Received: 19 May 2019 | Revised: 19 October 2020 | Accepted: 6 November 2020

DOI: 10.1111/mec.15731

ORIGINAL ARTICLE

MOLECULAR ECOLOGY WILEY

Landscape genomics of *Quercus lobata* reveals genes involved in local climate adaptation at multiple spatial scales

Paul F. Gugger[1,2] | Sorel T. Fitz-Gibbon[1] | Ana Albarrán-Lara[1] | Jessica W. Wright[3] | Victoria L. Sork[1,4]

[1]Ecology and Evolutionary Biology, University of California, Los Angeles, CA, USA

[2]Appalachian Laboratory, University of Maryland Center for Environmental Science, Frostburg, MD, USA

[3]USDA Forest Service, Pacific Southwest Research Station, Davis, CA, USA

[4]Institute of the Environment and Sustainability, University of California, Los Angeles, CA, USA

Correspondence
Paul F. Gugger and Victoria L. Sork, Ecology and Evolutionary Biology, University of California, Los Angeles, CA 90095-7239, USA.
Emails: pgugger@gmail.com (P.F.G) and vlsork@ucla.edu (V.L.K.)

Funding information
UC-MEXUS/CONACyT; University of California, Los Angeles; Division of Integrative Organismal Systems, Grant/Award Number: 1444661; U.S. Forest Service

Abstract

Understanding how the environment shapes genetic variation provides critical insight about the evolution of local adaptation in natural populations. At multiple spatial scales and multiple geographic contexts within a single species, such information could address a number of fundamental questions about the scale of local adaptation and whether or not the same loci are involved at different spatial scales or geographic contexts. We used landscape genomic approaches from three local elevational transects and rangewide sampling to (a) identify genetic variation underlying local adaptation to environmental gradients in the California endemic oak, *Quercus lobata*; (b) examine whether putatively adaptive SNPs show signatures of selection at multiple spatial scales; and (c) map putatively adaptive variation to assess the scale and pattern of local adaptation. Of over 10 k single-nucleotide polymorphisms (SNPs) generated with genotyping-by-sequencing, we found signatures of natural selection by climate or local environment at over 600 SNPs (536 loci), some at multiple spatial scales across multiple analyses. Candidate SNPs identified with gene–environment tests (LFMM) at the rangewide scale also showed elevated associations with climate variables compared to the background at both rangewide and elevational transect scales with gradient forest analysis. Some loci overlap with those detected in other oak species, raising the question of whether the same loci might be involved in local climate adaptation in different congeneric species that inhabit different geographic contexts. Mapping landscape patterns of adaptive versus background genetic variation identified regions of marked local adaptation and suggests nonlinear association of candidate SNPs and environmental variables. Taken together, our results offer robust evidence for novel candidate genes for local climate adaptation at multiple spatial scales.

KEYWORDS

genotyping by sequencing, landscape genomics, local adaptation, natural selection, *Quercus lobata*

1 | INTRODUCTION

Understanding how the environment shapes genetic variation provides critical insight about the evolution of local adaptation in natural populations. Spatial genetic variation can occur as a result of natural selection by local environmental conditions in populations inhabiting heterogeneous environments or as a result of neutral processes, such as genetic drift in structured populations, limited gene flow that isolates populations by distance, biased dispersal related to environmental factors, or demographic history linked or correlated with environmental variation (Sork et al., 2013; Wang & Bradburd, 2014). When selection acts on specific phenotypes that enhance local survival, fecundity, or other fitness components, allele frequencies at specific loci underlying those phenotypes will vary across the landscape in relation to the environmental gradients imposing selective pressure (Endler, 1986; Savolainen et al., 2013). This process leads to local adaptation, which arises when the fitness of local genotypes is higher in their home habitat than the fitness of genotypes from other habitats (Savolainen et al., 2007). The genomics of local adaptation has been of increasing interest in ecologically important, long-lived species such as trees for its potential value in optimizing seed sourcing, reforestation efforts, and other management applications (Aitken & Bemmels, 2016; Aitken & Whitlock, 2013).

Genomic tools coupled with environmental or phenotypic data have led to advances in our understanding of adaptation in natural, non-model systems because they allow the screening of thousands to millions of sequence variants for signatures of natural selection that are consistent with local adaptation (Rellstab et al., 2015; Sork et al., 2013). Landscape genomic methods offer one approach to understanding local adaptation and fall under two broad classes: genotype–environment association and differentiation-based outlier tests. These tests operate under the premise that spatially varying natural selection along environmental gradients will lead to significant genotype–environment associations or exceptionally high genetic differentiation among populations at locally adaptive loci, after controlling for background levels of association or differentiation across the genome. These methods aim to account for the fact that variation throughout the genome can correlate with environmental variation or vary spatially due to numerous neutral processes (Wang & Bradburd, 2014). Thus, appropriate sampling and proper statistical approaches can separate loci that have exceptional environmental associations from the background associations generated by neutral processes, while minimizing false positives (De Mita et al., 2013; Lotterhos & Whitlock, 2015). Landscape genomics offers one view of the effects of local adaptation at the molecular level that complements studies of fitness and phenotypic variation in common garden experiments (Sork et al., 2013).

With knowledge of the genes underlying local adaptation, spatial patterns of adaptive genetic variation as well as background genetic variation can be mapped to identify geographic regions with similar genetic backgrounds (Fitzpatrick & Keller, 2015). As a result, it is possible to evaluate the scale and pattern of local adaptation in natural populations distributed across a landscape. For example, using a machine-learning approach demonstrated by Fitzpatrick and Keller (2015), Martins et al. (2018) demonstrated regional geographical gradients of adaptive genetic variation and identified nonlinear relationships between adaptive variation and geographic and environmental variables among individuals of *Quercus rugosa* in Mexico. Comparing mapped landscapes of background genetic variation to adaptive variation, they identified regions that may harbour especially strong adaptive differences, which may also have applications in forest management.

Previous studies in trees have used landscape genomic approaches to identify putative candidate genes underlying local adaptation at broad and fine spatial scales. For example, *Pinus taeda* exhibits signatures of local adaptation to an aridity gradient spanning the southeastern USA in several genes with abiotic stress response functions (Eckert et al., 2010), and *Populus balsamifera* exhibits evidence of local adaptation reflected in variation in flowering phenology genes across a continent-wide climate gradient in northern North America (Keller et al., 2012). At fine spatial scales (~20 km), *Pinus lambertiana* is locally adapted to precipitation gradients as demonstrated both by landscape genomic and common garden approaches (Eckert et al., 2015). However, it is still unknown whether the same genes and processes underlie species adaptation at broad and fine spatial scales.

Evaluating landscape patterns of adaptive variation at multiple spatial scales or geographic contexts within a single species, could help address fundamental questions about local adaptation and evolution. For example, SNPs found to be under selection at multiple scales can reflect either independent selective processes in each context or a global process that was detected in multiple ways. If different SNPs within the same loci are under selection by the same climate variables at the different scales or contexts, then independent parallel selective processes may be occurring. Furthermore, finding the same gene–environment associations in multiple contexts might increase our confidence in attributing them to underlying local adaptation because the randomness of drift is unlikely to lead to the same associations in multiple contexts (Rellstab et al., 2017).

Here, we take first steps towards identifying genetic variation underlying local adaptation to environmental gradients and the spatial patterns of adaptive variation at multiple scales. We focus on *Quercus lobata* (valley oak; *Quercus* sect. *Quercus*), which occupies a heterogeneous environment from highly seasonal habitat in the southeastern part of its distribution to mild and cool habitat near coastal regions (Figures 1 and S1). *Q. lobata* may be particularly amenable to landscape genomic approaches because it did not undergo large-scale range expansion in response to Quaternary climate change (Gugger et al., 2013), and thus there may be less confounding of selective and neutral signatures than in many other temperate tree species. Furthermore, there are now substantial genomic resources available, including an annotated reference genome (Sork, Fitz-Gibbon, et al., 2016) and transcriptome (Cokus et al., 2015). In addition, surveys of trait variation in the field (e.g., Albarrán-Lara et al., 2015) and ongoing research at a large-scale common garden that includes samples at multiple spatial scales (Delfino Mix et al., 2015) suggests variation in fitness as measured by leaf and growth-related traits consistent with local adaptation to current or

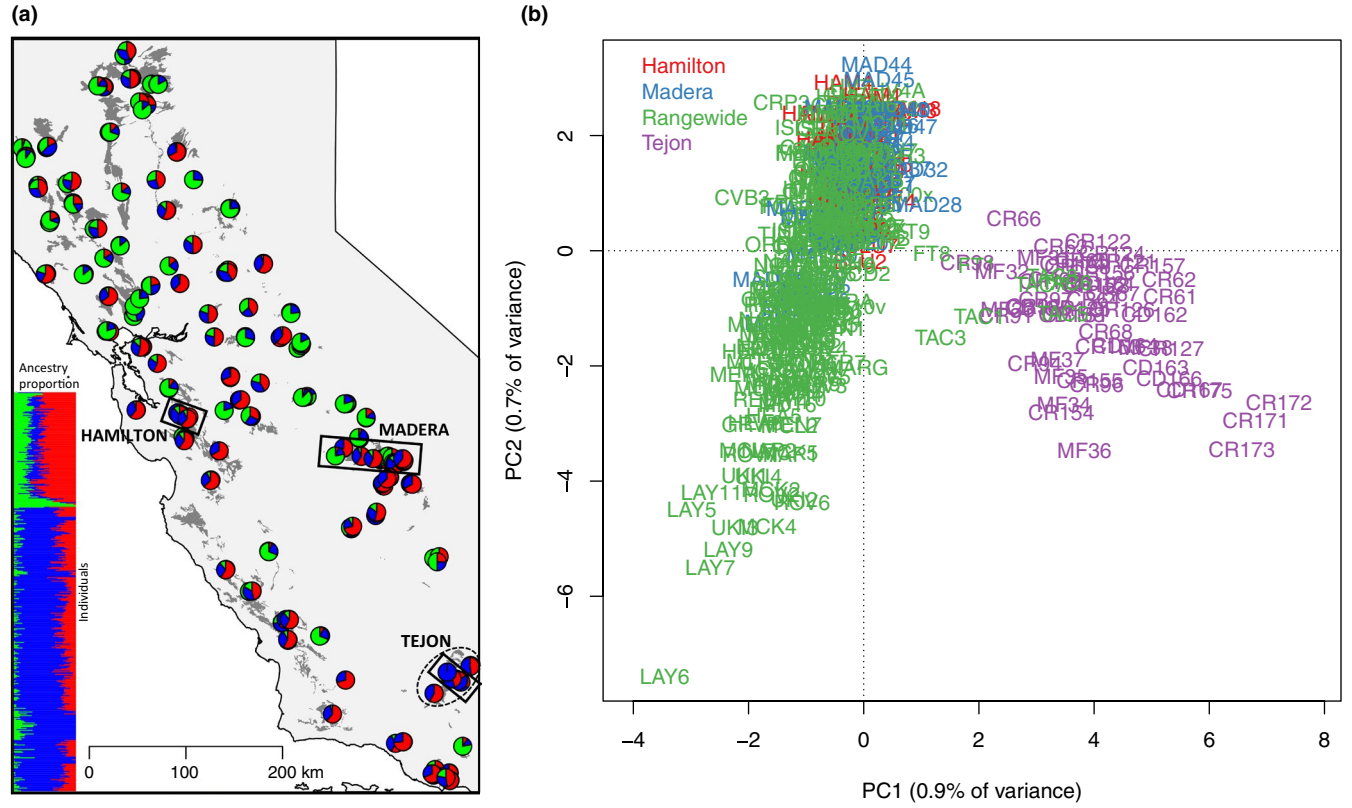

FIGURE 1 Genetic structure as inferred from (a) sNMF and (b) principal components analysis (PCA). In (a) pie charts represent population ancestry proportions for individuals, which are also shown in bar chart format to the lower left. Some pies overlap obscuring others, but the general pattern is not affected. Black boxes delimit the three elevational transects (see Figure S2 for detailed view). Dark grey background marks the distribution of valley oak as extracted from CA GAP data (Davis et al., 1998)

past environmental gradients (Browne et al., 2019). Furthermore, Browne et al. (2019) demonstrated that SNPs can be used to identify genotypes associated with higher growth rates under warmer temperatures in this species. Previous landscape genomic studies in *Q. lobata* were limited in sample size (Gugger et al., 2016) or the number of genetic loci (Sork, Squire, et al., 2016), necessitating further study, but provide some candidates for comparison along with those from a water stress gene expression experiment (Gugger et al., 2017).

Quercus is an ecologically and economically important taxon in northern temperate forests (Cavender-Bares, 2016, 2019). Understanding the molecular basis of local adaptation in this group is an important goal both for our basic understanding of evolution and genomics, as well as for our ability to effectively manage oak species for conservation and natural products in a world in which climate change imposes novel stresses (Kremer et al., 2012). Landscape genomic analyses in Mexican (Martins et al., 2018) and European oaks (Pina-Martins et al., 2019; Rellstab et al., 2016), and gene expression experiments in European oaks (Lesur et al., 2015; Porth et al., 2005) also provide some candidate genes for local adaptation in the same taxonomic section. Since the publication of most of the studies above, reference genome sequences and annotations have become available (Plomion et al., 2016; Sork, Fitz-Gibbon, et al., 2016), greatly enhancing the potential to make inference from genomic association studies.

In this study, we generate a large single-nucleotide polymorphism (SNP) data set using genotyping-by-sequencing (GBS; Elshire et al., 2011) of samples collected at multiple spatial scales to (i) identify genetic variation underlying local adaptation with multiple landscape genomic analyses; (ii) examine whether putatively adaptive SNPs show signatures of selection at both rangewide and local transect spatial scales; and (iii) map putatively adaptive variation to assess the scale and pattern of local adaptation. We employ genotype–environment association and differentiation-based outlier tests to identify putatively adaptive SNPs and then use a machine-learning approach to collectively evaluate support for putatively adaptive SNPs and map adaptive allele frequency changes on the landscape at multiple spatial scales allowing for possible nonlinear patterns. We focus our local-scale analyses on elevational gradients because the environment changes rapidly over short distances, thus enhancing the opportunity for local adaptation at this scale. Given the relatively stable distribution of *Q. lobata* and evidence of local adaptation at multiple spatial scales in previous studies, we expect to identify SNPs under natural selection by environmental factors at both rangewide and local spatial scales along elevational transects, reflecting either independent selective processes or a global selective process and allowing us to evaluate, in part, whether independent parallel selective processes may be occurring. We discuss our findings in comparison to SNPs detected in a

study of the Mexican oak, *Q. rugosa*, that employed similar methods. Through discovery-based tests of selection and spatial analysis of adaptive variation, we aim to provide insights into the mechanisms and patterns of local adaptation.

2 | MATERIALS AND METHODS

2.1 | Sampling and processing

In fall of 2012, leaf samples from 436 adult trees were collected from throughout the range of valley oak and separated from each other by at least 50 m (Figure 1; Table S1). Within this rangewide sampling, three areas were more intensively sampled along elevational gradients (subsets of the 436 total samples). The Hamilton transect up Mt. Hamilton spans 12 km and includes 30 samples from 383–1,112 m elevation; the Madera transect from the eastern Central Valley near the town of Madera to the Sierra Nevada foothills spans 69 km and includes 45 samples from 78–607 m; and the Tejon transect from the southern Central Valley up Cordon Ridge in Tejon Ranch spans 17 km and includes 52 samples from 361–1,766 m elevation (Figures 1, S1, and S2). Fresh leaf samples were stored on ice until arriving at UCLA, where they were stored at −80°C.

Total genomic DNA was extracted from approximately 50 mg of frozen tissue using the Qiagen DNeasy Plant Mini Kit with an added "prewash" step to minimize secondary compounds in the final product (Gaddis et al., 2014). The prewash was as follows: after bead-based tissue grinding under liquid N_2, 1 ml of prewash buffer was added, the mixture was shaken for 10–20 s at 30 Hz and then spun for 10 min at 10,000 rpm in a microcentrifuge, the supernatant was discarded, and finally the pellet was processed with the standard DNeasy protocol starting at the Buffer AP1/RNase A addition step. The prewash buffer was composed of 100 mM Tris-HCl (from pH 8.0 stock), 50 mM EDTA, 1 M NaCl, 0.01 g/ml polyvinylpyrrolidone (PVP; k-30, S_{ABC}), and 4.5 μl/ml 2-mercaptoethanol (added immediately prior to use).

2.2 | Genotyping by sequencing

Total genomic DNA was prepared for sequencing using an efficient restriction-enzyme-based approach commonly known as genotyping by sequencing (GBS; Elshire et al., 2011). GBS is a cost-effective method that has proven fruitful in population and landscape genomic studies with related goals (Gugger et al., 2018; Martins et al., 2018; Parchman et al., 2018). Briefly, DNA was digested with a restriction enzyme, common and unique barcoded adapters with overhangs complementary to the cut site were ligated to each sample, samples were pooled in equimolar ratios, and the pooled library was PCR-amplified and sent for Illumina sequencing. We largely followed the original protocol, including using the same adapter sequences, adapter concentration (0.036 ng/μl of each adapter), and restriction enzyme (*Ape*KI). However, we pooled 48 samples per library prep/sequencing lane rather than 96; adapters were added during the ligation step rather than prior to restriction digestion; AMPure XP bead-based size selection/purification steps were added after the ligation step and repeated after the PCR step to ensure a consistent distribution of fragment sizes between 200 and 500 bp (including adapters) among all preps; and we reduced the number PCR cycles to 16 from 18. Final libraries were checked for the proper size distribution on an Agilent BioAnalyzer with the High Sensitivity DNA assay and quantified using a Qubit fluorometer. Libraries were sent to the UCLA Broad Stem Cell Research Center for single-end, 100 bp sequencing on an Illumina HiSeq2000 v3.

GBS provides thousands of genome-wide SNPs for analysis but these represent a small fraction of the total oak genome (850 Mb; Sork, Fitz-Gibbon, et al., 2016). Thus, we are likely to miss many relevant adaptive loci (Lowry et al., 2017), but the ones detected in this study are likely to be in, or near, causal loci (see below), because linkage disequilibrium decays rapidly along the oak genome (<1 kb; Sork, Squire, et al., 2016), as in other highly outcrossing wind-pollinated trees (Brown et al., 2004). In addition, due to the fragmentary evidence provided by GBS and association tests, we cannot determine whether we have captured loci causally related to local adaptation or linked loci. Nonetheless, the approach is conducive to identifying some of the adaptive variation within a species and for evaluating spatial patterns of that adaptive variation in comparison with background variation, as demonstrated in related studies (Gugger et al., 2018; Martins et al., 2018; Parchman et al., 2018).

2.3 | SNP calling

Illumina reads were demultiplexed and quality filtered using STACKS 1.28 to 1.41 (Catchen et al., 2011) `process_radtags`, which removed adapter sequence with up to two mismatches (`adapter_mm`), recovered barcodes with up to one mismatch to the expected barcodes (`r`), removed any read with an uncalled base (`c`), discarded low quality reads as defined by default settings (`q`), and trimmed all reads to 92 bp (`t`). Using BWA 0.7.12 (Li & Durbin, 2010), the filtered reads were aligned to the *Quercus lobata* reference genome version 0.5 (NCBI accession # ID 308,314, also available at http//valleyoak.ucla.edu; [Sork, Fitz-Gibbon, et al., 2016]). We used GATK 3.7 (DePristo et al., 2011) to identify SNPs in each aligned sample using a minimum Phred-scaled confidence threshold of 30. We then used "VariantFiltration" and "SelectVariants" tools in GATK to exclude low quality variants, applying the following filters: QD < 10.0 (quality by depth), DP < 4 (genotype read depth), and ExcessHet > 100. We used VCFTOOLS 0.1.15 (Danecek et al., 2011) to filter the SNPs to include only diallelic sites, present in at least 90% of individuals, minor allele frequency (MAF) ≥ 0.10, and mean depth of coverage across all samples <80× to filter organellar DNA and sites aligning to genomic regions that may have been overcollapsed in the reference genome. Finally, SNPs were pruned in PLINK 1.90b4.9 (Chang et al., 2015) to remove those exhibiting high linkage disequilibrium defined by using a 5 kb window, sliding five SNPs, and removing a SNP from all pairs

with $r^2 > .5$. The above filtering was performed four times: once for all samples together to generate a rangewide data set and then once for each set of samples belonging to a transect. Our filtering strategy maximizes the number of high-quality SNPs that can be meaningfully analysed in each data set.

2.4 | Genetic structure and outliers

Genetic structure of rangewide and each transect data set was assessed with principal components analysis (PCA) to capture continuous variation and sparse non-negative matrix factorization (sNMF) to estimate the number of discrete genetic clusters (K) and admixture coefficients (Frichot et al., 2014), similar to Bayesian clustering methods (Pritchard et al., 2000). The optimal number of sNMF clusters was selected according to the lowest cross-entropy value, or where the cross-entropy value begins to plateau. The sNMF clusters were then used as the basis for testing for SNPs that have significantly elevated differentiation among clusters consistent with natural selection for local adaptation among clusters (Martins et al., 2016), analogous to F_{ST}-outlier approaches (Lotterhos & Whitlock, 2014) but shown to be among the best performing methods. Differentiation-based tests of selection may capture loci related to environmental variables that we could not or did not measure, such as soil or the biotic environment (see below). PCA and sNMF were performed in R 3.3.2 (R Core Team, 2019) using the "vegan" 2.5–1 (Oksanen et al., 2018) and "LEA" 2.0.0 packages (Frichot & François, 2015a), respectively. The resulting p-values for the outlier tests were adjusted for multiple testing to Q-values using the false-discovery rate method (Benjamini & Hochberg, 1995) as implemented in the "qvalue" 2.6.0 package. $Q < 0.1$ was considered significant.

2.5 | Environmental association analyses

To identify candidate SNPs that may be under the influence of natural selection by climate stressors, we used latent factor mixed modelling (LFMM; Frichot et al., 2013) as implemented in the LEA package (Frichot & François, 2015a) in R to test for significant associations of climate variables with SNP allele frequencies after accounting for background genetic structure. LFMM is among the most powerful models for detecting loci under selection for individual-based sampling designs (Forester et al., 2018; Lotterhos & Whitlock, 2015). Background structure was modelled in LFMM through the association of a predefined number of latent factors. To select the "optimal" number of latent factors, we used the number of clusters inferred from sNMF (see above) and checked histograms of p-values from the LFMM results to ensure that the false-positive rate was well controlled. LFMM does not depend on accurately inputting the "true" number of clusters as the number of latent factors; rather, the goal is to control the statistical error rates. Missing SNP data were imputed using built-in LEA functions based on the inferred genetic structure. LFMM was run with 10 repetitions and burnin of 5,000 steps followed by 10,000 iterations. The resulting z-scores were combined across repetitions, and p-values were recalculated following the program developers' guidance. The resulting p-values were adjusted for multiple testing using the false-discovery rate method and significance threshold of $Q < 0.1$ as above. These analyses were performed separately for the rangewide data set and each elevational transect.

Climate variables (1981–2010 averages) included growing degree-days above 5°C (GDD; a measure of the amount of energy a plant receives), which was derived from PRISM data with 800 m resolution (Coop, 2017); and mean minimum temperature (Tmin), mean maximum temperature (Tmax), climatic water deficit (CWD; an integrated measure of water availability), and actual evapotranspiration (AET), which were extracted from the 2014 California Basin Characterization Model at 270 m resolution (Flint & Flint, 2007; Flint et al., 2013). These variables were chosen because they are expected to relate to stresses, and thus selective pressures, that valley oaks face in California and because they have been linked to genetic variation or limiting distribution in other studies (Gugger, Fitz-Gibbon, et al., 2016; Gugger et al., 2013; McLaughlin & Zavaleta, 2012; Sork et al., 2010). Because LFMM does not account for correlation structure among environmental variables, we followed the developers' guidance to distill the five climate variables to synthetic variables based on principal components axes: three for the rangewide data set and two for each of the transects. We chose to run separate PCAs for each geographic context because we wanted to characterize the primary patterns of environmental change in each context. Using PC scores from a single rangewide analysis at the transect scale would not be efficient and would necessitate including many axes. As a result of our decision, we focus on assessing general environmental associations rather than associations related to specific climate variables. We believe this decision is reasonable because climate variables, both those used in the analysis and those not included, are correlated with each other in each context, thus confounding our ability to attribute associations causally to any specific variable. Moreover, different variables can lead to the same stress, for example, high temperature and low water availability can both lead to water stress. Indeed, the LFMM developers suggest using PCs in their documentation (see below; Frichot & François, 2015b).

2.6 | Genomic context and annotation

The genomic context and functional annotations for SNPs with significant association in LFMM associations or sNMF outlier SNPs were drawn from the annotated genome assembly version 0.5 for *Q. lobata* (Sork, Fitz-Gibbon, et al., 2016), which are based on the alignment of an annotated reference transcriptome for *Q. lobata* (Cokus et al., 2015), whose annotation is based on TAIR 10 (Lamesch et al., 2012) and Pfam (Finn et al., 2016) for sequences with high-quality matches to *Arabidopsis*. For loci without annotations in our reference, we considered BLASTx hits to all nonredundant GenBank coding regions (translations + PDB + SwissProt + PIR + PRF). We focus on genes within which the SNP is located or genes that are less than 10 kb away

in cases in which the SNP is not located within a gene. Although linkage disequilibrium is thought to decay within 1 kb based on a small number of genes, we chose 10 kb as the upper threshold for genes to consider because patterns of linkage disequilibrium have not been evaluated in most of the genome and thus may extend longer than 1 kb in some regions. To evaluate overlap among sets of significant SNPs in different analyses or spatial contexts, we considered whether SNPs were within 2.5 kb of each other in either direction, creating an approximately 5 kb window, in addition to whether the exact same SNP was found in each set. Multiple SNPs within this window are referred to here as being within the same locus. We performed the same comparison among sets using the top 100 SNPs (100 lowest *Q*-values) from each set, regardless of statistical significance, as a means of providing a more equal basis for comparison when the number of significant SNPs differed greatly across sets. Different SNPs from the same locus might be significant in different analyses due to differences in the specific SNPs that passed filtering in each geographic context, statistical artefacts due to linkage, or parallel selection at different SNPs within a locus in different geographic contexts.

2.7 | Comparing and mapping climate associations in candidate and overall genetic variation

We ran gradient forest (GF; Ellis et al., 2012) as implemented in "gradientForest" 0.1–17 (http://gradientforest.r-forge.r-project.org/) in R to (a) quantify and characterize the shape of the gradient in the association of climate variables with candidate SNPs collectively and with all SNPs; (b) map patterns of variation in candidate SNPs and all SNPs; and (c) evaluate whether LFMM candidate SNPs are also among the top associations in gradient forest (Fitzpatrick & Keller, 2015). We, thus, used this approach as a "validation" of LFMM and sNMF findings, to provide more detailed quantification of climate–SNP association in each geographic context, and as a basis for mapping adaptive and background genetic variation allowing for nonlinear patterns that would not otherwise be apparent from LFMM. Gradient forest is a machine learning, regression tree approach that allows for exploration of nonlinear associations of spatial, environmental, and allelic variables. The approach partitions the allele frequency data at "split values" along the environmental gradients. Split importance, a measure of the amount of variation explained, is high in positions along the gradient where allelic change is large. Moving along the gradient, the split importance values are summed cumulatively to produce a step-like function for allele frequency change along the environmental gradient. These functions are generated (a) for each SNP with each climate and spatial variable, which enables comparison of the SNPs with the highest climate associations here to those detected in LFMM; and (b) summarized across all included SNPs to provide a composite view of SNP–climate associations, which enables comparison of a model with all SNPs versus one with candidate SNPs only. To this end, we ran gradient forest up to three times for each data set using individuals as data points: once based on all rangewide SNPs, once based on LFMM candidate SNPs only, and once based on sNMF candidate SNPs only (rangewide only) as dependent variables. Missing SNP data were not imputed. The independent variables included the five climate variables described above, as well as spatial variables defined using principal coordinates of neighbourhood matrices (PCNM), also known as Moran's eigenvector maps (MEM), which were calculated from the geographic coordinates in decimal degrees using the `pcnm` function in "vegan" (Oksanen et al., 2018). PCNMs are a set of orthonormal variables calculated through eigenvalue decomposition of a spatial weighting matrix of coordinates (Dray et al., 2006) and are included here to account for spatial genetic processes that lead to spatial autocorrelation and unmeasured environmental variation. We retained half of the PCNM variables with positive eigenvalues, as has been suggested in similar contexts (Fitzpatrick & Keller, 2015; Manel et al., 2012): 55 in the rangewide data set, three in Hamilton Transect, eight in Madera, and six in Tejon. We ran gradient forest with 500–2000 regression trees per SNP, `maxLevel` = $\log_2(0.368n)/2$, and a variable correlation threshold of 0.5 to calculate conditional variable importances as recommended (Ellis et al., 2012; Strobl et al., 2008). Other parameters were set to default values. Pearson correlations among climate variables, PCNMs, and sample coordinates are given in Table S2.

For mapping gradient forest results, we first used the `predict` function to predict genetic variation across all grid cells in the study region within California, once for each SNP set. We then mapped the resulting predictions across the landscape using principal components of the predictions to generate a red–green–blue colour scale according to the first three axes for each set (see Fitzpatrick & Keller, 2015 for more details on the approach). The results from all SNPs versus candidate SNPs were compared by performing a Procrustes superimposition (Gower, 1971) in "vegan" on the above principal components and then mapping the Procrustes residuals on a colour scale to represent the absolute distance in genetic composition between candidate and all SNPs for each point location.

Finally, we searched for LFMM candidates among the top 10 climate-associated SNPs in GF to provide further support for their putative role in climate adaptation. We used the top ten from GF because GF does not provide a significance test.

3 | RESULTS

3.1 | SNP data

The final rangewide data set for 436 samples contains 11,019 SNPs with median coverage depth of 36 × across loci and 35 × across individuals (Figure S3). Applying the same filters to each transect, we retained 11,786 SNPs in the Hamilton transect, 13,262 in the Madera transect, and 10,159 in the Tejon transect.

3.2 | Genetic structure and outliers

SNP-based PCA of the rangewide samples suggests two groups of samples: those within and near the Tejon Transect and those from

the rest of the species' range (Figure 1, also supported by PCA of transect samples alone, Figure S4). sNMF suggests either six clusters if choosing the lowest cross-entropy value or three if choosing where the values begin to plateau (Figure S5). We chose three because they also produced reasonable clusters that have some geographic signal: one most common around Tejon in the south, one most common in the northwest, and the third most common in between (Figure 1). However, most samples were inferred to be admixed, containing ancestry from two or three of these clusters. Based on these clusters, sNMF identified 577 significant outlier SNPs with $Q < 0.1$ (Tables 1, S3, and S4) that fall within 462 loci defined by 5 kb windows. A total of 195 SNPs fall within known oak genes, and of the remaining, 92 have a gene within 10 kb downstream, and 78 have a gene within 10 kb upstream. Functional annotations are diverse and are detailed in Table S3 by SNP.

The Hamilton and Madera transects have no genetic structure ($K = 1$), and the Tejon transect has three clusters ($K = 3$; Figure S4) according to the lowest cross-entropy value (Figure S5). As a result, we opted to perform structure-based sNMF outlier tests only in the Tejon transect. We detected 36 significant SNPs within 32 loci in the Tejon transect (Tables 1, S2, and S5). One SNP is in a gene that encodes a protein whose sequence is similar to pinoresinol-lariciresinol reductase from pine (oak gene: m01oak08871cC; TAIR ID: AT4G34540) and was also significant at the rangewide scale (Table 2).

3.3 | Environmental associations

LFMM revealed 34 total significant climate-associated SNPs from 21 loci: 14 SNPs associated with the first principal component of climate variation (PC1), 11 with PC2, and nine with PC3 (Tables 1, S3, and S4). Climate PC1 is strongly associated with AET, CWD, GDD, and Tmax (49% of variance); PC2 is strongly associated with AET and Tmax (26%); and PC3 is strongly associated with Tmin (22%; Figure S6). Seven SNPs are within genes (phosphoglycerate mutase family protein, protein serine/threonine kinase, snurportin-1 protein, and a transmembrane amino acid transporter protein), six have genes within 10 kb upstream (Sec14p-like phosphatidylinositol transfer family protein, 1,8-cineole synthase, NIN-like protein 7, and α/β-hydrolases superfamily protein), and two with a gene downstream within 10 kb (rcd1-like protein). Five SNPs with significant associations in LFMM are in common with sNMF outliers, but none of these have known functions or are located within 10 kb of annotated genes (Tables 2 and S3; Figure 2). Overlap among top 100 SNPs at the rangewide scale for LFMM and sNMF consists of three loci: m01oak14993cC or Y-family DNA polymerase H, m01oak12927cC or TRFL10, and m01oak05143CC or quinoprotein amine dehydrogenase (Tables 2 and S6).

For the elevational transects, only three significant climate associations were detected, all in Madera: one of unknown function with PC1 and two with PC2 (Tables 1, S3, S5, S7, and S8). PC1 is strongly associated with all the climate variables and PC2 is most associated with GDD. The SNPs associated with PC2 are 3.6 kb from oak gene m01oak02014cC, which is an auxin-induced transcription factor (Pfam: AUX_IAA, PF02309) and 6.2 kb from α/β-hydrolases superfamily protein (m01oak02652cC, AT3G60340). Eight loci are in the top 100 in both the Tejon LFMM and sNMF analyses (Table S6). Comparing the loci surrounding the top 100 SNPs in each LFMM analysis, six loci overlap among transects and rangewide (Figure 2), five of which are in protein-coding regions: Y-family DNA polymerase H, plasma-membrane choline transporter family protein, ankyrin repeat-containing protein, and two in uncharacterized proteins (Tables 1 and S3).

Histograms of *p*-values from the LFMM results suggest that the false-positive rate was well controlled in tests from the rangewide and transect data sets (Figure S7).

TABLE 1 Number of SNPs that are significant for each test (LFMM, sNMF) for each geographic context

	LFMM			
	PC1	PC2	PC3	sNMF
Rangewide	14	11	9	577
Hamilton	0	0	NA	NA
Madera	1	2	NA	NA
Tejon	0	0	NA	36

3.4 | Comparing and mapping climate associations in candidate and overall genetic variation

At the rangewide scale, climate variables exhibit moderate R^2-weighted importance in the gradient forest model with all SNP data, all ranking between 20 and 35 of 60 total variables and in the following order: CWD, GDD, Tmax, AET, Tmin (Figures 3 and S8). Nineteen of the PCNMs rank above the climate variables. The relationship of SNP allelic variation with each of the climate variables appears nearly linear but more step-like for several of the top PCNMs (one representative shown; Figure 3). The 577 sNMF candidate SNPs do not appear substantially more associated with climate variables than the background, although subtle differences can be observed (Figures 3 and S7). In contrast, gradient forest associations of the 34 LFMM candidate SNPs with all climate variables and most PCNMs achieve greater R^2-weighted importance than the all-SNP model (0.002–0.008 vs. <0.0015). Cumulative importance plots show steep gradients in LFMM candidate allelic composition around CWD = 1,000, GDD = 7,500, and Tmax = 34, and more gradual change with AET and Tmin.

Among the transects, gradient forest models with all SNPs indicate modest differences in magnitudes of climate association and different rank orders of importance among climate variables (Figures 4 and S8). However, in all cases the shapes of the associations appear roughly linear, and the magnitudes of cumulative importance among variables within models are similar. Because only three SNPs are

TABLE 2 SNPs that are significant in LFMM or in more than one geographic context for either LFMM or sNMF organized by locus and including available functional annotation information

SNP (contig:position)*	# Sets	Sets	BLASTX ($E < 1e\text{-}10$)	Oak gene	TAIR ID	TAIR common name	Distance (bp; upstream negative)
1:11554759, 11554828	1	**LFMM_R**	**Uncharacterized protein**	—	—	—	—
1:15423811	2	LFMM_T, sNMF_T	Auxin-regulated gene involved in organ size	m01oak08988cC	AT3G59910	Ankyrin repeat family protein	3538
1:18970751	1	**LFMM_R**	—	—	—	—	—
2:9221847	1	**LFMM_R**	Reverse transcriptase	—	—	—	—
2:24885589	1	**LFMM_M**	—	**m01oak02014cC**	—	—	**−3574**
2:29606241	1	**LFMM_R**	Uncharacterized protein	**m01oak07773cC**	**AT1G22170**	**Phosphoglycerate mutase family protein**	**0**
2:45001581, 45001588	3	LFMM_H, LFMM_T, sNMF_T	Uncharacterized protein	—	—	—	—
2:49247688, 49248110, 49248378	2	LFMM_T, sNMF_T	—	m01oak15757cC	AT5G28150	Plant protein of unknown function	0
3:18261936	1	**LFMM_R**	Uncharacterized mitochondrial protein	—	—	—	—
3:49429610, 49429721	2	LFMM_M, sNMF_R	Uncharacterized protein	m01oak04218cC	—	—	898
3:68485717, **68485610**	2	LFMM_H, LFMM_R	DNA polymerase η	m01oak14993cC	AT5G44740	POLH: Y-family D polymerase H	0
3:94030726	2	LFMM_M, sNMF_T	—	—	—	—	—
3:96376299, 96376268	2	LFMM_T, sNMF_T	—	—	—	—	—
3:97288010, 97288082	2	LFMM_M, sNMF_R	Glutamate receptor 3.7	—	—	—	460
4:28301256	1	**LFMM_R**	—	**m01oak11464Ct**	**AT4G24880**	**Unknown protein**	**0**
4:30601425	2	LFMM_T, sNMF_R	Hypothetical protein	m01oak08175cM	—	—	0
4:32561505, 32561633	2	**LFMM_R, sNMF_R**	**Pentatricopeptide repeat-containing protein**	—	—	—	—
4:51975468	2	**sNMF_R, sNMF_T**	**Pinoresinol-lariciresinol reductase 3**	**m01oak08871cC**	**AT4G34540**	**NmrA-like negative transcriptional regulator family protein**	**0**
4:53109430, 53109358, 53109591	2	LFMM_M, LFMM_R	Ankyrin repeat-containing protein	—	—	—	—
4:54060719	1	**LFMM_R**	—	**m01oak06083cC**	**AT5G04990**	**SUN1 SAD1/UNC−84 domain protein 1**	**4570**
4:72987385	2	LFMM_T, sNMF_T	Serine/threonine-protein phosphatase PP2A−4 catalytic subunit	m01oak01112cC	—	—	0
4:91136324	1	**LFMM_R**	Uncharacterized protein	—	—	—	—

(Continues)

TABLE 2 (Continued)

SNP (contig:position)*	# Sets	Sets	BLASTX ($E < 1e\text{-}10$)	Oak gene	TAIR ID	TAIR common name	Distance (bp; upstream negative)
5:5382352	2	LFMM_M, LFMM_T	—	m01oak08563CC	AT5G13760	Plasma-membrane choline transporter family protein	−6542
5:23114306	1	**LFMM_R**	—	—	—	—	—
5:73671915	1	**LFMM_R**	—	**m01oak22481cC**	—	**Unknown protein**	**−7828**
5:82138776, 82138796	1	**LFMM_R**	**G-type lectin S-receptor-like serine/threonine-protein kinase**	—	—	—	—
6:4659914	2	LFMM_T, sNMF_T	LysM domain receptor-like kinase 4	m01oak42472jC	AT2G23770	Protein kinase family protein/ peptidoglycan-binding LysM domain-containing protein	2561
6:7264652	2	LFMM_T, sNMF_T	Zinc finger CCCH domain-containing protein 24	m01oak05677CC	AT2G28450	Zinc finger (CCCH-type) family protein	0
6:17934100	1	**LFMM_R**	GATA transcription factor 12-like	—	—	—	—
6:23071769	1	**LFMM_R**	—	**m01oak11154JC**	—	—	—
6:55544654	1	**LFMM_M**	Uncharacterized protein	**m01oak02652cC**	**AT3G60340**	**α/β-hydrolases superfamily protein**	**6217**
6:56230660	1	**LFMM_R**	Auxin transporter protein 1-like	**m01oak16771CC**	—	—	**686**
6:71077555	1	**LFMM_R**	Cell differentiation protein RCD1 homologue	**m01oak04390CC**	**AT3G20800**	**Cell differentiation, Rcd1-like protein**	**7990**
7:13857852	1	**LFMM_R**	Bidirectional sugar transporter SWEET4	**m01oak16163CC**	**AT4G24020**	**NLP7: NIN like protein 7**	**−1777**
7:20294827	2	**LFMM_R, sNMF_R**	**Hypothetical protein**	**m01oak18542CC**	—	—	**−5064**
7:40995197	1	**sNMF_R**	**AAA-ATPase/hypersensitivty related**	**m01oak10076cC**	**AT4G34570**	**THY-2: thymidylate synthase 2**	**1519**
7:42572957	2	LFMM_T, sNMF_T	—	—	—	—	—
7:50823065	3	LFMM_H, LFMM_R, LFMM_T	—	—	—	—	—
7:66500693, 66500692	2	LFMM_H, LFMM_R	Uncharacterized protein	—	—	—	—
7:74552124	1	**LFMM_R**	Uncharacterized protein	—	—	—	—
7:78375890, 78375914, 78375943	1	**LFMM_R**	Carbon catabolite repressor protein 4 homologue	**m01oak11974CC**	-	**α/β-hydrolases superfamily protein**	**−7574**
8:19843994	2	LFMM_R, sNMF_R	—	m01oak12927cC	AT5G03780	TRFL10	342
8:27287480	1	**LFMM_R**	Zinc finger CCCH domain-containing protein 29-like	—	—	—	—
8:29021994	2	**LFMM_R, sNMF_R**	**ETO1-like protein 1**	—	—	—	—

(Continues)

TABLE 2 (Continued)

SNP (contig:position)*	# Sets	Sets	BLASTX ($E < 1e\text{-}10$)	Oak gene	TAIR ID	TAIR common name	Distance (bp; upstream negative)
8:29424317	1	**LFMM_R**	Uncharacterized protein	–	–	–	–
8:31873318	2	**LFMM_R, sNMF_R**	**Ribosomal protein S3**	–	–	–	
8:35468633	1	**LFMM_R**	Hypothetical protein	**m01oak00797cC**	–	–	**0**
8:38378045	1	**LFMM_R**	Uncharacterized protein	**m01oak07273cC**	**AT3G51270**	**Protein serine/threonine kinases; ATP binding; catalytics**	**0**
8:39960320	1	**LFMM_R**	–	–	–	–	–
8:52211286	1	**LFMM_R**	Uncharacterized protein	–	–	–	–
8:56967059, 56967217	2	LFMM_R, sNMF_R	–	m01oak05143CC	AT5G28350	Quinoprotein amine dehydrogese, β chain-like; RIC1-like guanyl-nucleotide exchange factor	0

Abbreviations: H, Hamilton; M, Madera; R, rangewide; T, Tejon.

*Bold rows are significant with $Q < 0.1$.

LFMM candidates in the transects, we opted to forgo modelling these candidate SNPs in the transects with GF. Instead, we used the 34 LFMM candidate SNPs from the rangewide data set and examined whether they collectively showed evidence of elevated climate association relative to all SNPs in the transects using GF. Six cases of SNP–climate variable associations in GF achieve greater importance for the LFMM candidate genes than the complete SNP set: Tmin in Hamilton (0.020 vs. 0.015), Tmax in Hamilton (0.032 vs. 0.018) and in Madera (0.027 vs. 0.007), CWD in Madera (0.018 vs. 0.008), and to some extent Tmin and CWD in Tejon (0.011 vs. 0.007; Figure 4).

Projecting the rangewide climate–SNP relationships onto maps revealed moderate changes in allelic composition from north to south (or northwest to southeast) in both the complete SNP data set and the LFMM candidate SNP data set (Figure 5; map for sNMF candidates depicted in supplement Figure S9 because they follow the overall SNP pattern as shown in Figure 3). Comparing the map based on candidate SNPs to that with all SNPs, the largest differences between putatively adaptive genetic variation and overall variation is in the southeast and select points in the north.

Map projections for the transects show that both overall and putatively adaptive genetic variation are aligned with the environmental, and elevational, gradient (Figure 6; also see Figures S1 and S2). The relationship in the Hamilton transect is more complex, however, and differences between candidate SNP and overall genetic variation do not follow a clear pattern. In Madera, candidate SNP variation deviates from the background most at higher

FIGURE 2 Venn diagrams showing (a) overlap of SNPs and loci among rangewide LFMM and sNMF tests and (b) overlap of loci surrounding top 100 SNPs in LFMM tests for each geographic context. Loci are defined as a 5 kb window centred on a SNP [Colour figure can be viewed at wileyonlinelibrary.com]

elevations. In Tejon, candidate SNP variation deviates most at the lowest and a middle elevation site and to some extent at a high elevation site.

Finally, we found that six LFMM candidate SNPs are among the strongest climate-associated SNPs in the GF rangewide model (contig:position: 4:32561505, 4:32561633, 7:20294827, 7:74552124, 8:29021994, 8:31873318; Table 2, Figure S10). None of these SNPs are in or near genes of known function, although one is 5 kb from oak gene m01oak18542CC. Three of these SNPs are also sNMF outliers.

4 | DISCUSSION

4.1 | Evidence of natural selection

We found signatures of natural selection by climate or local environment at over 600 SNPs (536 loci), some at multiple spatial scales across multiple analyses (Tables 1, 2, and S2). At the rangewide scale, five loci (defined by 5 kb windows) emerged as candidates from both LFMM and sNMF, six from both LFMM and GF, and three from all three analyses (Table 1). Furthermore, two loci were identified as candidates in sNMF at both the rangewide and transect scales. Additional overlap among sets is apparent when comparing the top one hundred associations regardless of statistical significance. LFMM candidates also had stronger associations with climate variables than the background when analysed in the multivariate, nonlinear gradient forest framework at the rangewide scale and at the transect scale for several climate variables (Figures 3 and 4). Taken together, our results offer robust evidence for novel candidate genes underlying local climate adaptation or spatially varying fitness related to current or past local environments.

LFMM tests within the transects may have been underpowered due to small sample size, as only a few loci are significant. An alternative to low power is that local adaptation is not strong among sites within transects. Similarly, valley oak populations may not be locally adapted to current climate gradients, but rather past climate gradients, as several studies at the rangewide scale suggest a lag in local adaptation that may date to the Last Glacial Maximum (Browne et al., 2019; Gugger et al., 2013). However, these studies, along with field-based trait data, still point to spatial variation in fitness that is at least correlated with present climate as well as growth rates of progeny of the trees in our study that were grown in a common garden (Albarrán-Lara et al., 2015; Browne et al., 2019; Delfino Mix et al., 2015). Furthermore, we present evidence that loci significant

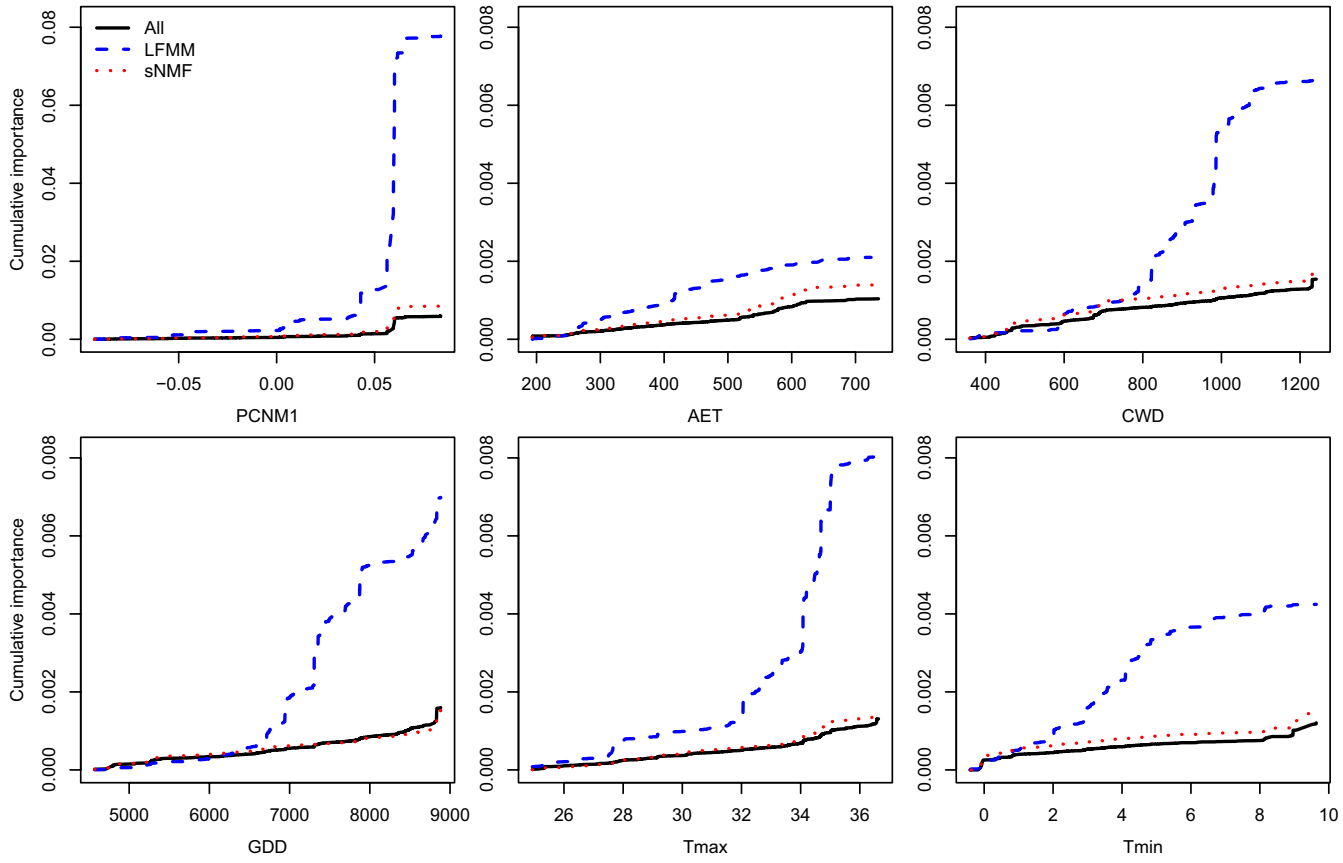

FIGURE 3 Cumulative importance of allelic change along environmental gradients at the rangewide scale for all SNPs (solid black line), LFMM candidates (blue dashed), and sNMF candidates (red dotted). Cumulative importance depicts the amount of allelic change along the environmental gradient (see Methods). Y-axes are on the same scale for climate variables but not PCNM1. PCNM, principal coordinate of neighbour matrices (spatial variable); AET, actual evapotranspiration; CWD, climatic water deficit; GDD, growing degree-days; Tmax, mean maximum annual temperature; Tmin, mean minimum annual temperature [Colour figure can be viewed at wileyonlinelibrary.com]

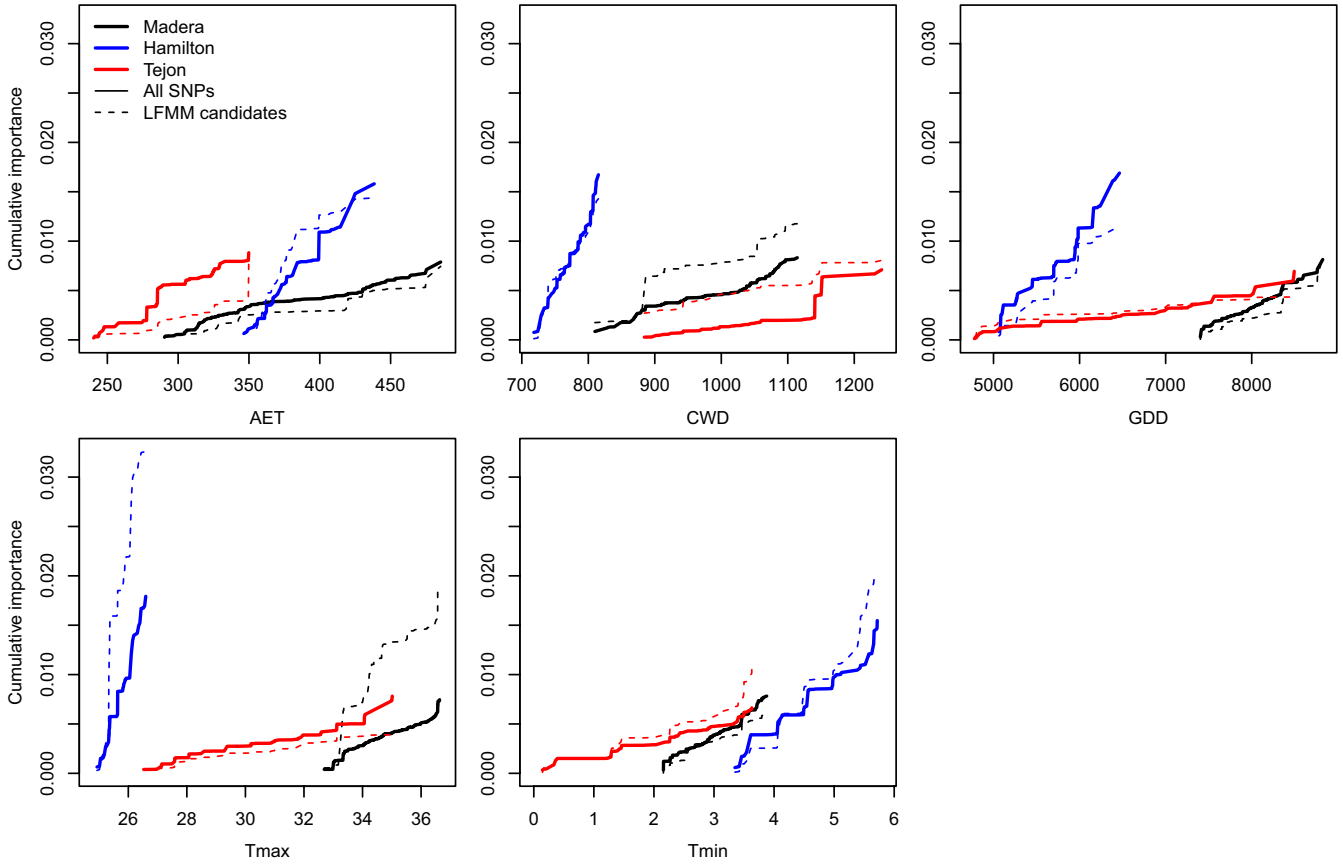

FIGURE 4 Cumulative importance of allelic change along environmental gradients for each transect (line colours) for all SNPs (solid line) versus LFMM candidates (dashed line). Cumulative importance depicts the amount of allelic change along the environmental gradient (see Methods). AET, actual evapotranspiration; CWD, climatic water deficit; GDD, growing degree-days; Tmax, mean maximum annual temperature; Tmin, mean minimum annual temperature [Colour figure can be viewed at wileyonlinelibrary.com]

at the rangewide scale exhibit elevated association at the transect scale when analysed collectively (see below).

The genomic context and gene functions of the candidate loci vary substantially and include some interesting examples that are plausibly related to local adaptation. From rangewide LFMM, NIN-like protein 7 modulates nitrate sensing and metabolism, and *Arabidopsis* mutants show enhanced drought tolerance under nitrogen starvation (Castaings et al., 2009). An auxin transporter, which may be involved in growth regulation, is also implicated. From sNMF analysis of the Tejon transect, we find a protein involved in embryogenesis and seed storage protein expression (scarecrow-like protein; Hamilton and Tejon), and a putative pathogen resistance gene (rust resistance kinase Lr10-like). A pair of other disease resistance proteins are also significant in Madera and Hamilton, respectively, in LFMM analyses. Pathogen response has been suggested as an important mechanism in the diversification of California oaks (Cokus et al., 2015) and more broadly in the evolution of oaks as long-lived species (Plomion et al., 2018; Sork et al., in prep.). Although these functions are based on orthology with *Arabidopsis* genes, they are suggestive of the functional relevance for adaptation in oaks.

A number of candidate loci in this study overlap with previous studies in oaks. For example, a protein containing methyl-CpG-binding domain (m01oak10601Ct) was associated with Tmin in a transcriptome-wide SNP–climate association study of *Q. lobata* (Gugger et al., 2016) and is an sNMF outlier in this study. Methyl-CpG has several possible functions in gene regulation or transposon silencing and has itself been shown to correlate with environment and patterns of local adaptation in valley oak (Gugger, Fitz-Gibbon, et al., 2016; Platt et al., 2015). A protein with an F-box domain, which can be involved in growth or pathogen defence, showed significant population by water stress treatment (~genotype × environment) in a gene expression experiment and is about 1.5 kb from another significant sNMF outlier (Gugger et al., 2017). Another sNMF SNP is 4 kb from the triophosphate isomerase gene (m01oak01126cC), which was expressed in a European oak species during bud burst—a trait known to be controlled by climate and thought to be under selection for local adaptation in oaks (Alberto et al., 2013; Derory et al., 2010; Savolainen et al., 2007). Three other candidate genes from European species (www.evoltree.eu; (Alberto et al., 2013) are near or overlap with significant SNPs in this study: one detected in LFMM at the rangewide scale (3.5 kb from m01oak02014cC) and one from rangewide sNMF (m01oak02471jC) that functions in embryogenesis, pollen tube growth, and mitochondrial development; and one from Tejon sNMF (2.4 kb from m01oak10456CC). However, we did not find overlap of significant loci between our study and

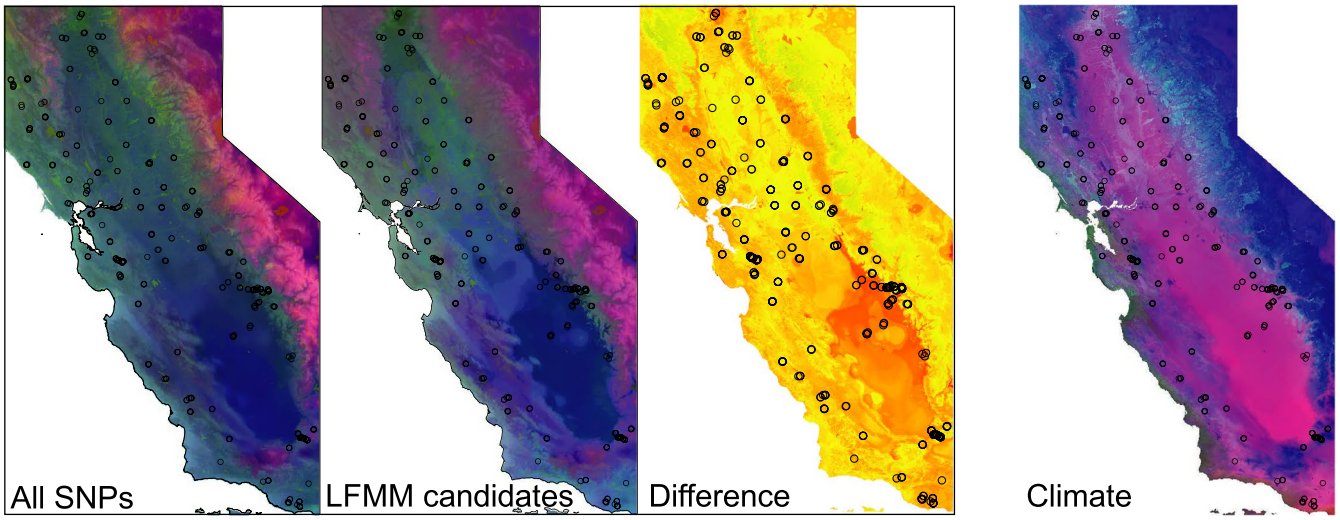

FIGURE 5 Mapped patterns of allelic composition (genetic variation) as predicted from transformed environmental and spatial variables from gradient forest models based on (a) all SNPs and (b) LFMM candidate SNPs. Within each of these panels, similar colours represent similar expected genetic compositions (colours are not necessarily comparable between these panels). Both show moderate north to south changes in allelic composition within the sampled area (open points). (c) Procrustes residuals from comparison of (a) and (b) showing that the largest differences (redder colour) between putatively adaptive genetic variation and overall variation is in the southeast and select points in the north. (d) The first three principal components of climate variation are shown in red, green blue colour scales, respectively, for general comparison with mapped patterns of genetic variation and differences in the other panels

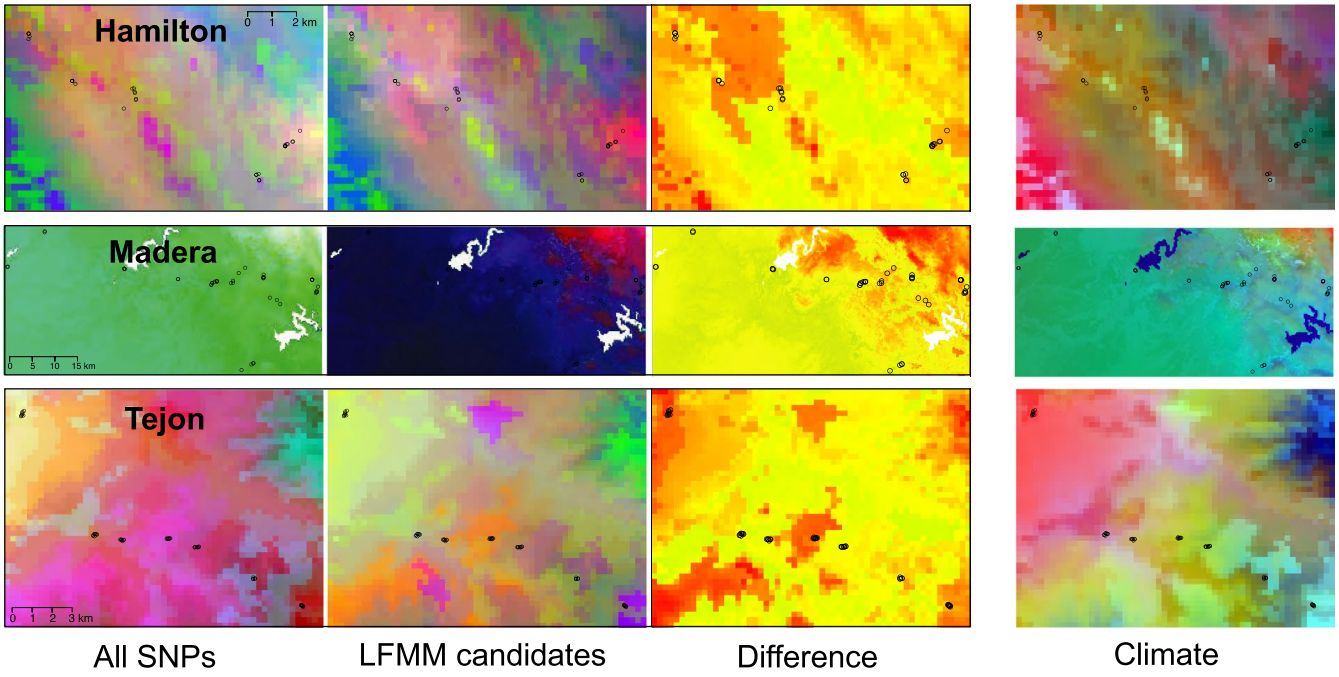

FIGURE 6 Mapped patterns of allelic composition (genetic variation) as predicted from transformed environmental and spatial from gradient forest models based on all SNPs (left) and LFMM candidate SNPs (centre) and the difference between them based on Procrustes residuals (right) for the (a) Hamilton; (b) Madera; and (c) Tejon transects. Within each of the left two panels, similar colours represent similar expected genetic compositions; colours are not necessarily comparable between these panels. In the third column of panels, larger differences between overall genetic variation and putative adaptive variation are redder. The right column of images depicts the first three principal components of climate variation. Sample coordinates are shown as open circles

two recent landscape genomic papers of European oak species (Pina-Martins et al., 2019; Rellstab et al., 2016). Lack of overlap is not surprising because the methods, and thus genomic loci that were analysed, varied greatly across studies.

Perhaps most interestingly, several SNPs from our study in *Q. lobata* overlap with those detected in the Mexican oak, *Q. rugosa*, using the same GBS protocol and similar association methods (Martins et al., 2018). Six loci are differentiation-based outliers at

the rangewide scale in both studies, including a MyB DNA binding protein (m01oak09781CC); O-fucosyltransferase family protein (m01oak07807cC); P-loop containing nucleoside triphosphate hydrolases superfamily protein involved in nucleotide binding and metabolic processes (m01oak09105CC); a hydroxyproline-rich glycoprotein family protein that controls plant development and flowering through a signal transduction pathway and is involved in stress response and circadian rhythm (m01oak08883CC), and two others of unknown function (m01oak06594CC, m01oak52010cC; Table 2). A rangewide LFMM candidate is 4.5 kb from an unknown protein and is a differentiation-based outlier in *Q. rugosa* (m01oak45522cC). Finally, SNPs within a few hundred bases of a protein kinase (m01oak13604CC) are associated with temperature seasonality in *Q. rugosa* and is an sNMF candidate in the Tejon transect in this study. The overlap among these species raises the question of whether the same loci might be involved in local climate adaptation in different congeneric species that inhabit different geographic contexts.

To the extent that the candidate loci identified in this study are adaptive, we cannot say whether these loci would be causally involved in local adaptation at different spatial scales because GBS only provides fragmentary evidence of variation along the genome and may or may not include the causal variants. Nonetheless, many of the significant loci are likely to be in or near causal loci because linkage disequilibrium decays rapidly along the oak genome (Sork, Squire, et al., 2016).

4.2 | Natural selection at multiple spatial scales

A key finding of our analysis is the evidence for selection on some loci at multiple spatial scales (Figures 3 and 4; Table 2). When modelling SNP–climate associations at the transect scale with SNPs that were significant at the rangewide scale, we found six cases of multi-SNP association elevated above the background association (Figure 4). In Hamilton and Madera, Tmax is especially strongly associated with rangewide LFMM candidates, and in Tejon and Madera, the relationship with CWD (an integrated measure of water availability) is particularly strong. Furthermore, in a few cases associations with PCNM spatial variables are high, suggestive of exceptionally strong spatial structuring or association with unmeasured environmental variation that PCNMs might capture (Figure S8; Manel et al., 2010). In addition, seven specific loci overlap in the top one hundred LFMM associations at both rangewide and transect scales. The functions of several genes implicated in multiple spatial contexts also have plausible functions for local adaptation. For example, Y-family DNA polymerase H (significant in multiple LFMM contexts) catalyses reactions in response to damage from ultraviolet light, and several genes encode transcription regulators (significant in multiple sNMF contexts). The fact that loci, individually and/or collectively, exhibit elevated climate associations at multiple spatial scales suggests that in general these associations are unlikely to be spurious and unlikely to be the product of the stochastic effects of genetic drift (Rellstab et al., 2017).

Significant associations of the same loci at multiple spatial scales can reflect independent selective processes on the same loci in each context or a global selective process that we have merely detected in multiple ways. Our observation that some loci in this study are also involved in local adaptation in another oak species (*Q. rugosa*) supports the possibility that parallel adaptive processes are at play within *Q. lobata*. This possibility further raises the interesting question of parallel adaptation via different variants at the same causal loci or selection on and sorting of the same variants. Most of the SNPs analysed are shared across each spatial context, and we found a number of cases in which the same SNPs appear significant in more than one context (Tables 1, 2, S2, and S6). However, it was not always possible to compare exactly the same SNPs due to data filtering criteria. In a few cases, we find different SNPs at the same locus in different contexts are significant in our tests (Tables 2, S2, and S6). The fragmentary nature of GBS data limits our ability to resolve this question here, motivating future studies using whole-genome sequencing.

Finally, some SNPs/loci in our study were uniquely identified as candidates in only one context or spatial scale (Tables 1 and S3) and most of the top 100 SNPs from LFMM differed among geographic contexts (Figure 2b; Table S6). Future studies with increased genomic coverage and increased sampling at the local spatial scales will be needed to determine the significance of these differences. Broadly, however, our findings suggest the importance of considering spatial scale and geographic context in evaluating patterns of local climate adaptation, consistent with findings in *Arabidopsis* of unique patterns of climate adaptation at various spatial scales (Brachi et al., 2013).

4.3 | Genetic structure and climate association

The genetic structure of valley oak has been discussed in a number of other previous publications, and the patterns in our study mirror the north–south patterns observed in recent papers (Ashley et al., 2015; Gugger, Cokus, et al., 2016) more so than the east–west observed in earlier studies (Grivet et al., 2006; Gugger et al., 2013; Sork et al., 2010). However, a more general northwest to southeast pattern is visible (Figures 1 and 5), and we suggest that the different markers used in previous studies may have been sensitive to different components of this structure.

In past studies and the present one background genetic structure has been shown to be associated with climate variables, probably due to the influence of neutral forces such as gene flow that aligns with climate gradients (e.g., via flowering phenology) or the influence of climate on local demographics. As a result, we expected that genetic structure-based outliers from sNMF would also probably show elevated climate associations. Interestingly, however, sNMF candidates are not climate-associated on average beyond the overall background association of genetic and climate variation (Figure 3), though a few loci are significant in LFMM analyses. In some cases, sNMF candidates exhibit elevated associations with spatial variables

in the gradient forest analyses (e.g., PCNM1, PCNM2, Figure S8), further supporting their exceptionally strong spatial structure and offering the possibility of a relationship with unmeasured environmental variables captured by PCNMs. Thus, the sNMF candidates might include many false positives reflecting drift-related climate association, or they might represent loci that are under diversifying selection by forces unrelated to the climate variables that we measured.

4.4 | Mapping local adaptation

Mapping putatively adaptive SNPs extends our observations to unsampled regions of the species distribution and reveals where geographic patterns of adaptive variation deviate from the background genetic structure (Figures 5 and 6). Putatively adaptive and background genetic variation are generally correlated broadly and within transects. However, populations in the southeastern portion of the distribution, such as the southern Central Valley, as well as a few small pockets in the far north and south, show marked deviation. Within transects, the patterns are patchy and unique. One might posit that these regions are where local climate adaptation is particularly strong, which can be tested in ongoing common garden experiments (Delfino Mix et al., 2015). Indeed, these regions at the rangewide scale represent some of the most extreme climates in the distribution. Similar patterns in which climatically unique or extreme regions show the strongest deviation of putatively adaptive and neutral genetic variation were also observed in *Populus* and *Q. rugosa* (Fitzpatrick & Keller, 2015; Martins et al., 2018), though the scale differs in each study. Collectively, these maps at multiple spatial scales highlight the potential importance of these loci in adaptation. More generally, the maps of adaptive variation can provide information for choosing seed sources for conservation management or other planting initiatives (Aitken & Bemmels, 2016; Fitzpatrick & Keller, 2015; Gugger et al., 2018).

4.5 | Conclusion

In conclusion, this study provides evidence of both novel and previously implicated candidate loci for local adaptation. Many of the same loci, and perhaps the same SNPs, may underlie local adaptation at multiple spatial scales or in multiple geographic contexts. Candidate SNPs or loci identified in multiple contexts offer strong evidence for their importance in local adaptation and point to the reuse of the same loci for local adaptation to environment in different regions. Interestingly, the same loci may even be involved in local adaptation in different *Quercus* species, such as *Q. rugosa*. This observation raises the question of whether recurrent mutations in the same loci have been independently selected on or whether ancestral variation is maintained and selected. On the other hand, some SNPs/loci in our study were uniquely identified as candidates in only one context or spatial scale, warranting further investigation with whole-genome data and common garden evidence to determine whether these represent false-positives or unique mechanisms underlying local adaptation in these contexts. It is likely that whole-genome sequences would help find even more candidate loci and provide greater clarity in answering some of the unresolved questions in this study. However, this GBS study illustrates that local adaptation may occur through multiple mechanisms. Nonetheless, an advantage of a landscape genomic approach is that spatial patterns of adaptive genetic variation can be mapped to address fundamental questions in evolutionary biology and to facilitate the development of conservation strategies that take into account adaptive genetic variation.

ACKNOWLEDGEMENTS

We thank J.M. Peñaloza-Ramírez, A. Lentz, S. Lentz, S. Ratay, P. Prather, P.D. Hodgskiss, I. Pearse, K. Gaddis, P. Thompson, K. Zecker, and other volunteers for assistance sampling; H. Rai for help with genotyping by sequencing methods; and J. Chen, J. Zhao, K. Beckley, J. Tang, A. Hovaspasian, F. Ashrafi, M. Suiter, C. Huynh, T. Percival for help with laboratory work. The California Department of Parks and Recreation, Cosumnes River Preserve, Kaweah Oaks Preserve, National Park Service, USDA Forest Service, Tejon Ranch Conservancy, and the University of California Natural Reserve system provided permission to sample where applicable. This work was supported by an NSF PGRP grant (IOS-1444661) awarded to V.L.S., P.F.G., and colleagues, USDA Forest Service funds to J.W.W., and UCLA seed money to V.L.S. A.A.L. was supported by a UC-MEXUS/CONACyT Postdoctoral Research Fellowship. Any use of product names is for informational purposes only and does not imply endorsement by the US Government.

AUTHOR CONTRIBUTIONS

P.F.G., J.W.W., and V.L.S. designed the research; P.F.G., A.A.L., J.W.W., and V.L.S. performed the field work; P.F.G. and A.A.L. performed the lab work; P.F.G. and S.F.G. analyzed the data; and P.F.G. wrote the paper with input from all coauthors.

DATA AVAILABILITY STATEMENT

Illumina data are publicly available at the NCBI BioProject PRJNA587528, and filtered SNP data for each geographic context are available through Dryad https://doi.org/10.5061/dryad.5dv41ns4n.

ORCID

Paul F. Gugger https://orcid.org/0000-0002-4464-8453
Victoria L. Sork https://orcid.org/0000-0003-2677-1420

REFERENCES

Aitken, S. N., & Bemmels, J. B. (2016). Time to get moving: Assisted gene flow of forest trees. *Evolutionary Applications*, *9*(1), 271–290. https://doi.org/10.1111/eva.12293

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. https://doi.org/10.1146/annurev-ecolsys-110512-135747

Albarrán-Lara, A. L., Wright, J. W., Gugger, P. F., Delfino-Mix, A., Peñaloza-Ramírez, J., & Sork, V. L. (2015). Phenotypic variation in California populations of valley oak (Quercus lobata Née) sampled along elevational gradients. In *Proceedings of the seventh oak symposium: managing oak woodlands in a dynamic world: Vol. General Technical Report PSW-GTR-251* (pp. 433–445). : USDA Forest Service, Pacific Southwest Research Station, General Technical Report PSW-GTR-251.

Alberto, F. J., Derory, J., Boury, C., Frigerio, J.-M., Zimmermann, N. E., & Kremer, A. (2013). Imprints of natural selection along environmental gradients in phenology-related genes of *Quercus petraea*. *Genetics*, *195*, 495–512. https://doi.org/10.1534/genetics.113.153783

Ashley, M. V., Abraham, S. T., Backs, J. R., & Koenig, W. D. (2015). Landscape genetics and population structure in Valley Oak (*Quercus lobata* Née). *American Journal of Botany*, *102*(12), 2124–2131. https://doi.org/10.3732/ajb.1500182

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*(1), 289–300. https://doi.org/10.2307/2346101

Brachi, B., Villoutreix, R., Faure, N., Hautekèete, N., Piquot, Y., Pauwels, M., Roby, D., Cuguen, J., Bergelson, J., & Roux, F. (2013). Investigation of the geographical scale of adaptive phenological variation and its underlying genetics in Arabidopsis thaliana. *Molecular Ecology*, *22*(16), 4222–4240. https://doi.org/10.1111/mec.12396

Brown, G. R., Gill, G. P., Kuntz, R. J., Langley, C. H., & Neale, D. B. (2004). Nucleotide diversity and linkage disequilibrium in loblolly pine. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(42), 15255–15260. https://doi.org/10.1073/pnas.0404231101

Browne, L., Wright, J. W., Fitz-Gibbon, S., Gugger, P. F., & Sork, V. L. (2019). Adaptational lag to temperature in valley oak (*Quercus lobata*) can be mitigated by genome-informed assisted gene flow. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(50), 25179. https://doi.org/10.1073/pnas.1908771116

Castaings, L., Camargo, A., Pocholle, D., Gaudon, V., Texier, Y., Boutet-Mercey, S., Taconnat, L., Renou, J.-P., Daniel-Vedele, F., Fernandez, E., Meyer, C., & Krapp, A. (2009). The nodule inception-like protein 7 modulates nitrate sensing and metabolism in *Arabidopsis*. *The Plant Journal*, *57*(3), 426–435. https://doi.org/10.1111/j.1365-313X.2008.03695.x

Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., & Postlethwait, J. H. (2011). Stacks: Building and genotyping loci de novo from short-read sequences. *G3: Genes, Genomes, Genetics*, *1*(3), 171–182. https://doi.org/10.1534/g3.111.000240

Cavender-Bares, J. (2016). Diversity, distribution and ecosystem services of the North American oaks. *International Oaks*, *27*, 37–49.

Cavender-Bares, J. (2019). Diversification, adaptation, and community assembly of the American oaks (*Quercus*), a model clade for integrating ecology and evolution. *New Phytologist*, *221*(2), 669–692. https://doi.org/10.1111/nph.15450

Chang, C., Chow, C., Tellier, L., Vattikuti, S., Purcell, S., & Lee, J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience*, *4*(1), 7. https://doi.org/10.1186/s13742-015-0047-8

Cokus, S. J., Gugger, P. F., & Sork, V. L. (2015). Evolutionary insights from de novo transcriptome assembly and SNP discovery in California white oaks. *BMC Genomics*, *16*(1), 552. https://doi.org/10.1186/s12864-015-1761-4

Coop, L. B. (2017). *U.S. degree-day mapping calculator, version 4.0*. Retrieved from Oregon State University Integrated Plant Protection Center Web Site Publication E.10-04-1 website: http://uspest.org/cgi-bin/usmapmaker.pl

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., McVean, G., Durbin, R., & 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Davis, F. W., Stoms, D. M., Hollander, A. D., Thomas, K. A., Stine, P. A., Odion, D., Borchert, M. I., Thorne, J. H., Gray, M. V., Walker, R. E., Warner, K., & Graae, J. (1998). *The California gap analysis project-final report* (p. 255). University of California.

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*(5), 1383–1399. https://doi.org/10.1111/mec.12182

Delfino Mix, A., Wright, J. W., Gugger, P. F., Liang, C. F., & Sork, V. L. (2015). Establishing a range-wide provenance test in valley oak (Quercus lobata Née) at two California sites. In *Proceedings of the seventh oak symposium: managing oak woodlands in a dynamic world* (pp. 413–424). : USDA Forest Service, Pacific Southwest Research Station, General Technical Report PSW-GTR-251.

DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., Philippakis, A. A., del Angel, G., Rivas, M. A., Hanna, M., McKenna, A., Fennell, T. J., Kernytsky, A. M., Sivachenko, A. Y., Cibulskis, K., Gabriel, S. B., Altshuler, D., & Daly, M. J. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, *43*(5), 491–498. https://doi.org/10.1038/ng.806

Derory, J., Scotti-Saintagne, C., Bertocchi, E., Le Dantec, L., Graignic, N., Jauffres, A., Casasoli, M., Chancerel, E., Bodénès, C., Alberto, F., & Kremer, A. (2010). Contrasting relationships between the diversity of candidate genes and variation of bud burst in natural and segregating populations of European oaks. *Heredity*, *104*(5), 438–448. https://doi.org/10.1038/Hdy.2009.134

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*(3–4), 483–493. https://doi.org/10.1016/j.ecolmodel.2006.02.015

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*(3), 42. https://doi.org/10.1007/s11295-015-0863-0

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. https://doi.org/10.1534/genetics.110.115543

Ellis, N., Smith, S. J., & Pitcher, C. R. (2012). Gradient forests: Calculating importance gradients on physical predictors. *Ecology*, *93*(1), 156–168. https://doi.org/10.1890/11-0252.1

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One*, *6*(5), e19379. https://doi.org/10.1371/journal.pone.0019379

Endler, J. A. (1986). *Natural selection in the wild*. Princeton University Press.

Finn, R. D., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., Potter, S. C., Punta, M., Qureshi, M., Sangrador-Vegas, A., Salazar, G. A., Tate, J., & Bateman, A. (2016). The Pfam protein families database: Towards a more sustainable future. *Nucleic Acids Research*, *44*(D1), D279–D285. https://doi.org/10.1093/nar/gkv1344

Fitzpatrick, M. C., & Keller, S. R. (2015). Ecological genomics meets community-level modelling of biodiversity: Mapping the genomic landscape of current and future environmental adaptation. *Ecology Letters*, *18*(1), 1–16. https://doi.org/10.1111/ele.12376

Flint, A. L., & Flint, L. E. (2007). *Application of the basin characterization model to estimate in-place recharge and runoff potential in the Basin and Range carbonate-rock aquifer system, White Pine County, Nevada, and adjacent areas in Nevada and Utah*.

Flint, L., Flint, A., Thorne, J., & Boynton, R. (2013). Fine-scale hydrologic modeling for regional landscape applications: The California Basin Characterization Model development and performance. *Ecological Processes*, *2*(1), 25. https://doi.org/10.1186/2192-1709-2-25

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*(9), 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 in Ecology and Evolution*, *6*(8), 925–929. https://doi.org/10.1111/2041-210X.12382

Frichot, E., & François, O. (2015b). *A short manual for LFMM version 1.4 (command line version)*. Retrieved from http://membres-timc.imag.fr/Olivier.Francois/lfmm/files/note.pdf

Frichot, E., Mathieu, F., Trouillon, T., Bouchard, G., & François, O. (2014). Fast and efficient estimation of individual ancestry coefficients. *Genetics*, *196*(4), 973. https://doi.org/10.1534/genetics.113.160572

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*(7), 1687–1699. https://doi.org/10.1093/molbev/mst063

Gaddis, K. D., Zukin, H. L., Dieterich, I. A., Braker, E., & Sork, V. L. (2014). Effect of clonal reproduction on genetic structure in *Pentaclethra macroloba* (Fabaceae: Mimosoideae). *Revista de Biología Tropical*, *62*(2), 443–454.

Gower, J. C. (1971). General coefficient of similarity and some of its properties. *Biometrics*, *27*(4), 857–871. https://doi.org/10.2307/2528823

Grivet, D., Deguilloux, M.-F., Petit, R. J., & Sork, V. L. (2006). Contrasting patterns of historical colonization in white oaks (*Quercus* spp.) in California and Europe. *Molecular Ecology*, *15*(13), 4085–4093. https://doi.org/10.1111/j.1365-294X.2006.03083.x

Gugger, P. F., Cokus, S. J., & Sork, V. L. (2016). Association of transcriptome-wide sequence variation with climate gradients in valley oak (*Quercus lobata*). *Tree Genetics & Genomes*, *12*(2), 15. https://doi.org/10.1007/s11295-016-0975-1

Gugger, P. F., Fitz-Gibbon, S., Pellegrini, M., & Sork, V. L. (2016). Species-wide patterns of DNA methylation variation in *Quercus lobata* and their association with climate gradients. *Molecular Ecology*, *25*(8), 1665–1680. https://doi.org/10.1111/mec.13563

Gugger, P. F., Ikegami, M., & Sork, V. L. (2013). Influence of late Quaternary climate change on present patterns of genetic variation in valley oak. *Quercus Lobata Née*. *Molecular Ecology*, *22*(13), 3598–3612. https://doi.org/10.1111/mec.12317

Gugger, P. F., Liang, C. T., Sork, V. L., Hodgskiss, P., & Wright, J. W. (2018). Applying landscape genomic tools to forest management and restoration of Hawaiian koa (*Acacia koa*) in a changing environment. *Evolutionary Applications*, *11*(2), 231–242. https://doi.org/10.1111/eva.12534

Gugger, P. F., Peñaloza-Ramírez, J. M., Wright, J. W., & Sork, V. L. (2017). Whole-transcriptome response to water stress in a California endemic oak, *Quercus lobata*. *Tree Physiology*, *37*(5), 632–644. https://doi.org/10.1093/treephys/tpw122

Keller, S. R., Levsen, N., Olson, M. S., & Tiffin, P. (2012). Local adaptation in the flowering-time gene network of balsam poplar, *Populus balsamifera* L. *Molecular Biology and Evolution*, *29*(10), 3143–3152. https://doi.org/10.1093/molbev/mss121

Kremer, A., Ronce, O., Robledo-Arnuncio, J. J., Guillaume, F., Bohrer, G., Nathan, R., Bridle, J. R., Gomulkiewicz, R., Klein, E. K., Ritland, K., Kuparinen, A., Gerber, S., & Schueler, S. (2012). Long-distance gene flow and adaptation of forest trees to rapid climate change. *Ecology Letters*, *15*(4), 378–392. https://doi.org/10.1111/j.1461-0248.2012.01746.x

Lamesch, P., Berardini, T. Z., Li, D., Swarbreck, D., Wilks, C., Sasidharan, R., Muller, R., Dreher, K., Alexander, D. L., Garcia-Hernandez, M., Karthikeyan, A. S., Lee, C. H., Nelson, W. D., Ploetz, L., Singh, S., Wensel, A., & Huala, E. (2012). The Arabidopsis Information Resource (TAIR): Improved gene annotation and new tools. *Nucleic Acids Research*, *40*(D1), D1202–D1210. https://doi.org/10.1093/nar/gkr1090

Lesur, I., Le Provost, G., Bento, P., Da Silva, C., Leplé, J.-C., Murat, F., Ueno, S., Bartholomé, J., Lalanne, C., Ehrenmann, F., Noirot, C., Burban, C., Léger, V., Amselem, J., Belser, C., Quesneville, H., Stierschneider, M., Fluch, S., Feldhahn, L., … Plomion, C. (2015). The oak gene expression atlas: Insights into Fagaceae genome evolution and the discovery of genes regulated during bud dormancy release. *BMC Genomics*, *16*(1), 1–23. https://doi.org/10.1186/s12864-015-1331-9

Li, H., & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*, *26*(5), 589–595. https://doi.org/10.1093/bioinformatics/btp698

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*(9), 2178–2192. https://doi.org/10.1111/mec.12725

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*(5), 1031–1046. https://doi.org/10.1111/mec.13100

Lowry, D. B., Hoban, S., Kelley, J. L., Lotterhos, K. E., Reed, L. K., Antolin, M. F., & Storfer, A. (2017). Breaking RAD: An evaluation of the utility of restriction site-associated DNA sequencing for genome scans of adaptation. *Molecular Ecology Resources*, *17*(2), 142–152. https://doi.org/10.1111/1755-0998.12635

Manel, S., Gugerli, F., Thuiller, W., Alvarez, N., Legendre, P., Holderegger, R., Gielly, L., Taberlet, P., & Intra Bio Div Consortium. (2012). Broad-scale adaptive genetic variation in alpine plants is driven by temperature and precipitation. *Molecular Ecology*, *21*(15), 3729–3738. https://doi.org/10.1111/j.1365-294X.2012.05656.x

Manel, S., Poncet, B. N., Legendre, P., Gugerli, F., & Holderegger, R. (2010). Common factors drive adaptive genetic variation at different spatial scales in *Arabis alpina*. *Molecular Ecology*, *19*(17), 3824–3835. https://doi.org/10.1111/j.1365-294X.2010.04716.x

Martins, H., Caye, K., Luu, K., Blum, M. G. B., & François, O. (2016). Identifying outlier loci in admixed and in continuous populations using ancestral population differentiation statistics. *Molecular Ecology*, *25*(20), 5029–5042. https://doi.org/10.1111/mec.13822

Martins, K., Gugger, P. F., Llanderal-Mendoza, J., González-Rodríguez, A., Fitz-Gibbon, S. T., Zhao, J.-L., Rodriguez-Correa, H., Oyama, K., & Sork, V. L. (2018). Landscape genomics provides evidence of climate-associated genetic variation in Mexican populations of *Quercus rugosa*. *Evolutionary Applications*, *11*(10), 1842–1858. https://doi.org/10.1111/eva.12684

McLaughlin, B. C., & Zavaleta, E. S. (2012). Predicting species responses to climate change: Demography and climate microrefugia in California valley oak (*Quercus lobata*). *Global Change Biology*, *18*(7), 2301–2312. https://doi.org/10.1111/j.1365-2486.2011.02630.x

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., Szoecs, E., & Wagner, H. (2018). *"vegan": Community Ecology Package (Version 2.5-1)*. Retrieved from https://cran.r-project.org

Parchman, T. L., Jahner, J. P., Uckele, K. A., Galland, L. M., & Eckert, A. J. (2018). RADseq approaches and applications for forest tree genetics. *Tree Genetics & Genomes*, *14*(3), 39. https://doi.org/10.1007/s11295-018-1251-3

Pina-Martins, F., Baptista, J., Pappas, G. Jr, & Paulo, O. S. (2019). New insights into adaptation and population structure of cork oak using genotyping by sequencing. *Global Change Biology*, *25*(1), 337–350. https://doi.org/10.1111/gcb.14497

Platt, A., Gugger, P. F., Pellegrini, M., & Sork, V. L. (2015). Genome-wide signature of local adaptation linked to variable CpG methylation in oak populations. *Molecular Ecology*, *24*(15), 3823–3830. https://doi.org/10.1111/mec.13230

Plomion, C., Aury, J.-M., Amselem, J., Alaeitabar, T., Barbe, V., Belser, C., Bergès, H., Bodénès, C., Boudet, N., Boury, C., Canaguier, A., Couloux, A., Da Silva, C., Duplessis, S., Ehrenmann, F., Estrada-Mairey, B., Fouteau, S., Francillonne, N., Gaspin, C., … Kremer, A. (2016). Decoding the oak genome: Public release of sequence data, assembly, annotation and publication strategies. *Molecular Ecology Resources*, *16*(1), 254–265. https://doi.org/10.1111/1755-0998.12425

Plomion, C., Aury, J.-M., Amselem, J., Leroy, T., Murat, F., Duplessis, S., Faye, S., Francillonne, N., Labadie, K., Le Provost, G., Lesur, I., Bartholomé, J., Faivre-Rampant, P., Kohler, A., Leplé, J.-C., Chantret, N., Chen, J., Diévart, A., Alaeitabar, T., … Salse, J. (2018). Oak genome reveals facets of long lifespan. *Nature Plants*, *4*(7), 440–452. https://doi.org/10.1038/s41477-018-0172-3

Porth, I., Koch, M., Berenyi, M., Burg, A., & Burg, K. (2005). Identification of adaptation-specific differences in mRNA expression of sessile and pedunculate oak based on osmotic-stress-induced genes. *Tree Physiology*, *25*(10), 1317–1329. https://doi.org/10.1093/treephys/25.10.1317

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*(2), 945–959.

R Core Team (2019). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/

Rellstab, C., Fischer, M. C., Zoller, S., Graf, R., Tedder, A., Shimizu, K. K., Widmer, A., Holderegger, R., & Gugerli, F. (2017). Local adaptation (mostly) remains local: Reassessing environmental associations of climate-related candidate SNPs in *Arabidopsis halleri*. *Heredity*, *118*, 193–201. https://doi.org/10.1038/hdy.2016.82

Rellstab, 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*(17), 4348–4370. https://doi.org/10.1111/mec.13322

Rellstab, C., Zoller, S., Walthert, L., Lesur, I., Pluess, A. R., Graf, R., Bodénès, C., Sperisen, C., Kremer, A., & Gugerli, F. (2016). Signatures of local adaptation in candidate genes of oaks (*Quercus* spp.) with respect to present and future climatic conditions. *Molecular Ecology*, *25*(23), 5907–5924. https://doi.org/10.1111/mec.13889

Savolainen, O., Lascoux, M., & Merila, J. (2013). Ecological genomics of local adaptation. *Nature Reviews Genetics*, *14*(11), 807–820. https://doi.org/10.1038/nrg3522

Savolainen, O., Pyhajarvi, 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

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*(4), 901–911. https://doi.org/10.1007/s11295-013-0596-x

Sork, V. L., Davis, F. W., Westfall, R., Flint, A., Ikegami, M., Wang, H., & Grivet, D. (2010). Gene movement and genetic association with regional climate gradients in California valley oak (*Quercus lobata* Née) in the face of climate change. *Molecular Ecology*, *19*(17), 3806–3823. https://doi.org/10.1111/j.1365-294X.2010.04726.x

Sork, V. L., Fitz-Gibbon, S. T., Puiu, D., Crepeau, M., Gugger, P. F., Sherman, R., Stevens, K., Langley, C. H., Pellegrini, M., & Salzberg, S. L. (2016). First draft assembly and annotation of the genome of a California endemic oak *Quercus lobata* Née (Fagaceae). *G3: Genes|Genomes|Genetics*, *6*(11), 3485. https://doi.org/10.1534/g3.116.030411

Sork, V. L., Squire, K. C., Gugger, P. F., Levy, E., Steele, S., & 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*(1), 33–46. https://doi.org/10.3732/ajb.1500162

Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T., & Zeileis, A. (2008). Conditional variable importance for random forests. *BMC Bioinformatics*, *9*(1), 307. https://doi.org/10.1186/1471-2105-9-307

Wang, I. J., & Bradburd, G. S. (2014). Isolation by environment. *Molecular Ecology*, *23*(23), 5649–5662. https://doi.org/10.1111/mec.12938

SUPPORTING INFORMATION

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

How to cite this article: Gugger PF, Fitz-Gibbon ST, Albarrán-Lara A, Wright JW, Sork VL. Landscape genomics of *Quercus lobata* reveals genes involved in local climate adaptation at multiple spatial scales. *Mol Ecol*. 2021;30:406–423. https://doi.org/10.1111/mec.15731