

ORIGINAL RESEARCH

Association of genetic and climatic variability in giant sequoia, *Sequoiadendron giganteum*, reveals signatures of local adaptation along moisture-related gradients

Rainbow DeSilva  | Richard S. Dodd

Department of Environmental Science, Policy, and Management, University of California at Berkeley, Berkeley, California

Correspondence

Rainbow DeSilva, Department of Environmental Science, Policy, and Management, University of California at Berkeley, Berkeley, CA 94720.
Email: rainbow222@berkeley.edu

Funding information

U.S. Department of Agriculture, Grant/Award Number: CA-B-ECO-0165-MS

Abstract

Uncovering the genetic basis of local adaptation is a major goal of evolutionary biology and conservation science alike. In an era of climate change, an understanding of how environmental factors shape adaptive diversity is crucial to predicting species response and directing management. Here, we investigate patterns of genomic variation in giant sequoia, an iconic and ecologically important tree species, using 1,364 bi-allelic single nucleotide polymorphisms (SNPs). We use an F_{ST} outlier test and two genotype–environment association methods, latent factor mixed models (LFMMs) and redundancy analysis (RDA), to detect complex signatures of local adaptation. Results indicate 79 genomic regions of potential adaptive importance, with limited overlap between the detection methods. Of the 58 loci detected by LFMM, 51 showed strong correlations to a precipitation-driven composite variable and seven to a temperature-related variable. RDA revealed 24 outlier loci with association to climate variables, all of which showed strongest relationship to summer precipitation. Nine candidate loci were indicated by two methods. After correcting for geographic distance, RDA models using climate predictors accounted for 49% of the explained variance and showed significant correlations between SNPs and climatic factors. Here, we present evidence of local adaptation in giant sequoia along gradients of precipitation and provide a first step toward identifying genomic regions of adaptive significance. The results of this study will provide information to guide management strategies that seek to maximize adaptive potential in the face of climate change.

KEYWORDS

climate change, genotype by sequencing, giant sequoia, landscape genomics, local adaptation

1 | INTRODUCTION

In an era of unprecedented climate change, the adaptive potential of populations has become an increasingly important topic to

conservation biologists, raising questions of landscape partitioning of adaptive variation and management strategies to maintain population viability. Given the rapid rate of climate change, new beneficial mutations are expected to play a limited role for species with

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2020 The Authors. *Ecology and Evolution* published by John Wiley & Sons Ltd

low mutation rates and long generation times. Therefore, adaptive evolution under climate change for many species will depend on standing genetic variation (Aitken, Yeaman, Holliday, Wang, & Curtis-McLane, 2008; Barrett & Schluter, 2008) that may vary across the landscape and include alleles that gain adaptive value as selection pressures change (Olson-Manning, Wagner, & Mitchell-Olds, 2012). Reliance on standing genetic variation is likely to be particularly true for long-lived sedentary species, such as forest trees that are characterized by adaptive constraints that can limit their evolutionary response to rapid environmental change: extended generation times that result in local persistence, increased rates of genetic drift associated with overlapping generations (Rogers & Prügel-Bennett, 2000), and limited rates of migration due to long generation times. For these species, understanding the distribution of adaptive diversity in relation to recent or past climatic gradients is a critical first step in promoting the adaptive potential of populations in hopes of maintaining future viability (Aitken & Whitlock, 2013; Holderegger, Kamm, & Gugerli, 2006).

The rich history of field research on phenotypic traits in plants (common gardens and reciprocal transplant studies) provides evidence for abundant heritable variation for quantitative traits that are organized along environmental clines (Morgenstern, 1996; Savolainen, Pyhäjärvi, & Knurr, 2007). Until recently, determining the molecular basis of this variation has been less tractable. However, the rapid advancement of genome sequencing, including methods that use reduced genomic complexity (e.g., genotyping by sequencing (GBS), restriction-site associated DNA sequencing (RADseq)), has opened the door to more comprehensive assessments of population-level diversity and allowed for the detection of regions under selection. Although some instances of strong selection on single or few gene loci have been noted (Akey, 2009; Linnen, Kingsley, Jensen, & Hoekstra, 2009; Sella, Petrov, Przeworski, & Andolfatto, 2009), many traits of adaptive importance in plants are believed to be polygenic in nature (Holland, 2007; Le Corre & Kremer, 2012; Pritchard & Di Rienzo, 2010; Yeaman et al., 2016). Under selection, these traits can exhibit subtle changes in frequency across many loci of small effect. Further, demographic processes can shape genetic diversity in ways that mimic selective gradients, as geographic distance and climatic gradients are often autocorrelated. As a result, imprints of selection within the genome can be difficult to detect (Yeaman, 2015), and it is necessary to parcel out the contribution of geographic space in order to successfully identify regions of functional importance (Excoffier, Hofer, & Foll, 2009; Rellstab, Gugerli, Eckert, Hancock, & Holderegger, 2015).

By coupling genome-wide markers with landscape genomics analyses, many researchers have successfully uncovered patterns of adaptive variation and identified potential genomic regions under selection across a wide variety of species (Benestan et al., 2016; De Kort et al., 2014; Dudaniec, Yong, Lancaster, Svensson, & Hansson, 2018; Harrison et al., 2017; Lind et al., 2017; Pais, Whetten, & Xiang, 2016). F_{ST} outlier tests, that scan for highly differentiated loci as candidates for divergent selection, have proven useful in detecting regions under selection but often cannot detect weak or polygenic selection (Lotterhos & Whitlock, 2015; Narum & Hess, 2011; Pritchard & Di Rienzo, 2010). Genotype–environment association (GEA)

tests have demonstrated high power to detect signals of adaptive evolution under varying demographic scenarios (Forester, Lasky, Wagner, & Urban, 2018; Lotterhos & Whitlock, 2015; de Villemereuil, Frichot, Bazin, François, & Gaggiotti, 2014). Univariate association methods that test for single-locus–single-predictor correlation after accounting for population structure are powerful tools to accurately detect even weak signatures of adaptation (Frichot, Schoville, Bouchard, & François, 2013; Gunther & Coop, 2013; Lotterhos & Whitlock, 2015; Rellstab et al., 2015; de Villemereuil et al., 2014). However, a shortcoming of assessing each locus independently is a potential failure to detect signals of polygenic selection (Forester et al., 2018). Multivariate approaches can fill this gap by assessing the combined effects of multiple loci and predictors (Capblancq, Luu, Blum, & Bazin, 2018; Forester et al., 2018; Rellstab et al., 2015), which is perhaps more reflective of real-life evolutionary pressures. Given the advantages of each method, combining outlier tests with GEA can increase the likelihood of detecting complex patterns of selection (Rellstab et al., 2015).

Determining the presence of adaptively important genetic variation and its distribution across a species range is crucial to predicting species' responses to global climate change and directing biodiversity conservation and management efforts (Aitken & Whitlock, 2013; Alberto et al., 2013; Funk, McKay, Hohenlohe, & Allendorf, 2012; Sgrò, Lowe, & Hoffmann, 2011; Sork et al., 2013). This has become an urgent challenge in California, where a protracted drought has resulted in massive tree mortality (USDA, 2016). The Sierra Nevada of California is a high mountain range that collects precipitation from the Pacific Ocean mostly in the form of winter rain and snowfall. The slow release of water from snowmelt in the spring is an important source of moisture for seedling growth and establishment. Sierra snowpack has declined in recent years (Fyfe et al., 2017) and high-resolution regional climate models suggest that spring snow water equivalent will decline by 73% by the end of the century, with midelevations (1,500–2,500 m) experiencing the greatest declines (Sun, Berg, Hall, Schwartz, & Walton, 2018).

This elevational range includes the extant groves of the iconic, long-lived conifer, giant sequoia (*Sequoiadendron giganteum* [Lindl.] Buchholz) that occur in a highly disjunct range consisting of ~70 groves spanning approximately 400 km north to south (Figures 1 and 2). Currently, most giant sequoia populations are in protected areas as this species is valued both culturally and for ecotourism. However, despite this protected status, a key question is whether populations of giant sequoia will remain viable under changing climate. Our previous work has shown very restricted gene flow (DeSilva & Dodd, 2020), suggesting that natural dispersal outside of existing groves will be unlikely. Long generation times (~305 years; Dodd & DeSilva, 2016) will slow the expansion of new variants that may arise through mutation, which underscores the role of standing genetic variation in determining the future viability of giant sequoia populations. In our landscape genetics study of microsatellite variation, we found some evidence for isolation by environment (IBE), linking genetic divergence at putatively neutral loci to dissimilarity in precipitation, and temperature-related variables (DeSilva & Dodd, 2020). Although IBE is consistent with local adaptation, it is dependent on



FIGURE 1 A giant sequoia tree in Giant Forest, Sequoia and Kings Canyon National Park, CA, USA

a reduction of gene flow from divergent habitats due to selection against nonadapted immigrants, and therefore, patterns of IBE may not be reflective of local adaptation when gene flow is low or absent (Nosil, Vines, & Funk, 2005; Wang & Bradburd, 2014), as is likely the case in sections of giant sequoia range (DeSilva & Dodd, 2020). A recent common garden study reported provenance variation in growth performance, providing support for the existence of adaptive genetic variation across the species' range (Valness, 2016). Yet, to date, no studies have investigated local adaptation in giant sequoia using genomic data. Our ultimate goal was to detect populations that may be genetically responsive to anticipated climate change, so here, we build upon this earlier work by reporting on genomic signatures of selection using a range-wide genotyping-by-sequencing dataset. Specifically, we utilize an F_{ST} outlier test and gene–environment association methods (LFMM and RDA), to find signatures of local adaptation among giant sequoia populations and locate potential genomic regions under selection.

2 | METHODS

2.1 | DNA extraction, GBS library preparation, and data processing

Foliage was collected from 6 to 9 trees within each of 18 populations of giant sequoia distributed throughout the range (Figure 2).

To reduce the potential of sampling related individuals, we aimed to sample individual trees >40 m apart. However, this was not possible in some small and highly clustered populations. In this latter case, we attempted to maximize the distance between sampled individuals, with the exception of the PLAC population, where all individuals were sampled. Our goal was to maximize the capture of variation across the range of our study species. Thus, we prioritize increased sampling of populations across the *S. giganteum* range, with the trade-off of limited sampling within each population. Appropriate permits were obtained for all sampling.

High-purity genomic DNA from 143 individuals was isolated from leaf tissue using Plant/Fungi DNA Isolation kits (Norgen Biotek). We constructed three sequencing libraries using a double-digest restriction enzyme-associated genotyping-by-sequencing (GBS) protocol outlined in Peterson, Dong, Horback, and Yong-Bi (2014). Genomic DNA was digested using SbfI and EcoRI restriction enzymes (New England Biolabs). The resulting product was ligated to barcoded adapters and purified, and 46–48 individuals per library were then pooled and subjected to PCR amplification using Phusion High-Fidelity PCR Kit (New England Biolabs) and an automated size selection for fragments between 430 and 570 bp using Pippin Prep. The resulting three libraries were sequenced on an Illumina HiSeq 4000 platform using 150 bp pair-end reads. Sequence data were then demultiplexed using the `process_radtags` module within the STACKS pipeline (Catchen, Hohenlohe, Bassham, Amores, & Cresko, 2013), during which reads with a phred quality score < 10 were removed. Sequences were then aligned to the giant sequoia reference genome v1.0 (Redwood Genome Project, 2019), using the software Bowtie 2 and SAMtools (Langmead & Salzberg, 2012; Li et al., 2009). Variable sites were called using FreeBayes (Garrison & Marth, 2012) and filtered to remove low-quality reads, potential sequencing errors, and paralogs. Data filtering steps included removing loci with uneven mapping quality and those with average read depth >200, requiring a minimum read depth of 5 \times and a minor allele count >3, removal of loci with more than 80% missing data, and a thinning step that retains one SNP per DNA fragment to remove potentially linked loci (Appendix S1). This filtering protocol resulted in a final dataset of 1,364 bi-allelic SNPs used for outlier tests and environmental association analyses and to obtain genetic diversity statistics.

2.2 | Environmental data

To characterize the climatic conditions for each population, we used the spatial centroid of each population to extract and compile twenty-one environmental variables at a spatial resolution of approximately 1 km². Nineteen climate variables were obtained from the WorldClim database (Fick & Hijmans, 2017), and elevation and climate water deficit (CWD) were obtained from the California Basin Characterization Model (Flint, Flint, Thorne, & Boynton, 2013). CWD provides an indication of aridity that is important for Mediterranean climate systems, such as in California (Stephenson, 1998). We conducted a principal component analysis (PCA) on the full environmental dataset (21

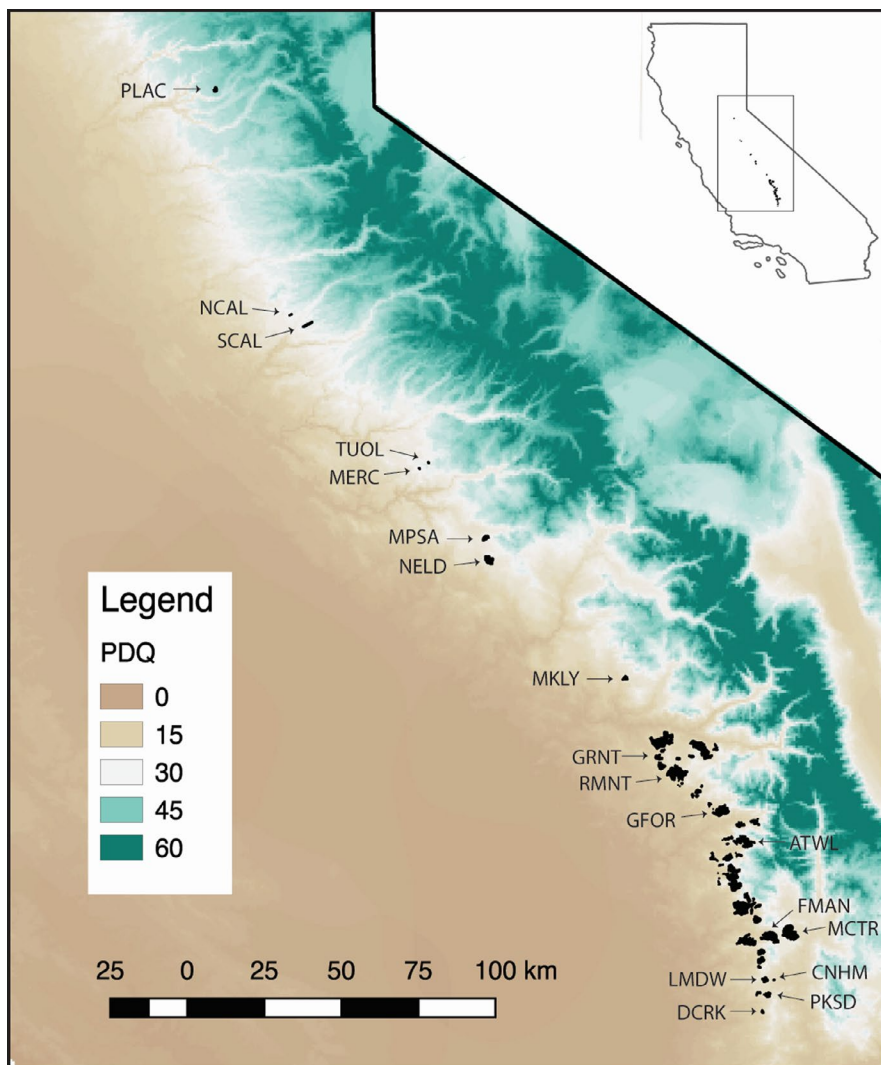


FIGURE 2 Range map of giant sequoia (black) showing the gradient of precipitation of driest quarter (mm) across a section of California. Sampled populations indicated by a population code

variables) after standardization, to reduce dimensionality in the climate data. We retained the first two axes (hereafter PC1 and PC2), which together explained 82% of the climate variation (Appendices S2 and S3). PC1 was driven predominantly by temperature and elevation variables, with a small contribution from annual and winter precipitation, whereas PC2 was determined mostly by precipitation-related variables and CWD with a minor contribution from variables related to temperature seasonality (Appendices S2 and S3).

2.3 | Genetic diversity

Genetic diversity and differentiation statistics were calculated using both the “diveRsity” package in R and GenoDive (Keenan, McGinnity, Cross, Crozier, & Prodöhl, 2013; Meirmans & Van Tienderen, 2004). Calculated statistics included observed and unbiased expected heterozygosity (H_o and uHe , respectively), the inbreeding coefficient (F_{IS}), and the pairwise fixation index (G'_{ST}). Since the removal of rare alleles (minor allele count filtering) can bias genetic diversity estimation, we also calculated genetic diversity statistics without this filtering step for comparison. To further investigate the partitioning

of genetic variation, we used AMOVA with 10,000 permutations to estimate F_{ST} across all populations as well as between northern and southern regions which previous evidence suggested were divergent (DeSilva & Dodd, 2020; Dodd & DeSilva, 2016). For regional diversity comparisons, groves north of GRNT were grouped as northern populations and groves from GRNT to the south as southern populations (Figure 2; DeSilva & Dodd, 2020).

2.4 | Genomic signatures of selection: F_{ST} outliers and gene–environment association tests

To detect F_{ST} outliers that are candidates for selection, we utilized the Bayesian likelihood approach implemented in BayeScan v.2.1 (Foll & Gaggiotti, 2008). This method scans the genome for highly differentiated SNPs that potentially have been subjected to divergent selection while accounting for neutral genetic structure (Narum & Hess, 2011). BayeScan was run using the false discovery rate (FDR) set to 0.05 under the following parameters: 20 pilot runs of 5,000 with an additional burn in of 50,000 iterations and a subsequent run with 5,000 iterations and a thinning interval of 10. The prior odds for the neutral

model were increased to 100 (default is 10) as raising this value has been shown to reduce false positives with little effect on false negatives (Lotterhos & Whitlock, 2014). Loci with \log_{10} values of the posterior odds >1.0 were retained, as the program documentation suggests these loci show “strong” evidence for selection (Foll, 2010).

To test for associations between genomic variation and environmental factors, we utilized latent factor mixed models (LFMMs; Frichot et al., 2013), as implemented in the LEA package in R (Frichot & François, 2015a). LFMM is a univariate approach that treats each individual locus as a response variable with climate data (PC1 and PC2 separately) as the explanatory variable, while incorporating neutral structure using latent factors (Frichot et al., 2013). In simulation studies, LFMM has demonstrated a good balance between high power and low false-positive rate (Frichot et al., 2013; de Villemereuil et al., 2014). As suggested by Frichot et al. (2013) and Frichot and François (2015a), we used two methods to determine the optimal number of latent factors (K value) that correct for the neutral genetic structure of our data. First, we ran a principal component analysis on the individual allele frequencies. We then determined the number of components that explain the genetic variance, based on the Tracy–Widom test on the eigenvalues, as an estimate of K (Frichot & François, 2015a). Second, we utilized the Bayesian clustering algorithm STRUCTURE that estimates the number of genetic clusters (K) without prior information about geographic origin (Pritchard, Stephens, & Donnelly, 2000). The best K value was determined using the ΔK statistic as suggested by Evanno, Regnaut, and Goudet (2005). We used four replicates and a burnin of 300,000 and 1,000,000 MCMC repeats after burnin for $K = 2$ –12.

We ran LFMM to test for associations between SNP's and two composite climate variables (PC1 and PC2) using ten independent replications at 50,000 iterations after a burnin period of 25,000 with the number of latent factors (K) ranging from 8 to 12, as the methods outlined above suggested K equal to 10 and 9, respectively. We chose high run length parameters because of the relatively small number of individuals and loci. LFMM uses the z -scores to indicate the strength of the gene–environment association (Frichot & François, 2015b). As suggested by the authors, we calculated the median z -score from ten replicate runs, re-adjusted the p -values, controlled for FDR using the q -value of 0.05, and determined candidate SNPs based on the Benjamini–Hochberg procedure (Frichot & François, 2015b).

We also utilized RDA, a multivariate GEA method, to test for more subtle polygenic signatures of adaptation and detect outlier loci as candidates of functional importance. Redundancy analysis (RDA) is an extension of multiple regression to multivariate response variables (Legendre & Legendre, 2012). In finding the ideal combination of predictor and explanatory variables, RDA has shown high power to detect potential signals of polygenic adaptation (Forester et al., 2018; Harrison et al., 2017). For these analyses, Hellinger-transformed allele frequencies (Legendre & Gallagher, 2001) were treated as response variables. Because RDA models do not allow missing data, we imputed allele frequency data using probabilistic principal component analysis (ppca) as implemented in the “pcaMethods” package in R (Stacklies, Redestig, Scholz, Walther, & Selbig, 2007).

Ppca uses a decomposition of SNP frequencies to create principal components; the components with the largest eigenvalues are then used to impute the missing data. We evaluated space and climate as explanatory variables. Space was defined by distance-based Moran's eigenvector maps (dbMEMs; Borcard & Legendre, 2002; Dray, Legendre, & Peres-Neto, 2006) based on Euclidean distances between all giant sequoia groves (sampled and unsampled) and extracting the values that correspond to our sample sites. Then, we conducted backward model selection, using the “ordistep” function within the vegan package for R (Oksanen et al., 2013), to reduce the number of dbMEM vectors. For climate, we reduced the twenty-one untransformed environmental variables described above, first by removing highly correlated environmental variables, ($|r| < 0.7$), and subsequently by using the “ordistep” function for backwards model selection to remove variables lacking explanatory power. The above process resulted in climate being represented by “isothermality” (ISO), a measure diurnal and annual temperature fluctuation, “precipitation of driest quarter” (PDQ), a measure of summer precipitation in Mediterranean climates, and “climate water deficit” (CWD), a measure of aridity, in all RDA models, and space represented by two dbMEM vectors, MEM3 and MEM5. All variables were centered and standardized before use in each model.

We set up multiple RDA models to determine the relative amount of variation in allele frequency explained by climate after correcting for geographic space as a signature of local adaptation (Harrison et al., 2017; Lasky et al., 2012; Sork et al., 2016). First, to elucidate the major factors shaping genetic variation and to detect potential signals of local adaptation, we set up three models for comparison: a full RDA model where allele frequencies were associated with both climate and spatial explanatory variables, a partial RDA in which the effects of climate were conditioned on geography (dbMEMs), and a second partial RDA, where the effects of geography were conditioned on climate. Next, to detect outlier loci, allele frequencies were associated with climate predictors after removing the effects of spatial predictors (Forester et al., 2018; Harrison et al., 2017; Lasky et al., 2012). Using the first constrained axis, we identified candidate SNPs of potential adaptive importance as those with loadings in the tails of a 95% confidence interval from the mean or 2.0SD from the mean loadings. One risk of using such a low cutoff is an elevated rate of false positives. However, we chose this to maximize the number of SNPs detected, as we did not expect to find single loci that would be under very strong selection for climate variation in the range of giant sequoia. Moreover, we also identified the climate predictor with the highest correlation to each indicated SNP. In all RDA models, we assessed model and constrained-axis significance using 999 permutations.

2.5 | Genomic context of outlier loci

To gain insights into the potential adaptive significance of outlier loci, we obtained the flanking sequence of each outlier SNP locus from the giant sequoia reference sequence (Redwood Genome Project, 2019). Since the giant sequoia reference genome is not

annotated, functional annotation was performed using the online BLAST (Basic Local Alignment Search Tool) database. Using a 601-bp sequence (300 bp upstream and downstream of the SNP site), we searched the NCBI database using BLASTn with an *e*-value cutoff set to 1×10^5 and the requirement of >70% sequence similarity.

3 | RESULTS

3.1 | Genetic diversity and differentiation

Genetic diversity and differentiation differed substantially across the eighteen sampled populations (Table 1). Observed heterozygosity (H_o) ranged from 0.09 to 0.17 and was lowest in PLAC and highest in ATWL and FMAN (Table 1). Unbiased expected heterozygosity (uHe) was also lowest in PLAC (0.07) and highest in ATWL (0.21) (Table 1). Average pairwise population differentiation (G'_{ST}) varied from 0.09 to 0.32 and was lowest for GFOR and highest for PLAC (Table 1). Average G'_{ST} was significantly higher in the northern populations than in the southern populations ($G'_{ST} N = 0.235$, $G'_{ST} S = 0.109$, Prob $G'_{ST} N \neq G'_{ST} S = 0.017$). Diversity analysis of SNPs without minor allele count (MAC) filtering yielded significantly different results: H_o and uHe were lower and ranged from 0.08 to 0.15 and 0.06 to 0.18, respectively ($p < .001$, $p = .01$ respectively, Table 1), whereas G'_{ST} was slightly higher in the dataset without MAC filtering ($G'_{ST} 0.09$ – 0.34 , $p = .003$, Table 1).

Hierarchical AMOVA found a small, but significant variance due to regions ($F_{ct} = 0.02$, $p = .000$) and a larger portion of genomic

variation distributed among populations ($F_{ST} = 0.15$, $p = .000$; Appendix S4). Population clustering (STRUCTURE) at $K = 9$ indicated strong differentiation among many of the northern populations (north of GRNT) with little admixture (Figure 3, Appendix S5). In addition, populations NELD, GFOR, ATWL, GRNT, MCTR, and FMAN were assigned to the same cluster and the four southernmost populations, LMDW, CNHM, PKSD, and DCRK, consisted of two clusters with PKSD as a transitional population exhibiting admixture from both clusters (Figure 3). Finally, RMNT and GRNT show admixture between the neighboring GFOR, ATWL, GRNT, MCTR, FMAN, cluster, and the geographically separate LMDW, CNHM, cluster (Figure 3).

3.2 | F_{ST} outliers

BayeScan indicated seven F_{ST} outliers, six demonstrating evidence for divergent selection with F_{ST} values ranging from 0.55 to 0.72 (Table 2, Appendix S6) and one showing signs of balancing selection (locus 1,114, $F_{ST} = 0.04$). Since our focus here is on patterns of spatially varying selection, no further discussion is presented for locus 1,114. A BLAST search found one of these loci exhibited functional significance (Locus 828; Table 2).

3.3 | Candidate genomic regions associated with climate variables

Univariate environmental association analyses (LFMM with $K = 9$) indicated a total of 58 loci with strong correlations to composite

TABLE 1 Population information and genetic diversity summary statistics calculated for each population. Diversity statistics calculated without minor allele filtering are noted within parentheses.

Grove name	Population code	GPS location	Sample size	H_o	uHe	F_{IS}	Mean G'_{st}
Placer	PLAC	39.06, -120.57	6	0.09 (0.08)	0.07 (0.06)	0.00 (0.01)	0.32 (0.34)
North Calaveras	CALN	38.28, -120.30	8	0.14 (0.13)	0.19 (0.16)	0.17 (0.17)	0.17 (0.17)
South Calaveras	CALS	38.24, -120.25	8	0.16 (0.14)	0.18 (0.16)	0.13 (0.13)	0.15 (0.15)
Tuolumne	TUOL	37.77, -119.81	8	0.16 (0.14)	0.17 (0.14)	-0.04 (-0.04)	0.24 (0.25)
Merced	MERC	37.75, -119.84	8	0.16 (0.14)	0.17 (0.15)	0.00 (-0.01)	0.21 (0.21)
Mariposa	MPSA	37.51, -119.60	8	0.13 (0.11)	0.14 (0.12)	0.11 (0.11)	0.20 (0.20)
Nelder	NELD	37.43, -119.59	8	0.16 (0.15)	0.20 (0.17)	0.13 (0.11)	0.13 (0.14)
McKinley	MKLY	37.03, -119.11	8	0.15 (0.13)	0.17 (0.15)	0.07 (0.06)	0.17 (0.17)
Grant	GRNT	36.75, -118.97	8	0.12 (0.10)	0.13 (0.11)	0.17 (0.18)	0.14 (0.15)
Redwood Mountain	RMNT	36.60, -118.92	8	0.12 (0.10)	0.11 (0.10)	0.17 (0.18)	0.13 (0.13)
Giant Forest	GFOR	36.57, -118.76	8	0.15 (0.14)	0.20 (0.18)	0.18 (0.17)	0.09 (0.09)
Atwell	ATWL	36.47, -118.67	8	0.17 (0.15)	0.21 (0.18)	0.08 (0.08)	0.11 (0.12)
Mcintyre	MCTR	36.13, -118.58	8	0.13 (0.11)	0.11 (0.09)	0.10 (0.11)	0.12 (0.13)
Freeman Creek	FMAN	36.14, -118.52	8	0.17 (0.12)	0.19 (0.16)	0.08 (0.07)	0.10 (0.11)
Long Meadow	LMDW	35.96, -118.60	8	0.12 (0.11)	0.11 (0.10)	0.16 (0.15)	0.15 (0.16)
Cunningham	CNHM	35.92, -118.57	9	0.16 (0.15)	0.17 (0.16)	0.00 (0.00)	0.16 (0.16)
Packsaddle	PKSD	35.93, -118.59	8	0.15 (0.13)	0.15 (0.13)	0.08 (0.08)	0.16 (0.16)
Deer Creek	DCRK	35.88, -118.61	8	0.16 (0.14)	0.18 (0.16)	0.05 (0.04)	0.16 (0.16)

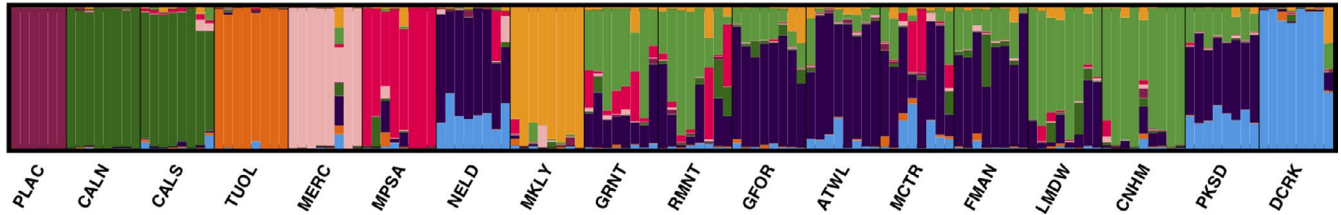


FIGURE 3 Results of population structure analyses from STRUCTURE. Vertical bars represent a sampled individual, color-coded for assigned cluster at $K = 9$

TABLE 2 Functional significance, detection method, and associated variable for highly supported outlier loci

Locus ID	Detection method	Associated variable* (F_{ST})	Adjusted p -value	Axis 1 Loading	Annotation
828	BayeScan	(0.58)			Unknown mRNA
1,123	BS, LFMM	PC1(0.66)	1.72E-04		—
186	LFMM	PC1	7.39E-06	—	Unknown mRNA
218	LFMM	PC1	5.50E-05	—	Unknown mRNA
452	LFMM	PC1	5.30E-05	—	Unknown mRNA
251	LFMM	PC2	1.91E-05	—	Unknown mRNA
722	LFMM	PC2	1.52E-03	—	Unknown mRNA
870	LFMM	PC2	9.48E-04	—	Unknown mRNA
1,029	LFMM	PC2	6.56E-06	—	Unknown mRNA
1,062	LFMM	PC2	1.66E-04	—	Unknown mRNA
1,214	LFMM	PC2	1.52E-03	—	Magnesium transporter MRS2-4-like
1,253	LFMM	PC2	2.91E-04	—	Unknown mRNA
1,313	LFMM	PC2	5.87E-06	—	Pleiotropic drug resistance protein 1-like
368	LFMM, RDA	PC2, PDQ	2.35E-05	0.21132	—
421	LFMM, RDA	PC2, PDQ	9.56E-05	0.19423	—
471	LFMM, RDA	PC2, PDQ	2.03E-05	0.19892	—
515	LFMM, RDA	PC2, PDQ	1.46E-03	0.21181	—
679	LFMM, RDA	PC2, PDQ	3.51E-04	0.22348	—
827	LFMM, RDA	PC2, PDQ	9.12E-04	0.20579	Kinesin-like protein KIN-13A mRNA
1,229	LFMM, RDA	PC2, PDQ	1.65E-03	0.17822	Pollen allergen gene
1,286	LFMM, RDA	PC2, PDQ	4.41E-07	0.26369	—
338	RDA	PDQ	—	0.18769	Wall-associated receptor kinase-like 1
612	RDA	PDQ	—	0.24267	Arogenate dehydratase gene
617	RDA	PDQ	—	0.17057	Unknown mRNA
940	RDA	PDQ	—	0.23536	Signal peptidase I AT2G30440 mRNA
1,066	RDA	PDQ	—	0.19520	Unknown mRNA

environmental variables (Figure 4; Appendix S6). Of these, 51 were correlated with PC2 that was predominantly driven by precipitation, and seven were correlated with the temperature-driven PC1 (Figure 4; Table 2; Appendix S6). An examination of the adjusted P-values from all runs (K set from 8–12) provided additional support for $K = 9$ (Appendix S7; Frichot & François, 2015b). Our BLAST analysis was successful in finding functional annotation for three loci that were strongly associated with PC1 and 11 loci that were associated with PC2 (Table 2).

We used the partial RDA model to detect outlier loci as candidates of importance in selection in a multivariate context, where allele frequencies were associated with climate after removing the effect of spatial predictors. Using the SNP loadings on the first RDA axis, we identified 24 outlier loci beyond the 95% confidence that demonstrated strong correlations to environmental variation, all of which were most correlated with precipitation of the driest quarter (PDQ) (Figure 5; Appendices S6 and S8). Our annotation procedure supported functional importance for seven of the 24 loci (Table 2).

3.4 | Concordance among tests for signatures of selection

Overall, eight loci were detected as outliers by both RDA and LFMM. All of these loci were most associated with precipitation-related variables (PC2 in LFMM and PDQ in RDA). Annotation through BLAST identified two of these loci as having a putative function (Table 2). Overlap was found between a BayeScan (F_{ST} outlier) and LFMM at one locus (1,123; Table 2).

3.5 | Partitioning variation between climate and geographic space

The full RDA model explained 45% of the total variation in allele frequency and supported an influence of climate and/or space in

shaping allelic variation ($p = .001$; adjusted $R^2 = .22$). The first two canonical axes from the full RDA model were significant ($p = .002$, and 0.013 respectively) and together accounted for 77% of the explained variation (Figure 5). The partial RDA model, with climate conditioned on space, was significant ($p = .019$; adjusted $R^2 = .09$) and constrained 49% of the variance explained by the full model. The first partial RDA axis was significant at the 0.1 level ($p = .098$) and accounted for 45% of the variation. The partial RDA model with space conditioned on climate accounted for 24% of the explained variation and was nonsignificant ($p = .207$). The remaining 27% of the explained variation was confounded between climate and geography.

4 | DISCUSSION

Giant sequoia is a paleoendemic of California that has likely suffered from a long-term demographic decline (Dodd & DeSilva, 2016). Today, it is limited to a number of restricted groves in the Sierra Nevada mountain chain. Small grove sizes and limited gene exchange among populations (DeSilva & Dodd, 2020) might be expected to limit its adaptive potential through inbreeding effects and genetic drift. However, through different approaches we have found evidence for a signal of spatially varying local adaptation associated with climate variables and, in particular, along gradients dominated by precipitation. We report here that population genetic structure in giant sequoia has been shaped by local adaptation overlain on historical population processes. From our study of genomic variation, we detected 79 loci as either F_{ST} outliers or loci with strong associations to climate as candidate regions of adaptive importance. Of these, we highlight 26 SNPs, found from multiple methods, or that correspond with functional annotation, as prime candidates for additional research. We emphasize that these outlier loci may include false positives and that experimental studies are needed to demonstrate functional significance of putative adaptive genomic regions (Barrett & Hoekstra, 2011; Kawecki & Ebert, 2004). Here, we present a first step toward understanding local adaptation in an iconic forest tree.

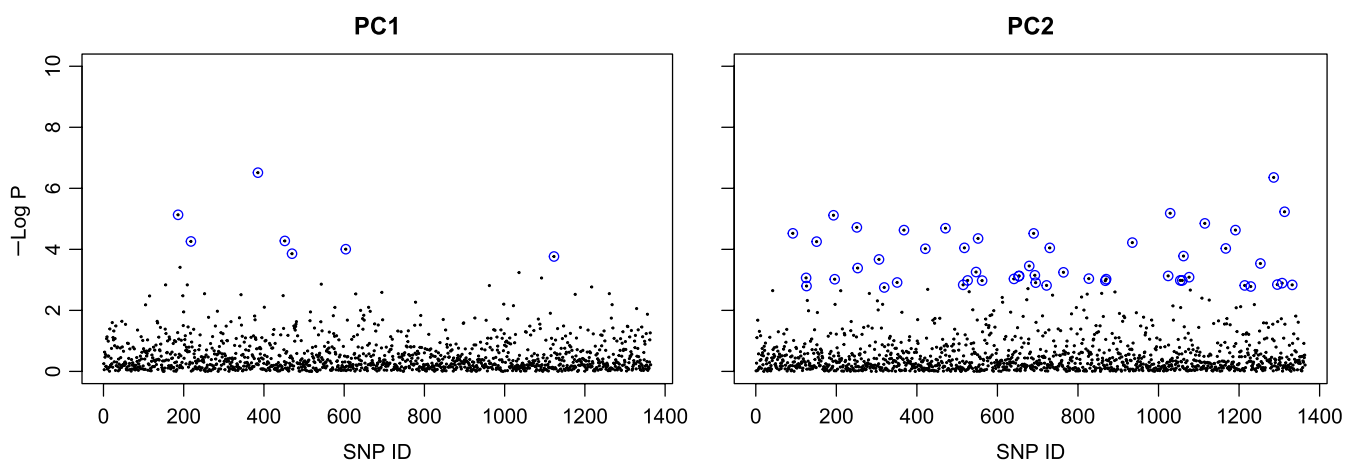
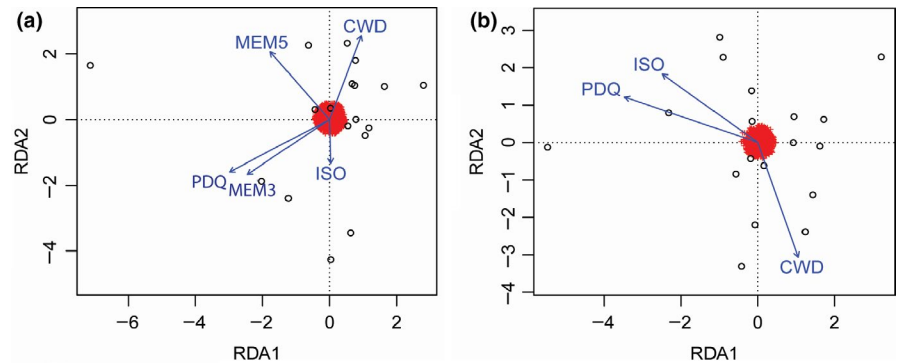


FIGURE 4 Adjusted p -values from LFMM for association with PC1 and PC2. Outliers are outlined in blue

FIGURE 5 Triplot from Redundancy analysis showing how each explanatory variable affects the RDA axis with (a) representing the full RDA model and (b) a partial RDA model with the effects of climate conditioned on geography



4.1 | Population divergence and structure

We found evidence for strong population differentiation in our genomic data ($F_{ST} = 0.15$), which was close to our earlier estimate of $F_{ST} = 0.14$ from microsatellite variation (DeSilva & Dodd, 2020). Such high levels of differentiation are unusual in wind-pollinated tree species, for which population differentiation is typically low and suggests that at least some populations have been isolated for a considerable time (McKay & Latta, 2002; Petit & Hampe, 2006). The 18 giant sequoia populations that we sampled, covering the range of the species, could be partitioned into nine clusters. Six of these clusters were restricted to northern isolated groves and the remaining three clusters include all southern populations that are somewhat more contiguous. Our results confirm the previously report of strong population structure among populations north of GRNT based on microsatellite data (DeSilva & Dodd, 2020; Dodd & DeSilva, 2016). Genetic diversity (unbiased heterozygosity) was no lower in the northern fragmented groves than in most populations within the range, supporting our earlier inference that northern groves have a long evolutionary history (Dodd & DeSilva, 2016). Moreover, estimates of uHe (calculated after removing putative adaptive sites; see Appendix S9) show consistent patterns across populations as previous estimates of He from microsatellite markers (DeSilva & Dodd, 2020), although uHe is lower in the SNP dataset. This pattern of population and genetic structure is unusual for north temperate conifers, for which higher latitude populations are commonly thought of as “leading edge” colonization following glacial retreat. Current groves extend above the lower extent of late Pleistocene glaciers (Moore and Moring 2013), so either some short distance upward colonization must have occurred, or pockets of unglaciated terrain may have served as very local refugia. Given the pattern of genomic diversity that we have detected and the long generation time (~300 years) of giant sequoia, it seems most likely that extant groves have either persisted through many generations or were colonized by short distance migrations.

4.2 | Evidence for local adaptation

Despite the strong structure among populations, analysis of our genomic data revealed a signal of divergent selection associated

with climatic variables. BayeScan detected few F_{ST} outliers due, in part, to the high population structure and increasing the prior odds for the neutral model to 100. However, each of these outliers exhibited high levels of differentiation (F_{ST} 0.55–0.72; Appendix S6). Univariate (LFMM) and multivariate (RDA) environmental association approaches identified more loci indicative of local adaptation. Although F_{ST} outlier approaches have been found to be more robust with respect to false positives than other methods (Lotterhos & Whitlock, 2014), environmental association studies are more successful in detecting loci under selection and can provide context for selective forces as well (De Mita et al. 2013). Our environmental association studies found climate to be an important predictor of allele frequency and accounted for the largest portion of explained variation after correcting for geographic space, a pattern consistent with local adaptation. Signatures of local adaptation to climate are prevalent in many tree species, including *Picea mariana*, *Alnus glutinosa*, *Populus trichocarpa*, *Cornus florida*, and *Quercus lobata* (De Kort et al., 2014; Geraldès et al., 2014; Pais et al., 2016; Prunier, Laroche, Beaulieu, & Bousquet, 2011; Sork et al., 2016). Moreover, the association of genomic variation with climate in giant sequoia is consistent with our previous work that found precipitation-related variables play a role in patterns of isolation by environment at neutral genetic markers (DeSilva & Dodd, 2020). Here, a genome-wide dataset that includes putative functional regions showed a signal of climatic factors shaping genomic variation, which suggests that local adaptation in situ, likely under conditions of limited gene flow, is important in this species.

Local adaptation is further supported by nine loci detected by multiple methods and 19 candidate loci with functional annotation (Table 2). Overlap in the detection of outlier loci has been reported in numerous field studies (Harrisson et al., 2017; Hess, Zendt, Matala, & Narum, 2016; Sork et al., 2016). A carefully designed simulation study demonstrated that overlap between GEA methods was found more often for actual targets of selection rather than false positives (Forester et al., 2018). In addition, two of the eight loci detected by both RDA and LFMM have relevant functional annotation, (loci 827 and 1,229; Table 2). A BLAST search suggests that Locus 827 is a kinesin-like protein, KIN-13A, which has been found to be involved in trichome morphogenesis (Lu, Lee, Pan, Maloof, & Liu, 2005). In plants, trichome occurrence and density are associated with increased drought resistance (Galdon-Armero et al., 2018;

Sletvold & Ågren, 2012). Locus 1,229 represents a potential pollen allergen gene, which is thought to be involved in plant responses to stress (Chen et al., 2016). Although our BLAST search suggested functional importance for 20 loci, many of the annotated regions are characterized only as mRNA with further functional roles yet to be determined (Table 2). The non-annotated outlier loci are promising candidates for future research as they may be of unknown importance, linked to adaptive genes, in regulatory regions, or represent false positives. Any future annotation of the giant sequoia genome will provide valuable clarity as to the specific role of all outlier loci. Yet, we emphasize the candidate loci noted in this study demonstrate only strong associations with climate and identifying the exact targets of selection involves rigorous experimental research. Taken together, outlier loci with functional annotation or those detected by multiple methods provide strong support for adaptive variation across the range of giant sequoia.

4.3 | Outlier SNPs driven by precipitation

Interestingly, variables associated with precipitation appeared to be the major drivers of local adaptation, which perhaps reflects the strong gradients of water relations on the western slope of the Sierra Nevada. Using gene–environment association methods (LFMM and RDA), we found evidence for adaptive differentiation across giant sequoia populations in response to gradients in precipitation and a more limited signal of local adaptation to temperature. LFMM analyses demonstrated seven times as many outliers correlated with precipitation-related PC2 than to temperature-related PC1 (Figure 4, Appendix S6). Although all RDA outliers were correlated with three environmental factors, PDQ, a measure of summer precipitation, CWD, a measure of aridity, and ISO, a measure diurnal and annual temperature fluctuation, outlier loci showed the strongest relationship to PDQ (Appendix S8). Thus, both LFMM and RDA indicate a subtle signal of adaptation to temperature and a stronger signature of divergent selection in response to gradients in water-related variables. Gradients of water availability are important selective agents for many tree species including *Picea mariana*, *Cornus florida*, *Fagus sylvatica*, *Quercus* spp., and *Pinus albicaulis* (Lind et al., 2017; Martins et al., 2018; Pais et al., 2016; Pluess et al., 2016; Prunier et al., 2011; Sork et al., 2016). In addition, many ecological studies have noted that giant sequoia is sensitive to water availability during its establishment phase (Hartesveldt, Harvey, Shellhammer, & Stecker, 1975; Rundel, 1972; Shellhammer & Shellhammer, 2006; York, Battles, & Heald, 2003). Giant sequoia is known to have bursts of reproduction after fire and subsequently experience high seedling mortality due to desiccation (Weatherspoon, 1990). Considering the reproductive biology of this species, water availability is a highly plausible selective agent. For Mediterranean type climates, PDQ is of particular relevance for desiccation sensitive species as it equates to summer precipitation, which may represent a vital water source for giant sequoia seedlings during a vulnerable establishment phase.

4.4 | Limitations, opportunities, and future implications

It is important to note that GEA methods can suffer from low power or high false-positive rate under some demographic scenarios. Although LFMM has been shown to be robust to various demographic scenarios, including those that create high levels of population structure, this method can have elevated false discovery rates (FDR) under scenarios that create IBD (Forester, Jones, Joost, Landguth, & Lasky, 2016; Lotterhos & Whitlock, 2015; de Villemerueil et al., 2014). Here, we do not find a significant signal of IBD in our data. Yet, the role of IBD or other neutral factors affecting population structure in giant sequoia has not been fully elucidated. Previous research has indicated isolation by distance (IBD) and/or ancient divergence separating the northern and southern populations of giant sequoia (DeSilva & Dodd, 2020; Dodd & DeSilva, 2016). Here, a large portion of the explained variance in our data (27%) was confounded between climatic variation and geographic space, which is perhaps due, to the strictly north-south range of giant sequoia, making it inherently difficult to decouple distance from environmental gradients that vary latitudinally. Therefore, LFMM results should be treated with some caution due to the potential contribution of IBD to population structure. In contrast to LFMM, RDA models show high power and low false-positive rates under IBD (Forester et al., 2016). Simulation studies indicate that the performance of RDA also remains high when population structure and selective gradients are explicitly correlated (Capblancq et al., 2018). Yet, RDA is not without limitations, as it can have low power under island demographic models (Forester et al., 2018) or when selective pressures are highly clustered (Capblancq et al., 2018). Given that each GEA method has particular limitations, we believe the outlier loci detected by both LFMM and RDA, as well as outliers with functional annotation remain strong candidate regions of adaptive importance.

There is an ongoing need for future studies to provide additional clarity on the distribution of adaptive variation and genetic architecture of local adaptation in giant sequoia. To our knowledge, this study represents the first investigation of adaptive variation using genome-wide data. Yet, the results presented here are based on a small subset of the genomic variation within the species, as the giant sequoia genome is very large (8.5 Gb, Redwood Genome Project). More comprehensive sampling of the genome as well as an incorporation of phenotypic information will greatly improve our understanding of local adaptation in this species. In addition, future annotation of the giant sequoia genome will provide opportunities to better understand the genetic underpinnings of many phenotypic traits.

A trend toward increased aridity along midelevation Sierra Nevada forests could undermine the long-term persistence of giant sequoia. With end-of-century predictions for this region that include decreasing snowfall and earlier snowmelt, forests of the Sierra Nevada mountains will likely experience an accentuation of the summer drought that is typical of Mediterranean climates (Fyfe et al., 2017; Stewart, Cayan, & Dettinger, 2004; Sun et al., 2018). Considering the evidence presented here, we highlight the potential

that increased water stress may create maladaptation of giant sequoia populations to their environment. It has been suggested that a long-term decline (over the last ~2My) of giant sequoia is tied to increasing aridity during the development of current climate regimes (Dodd & DeSilva, 2016). Moreover, some giant sequoia populations suffered extensive foliage die back during the drought period from 2012 to 2016 (Stephenson et al., 2018), providing further indication of sensitivity of giant sequoia to arid conditions.

5 | CONCLUSIONS

We provide evidence of local adaptation along gradients of precipitation and highlight genomic regions of potential adaptive importance for additional research. This information can aid in determining the best course of action to preserve giant sequoia into the future. Locally adapted populations of giant sequoia are facing an accentuation of summer drought to which they may be maladapted. Genomic variation currently present in more arid regions of the giant sequoia range could include “preadapted” variants that might enhance the adaptive response of nearby populations (Aitken & Bemmels, 2016; Kremer et al., 2012). Currently, DCRK, GRNT, MCTR, and RMNT inhabit areas experiencing the lowest levels of summer precipitation (PDQ) and thus may provide potential sources of adaptive alleles. Given the limited gene flow in much of giant sequoia range, it is unlikely this variation will spread quickly by natural means (DeSilva & Dodd, 2020). Thus, forest managers may consider assisting in the movement of genetic resources in order to enhance the adaptive potential of giant sequoia populations.

ACKNOWLEDGMENTS

We are grateful to Matthew Hughes for assistance with leaf collections and to Prahlada Papper for vital help with data processing. This research was partially funded by McIntire Stennis grant CA-B-ECO-0165-MS awarded to R.S.D. We also thank reviewers whose feedback helped improve the manuscript.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

AUTHOR CONTRIBUTION

Rainbow DeSilva: Conceptualization (lead); Data curation (lead); Formal analysis (lead); Methodology (lead); Writing-original draft (lead); Writing-review & editing (equal). **Richard S. Dodd:** Conceptualization (supporting); Formal analysis (supporting); Funding acquisition (lead); Methodology (supporting); Supervision (lead); Writing-original draft (supporting); Writing-review & editing (equal).

DATA AVAILABILITY STATEMENT

The full SNP data are available through DRYAD <https://doi.org/10.6078/D1GT4D>

ORCID

Rainbow DeSilva  <https://orcid.org/0000-0002-2223-345X>

REFERENCES

- Aitken, S. N., & Bemmels, J. B. (2016). Time to get moving: Assisted gene flow of forest trees. *Evolutionary Applications*, *9*, 271–290.
- Aitken, S. N., & Whitlock, M. C. (2013). Assisted gene flow to facilitate local adaptation to climate change. *Annual Review of Ecology, Evolution, and Systematics*, *44*, 367–388.
- Aitken, S. N., Yeaman, S., Holliday, J. A., Wang, T. L., & Curtis-McLane, S. (2008). Adaptation, migration or extirpation: Climate change outcomes for tree populations. *Evolutionary Applications*, *1*, 95–111.
- Akey, J. M. (2009). Constructing genomic maps of positive selection in humans: Where do we go from here? *Genome Research*, *19*, 711–722.
- Alberto, F. J., Aitken, S. N., Alía, R., González-Martínez, S. C., Hänninen, H., Kremer, A., ... Savolainen, O. (2013). Potential for evolutionary responses to climate change—evidence from tree populations. *Global Change Biology*, *19*, 1645–1661. <https://doi.org/10.1111/gcb.12181>
- Barrett, R. D. H., & Hoekstra, H. (2011). Molecular spandrels: Tests of adaptation at the genetic level. *Nature Reviews Genetics*, *12*, 767–780. <https://doi.org/10.1038/nrg3015>
- Barrett, R. D. H., & Schluter, D. (2008). Adaptation from standing genetic variation. *Trends in Ecology and Evolution*, *23*, 38–44.
- Benestan, L., Quinn, B. K., Maaroufi, H., Laporte, M., Clark, F. K., Greenwood, S. J., ... Bernatchez, L. (2016). Seascape genomics provides evidence for thermal adaptation and current-mediated population structure in American lobster (*Homarus americanus*). *Molecular Ecology*, *25*, 5073–5092. <https://doi.org/10.1111/mec.13811>
- Borcard, D., & Legendre, P. (2002). All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. *Ecological Modelling*, *153*, 51–68. [https://doi.org/10.1016/S0304-3800\(01\)00501-4](https://doi.org/10.1016/S0304-3800(01)00501-4)
- Capblancq, T., Luu, K., Blum, M. G. B., & Bazin, E. (2018). Evaluation of redundancy analysis to identify signatures of local adaptation. *Molecular Ecology Resources*, *18*(6), 1223–1233. <https://doi.org/10.1111/1755-0998.12906>
- Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: An analysis tool set for population genomics. *Molecular Ecology*, *22*, 3124–3140. <https://doi.org/10.1111/mec.12354>
- Chen, M., Xu, J., Devix, D., Shi, J., Ren, K., Searle, I., & Zhang, D. (2016). Origin and functional prediction of pollen allergens in plants. *Plant Physiology*, *172*, 341–357. <https://doi.org/10.1104/pp.16.00625>
- De Kort, H., Vandepitte, K., Bruun, H. H., Closset-Kopp, D., Honnay, O., & Mergeay, J. (2014). Landscape genomics and a common garden trial reveal adaptive differentiation to temperature across Europe in the tree species *Alnus glutinosa*. *Molecular Ecology*, *23*, 4709–4721.
- de Villemereuil, P., Frichot, É., Bazin, É., François, O., & Gaggiotti, O. E. (2014). Genome scan methods against more complex models: When and how much should we trust them? *Molecular Ecology*, *23*, 2006–2019. <https://doi.org/10.1111/mec.12705>
- DeSilva, R., & Dodd, R. S. (2020). Fragmented and isolated: Limited gene flow coupled with weak isolation by environment in the paleoendemic giant sequoia (*Sequoiadendron giganteum*). *American Journal of Botany*, *107*(1), 45–55.
- De Mita, S., Thuillet, A.-C., Gay, L., Ahmadi, N., Manel, S., Ronfort, J., & Vigouroux, Y. (2013). Detecting selection along environmental gradients: analysis of eight methods and their effectiveness for outbreeding and selfing populations. *Molecular Ecology*, *22*, 1383–1399. <https://doi.org/10.1111/mec.12182>
- Dodd, R. A., & DeSilva, R. (2016). Long-term demographic decline and late glacial divergence in a Californian Paleoendemic: *Sequoiadendron giganteum*, giant sequoia. *Ecology and Evolution*, *6*, 3342–3355.

- Dray, S., Legendre, P., & Peres-Neto, P. R. (2006). Spatial modelling: A comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). *Ecological Modelling*, 196, 483–493. <https://doi.org/10.1016/j.ecolmodel.2006.02.015>
- Dudaniec, R. Y., Yong, C. J., Lancaster, L. T., Svensson, E. I., & Hansson, B. (2018). Signatures of local adaptation along environmental gradients in a range-expanding damselfly (*Ischnura elegans*). *Molecular Ecology*, 27, 2576–2593. <https://doi.org/10.1111/mec.14709>
- Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software structure: A simulation study. *Molecular Ecology*, 14, 2611–2620. <https://doi.org/10.1111/j.1365-294X.2005.02553.x>
- Excoffier, L., Hofer, T., & Foll, M. (2009). Detecting loci under selection in a hierarchically structured population. *Heredity*, 103, 285–298. <https://doi.org/10.1038/hdy.2009.74>
- Fick, S. E., & Hijmans, R. J. (2017). Worldclim 2: New 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, 37(12), 4302–4315. <https://doi.org/10.1002/joc.5086>
- Flint, L. E., Flint, A. L., Thorne, J. H., & Boynton, R. (2013). Fine-scale hydrologic modeling for regional landscape applications: The California Basin Characterization Model development and performance. *Ecological Processes*, 2, 25. <https://doi.org/10.1186/2192-1709-2-25>
- Foll, M. (2010). *BayeScan v2.0 user manual*. Retrieved from http://cmpg.unibe.ch/software/BayeScan/files/BayeScan2.0_manual.pdf
- Foll, M., & Gaggiotti, O. (2008). A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. *Genetics*, 180, 977–993. <https://doi.org/10.1534/genetics.108.092221>
- Forester, B. R., Jones, M. R., Joost, S., Landguth, E. L., & Lasky, J. R. (2016). Detecting spatial genetic signatures of local adaptation in heterogeneous landscapes. *Molecular Ecology*, 25, 104–120. <https://doi.org/10.1111/mec.13476>
- Forester, B. R., Lasky, J. R., Wagner, H. H., & Urban, D. L. (2018). Comparing methods for detecting multilocus adaptation with multivariate genotype–environment associations. *Molecular Ecology*, 27, 2215–2233. <https://doi.org/10.1111/mec.14584>
- Frichot, E., & François, O. (2015a). LEA: An R package for landscape and ecological association studies. *Methods Ecology and Evolution*, 6, 925–929. <https://doi.org/10.1111/2041-210X.12382>
- Frichot, E., & François, O. (2015b). *A short manual for LFMM version 1.4*. Retrieved from <http://membres-timc.imag.fr/Olivier.Francois/lfmm/files/note.pdf>
- Frichot, E., Schoville, S. D., Bouchard, G., & François, O. (2013). Testing for associations between loci and environmental gradients using latent factor mixed models. *Molecular Biology and Evolution*, 30, 1687–1699. <https://doi.org/10.1093/molbev/mst063>
- Funk, W. C., McKay, J. K., Hohenlohe, P. A., & Allendorf, F. W. (2012). Harnessing genomics for delineating conservation units. *Trends in Ecology & Evolution*, 27, 489–496. <https://doi.org/10.1016/j.tree.2012.05.012>
- Fyfe, J., Derksen, C., Mudryk, L., Flato, G. M., Santer, B. D., Swart, N. C., ... Jiao, Y. (2017). Large near-term projected snowpack loss over the western United States. *Nature Communications*, 8, 14996. <https://doi.org/10.1038/ncomms14996>
- Galdon-Armero, J., Fullana-Pericas, M., Mulet, P. A., Conesa, M. A., Martin, C., & Galmes, J. (2018). The ratio of trichomes to stomata is associated with water use efficiency in *Solanum lycopersicum* (tomato). *Plant Journal*, 96, 607–619. <https://doi.org/10.1111/tpj.14055>
- Garrison, E., & Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. *arXiv*, arXiv preprint arXiv:1207.3907 [q-bio.GN].
- Geraldes, A., Farzaneh, N., Grassa, C. J., McKown, A. D., Guy, R. D., Mansfield, S. D., ... Cronk, Q. C. B. (2014). Landscape genomics of *Populus trichocarpa*: The role of hybridization, limited gene flow, and natural selection in shaping patterns of population structure. *Evolution*, 68(11), 3260–3280. <https://doi.org/10.1111/evo.12497>
- Gunther, T., & Coop, G. (2013). Robust identification of local adaptation from allele frequencies. *Genetics*, 195, 205–220. <https://doi.org/10.1534/genetics.113.152462>
- Harrisson, K. A., Amish, S. J., Pavlova, A., Narum, S. R., Telonis-Scott, M., Rourke, M. L., ... Sunnucks, P. (2017). Signatures of polygenic adaptation associated with climate across the range of a threatened fish species with high genetic connectivity. *Molecular Ecology*, 26, 6253–6269. <https://doi.org/10.1111/mec.14368>
- Hartesveldt, R. J., Harvey, H. T., Shellhammer, H. S., & Stecker, R. E. (1975). *The giant sequoia of the Sierra Nevada*. Washington, DC: U.S. Department of the Interior, National Park Service.
- Hess, J. E., Zandt, J. S., Matala, A. R., & Narum, S. R. (2016). Genetic basis of adult migration timing in anadromous steelhead discovered through multivariate association testing. *Proceedings of the Royal Society. B, Biological Sciences*, 283, 20153064.
- Holderegger, R., Kamm, U., & Gugerli, F. (2006). Adaptive vs. neutral genetic diversity: Implications for landscape genetics. *Landscape Ecology*, 21(6), 797–807. <https://doi.org/10.1007/s10980-005-5245-9>
- Holland, J. (2007). Genetic architecture of complex traits in plants. *Current Opinion in Plant Biology*, 10, 156–161. <https://doi.org/10.1016/j.pbi.2007.01.003>
- Kawecki, T. J., & Ebert, D. (2004). Conceptual issues in local adaptation. *Ecology Letters*, 7, 1225–1241. <https://doi.org/10.1111/j.1461-0248.2004.00684.x>
- Keenan, K., McGinnity, P., Cross, T. F., Crozier, W. W., & Prodöhl, P. A. (2013). diveRsity: An R package for the estimation of population genetics parameters and their associated errors. *Methods in Ecology and Evolution*, 4, 782–788.
- Kremer, A., Ronce, O., Robledo-Arnuncio, J. J., Guillaume, F., Bohrer, G., Nathan, R., ... Schueler, S. (2012). Long-distance gene flow and adaptation of forest trees to rapid climate change. *Ecology Letters*, 15, 378–392. <https://doi.org/10.1111/j.1461-0248.2012.01746.x>
- Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9, 357–360. <https://doi.org/10.1038/nmeth.1923>
- Lasky, J. R., Des Marais, D. L., McKay, J. K., Richards, J. H., Juenger, T. E., & Keitt, T. H. (2012). Characterizing genomic variation of *Arabidopsis thaliana*: The roles of geography and climate. *Molecular Ecology*, 21, 5512–5529. <https://doi.org/10.1111/j.1365-294X.2012.05709.x>
- Le Corre, V., & Kremer, A. (2012). The genetic differentiation at quantitative trait loci under local adaptation. *Molecular Ecology*, 21, 1548–1566. <https://doi.org/10.1111/j.1365-294X.2012.05479.x>
- Legendre, P., & Gallagher, E. D. (2001). Ecologically meaningful transformations for ordination of species data. *Oecologia*, 129, 271–280. <https://doi.org/10.1007/s004420100716>
- Legendre, P., & Legendre, L. (2012). *Numerical ecology* (Vol. 24, 3rd English ed, p. 990). Boston, MA: Elsevier.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., ... Durbin, R. (2009). 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25, 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
- Lind, B. M., Friedline, C. J., Wegrzyn, J. L., Maloney, P. E., Vogler, D. R., Neale, D. B., & Eckert, A. J. (2017). Water availability drives signatures of local adaptation in whitebark pine (*Pinus albicaulis* Engelm.) across fine spatial scales of the Lake Tahoe Basin, USA. *Molecular Ecology*, 26(12), 3168–3185.
- Linnen, C. R., Kingsley, E. P., Jensen, J. D., & Hoekstra, H. E. (2009). On the origin and spread of an adaptive allele in deer mice. *Science*, 325, 1095–1098. <https://doi.org/10.1126/science.1175826>
- Lotterhos, K. E., & Whitlock, M. C. (2014). Evaluation of demographic history and neutral parameterization on the performance of F_{ST} outlier tests. *Molecular Ecology*, 23, 2178–2192.
- 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. <https://doi.org/10.1111/mec.13100>
- Lu, L., Lee, Y. J., Pan, R., Maloof, J. N., & Liu, B. (2005). An internal motor kinesin is associated with the golgi apparatus and plays a role in trichome morphogenesis in *Arabidopsis*. *Molecular Biology of the Cell*, 16, 811–823.
- Martins, K., Gugger, P. F., Llanderal-Mendoza, J., Gonzalez-Rodriguez, A., Fitz-Gibbon, S. T., Zhao, J., ... Sork, V. L. (2018). Landscape genomics provides evidence of climate-associated genetic variation in Mexican populations of *Quercus rugosa*. *Evolutionary Applications*, 11, 1842–1858. <https://doi.org/10.1111/eva.12684>
- McKay, J. K., & Latta, R. G. (2002). Adaptive population divergence: Markers, QTL and traits. *Trends in Ecology & Evolution*, 17, 285–291. [https://doi.org/10.1016/S0169-5347\(02\)02478-3](https://doi.org/10.1016/S0169-5347(02)02478-3)
- Meirmans, P. G., & Van Tienderen, P. H. (2004). GENOTYPE and GENODIVE: Two programs for the analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes*, 4, 792–794.
- Moore, J. G., & Morning, B. C. (2013). Rangewide glaciation in the Sierra Nevada. *California Geosphere*, 9(6), 1804–1818. <https://doi.org/10.1130/GES00891.1>
- Morgenstern, E. K. (1996). *Geographic variation in forest trees: Genetic basis and application of knowledge in silviculture*. Vancouver, BC: UBC Press.
- Narum, S. R., & Hess, J. E. (2011). Comparison of F_{ST} outlier tests for SNP loci under selection. *Molecular Ecology Resources*, 11, 184–194. <https://doi.org/10.1111/j.1755-0998.2011.02987.x>
- Nosil, P., Vines, T., & Funk, D. (2005). Perspective: Reproductive isolation caused by natural selection against immigrants from divergent habitats. *Evolution; International Journal of Organic Evolution*, 59, 705–719. <https://doi.org/10.1111/j.0014-3820.2005.tb01747.x>
- Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B., ... Wagner, H. (2013). *vegan: Community ecology package*. Retrieved from <https://cran.r-project.org/web/packages/vegan/vegan.pdf>
- Olson-Manning, C. F., Wagner, M. R., & Mitchell-Olds, T. (2012). Adaptive evolution: Evaluating empirical support for theoretical predictions. *Nature Review Genetics*, 13(12), 867–877. <https://doi.org/10.1038/nrg3322>
- Pais, A. L., Whetten, R. W., & Xiang, Q. J. (2016). Ecological genomics of local adaptation in *Cornus florida* L. by genotyping by sequencing. *Ecology and Evolution*, 7(1), 441–465. <https://doi.org/10.1002/ece3.2623>
- Peterson, G. W., Dong, Y., Horback, C., & Yong-Bi, F. (2014). Genotyping-by-sequencing for plant genetic diversity analysis: A lab guide for SNP genotyping. *Diversity*, 6(4), 665–680. <https://doi.org/10.3390/d6040665>
- Petit, R. J., & Hampe, A. (2006). Some evolutionary consequences of being a tree. *Annual Review of Ecology, Evolution, and Systematics*, 37(1), 187–214. <https://doi.org/10.1146/annurev.ecolsys.37.091305.110215>
- Pluess, A. R., Frank, A., Heiri, C., Lalagüe, H., Vendramin, G. G., & Oddou-Muratorio, S. (2016). Genome–environment association study suggests local adaptation to climate at the regional scale in *Fagus sylvatica*. *New Phytologist*, 210, 589–601. <https://doi.org/10.1111/nph.13809>
- Pritchard, J. K., & Di Rienzo, A. (2010). Adaptation—not by sweeps alone. *Nature Reviews Genetics*, 11, 665–667. <https://doi.org/10.1038/nrg2880>
- Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155, 945–959.
- Prunier, J., Laroche, J., Beaulieu, J., & Bousquet, J. (2011). Scanning the genome for gene SNPs related to climate adaptation and estimating selection at the molecular level in boreal black spruce. *Molecular Ecology*, 20, 1702–1716. <https://doi.org/10.1111/j.1365-294X.2011.05045.x>
- Redwood Genome Project (2019). *Sequoiadendron giganteum reference genome 1.0*. Retrieved from <https://nealelab.ucdavis.edu/redwood-genome-project-rgp/>
- Relstab, C., Gugerli, F., Eckert, A. J., Hancock, A. M., & Holderegger, R. (2015). A practical guide to environmental association analysis in landscape genomics. *Molecular Ecology*, 24, 4348–4370. <https://doi.org/10.1111/mec.13322>
- Rogers, A., & Prügel-Bennett, A. (2000). Evolving populations with overlapping generations. *Theoretical Population Biology*, 57, 121–129. <https://doi.org/10.1006/tpbi.1999.1446>
- Rundel, P. (1972). Habitat restriction in giant sequoia – Environmental control of grove boundaries. *American Midland Naturalist*, 87, 81–99.
- Savolainen, O., Pyhäjärvi, T., & Knurr, T. (2007). Gene flow and local adaptation in trees. *Annual Review of Ecology, Evolution and Systematics*, 38, 595–619. <https://doi.org/10.1146/annurev.ecolsys.38.091206.095646>
- Sella, G., Petrov, D. A., Przeworski, M., & Andolfatto, P. (2009). Pervasive natural selection in the *Drosophila* genome? *PLoS Genetics*, 5, e1000495. <https://doi.org/10.1371/journal.pgen.1000495>
- Sgrò, C. M., Lowe, A. J., & Hoffmann, A. A. (2011). Building evolutionary resilience for conserving biodiversity under climate change. *Evolutionary Applications*, 4, 326–337. <https://doi.org/10.1111/j.1752-4571.2010.00157.x>
- Shellhammer, H., & Shellhammer, T. (2006). Giant Sequoia (*Sequoiadendron giganteum* [Taxodiaceae]) seedling survival and growth in the first four decades following managed fires. *Madroño*, 53(4), 342–350. [https://doi.org/10.3120/0024-9637\(2006\)53\[342:GSSGTS\]2.0.CO;2](https://doi.org/10.3120/0024-9637(2006)53[342:GSSGTS]2.0.CO;2)
- Sletvold, N., & Ågren, J. (2012). Variation in tolerance to drought among Scandinavian populations of *Arabidopsis lyrata*. *Evolutionary Ecology*, 26, 559–577. <https://doi.org/10.1007/s10682-011-9502-x>
- Sork, V., Aitken, S., Dyer, R., Eckert, A., Legendre, P., & Neale, D. (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. <https://doi.org/10.1007/s11295-013-0596-x>
- Sork, V. L., Squire, K., Gugger, P. F., Steele, S. E., Levy, E. D., & Eckert, A. J. (2016). Landscape genomic analysis of candidate genes for climate adaptation in a California endemic oak, *Quercus lobata*. *American Journal of Botany*, 103, 33–46.
- Stackl, W., Redestig, H., Scholz, M., Walther, D., & Selbig, J. (2007). pcaMethods—A bioconductor package providing PCA methods for incomplete data. *Bioinformatics*, 23, 1164–1167. <https://doi.org/10.1093/bioinformatics/btm069>
- Stephenson, N. L. (1998). Actual evapotranspiration and deficit: Biologically meaningful correlates of vegetation distribution across spatial scales. *Journal of Biogeography*, 25, 855–870. <https://doi.org/10.1046/j.1365-2699.1998.00233.x>
- Stephenson, N. L., Das, A. J., Ampersee, N. J., Cahill, K. G., Caprio, A. C., Sanders, J. E., & Williams, A. P. (2018). Patterns and correlates of giant sequoia foliage dieback during California's 2012–2016 hotter drought. *Forest Ecology and Management, Volumes*, 419–420, 268–278. <https://doi.org/10.1016/j.foreco.2017.10.053>
- Stewart, I., Cayan, D., & Dettinger, M. (2004). Changes in snowmelt runoff timing in western North America under a 'business as usual' climate change scenario. *Climatic Change*, 62, 217–232. <https://doi.org/10.1023/B:CLIM.0000013702.22656.e8>
- Sun, F., Berg, N., Hall, A., Schwartz, M., & Walton, D. (2018). Understanding end-of-century snowpack changes over California's Sierra Nevada. *Geophysical Research Letters*, 46, 933–943. <https://doi.org/10.1029/2018GL080362>
- US Department of Agriculture-Forest Service [USDA-FS] (2016). *New aerial survey identifies more than 100 million dead trees in California*. Retrieved from <https://www.fs.usda.gov/news/releases/new-aerial-survey-identifies-more-100-million-dead-trees-california>

- Valness, C. M. (2016). *Performance and morphology in Sequoiadendron genotypes outside of their range*. Master's Thesis. Arcata, CA: Humboldt State University.
- Wang, I. J., & Bradburd, G. S. (2014). Isolation by environment. *Molecular Ecology*, 23, 5649–5662. <https://doi.org/10.1111/mec.12938>
- Weatherspoon, C. P. (1990). *Sequoiadendron giganteum* (Lindl.) Buchholz Giant Sequoia. In R. M. Burns, & B. H. Honkala (Eds.), *Silvics of North America, USDA Forest Service Agricultural Handbook 654* (pp. 552–562). Washington, DC: US Department-of-Agriculture-Forest-Service.
- Yeaman, S. (2015). Local adaptation by alleles of small effect. *American Naturalist*, 186(S1), S74–S89. <https://doi.org/10.1086/682405>
- Yeaman, S., Hodgins, K. A., Lotterhos, K. E., Suren, H., Nadeau, S., Degner, J. C., ... Aitken, S. N. (2016). Convergent local adaptation to climate in distantly related conifers. *Science*, 353, 1431–1433. <https://doi.org/10.1126/science.aaf7812>
- York, R., Battles, J., & Heald, R. (2003). Edge effects in mixed conifer group selection openings: Tree height response to resource gradients. *Forest Ecology and Management*, 179, 107–121. [https://doi.org/10.1016/S0378-1127\(02\)00487-5](https://doi.org/10.1016/S0378-1127(02)00487-5)

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: DeSilva R, Dodd RS. Association of genetic and climatic variability in giant sequoia, *Sequoiadendron giganteum*, reveals signatures of local adaptation along moisture-related gradients. *Ecol Evol.* 2020;10:10619–10632. <https://doi.org/10.1002/ece3.6716>