pinyon (*Pinus edulis*, Pinaceae), associated with soil moisture. American Journal of Botany **85**: 1262–1265.

Mitton J. B., Stutz H. P., Schuster W. S., 1989 Genotypic differentiation at PGM in Engelmann spruce from wet and dry sites. Silvae Genetica **38**: 217–221.

bioRxiv preprint doi: https://doi.org/10.1101/056317; this version posted May 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Moser G., Lee S. H., Hayes B. J., Goddard M. E., Wray N. R., Visscher P. M., 2015 Simultaneous Discovery, Estimation and Prediction Analysis of Complex Traits Using a Bayesian Mixture Model. PLoS Genetics **11**: e1004969.

Nowak C. L., Nowak R. S., Tausch R. J., Wigand P. E., 1994 A 30000 year record of vegetation dynamics at a semi-arid locale in the Great Basin. Journal of Vegetation Science **5**: 579-590.

Neale D. B., Savolainen O., 2004 Association genetics of complex traits in conifers. Trends in Plant Science **9**: 325–330.

Nilsson-Ehle H., 1909 Kreuzungsuntersuchungen an Hafer und Weizen. Lunds Universitets Arsskrift **5**: 1–122.

Ohta T., 1982 Linkage disequilibrium with the island model. Genetics **101**: 139–155.

Parchman T. L., Gompert Z., Mudge J., Schilkey F. D., Benkman C. W., Buerkle C. A., 2012 Genome-wide association genetics of an adaptive trait in lodgepole pine. Molecular Ecology **21**: 2991–3005.

Pérez F., Granger B. E., 2007 IPython: A system for interactive scientific computing. Computing in Science and Engineering **9**: 21-29.

Patterson N., Price A. L., Reich D., 2006 Population structure and eigenanalysis. PLoS Genetics **2**: e190.

Peterson B. K., Weber J. N., Kay E. H., Fisher H. S., Hoekstra H. E., 2012 Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PLoS ONE **7**: e37135.

bioRxiv preprint doi: https://doi.org/10.1101/056317; this version posted May 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Prasolova N. V., Xu Z. H., Lundkvist K., 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.

Pritchard J. K., Di Rienzo A., 2010 Adaptation – not by sweeps alone. Nature Reviews Genetics **11**: 665–667.

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/.

Richardson B. A., Brunsfeld S. J., Klopfenstein N. B., 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 J. L., Urban M. C., Bolnick D. I., Skelly D. K., 2014 Microgeographic adaptation and the spatial scale of evolution. Trends in Ecology & Evolution **29**: 165–176.

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.

Schoville S. D., Bonin A., Francois O., Lobreaux S., Melodelima C., Manel S., 2012 Adaptive Genetic Variation on the Landscape: Methods and Cases. Annual Review of Ecology, Evolution, and Systematics **43**: 23–43.

Seiler J. R., Johnson J. D., 1988 Physiological and morphological responses of three half-sib families of loblolly pine to water-stress conditioning. Forest Science **34**: 487–495.

bioRxiv preprint doi: https://doi.org/10.1101/056317; this version posted May 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Sork V. L., Aitken S. N., Dyer R. J., Eckert A. J., Legendre P., Neale D. B., 2013 Putting the landscape into the genomics of trees: approaches for understanding local adaptation and population responses to changing climate. Tree Genetics & Genomes **9**: 901–911.

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.

Storey, J. D., Bass A. J., Dabney A., Robinson D., 2015 qvalue: Q-value estimation for false discovery rate control. R package version 2.4.2.

Storz J. F., 2005 Using genome scans of DNA polymorphism to infer adaptive population divergence. Molecular Ecology **14**: 671–688.

Storz J. F., Kelly J. K., 2008 Effects of spatially varying selection on nucleotide diversity and linkage disequilibrium: Insights from deer mouse globin genes. Genetics **180**: 367–379.

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 V. L., 2016 Genomics of local adaptation with gene flow. Molecular Ecology AOP: 1-21.

Tomback D. F., 1982 Dispersal of whitebark pine seeds by Clark's nutcracker: a mutualism hypothesis. The Journal of Animal Ecology **51**: 451–467.

Tomback D. F., Achuff P., 2010 Blister rust and western forest biodiversity: ecology, values and outlook for white pines. Forest Pathology **40**: 186-225.

Tomback D. F., Arno, S.F, Keane R. E., 2001 Whitebark pine communities: ecology and restoration. Island Press, Washington, DC.

bioRxiv preprint doi: https://doi.org/10.1101/056317; this version posted May 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Tomback D. F., Resler L. M., Keane R. E., Pansing E. R., Andrade A. J., Wagner A. C., 2016 Community structure, biodiversity, and ecosystem services in treeline whitebark pine communities: potential impacts from a non-native pathogen. Forests **7**: 1-8.

Turchin M. C., Chiang C. W., Palmer C. D., Sankararaman S., Reich D., Hirschhorn J. N., 2012 Evidence of widespread selection on standing variation in Europe at height-associated SNPs. Nature Genetics **44**: 1015–1019.

USDA Forest Service, *Forest Health Technology Enterprise Team*. Ft. Collins, CO.

USDA NRCS, 2007 *Soil survey of the Tahoe Basin area, California and Nevada*.

van Mantgem P. J., Stephenson N. L., Byrne J. C., Daniels L. D., Franklin J. F., Fulé P. Z., Harmon M. E., Larson A. J., Smith J. M., Taylor A. H., Veblen T. T., 2009 Widespread increase of tree mortality in the western United States. Science **323**: 521-524.

Vicente-Serrano S. M., Gouveia C., Camarero J. J., Beguería S., Trigo R., López-Moreno J. I., 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.

Vizcaíno-Palomar N., Revuelta-Eugercios B., Zavala M. A., Alia R., González-Martínez S. C., 2014 The Role of Population Origin and Microenvironment in Seedling Emergence and Early Survival in Mediterranean Maritime Pine (*Pinus pinaster* Aiton) (S Delzon, Ed.). PLoS ONE **9**: e109132.

Weir B. C., Cockerham C. C., 1984 Estimating F-statistics for the analysis of population structure. Evolution: 1358-1370.

Whitlock M. C., Lotterhos K. E., 2015 Reliable detection of loci responsible for local adaptation:

bioRxiv preprint doi: https://doi.org/10.1101/056317; this version posted May 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Inference of a null model through trimming the distribution of FST. The American Naturalist **186**: S24–S36.

Yeaman S., 2015 Local adaptation by alleles of small effect. The American Naturalist **186**: S74-S89.

Yeaman S., Whitlock M. C., 2011 The genetic architecture of adaptation under migration-selection balance. Evolution **65**: 1897–1911.

Zerbino D. R., Birney E., 2008 Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. Genome Research **18**: 821–829.

Zhou X., Carbonetto P., Stephens M., 2013 Polygenic modeling with bayesian sparse linear mixed models. PLoS Genetics **9**: e1003264.

bioRxiv preprint doi: https://doi.org/10.1101/056317; this version posted May 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

TABLES

Table 1

	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 ($cmol_c \cdot kg^{-1}$)	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-⅓ bar (kPa)	9.7	8.0	8.4	7.3	9.8	11.4	14.4	6.2

TABLE 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-⅓bar – water capacity at -⅓bar (field capacity). Asterisks indicates populations from which seeds sampled from cones were planted in a common garden. Environmental variables are averaged across subplots.*

bioRxiv preprint doi: https://doi.org/10.1101/056317; this version posted May 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 2

Environmental Variable	N	$\overline{BF}$ (range)	$\overline{\rho_S}$ (range)	Median Focal $\widehat{D}_{a(ij)}$	Max. Random $\widehat{D}_{a(ij)}$	Perc. Focal < Max. Rand.
Annual Precipitation**	49	0.086–0.173	0.159–0.311	5.94e-04	1.17e-04	1.4 ($\binom{N}{2}$ = 1176)
AWS0-25**	95	0.088–0.158	0.159–0.267	2.05e-04	6.84e-05	11.9 (4465)
AWS0-50**	147	0.089–0.216	0.162–0.276	1.97e-04	6.70e-05	12.6 (10731)
CEC**	14	0.086–0.152	0.228–0.345	3.75e-04	1.63e-04	8.8 (91)
Clay**	22	0.086–0.186	0.224–0.296	1.49e-04	1.27e-04	43.7 (231)
Elevation**	143	0.096–0.325	0.173–0.269	2.60e-04	8.12e-05	6.4 (10153)
GDD-Aug**	157	0.096–0.286	0.190–0.283	4.72e-04	1.13e-04	5.9 (12246)
GDD-May**	80	0.096–0.344	0.172–0.282	2.40e-04	9.60e-05	15.6 (3160)
Latitude**	119	0.091–0.193	0.161–0.246	2.00e-04	8.05e-05	15.0 (7021)
Longitude**	67	0.087–0.175	0.138–0.255	2.52e-04	9.41e-05	17.6 (2211)
Max. solar rad. Input**	144	0.090–0.361	0.246–0.318	2.33e-04	1.03e-04	21.7 (10296)
Max. Temp – July**	50	0.088–0.178	0.222–0.328	3.02e-04	8.37e-05	5.8 (1225)
Min. Temp – Jan.**	116	0.092–0.289	0.177–0.280	6.47e-04	1.28e-05	3.0 (6670)
Rock coverage**	143	0.089–0.186	0.173–0.313	3.56e-04	9.38e-05	7.6 (10153)
Sand**	111	0.090–0.206	0.167–0.254	2.11e-04	8.62e-05	18.1 (6105)
Silt**	140	0.091–0.249	0.162–0.248	2.10e-04	7.89e-05	14.0 (9730)
WC-15bar**	86	0.089–0.297	0.150–0.242	1.92e-04	8.07e-05	17.7 (3655)
WC-⅓bar**	97	0.088–0.247	0.147–0.251	2.17e-04	1.03e-05	22.7 (4656)

TABLE 2 *Results from genotype-environment association with* `bayenv2`*. Environmental variables as in Table 1. N = number of outlier loci identified by the top 0.5% harmonic mean Bayes factor ($\overline{BF}$) and top 1.0% values for harmonic mean Spearman's rho ($\overline{\rho_S}$); range of $\overline{BF}$ or $\overline{\rho_S}$ refer to minimum and maximum values from outlier loci; Median Focal $\widehat{D}_{a(ij)}$ = median $\widehat{D}_{a(ij)}$ among outlier loci; Max. Random $\widehat{D}_{a(ij)}$ = greatest value of median $\widehat{D}_{a(ij)}$ values calculated from 1000 sets of random non-outlier SNPs of equal N and similar H_E; Perc. Focal < Max. Rand. = percentile of $\binom{N}{2}$ comparisons at which the Median Focal $\widehat{D}_{a(ij)}$ is less than Max. Random $\widehat{D}_{a(ij)}$. ** focal $\widehat{D}_{a(ij)}$> 100th percentile of random $\widehat{D}_{a(ij)}$.*

bioRxiv preprint doi: https://doi.org/10.1101/056317; this version posted May 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1086 ***Table 3***

Measure	Comparison		Mantel's *r*	*p*-value
Environment	AWS0-50	AWS0-25	0.9256	0.0020
	Clay	CEC	0.9424	0.0040
	Latitude	Elevation	0.3988	0.0490
	Longitude	Ann-ppt	0.7145	0.0030
	Max-rad-input	Latitude	0.4629	0.0370
	Sand	AWS0-50	0.4393	0.0170
	Sand	Clay	0.3723	0.0360
	Silt	AWS0-25	0.5691	0.0280
	Silt	AWS0-50	0.7552	0.0060
	Silt	Sand	0.8395	0.0010
	Tmin-Jan	GDD-Aug	0.4545	0.0470
	WC-15Bar	AWS0-25	0.6722	0.0170
	WC-15Bar	AWS0-50	0.8511	0.0030
	WC-15Bar	Silt	0.6807	0.0250
	WC-⅓bar	AWS0-25	0.6093	0.0190
	WC-⅓bar	AWS0-50	0.7761	0.0060
	WC-⅓bar	Silt	0.5582	0.0290
	WC-⅓bar	WC-15Bar	0.9423	0.0020
Phenotype	$\delta^{13}C$	$\delta^{15}N$	0.5278	0.0238
	$\delta^{15}N$	Height	0.6327	0.0198

1087 **TABLE 3** Environmental and phenotypic correlations. *Significant results from Mantel tests (9999 iterations) between pairwise environmental or*
1088 *phenotypic distance matrices calculated from centered and standardized measures. All other pairwise combinations not listed were found to have*
1089 *insignificant associations (p > 0.05).*

bioRxiv preprint doi: https://doi.org/10.1101/056317; this version posted May 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 4

Selection Criterion	Phenotype	N Loci	Median Focal $\widehat{D}_{a(ij)}$	Max. Random $\widehat{D}_{a(ij)}$	Perc. Focal < Max. Rand.
≥99.9th percentile of $\overline{PIP}$	Bud Flush*	118	0.001398	0.001526	53.2 ($\binom{N}{2}$ = 6903)
	δ^{13}C	122	0.000646	0.000833	58.6 (7381)
	Height	117	0.000512	0.000609	56.4 (6786)
	δ^{15}N	117	0.000583	0.000709	56.9 (6786)
	Root:Shoot**	119	0.001868	0.001720	46.6 (7021)
≥99.8th percentile of $\overline{PIP}$	Bud Flush*	232	0.001433	0.001568	53.7 (26796)
	δ^{13}C*	232	0.001490	0.001583	52.4 (26796)
	Height	232	0.001457	0.001650	54.9 (26796)
	δ^{15}N**	232	0.001566	0.001564	47.7 (26796)
	Root:Shoot**	232	0.001820	0.001610	45.3 (26796)

TABLE 4 Results from genotype-phenotype associations. Covariation among *N* adaptive loci identified from the top percentiles of the harmonic mean posterior inclusion probability ($\overline{PIP}$) estimated from GEMMA. Column names analogous to those as in Table 2. ** focal $\widehat{D}_{a(ij)}$ > 95th percentile of random $\widehat{D}_{a(ij)}$; ** focal $\widehat{D}_{a(ij)}$ > 100th percentile of random $\widehat{D}_{a(ij)}$.*

bioRxiv preprint doi: https://doi.org/10.1101/056317; this version posted May 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

1094 ***Table 5***

$pw\widehat{D}_{a(ij)}$	Environmental Euclidian Distance	Mantel's *r*	*p*-value
Ann-ppt	Ann-ppt	0.7135	0.0027
GDD-May	GDD-May	0.8480	0.0013
Longitude	Longitude	0.6522	0.0024
Rock-cov	Rock-cov	0.5124	0.0145
Sand	Sand	0.5574	0.0046
Tmin-Jan	Tmin-Jan	0.5791	0.0137
WC-⅓bar	WC-⅓bar	0.4806	0.0361
Longitude	Ann-ppt	0.7716	0.0016
Rock-cov	Ann-ppt	0.5542	0.0221
Tmin-Jan	Ann-ppt	0.5765	0.0132
Longitude	Latitude	-0.4257	0.0347
Rock-cov	Longitude	0.5566	0.0284
Tmin-Jan	Longitude	0.4822	0.0273
Sand	Clay	0.5345	0.0232
Silt	Sand	0.4408	0.0238
WC-15bar	Tmax-July	0.3490	0.0309
WC-⅓bar	Tmax-July	0.4539	0.0037
WC-⅓bar	AWS0-25	0.4329	0.0384
WC-⅓bar	AWS0-50	0.4538	0.0464
WC-⅓bar	WC-15bar	0.5126	0.0335

1095 **TABLE 5** Signatures of allele frequency shifts associated with environmental distance. *Significant Mantel tests (9999 permutations) from*
1096 *comparisons among* $pw\widehat{D}_{a(ij)}$ *matrices from SNPs associated with environment against environmental Euclidian distance. Above the grey line are*
1097 *significant associations among eponymous comparisons, while below the gray line are significant associations among the remaining permutations.*
1098 *Environmental variables as in Table 1.*

bioRxiv preprint doi: https://doi.org/10.1101/056317; this version posted May 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 6

Selection criterion	$pw\widehat{D}_{a(ij)}$	Environmental Euclidian Distance	Mantel's *r*	*p*-value
99.9th $\overline{PIP}$	Bud flush	GDD-Aug	0.5804	0.0181
	Bud flush	GDD-May	-0.5190	0.0458
	Bud flush	Max rad. input	-0.5486	0.0482
	Bud flush	Tmin-Jan	0.6984	0.0191
99.8th $\overline{PIP}$	Bud flush	Ann-ppt	0.4309	0.0140
	Bud flush	Longitude	0.5532	0.0405
	Bud flush	Max rad. input	-0.6312	0.0127
	$\delta^{15}N$	Elevation	0.5334	0.0246
	Height	Ann-ppt	0.7210	0.0320
	$\delta^{13}C$	Ann-ppt	0.5952	0.0195

TABLE 6 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 against environmental Euclidian distance. Selection Criterion refers to the process used to identify SNPs associated with phenotype.*

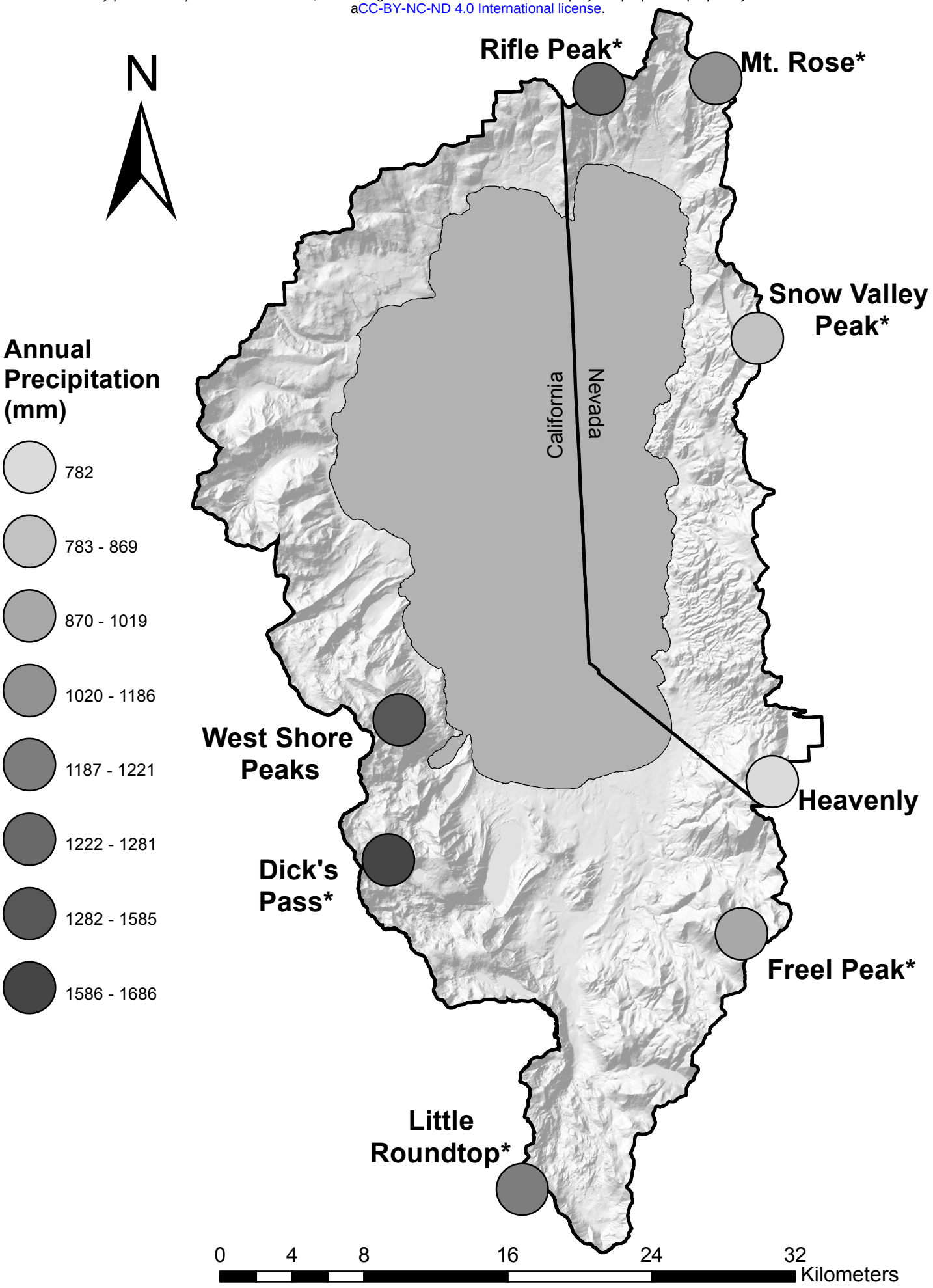
bioRxiv preprint doi: https://doi.org/10.1101/056317; this version posted May 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

FIGURE LEGENDS

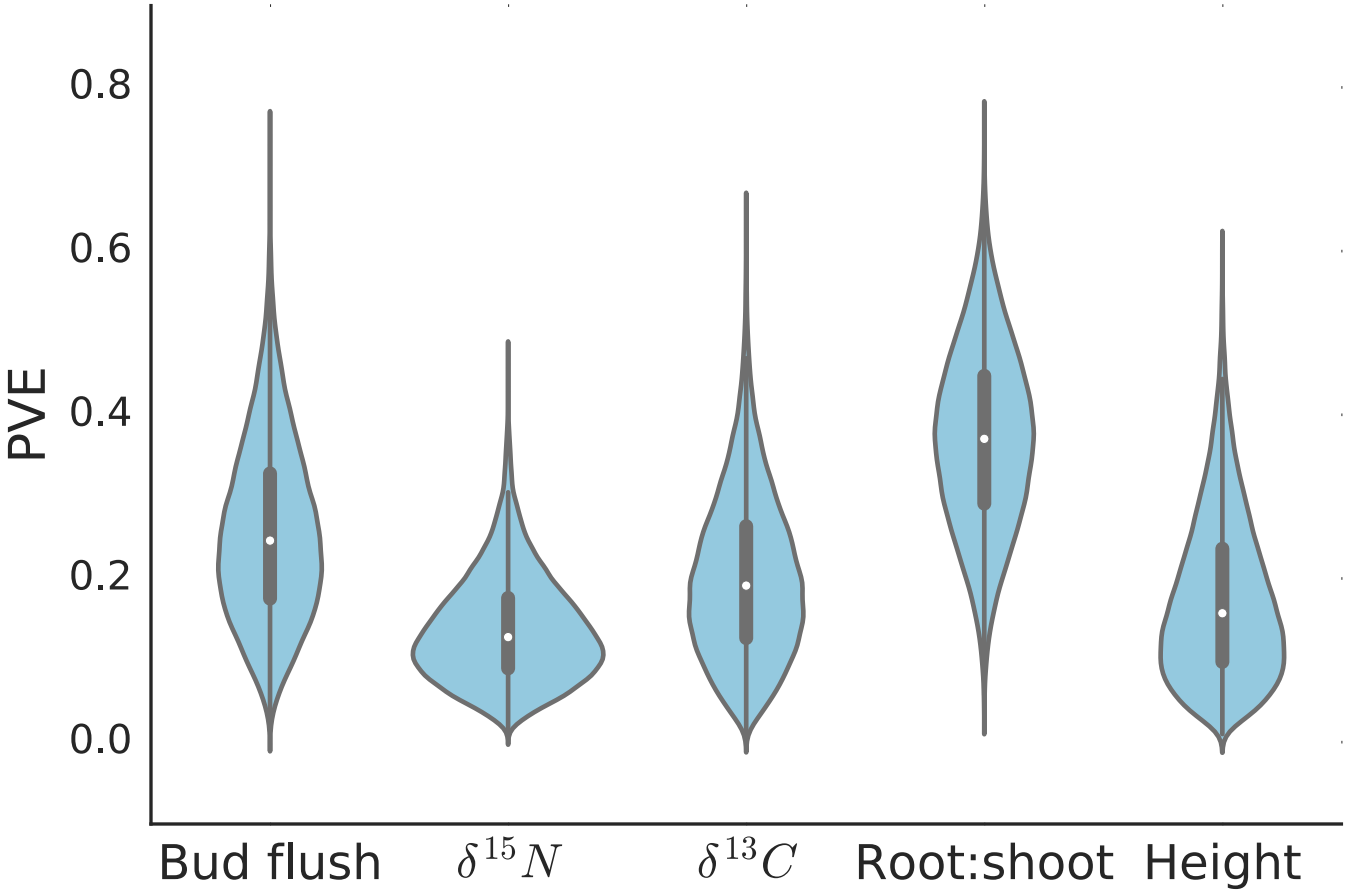
Figure 1 *Populations used for sampling P. albicaulis within the Lake Tahoe Basin (dark outline). Annual precipitation is given for each population to demonstrate the west-east rain shadow experienced across short spatial scales. Asterisks indicate populations in the common garden study.*

Figure 2 *Violin plots for the kernel density estimator of the posterior distributions (blue) taken from* `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. Grey vertical bars display the first through third interquartile range with the median represented by the white dot.*

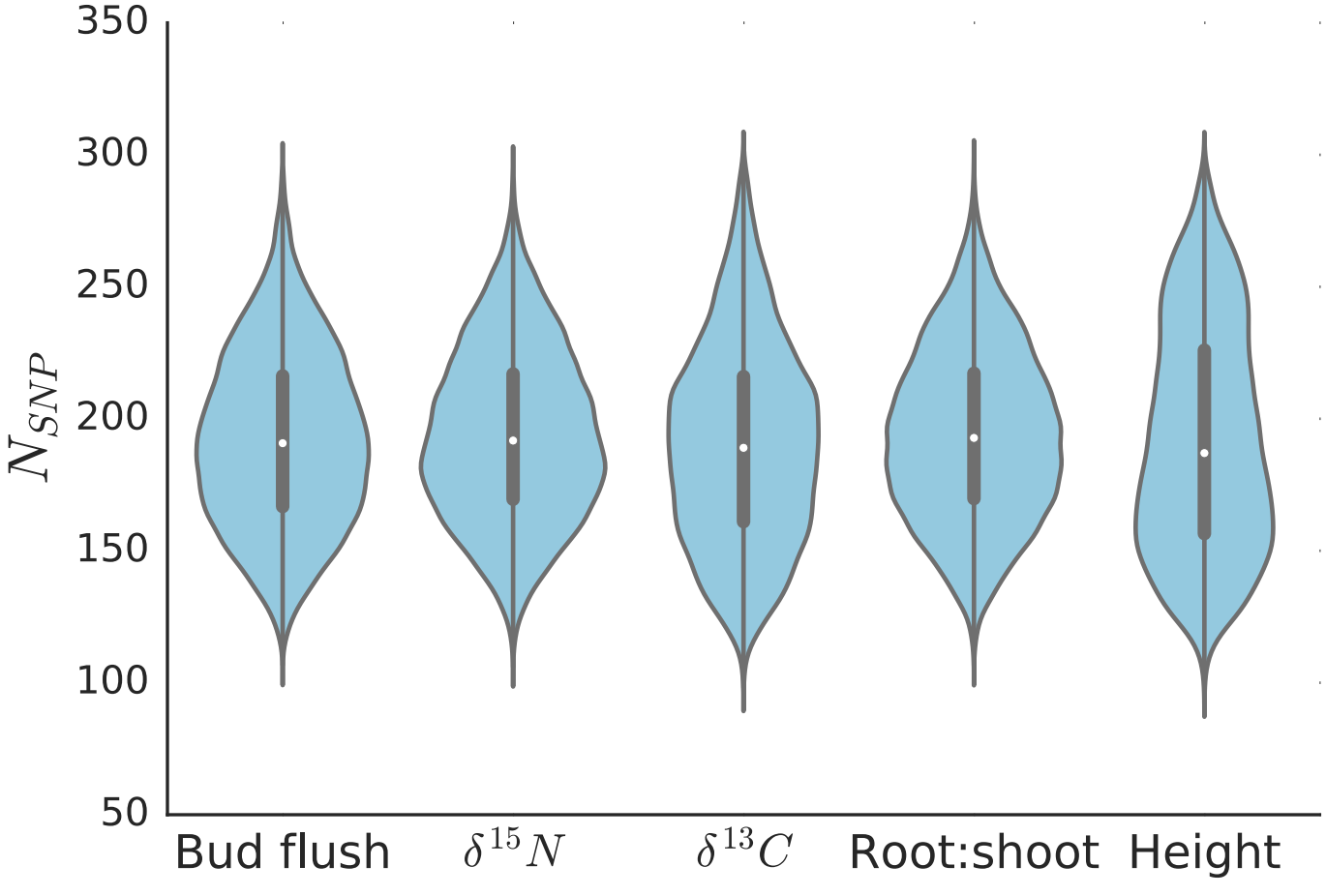
bioRxiv preprint doi: https://doi.org/10.1101/056317; this version posted May 31, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



A


PVE
0.0
0.2
0.4
0.6
0.8
Bud flush
$\delta^{15}N$
$\delta^{13}C$
Root:shoot
Height

B


350
300
250
200
150
100
50
N_{SNP}
Bud flush
$\delta^{15}N$
$\delta^{13}C$
Root:shoot
Height