

Gene movement and genetic association with regional climate gradients in California valley oak (*Quercus lobata* Née) in the face of climate change

VICTORIA L. SORK,* FRANK W. DAVIS,† ROBERT WESTFALL,‡ ALAN FLINT,§
MAKIHICO IKEGAMI,† HONGFANG WANG¶ and DELPHINE GRIVET**

*Department of Ecology and Evolutionary Biology and Institute of the Environment, University of California Los Angeles, Box 951606, Los Angeles, CA 90095-1606, USA, †Bren School of Environmental Science and Management, University of California Santa Barbara, Santa Barbara, CA 93106-5131, USA, ‡USDA Forest Service, PSW Research Station, PO Box 245, Berkeley, CA 94701, USA, §U.S. Geological Survey, California Water Sciences Center, Sacramento, CA 95819, USA, ¶State Key Laboratory of Earth Surface Processes and Resource Ecology & College of Life Sciences, Beijing Normal University, Beijing 100875, China, **Department of Ecology and Genetics, Center of Forest Research, CIFOR-INIA, Carretera de la Coruña km 7.5, 28040 Madrid, Spain

Abstract

Rapid climate change jeopardizes tree populations by shifting current climate zones. To avoid extinction, tree populations must tolerate, adapt, or migrate. Here we investigate geographic patterns of genetic variation in valley oak, *Quercus lobata* Née, to assess how underlying genetic structure of populations might influence this species' ability to survive climate change. First, to understand how genetic lineages shape spatial genetic patterns, we examine historical patterns of colonization. Second, we examine the correlation between multivariate nuclear genetic variation and climatic variation. Third, to illustrate how geographic genetic variation could interact with regional patterns of 21st Century climate change, we produce region-specific bioclimatic distributions of valley oak using Maximum Entropy (MAXENT) models based on downscaled historical (1971–2000) and future (2070–2100) climate grids. Future climatologies are based on a moderate-high (A2) carbon emission scenario and two different global climate models. Chloroplast markers indicate historical range-wide connectivity via colonization, especially in the north. Multivariate nuclear genotypes show a strong association with climate variation that provides opportunity for local adaptation to the conditions within their climatic envelope. Comparison of regional current and projected patterns of climate suitability indicates that valley oaks grow in distinctly different climate conditions in different parts of their range. Our models predict widely different regional outcomes from local displacement of a few kilometres to hundreds of kilometres. We conclude that the relative importance of migration, adaptation, and tolerance are likely to vary widely for populations among regions, and that late 21st Century conditions could lead to regional extinctions.

Keywords: chloroplast and nuclear microsatellite, climate change, climate envelope, environmental gradients, landscape genetics

Received 28 January 2010; revision received 19 May 2010; accepted 19 May 2010

Introduction

Climate change over the last 100 years has jeopardized species and ecosystems throughout the world (Parme-

san 2006; Parry *et al.* 2007; Solomon *et al.* 2007). This rapid climate change creates particular problems for tree species because they are long-lived and immobile once the seedlings are established. Tree populations must be able to tolerate changing climate, adapt to new local conditions through selection on local genetic variation, or migrate to new favourable locations (Jackson &

Correspondence: Victoria L. Sork, Fax: 1 310 206 0484; E-mail: vlsork@ucla.edu

Overpeck 2000; Davis *et al.* 2005; Savolainen *et al.* 2007; Aitken *et al.* 2008). Rapid environmental change, such as occurred during the most recent-interglacial period of rapid warming, probably caused the extinction of some tree species, but it is also possible that some populations survived those changes because local climates remained within the species physiological tolerance limits (Bennett 1997). Furthermore, as Davis *et al.* (2005) point out, evolutionary responses of local populations will include changes in the ability to migrate to and tolerate new sites. Thus, range shifts related to major past climate change have involved the interplay of dispersal and adaptive evolutionary responses. At issue is whether contemporary species can respond to the very rapid rates of ongoing climate change.

Populations in different parts of a species' range and in different microhabitats will experience and respond to climate change differently (Rehfeldt *et al.* 2002, 2006). This differential response is due to both the genetic composition of local populations and the magnitude of climate change, which will vary geographically as well. Some populations could remain under tolerable climatic conditions while others will experience unsuitable climates. Furthermore, in areas with steep spatial climatic gradients range adjustments could occur over relatively short distances compared to more uniform regions (Loarie *et al.* 2009), although potential migration rates will be highly sensitive to habitat fragmentation in the landscape (Ledig 1992; Holsinger 1993; Young *et al.* 1996; Bawa & Dayanandan 1998; Sork & Smouse 2006). Thus, understanding tree species' responses to modern climate change will require better knowledge of geographic patterns of climate adaptation. These geographical patterns have been shaped by a combination of isolation by distance, genetic drift, and selection by climate, along with other factors. Some local populations may have tremendous phenotypic plasticity that would allow tolerance to change, some may have sufficient genetic variation that will allow the next generation of trees to respond to change through natural selection (Hamrick 2004), and others may have neither. Spatial variation in genetic composition relative to vagility will be pivotal to the ability of tree populations to respond to rapid climate change. For most tree species, identifying this variation is hampered by a lack of common garden experiments and other more definitive methods to identify local adaptation to climate (Davis *et al.* 2005; Manel *et al.* 2010).

Landscape genetics offers a range of spatial methods for studying the influence of ecological processes on genetic variation (Manel *et al.* 2003, 2010; Storfer *et al.* 2007). Spatially explicit genetic and environmental information can be used to look for the impacts of selection (Manel *et al.* 2003; Storfer *et al.* 2007), and the

association between genetic and environmental gradients has been well established as evidence of natural selection (Endler 1986; Mitton 1997; Manel *et al.* 2010). Forest genetics studies have shown that environmental heterogeneity influences the genetic differentiation among tree populations, creating geographic genetic patterns that are consistent with phenotypic traits (see Savolainen *et al.* 2007). One can look at this association to detect climate variables that are shaping the genetic structure of populations (e.g. Foll & Gaggiotti 2006) or even identify which genes are under pronounced natural selection (e.g. Jump *et al.* 2006; Joost *et al.* 2008). If we are trying to understand the potential impact of climate change, analysing the genetic structure of populations overlaid on current and future climate gradients could help identify tree populations that are put at greatest risk by rapid climate change.

Evaluating species-climate relationships and assessing impacts of climate change requires the use of relatively fine-grained climate grids over large areas. High spatial resolution is especially important in mountainous and coastal regions where climates vary widely over short distances. In recent years climatologists have developed downscaling methods that incorporate information on topography and marine influences to improve high-resolution climate grids produced from sparse station data or from coarse global climate models (Daly *et al.* 2008). These 800–1000 m resolution grids have helped advance understanding of species-climate association and vulnerability to projected 21st century climate change. However, for many regions where local populations may experience climatic conditions on a smaller scale, such as rugged areas, even finer spatial resolution may be needed to represent biologically important climate variation (e.g. Trivedi *et al.* 2008; Randin *et al.* 2009).

The motivation of this paper is to understand the potential response of valley oak (*Quercus lobata* Née), a widely distributed California endemic tree species, to climate change by exploring the associations of genetic variation with locally resolved climatic variables. Valley oak is an ecologically important tree species in California that warrants evolutionary and conservation attention. This oak has experienced extensive habitat loss (Kelly *et al.* 2005; Whipple *et al.* 2010), exacerbating the impacts of ongoing environmental change (Sork *et al.* 2009). Kueppers *et al.* (2005) predicted that the suitable habitat for *Quercus lobata* will shift to favorable sites of higher latitude and elevation. Like nearly all climate-based species distribution models, their models assume that a single species-climate relationship can be applied throughout the range. However, provenance studies for many other species have documented systematic intraspecific variation in climate responses over tens to

thousands of kilometers (Aitken *et al.* 2008). In an earlier study of valley oak (Grivet *et al.* 2008), we observed a high degree of nuclear genetic variation across longitude, latitude, and elevation, which suggests that populations across the species' range will not respond similarly to climate change.

Here, we address three objectives. First, to understand the extent to which genetic lineages may contribute to the genetic footprint of contemporary populations, we examine historical patterns of colonization. Using maternally inherited chloroplast genotypes of *Quercus lobata*, we examine range-wide connectivity that has resulted from seed movement by constructing population networks. Second, to assess the extent to which current populations are climatically structured in ways that may influence their response to climate change, we examine the correlation between multivariate nuclear genetic variation and climatic variation using canonical correlation analysis. Our genetic analysis first transforms microsatellite markers, which behave in a neutral manner as single loci, into multivariate genotypes that provide a sensitive measure of genetic differentiation due to the combined effects of gene flow, genetic drift, and selection across loci (Smouse & Williams 1982; Westfall & Conkle 1992; Zaretto *et al.* 1994). Third, to explore variation in predicted 21st century spatial displacement of climate zones across the species' range of valley oak, we parameterize region-specific species distribution models using downscaled grids of historical climate and outputs from two climate models based on a moderate-high carbon emission scenario.

Materials and methods

Study species and study area

Quercus lobata (Née) is a dominant or co-dominant species in oak savannas, oak woodlands and riparian forests across the foothills of the Sierra Nevada, Coastal Ranges, and Transverse Ranges that surround the Central Valley of California (Griffin & Critchfield 1972; Fig. 1). The species is winter-deciduous, phreatophytic, and associated with deep loamy soils. This species has an extended latitudinal distribution (34–40° latitude) and altitudinal range (from sea level to 1700 m) resulting in *Q. lobata* populations that are spread across diverse climatic and geographical zones.

Gene movement occurs through wind-pollination and the mating system is predominantly outcrossing (Sork *et al.* 2002a). In estimates of contemporary gene flow through pollen, average pollen dispersal distance computed with indirect and direct approaches yielded an estimate of 65–114 m, with a high propensity for short

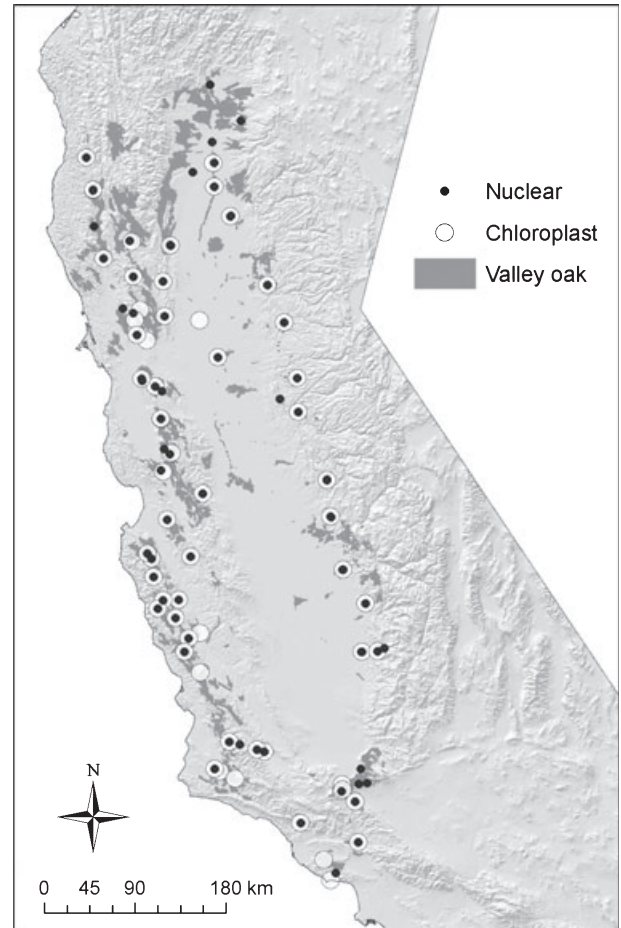


Fig. 1 Map of species range of *Quercus lobata* (shaded area) based on the California GAP Analysis database (Davis *et al.* 1998) and 80 sampling locations showing sites with nuclear genotypes (dots) and sites with chloroplast genotypes (circles).

distance dispersal (Smouse *et al.* 2001; Sork *et al.* 2002b; Pluess *et al.* 2009). However, these same populations also have a fat-tailed dispersal kernel that indicates long distance dispersal as well (Austerlitz *et al.* 2004; Pluess *et al.* 2009). Acorns mature in late September to early November of the year of pollination. Acorns lack enforced dormancy and typically germinate within 1–2 months after maturation. Dispersal by birds and rodents typically results in movement of seeds less than 150 m (Grivet *et al.* 2005; Scofield *et al.* 2010). Fine-scale genetic analysis of adults within a ~230-ha area indicates isolation by distance at ~350 m (σ) for seed and pollen together (Dutech *et al.* 2005). These studies of local historical and contemporary gene flow indicate that the scale of dispersal is on the range of 100–300 m, which allows for the opportunity for adaptation to local environmental conditions. Grivet *et al.* (2006) report regional scale of spatial autocorrelation in chloroplast genotypes up to 100 km.

Sampling

From 2003 through 2009, we sampled 330 *Q. lobata* individuals from 80 sites throughout the species range (Fig. 1), as part of an ongoing project on the phylogeography and conservation of California oaks (Grivet *et al.* 2006, 2008; Sork *et al.* 2009). For this study, our sample sizes include 190 chloroplast microsatellite haplotypes across 64 sites and 267 nuclear genotypes across 65 sites. The two data sets share 44 sites, and each samples the full species' range.

DNA extraction and genotyping

For the additional samples collected since our previous publications (Grivet *et al.* 2006, 2008; Sork *et al.* 2009), we analysed six chloroplast microsatellite genetic markers (Deguilloux *et al.* 2003) using the same DNA extraction method and PCR conditions. For the new samples and a subset of earlier samples, we measured the length of the amplified sequence by running an aliquot of each PCR product on an ABI 3700 capillary sequencer at the UCLA Sequencing & Genotyping Core Facility (<http://www.genetics.ucla.edu/sequencing/index.php>). The two methods were calibrated to ensure that we assigned the same haplotypes across primers.

The nuclear data set, which also augmented previous samples, was genotyped using the same methods described elsewhere (Grivet *et al.* 2008; Pluess *et al.* 2009) for six nuclear microsatellite primers: MSQ4, QpZAG1/5, QpZAG9, QpZAG36, QpZAG110, and QrZAG20. To verify repeatability, each sample was re-genotyped after repeating the PCR reactions. Again, we reran a subsample of previously genotyped individuals to ensure consistency in genotype assignments.

Data analyses

Chloroplast genotypes. To understand the extent to which seed movement has dispersed genes across the species range, we analysed the maternally inherited chloroplast genotypes through *PopGraph* (Dyer & Nason 2004; Garrick *et al.* 2009) to create a network of connections among sites. This software uses a graph-theoretical approach and analyses how genetic variation is distributed across the landscape. The software creates nodes based on the genetic variation within sites and identifies a network of connections among sites based on genetic covariation among all populations estimated simultaneously conditional on the entire data set. *PopGraph* identifies pairs of populations where long distance migration may have occurred by indicating population pairs with significantly greater inter-site distances than predicted by spatial separation based on a

chi-squared test (Garrick *et al.* 2009). *PopGraph* also identifies pairs of populations that are located significantly closer than predicted by inter-site separation, which suggests that a direct barrier might exist between the linear distances. We conducted analyses through the online version of Genetic Studio (<http://www.dyerlab.vmu/software>) to create the topology, and then viewed the population graph with a downloadable version of *Graph* (Dyer 2009). We plotted the geographical patterns using ArcMap 9.2 (ESRI Inc).

Climate data set. Grids of historical climate were produced by downscaling monthly 4 km PRISM climate data (Daly *et al.* 2008) for the period 1971–2000. Downscaling was accomplished with 90 m digital elevation grids using a modified gradient-inverse-distance square interpolation method (Flint & Flint 2007, 2010). The method preserves the structure of the coarser PRISM grids and uses 90 m digital elevation data for additional downscaling. Flint & Flint (2010) analysed predicted and observed air temperatures at more than 130 California Irrigation and Management Information System (CIMIS) stations and 80 National Weather Service stations in the rugged central Sierra Nevada to determine the reliability of downscaling from the 4 km PRISM cells to 90 m for California. They determined that there was little if any bias for air temperature and that the errors were the same in coastal as interior California. The downscaling algorithm either had no influence or improved interpolated estimates of temperature regimes at the validation sites.

We derived mean values from downscaled climate grids for the periods 1971–2000 and 2070–2100. We analysed mean annual precipitation and several temperature variables expected to be physiologically relevant and shown to be important in other studies of western tree distributions (Rehfeldt *et al.* 2006), including minimum temperature of the coldest month (T_{\min}), maximum temperature of the warmest month (T_{\max}), temperature seasonality (T_{seas}), and annual growing degree days above 5°C (GDD5):

$$GDD5 = \sum_{i=1}^{12} \left[\left(\frac{T_{\min_i} + T_{\max_i}}{2} \right) - 5 \right] D_i$$

where T_{\min} and T_{\max} are average minimum and maximum temperatures for month i and D is the number of days in month i . When $T_{\min_i} < 5^\circ\text{C} < T_{\max_i}$, we adjust the number of days in the month using $D^* = D(T_{\max} - 5)/(T_{\max} - T_{\min})$. Temperature seasonality was calculated as the standard deviation of monthly mean temperatures rescaled from 0 to 100 for the range of values observed in California. The climate variables

Table 1 Pearson's correlation coefficients of latitude and longitude spatial coordinates, and seven climate variables based on the 90 m resolution grids produced by downscaling of daily 800 m PRISM climate data (Daly *et al.* 2008)

	X	Y	GDD5	Tmax	Tmin	MAP
East-west (X)	1.000					
North-south (Y)	-0.585	1.000				
Growing degree days (GDD5)	0.430	-0.056	1.000			
Monthly maximum temperature (Tmax)	0.370	0.054	0.654	1.000		
Monthly minimum temperature (Tmin)	-0.009	0.074	0.435	-0.334	1.000	
Mean annual precipitation (MAP)	-0.499	0.684	-0.447	-0.232	-0.079	1.000
Temperature seasonality (Tseas)	0.533	0.167	0.467	0.709	-0.154	-0.095

are correlated with longitude (X) and latitude (Y) as well as with each other (Table 1).

Downscaled global climate model (GCM) predictions for the period 2070–2100 were produced by downscaling aggregated monthly outputs for the period 2070–2100 from the Parallel Climate Model (PCM1, developed by the National Center for Atmospheric Research and the Department of Energy) and the Goddard Fluid Dynamics Laboratory (GFDL) climate model. Downscaling involved initial downscaling of coarse PCM1 output using the constructed analogue method (Hidalgo *et al.* 2008). We are using a moderately high 21st Century carbon emission scenario (A2). The PCM-A2 and GFDL-A2 scenarios are two of four selected by the State of California for assessing climate change impacts, the others being PCM1 and GFDL with the lower B1 emission scenario (Cayan *et al.* 2008). Compared to the GFDL-A2 model runs, PCM1-A2 predicts lower late-century temperature increase (2.6°C vs. 4.5°C mean statewide annual temperature) and smaller differences in precipitation (e.g. -2% vs. -18% change in statewide mean annual precipitation).

Nuclear genotype data set. We used principal component analysis (PCA) to reduce dimensionality in the microsatellite allelic variables and create multivariate genetic variables in the same manner as described in Grivet *et al.* (2008). First, we transformed single locus genotypes into allelic variables by assigning a score of 0, 0.5, or 1, depending on whether the individual possesses a homozygous or heterozygous alleles at that locus (Westfall & Conkle 1992). The number of single variables created at each locus is the number of alleles minus one, which yielded 102 allelic variables. We then use PROC PRINCOM in SAS (SAS Institute 1989) to reduce the 102 variables into a smaller set of orthogonal axes.

Climate associations of genotypes. To test the relationship between genetic variation and climate, we conducted a canonical correlation analysis, which summarizes the relationship between two sets of variables by finding a

small number of linear combinations from each set that have the highest possible between-set correlations (Legendre & Legendre 1998). Using the PROC CANCORR function of SAS (SAS Institute 1989), we analysed the correlation between multivariate genotypic scores from the first 20 axes with five of the climate variables described above. To incorporate nonlinear structure in the climate variables, we used canonical trend surface analysis, which also includes squared terms and cross products among the five variables, which yielded a total of 20 climate variables. We tested the normality of residuals of the canonical models using SAS PROC UNIVARIATE. To diagnostically describe the forms of the equations (Box & Draper 1987), we analysed the observed scores for the first four canonical axes in SAS PROC RSREG. This procedure uses the method of least squares to fit quadratic response surface regression models and transforms that model to describe the shape of the response surface.

To assess the contribution of geography to the canonical correlation model, we partitioned out the spatial component of the climatic variation (Borcard *et al.* 1992), which was done in two steps. In the first step, using stepwise or sequential (Type I) sums of squares, we specified a quadratic model of geographic coordinates (x , y , and elevation, in metres) first, then the climatic model. In the second step, we specified the climatic model first, followed by the geographic model. In the first model, we have the proportion of the total sums of squares due to climate after removing geography; in the second, we have that proportion due to geography after removing that due to climate. From these, we can compute the proportion of total sums of squares that is collinear.

Modelling of future shifts in climate zones

To explore the extent to potential for spatial displacement of historical regional climates in the near future, we selected four sampling sites that were evenly distributed throughout the range of valley oak and also

spanned the range of genetic dissimilarity. A circular buffer with a radius of 75 km was created around each site.

We combined valley oak location information from our samples with occurrence data from a database of California vegetation plots (Seo *et al.* 2009) to generate a presence-only database of 529 valley oak locations that fell within the 4 subregions. Species distribution models were created using maximum entropy (MAXENT version 3.1.0) modelling (Phillips *et al.* 2006; Phillips & Dudík 2008) with mean annual precipitation, temperature seasonality and growing degree days above 5°. The MAXENT approach has proven to be very effective for bioclimatic modelling and performs better with presence-only data than most other available methods (Elith *et al.* 2006). Separate MAXENT models were produced based on nonoverlapping occurrence data in each of the four subregions. Model goodness-of-fit was evaluated using the area under the receiver operating characteristics curve (AUC).

To map late 21st Century distributions, each regional model produced using historical climate data was applied to state-wide grids of predicted future climate. To simplify presentation of the results, we show predicted presence or absence rather than probabilities using a threshold of $P = 0.5$ for species presence (Phillips & Dudík 2008).

Results

Both genetic markers included high genetic diversity. The chloroplast results indicate that *Quercus lobata* is variable across the six microsatellite markers with 2–5 haplotypes per marker with an overall average of 3.9 haplotypes, and a total of 45 unique multilocus haplotypes across the 64 sites and 190 individuals (see Table S1, Supporting information). The nuclear genetic diversity is greater than that of chloroplast, which is expected. The total number of alleles per nuclear microsatellite locus ranged from 10 to 29 alleles, with an average of 18 alleles across loci (see Table S1, Supporting information). The range of the average effective number of alleles per site was 2.5–4.7 alleles. The average number of alleles per site ranged from 2.9 to 5.5, which is fairly high given that most sites included only three individuals, and thus could not exceed 6 alleles per locus per site. We did not adjust these averages for sample size because the data are presented as background information and not for future statistical analysis.

Chloroplast genetic networks

The *PopGraph* network of populations based on chloroplast haplotypes indicates extensive historical connec-

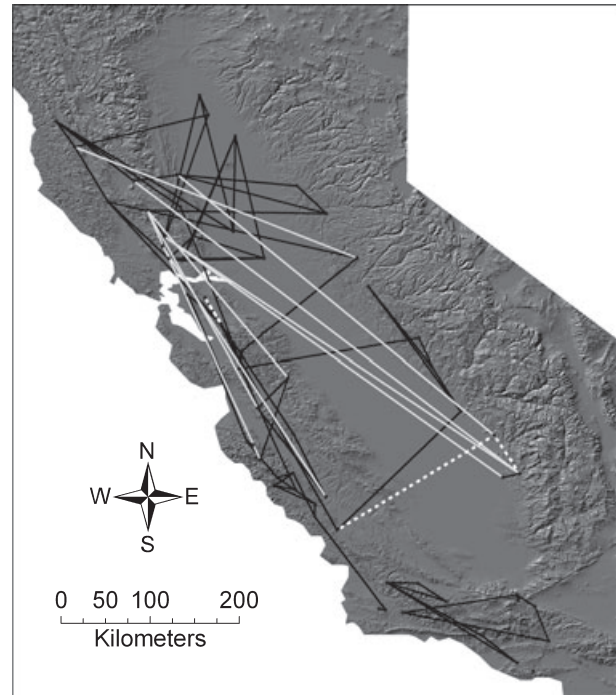


Fig. 2 Map of network showing 64 sites that are significantly connected with each other based on *PopGraph* topology (see text for details). Sites that are spatially more distant than expected from their genetic differences indicate long distance dispersal (white lines) and those that are spatially closer than expected from genetic differences indicate barriers (white dashed lines).

tivity among most populations (Fig. 2). The most common colonization patterns are north and south along the foothills of both the Coast Ranges and Sierra Nevada, but it is not uncommon to see east–west gene exchanges, especially in the northern part of the species range. Our data indicate occasional long distance gene exchange across the Central valley from the coastal northern site to an eastern southern site where the haplotypes are significantly more similar than expected based on spatial distance. Our analysis also identifies a network connection across the Central valley in southern California where the two sites are more genetically dissimilar than predicted by spatial distance.

Climate associations of genotypes

We used the multivariate scores from the first 20 principal component axes as multivariate genotypes for the canonical correlation analysis. We chose 20 PC's because this number captured a reasonable amount of the total genetic variation (the cumulative percentage was 41.52%, Table 2), and use of additional PC's improved the R^2 s in the subsequent correlation analysis marginally. The proportion of genetic variation attributed

Table 2 Eigenvalues of the Principal Component Axes (PCA) from the PC correlation matrix based on 103 allelic variables from 7 nuclear microsatellite loci

Prin	Eigenvalue	Proportion	Cumulative
Prin1	3.71933157	0.0365	0.0365
Prin2	3.04415982	0.0298	0.0663
Prin3	2.54074113	0.0249	0.0912
Prin4	2.43914075	0.0239	0.1151
Prin5	2.32291626	0.0228	0.1379
Prin6	2.23144517	0.0219	0.1598
Prin7	2.18288316	0.0214	0.1812
Prin8	2.09034455	0.0205	0.2017
Prin9	2.07807868	0.0204	0.2220
Prin10	1.98474470	0.0195	0.2415
Prin11	1.96732773	0.0193	0.2608
Prin12	1.88804604	0.0185	0.2793
Prin13	1.86301702	0.0183	0.2976
Prin14	1.79370240	0.0176	0.3152
Prin15	1.78359988	0.0175	0.3326
Prin16	1.74614542	0.0171	0.3498
Prin17	1.71333617	0.0168	0.3666
Prin18	1.68086312	0.0165	0.3830
Prin19	1.66361165	0.0163	0.3993
Prin20	1.61516464	0.0158	0.4152

to each ranged from 3.65% for the Prin1 to 1.58% for Prin20 (Table 2). For each of the 20 axes, multiple allelic variables loaded at some moderate level less than an absolute value of 0.20, and, in each of the PCA axes, there were an additional 4–7 variables with absolute value loadings between 0.20 and 0.38. This contribution of multiple alleles to the PCA's means that the principal component axes have reduced the initial 103 alleles into transformed genetic variables that are multivariate. The one notable exception is Prin1, which contained two loci each with two alleles, and these were highly loaded in reciprocal associations.

The canonical trend surface analysis tested the association between the 20 multilocus genetic variables with 20 climate variables that included the linear and nonlinear effects. The first six canonical axes were significant and the canonical correlations ranged from 0.766 to 0.434 (Table 3A), indicating a very strong association between the genetic and climatic variation across the species range. Focusing on the first four canonical axes, we see that Prin1 loads heavily with the first canonical axis (V1), while the other three axes include more balanced loadings of the 20 principal component axes (Table 3B). For the predicted scores of the first canonical vector (W1), temperature degree days above 5°C (GDD5) in its linear and squared form was slightly more highly correlated than other variables (Table 3C). For the second canonical vector (W2), temperature seasonality (*T*seas) correlated more highly than other

Table 3 Summary of Canonical Correlation Analysis based on nuclear microsatellite genetic markers. (A) Summary of the statistical tests of canonical correlation analysis for genetic variables vs. six climate variables, their quadratic forms, and their cross products. (B) Canonical correlations of the 20 PCAs of the 103 nuclear variables with the observed scores of the first four canonical axes. (C) Canonical correlations of the linear and higher order combinations of the climate variables for the first four canonical axes

Canonical axes	Canonical correlation	Pr > F		
(A) Statistical tests for first six canonical axes				
1	0.765617	<0.0001		
2	0.671367	<0.0001		
3	0.608128	<0.0001		
4	0.550027	<0.0001		
5	0.509182	<0.0028		
6	0.433851	0.1240		
	V1	V2	V3	V4
(B) Correlation of genotypic variables with observed scores for first four canonical axes				
Prin1	0.8920	-0.2027	-0.1921	-0.1195
Prin2	-0.1257	0.0223	-0.3730	0.1489
Prin3	-0.0760	0.0672	0.1665	-0.0410
Prin4	0.1320	-0.0183	-0.0793	-0.0936
Prin5	0.0686	-0.0985	-0.0877	0.2805
Prin6	-0.0876	-0.6196	0.1001	0.1530
Prin7	0.0597	0.0588	0.0863	-0.0083
Prin8	0.0080	-0.1703	-0.0063	0.2577
Prin9	0.0554	0.2863	0.2932	0.1181
Prin10	-0.2155	0.0421	-0.5973	0.1158
Prin11	-0.0836	-0.2273	-0.1960	0.1908
Prin12	0.1752	0.5170	-0.1708	0.3290
Prin13	0.0901	0.0579	0.1289	-0.2148
Prin14	-0.1740	0.0148	0.0447	-0.4573
Prin15	0.0662	0.1109	0.2110	0.2482
Prin16	0.0011	-0.1370	0.2672	-0.0261
Prin17	-0.0829	-0.1001	0.1389	0.0889
Prin18	-0.0535	-0.1266	0.1359	0.2369
Prin19	-0.0639	0.1911	-0.1010	-0.3095
Prin20	-0.0747	-0.1660	-0.2681	-0.3677
Climate variable	W1	W2	W3	W4
(C) Correlation of climate variables in linear, quadratic, and cross-product forms, with their canonical scores for four first axes. Within each canonical axis, selected high correlations are highlighted in bold. (See Table 1 for abbreviations.)				
GDD5	-0.3063	0.3755	-0.3768	-0.0524
MAP	0.2787	0.0227	0.5034	-0.3271
Tmax	-0.1586	0.3718	-0.1802	0.0695
Tmin	0.0260	0.1067	-0.0948	-0.1790
Tseas	-0.1903	0.6858	-0.0253	0.1719
GDD5 ²	-0.3224	0.4000	-0.3524	-0.0662
MAP ²	0.2011	-0.0156	0.3887	-0.3170
Tmax ²	-0.1573	0.3799	-0.1749	0.0623
Tmin ²	-0.0903	-0.0653	-0.0508	-0.0566
Tseas ²	-0.2501	0.7145	-0.0079	0.1732

Table 3 (Continued)

Climate variable	W1	W2	W3	W4
GDD5 × MAP	0.2440	0.1264	0.4703	-0.3596
GDD5 × Tmax	-0.2744	0.4409	-0.3152	-0.0186
GDD5 × Tmin	-0.0387	0.1410	-0.1146	-0.1595
GDD5 × Tseas	-0.2234	0.6875	-0.1033	0.1108
MAP × Tmax	0.2651	0.1032	0.4958	-0.3047
MAP × Tmin	0.1123	0.0118	0.0811	-0.3016
MAP × Tseas	0.1471	0.4277	0.4098	-0.0962
Tmax × Tmin	0.0097	0.1832	-0.1115	-0.1927
Tmax × Tseas	-0.2084	0.6674	-0.0393	0.1584
Tmin × Tseas	0.0970	0.4246	-0.1157	-0.1305

variables in its linear, quadratic, and cross product forms. In the third vector (W3), mean annual precipitation (MAP) was highly correlated in their linear, quadratic, and cross-product forms. In the fourth vector (W4), MAP is highly correlated. Overall, all of the vectors are correlated with more than one climate variable in linear and higher order forms, but GDD5, Tseas, and MAP, were the strongest climate variables for the first three canonical vectors, respectively. The residuals of the model for the first four canonical scores were normally distributed.

The RSREG procedure analysed the canonical trend surface for the first four canonical vectors (CV), and confirmed both that large proportions of variation in the models were nonlinear and also that substantial proportions were due to interactions among climate variables (Table 4). The forms of all models were saddles, where, from the middle (or stationary point), canonical scores rose or fell, moving off the stationary point. So depending on the signs of the relationships with each CV, PC scores rose or fell, as did frequencies of alleles associated with each PC, with variation in the CV. In the first two CVs, the stationary point was within the range of the climate data, whereas that point was outside the range of the data in CV3, representing a climate outside our data. Spatial patterns in precipitation are complex, not only dependent on orographic effects, but also on large-scale and local circulation pat-

terns. So that the stationary point is outside the data in CV3 is not unexpected. In CV4, there is little spatial association in plotted scores (not shown), which is indicated by the low collinearity with geography (next paragraph) and low correlations with climate. Instead, the low (albeit significant) association we did find was due to a few extreme values in the scores. This pattern suggests that trends in CV4 could be due to drift, rather than an influence of climate.

Our analysis of the collinearity between climate and spatial location indicated that 34% of the sums of squares (SS) for the first canonical axis were due to climate alone and 52% of the association between genetic and climate variation was due to the collinearity of climate and geography. In the second canonical axis, 43% of the SS was due to climate alone and 43% was due to collinearity with geography. In the third axis, 56% of SS was due to climate alone and 25% was due to the collinearity. In the fourth axis, 78% was due to climate alone and 4% due to collinearity.

To illustrate the spatial patterns in genetic and climate variables, we plotted the predicted canonical scores for the first three axes onto three climate gradients for GDD5, Tseas, and MAP, respectively. Because the scores are based on all five climate variables, we would not expect that the scores would precisely correspond to the temperate gradients, but these maps illustrate the trend in the data. Figure 3A illustrates the southern Sierra Nevada foothills and northern Sacramento Valley experience higher growing degree day sums than those sums elsewhere in the species' range (Fig. 3A). The canonical scores associated with the first axis (CCA1) were smaller in the areas in the southern Sierra Nevada foothills and very southern California. Temperature seasonality (Fig. 3B) is lower in the sites along the Coastal range foothills than along the Sierra-Nevada foothills, and here we see smaller canonical scores of the second axis (CCA2) located in the coastal area with low seasonality. Mean annual precipitation shows north south gradient (Fig. 3C), and the smaller CCA3 scores tend to be in areas with very little precipitation.

Table 4 Summary of Type I statistics from separate response surface regression models for scores of the four canonical axes of the CTSA reported in Table 3

	DF	V1		V2		V3		V4	
		R ²	Pr > F	R ²	Pr > F	R ²	Pr > F	R ²	Pr > F
Linear	5	0.2396	<0.0001	0.2474	<0.0001	0.1083	<0.0001	0.0719	0.0002
Quadratic	5	0.1264	<0.0001	0.0500	0.0006	0.1336	<0.0001	0.0458	0.0076
Cross-product	10	0.2201	<0.0001	0.1533	<0.0001	0.1279	<0.0001	0.1848	<0.0001
Total Model	20	0.5862	<0.0001	0.4507	<0.0001	0.3698	<0.0001	0.3025	<0.0001

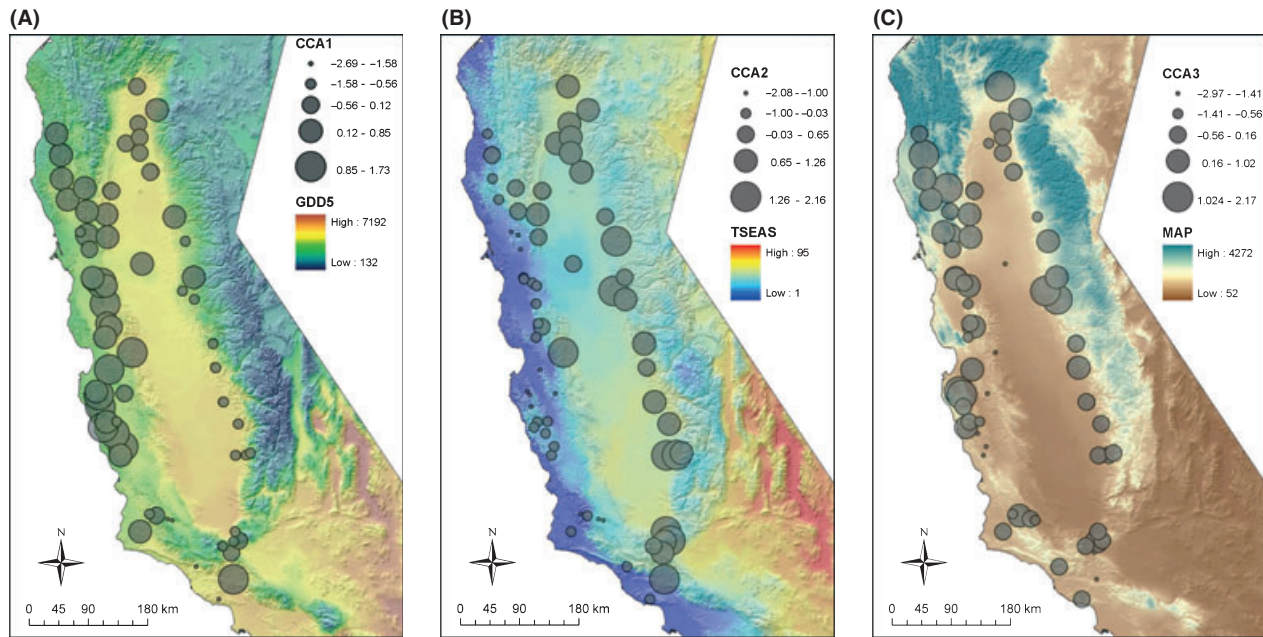


Fig. 3 Overlay of predicted canonical scores onto map of climate gradients. (A) Observed canonical scores for first vector (V1) vs. growing degree days (GDD5). (B) Observed canonical scores for second vector (V2) vs. Temperature Seasonality (Tseas). (C) Observed canonical scores for third vector (V3) vs. precipitation.

Future shifts in climate zones

Valley oaks occur in systematically different climates within the four test regions that span north–south and east–west extremes in the range (Table 5). The northern region and central Sierran foothill locations are generally wetter than locations in the southern range or the Central Coast. Northern sites experience lower GDD5 values while the central coastal and southern sites experience much lower temperature seasonality.

Both GFDL-A2 and PCM1-A2 climate scenarios predict large increases in GDD5 at all 4 test regions, although the GFDL-A2 forecasts consistently greater

increases. Both scenarios predict greatest warming in the southern and western subregions. Temperature seasonality is also predicted to increase in all regions, although the GFDL-A2 model predicts much larger increases. The models differ considerably in precipitation forecasts: the GFDL-A2 scenario predicts drier conditions in all four regions (more pronounced in the North and East) while the PCM1-A2 scenario includes modest increases in precipitation in all regions (less pronounced in the West; Table 5).

MAXENT models based on training data from the separate regions all have high values for model fit (AUC values ≥ 0.95) (Table 6). MAP contributes more

Table 5 Summary of downscaled climate values for valley oak localities in 4 circular regions 150 km in diameter within the current range. The number of nonoverlapping sites varies from 47 in the northernmost region to 198 in the southernmost region. Means and standard deviations are for the period 1971–2000. The values of Δ GFDL-A2 and Δ PCM-A2 are the differences of climate means for the period 2071–2100 minus the means for the period 1971–2000. Variables include mean annual precipitation in mm (MAP), temperature seasonality (the standard deviation of monthly mean temperatures, rescaled from 0 to 100 for minimum and maximum values in California), and growing degree days above 5° in °C (see text for explanation)

Region	No. of sites	MAP				Tseas				GDD5			
		Mean	SD	Δ GFDL-A2	Δ PCM-A2	Mean	SD	Δ GFDL-A2	Δ PCM-A2	Mean	SD	Δ GFDL-A2	Δ PCM-A2
North	138	766	173	-286	65	44	3	9	2	3365	583	1325	1096
East	47	874	232	-238	63	44	3	11	3	4058	298	1317	1069
West	198	439	64	-170	25	28	13	6	2	3959	736	1457	1159
South	146	535	128	-123	60	30	8	10	1	3880	263	1494	1201

Table 6 Summary of model fit (AUC, Area under the Receiver Operator Characteristics curve) and climate variable contributions for four climate-based species distribution models based on the Maximum Entropy (MAXENT) method

Region	N	AUC	Variable contributions (%)		
			MAP	T _{seas}	GDD5
North	138	0.991	49.5	11.4	39.1
East	47	0.946	52.8	19.7	27.5
West	198	0.987	18.8	44.1	37.1
South	146	0.983	42.4	32.9	24.9

than T_{seas} or GDD5 in three of the models but temperature variables contribute significantly to all models, especially the model for the western coastal region.

The predicted temperature changes lead to larger spatial shifts of suitable climate zones in some regions than others (Fig. 4). Climates similar to occupied sites in the western coastal region are predicted to be greatly reduced in extent (PCM-A2, Fig. 4b) or occur only in limited areas more than 100 km distant from current sites and nearer to the coast (GFDL-A2, Fig. 4a). Climatically suitable areas for trees in the northern region are predicted to expand greatly into surrounding foothill regions and highly suitable areas either partially overlap (GFDL-A2, Fig. 4c) or fully overlap with currently occupied sites (PCM1-A2, Fig. 4d). For interior sites currently located in the foothills of the Sierra Nevada, comparable climates are predicted to have little overlap with current sites but to shift locally upslope within 1–20 km of currently suitable climate locations (Fig. 4e,f). At the southern end of the range, areas of suitable climate comparable to currently occupied sites occur locally at higher elevations within a few km of currently occupied sites as well as at distant sites (Fig. 4g,h).

Discussion

The ability of a species to persist under the conditions of rapid climate change will be determined by the responses of local populations (Jackson & Overpeck 2000; Davis & Shaw 2001; Davis *et al.* 2005; Aitken *et al.* 2008). Will they be able to tolerate or adapt to changes in local conditions? Will they be able to track shifts in climate zones? Or, will they go extinct? Despite the fact that chloroplast markers show strong geographic structure (Grivet *et al.* 2008), here we show that historical seed movement of valley oak includes long distance colonization events. These events are mostly north and south, but we also observe east–west colonization that would have included the dispersal of nuclear genotypes

across the range. Thus, seed movement, when combined with gene flow through pollen movement, has resulted in continuously distributed genetic variation with a modest amount of nuclear geographic structure (Grivet *et al.* 2008). In this paper, we find that nuclear multilocus genetic structure shows a strong association with climatic gradients. This climatically associated genetic structure provides the opportunity for local adaptation to climate and may influence the response of local populations to future climate change.

Historical seed movement in Valley oak

The historical movement of valley oak over its evolutionary history appears to have been sufficient to allow gene exchange across its range. Valley oak has a relatively large amount of haplotype diversity, especially when contrasted with the European white oaks sampled on a similar spatial scale and genotyped with the same markers (Grivet *et al.* 2006). This contrast in haplotype diversity with European populations provides evidence that contemporary valley oak populations are older than the occurrence of the Last Glacial maxima. The lack of recent glaciation in California would have allowed populations to survive cold periods in multiple regions and retain their diversity. During the warm and cool periods of the Pleistocene, as populations expanded and contracted, they would have had ample opportunity to disperse and colonize many sites in California.

The network analysis of the chloroplast haplotypes demonstrates extensive gene movement through colonization. North/south movement along the foothills of the Coastal and Sierra Nevada Ranges seems particularly common. In the northern part of valley oak distribution, we see several connections across the Central Valley probably due to a strong riparian network (Grivet *et al.* 2008). In the southern part, long distance connectivity occurred from northwest to southeast, but it is difficult to distinguish whether the long distant events actually crossed the Central Valley diagonally or they moved in a stepping stone fashion across the northern Central Valley and then south along the foothills. The pattern of the population network suggests that the western region about the San Francisco Bay area may be a source region. In the south, we observed occasional east–west linkages across the southern Central valley, but with much less frequency. In one case, the inter-population genetic distance is significantly further than predicted by inter-site physical distance, which implies that those haplotypes may have taken an indirect route to arrive in the Sierra Nevada foothills.

The movement of seeds across the range would also have homogenized the nuclear genetic variation to

some extent. Moreover, given that gene flow in this species is even greater through pollen than seeds (Grivet *et al.* 2009a; Pluess *et al.* 2009), pollen-mediated gene movement should have also promoted gene exchange throughout the species range. We point out that the reduced seed movement across the southern Central valley between east and western populations would allow vicariance to contribute to the geographical genetic differences between those two regions. Future analyses can utilize models that will allow the separate contributions of phylogeographic structure and environmental variables that could help sort out the contribution of vicariance (Dyer *et al.* 2010). Nonetheless, we observe enough gene movement to deduce that the patterns of geographic structure we discuss below cannot be solely attributed to restricted gene movement.

Association of nuclear genetic variation with climatic variables

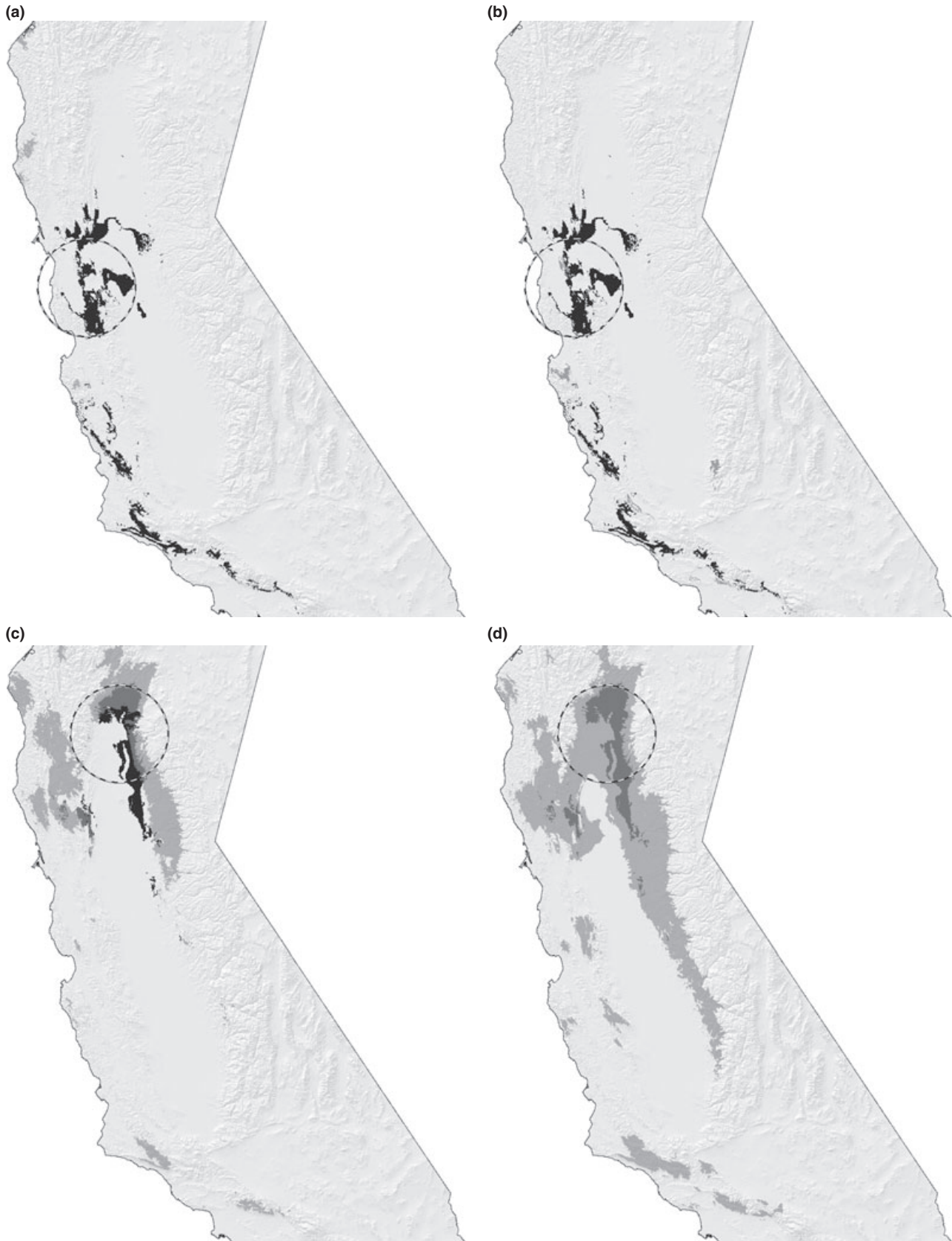
Multivariate genetic variation in valley oak is significantly correlated with multivariate climatic variation. It may seem paradoxical that markers such as nuclear microsatellites, which are putatively neutral as single loci, would show an association with climate; however, selection is not acting on these loci per se, but rather on the whole genotype. The accumulation of small effects across loci, especially linked loci, due to selection, along with gene flow and drift, creates a relatively sensitive measure of genetic differences among populations. The most conspicuous difference was between the eastern and western parts of the species range. However, for some of the canonical axes, those east–west differences are as strong in the northern part of the species range where connectivity is relatively high as in the southern part where the connectivity is less. Scores for the first canonical axis, which is most correlated with growing degree days, are highest in the central-western part of the range and the scores are lowest in the south and south eastern part of the range. Second axis scores, which are highly correlated with temperature seasonality, reveal genetic differentiation between populations in the Coast Range foothills vs. eastern populations adjoining the Sierra Nevada. For the third canonical axis, precipitation is a key climate variable, both linearly and in quadratic form.

These geographic patterns of climatically structured genetic variation may provide a useful indication of the geographic patterns of adaptive variation. They could have been shaped by selection on genes that are linked to the markers due to genetic hitchhiking (Barton 2000) or selective sweeps associated with polygenic traits (Pritchard *et al.* 2010). Recent work using genome scans to identify the association of AFLP loci with environmental gradients illustrates that selection can influence putatively neutral genetic variation through these processes (e.g. Freedman *et al.* 2010; Thomassen *et al.* 2010). Of course, isolation by distance and genetic drift could also shape geographic structure, but the correlation with environmental gradients indicates an influence of selection. In our study, we see evidence for the influence of selection when we examine colinearity between genetic variation, climate variation, and geography. Even when we partition out the effect of spatial location, genetic variation is strongly associated with climatic variation. While vicariance may contribute to the significance of the first canonical vector, the second and third canonical vectors provide evidence for the role of selection. Thus, the genetic structure of our populations illustrates the opportunity for adaptation to local climate and allows the formulation of hypotheses and identification of appropriate spatial scales for additional sampling. In the future, a landscape genomic approach, especially using candidate genes related to response to climate, will be a more direct way to identify geographic patterns of adaptive variation (Eckert *et al.* 2010; Manel *et al.* 2010; Thomassen *et al.* 2010).

Regional variation in distribution models under current and future climates

Maxent models based on mean annual precipitation, temperature seasonality and growing degree days provide good fits to observation data, with AUC values ranging from 0.95 to 0.99 for the different subregions. Consistent with theory (Davis *et al.* 2005) and empirical evidence (Rehfeldt *et al.* 2002), populations in different portions of the range experience systematically different climate regimes, so that climate models fitted to observations in one subregion have lower accuracy predicting the occurrence of valley oaks in more distant subregions (Fig. 4).

Fig. 4 Maps of climate-based species distribution models fitted using MAXENT and parameterized using locality data from each of four circled areas. Models were fitted using downscaled grids of mean annual precipitation (MAP), growing degree days above 5 degrees (GDD5), and temperature seasonality (T_{seas}) for the period 1971–2000. Black areas have probability of occurrence values exceeding 0.5. Gray areas are extrapolated distributions ($P > 0.5$) based on the same regional climate model but using climate forecasts based on GFDL-A2 (a, c, e, g) and PCM-A2 (b, d, f, h) models for the period 2071–2100. Dark gray areas are where $P > 0.5$ in both current and future models.



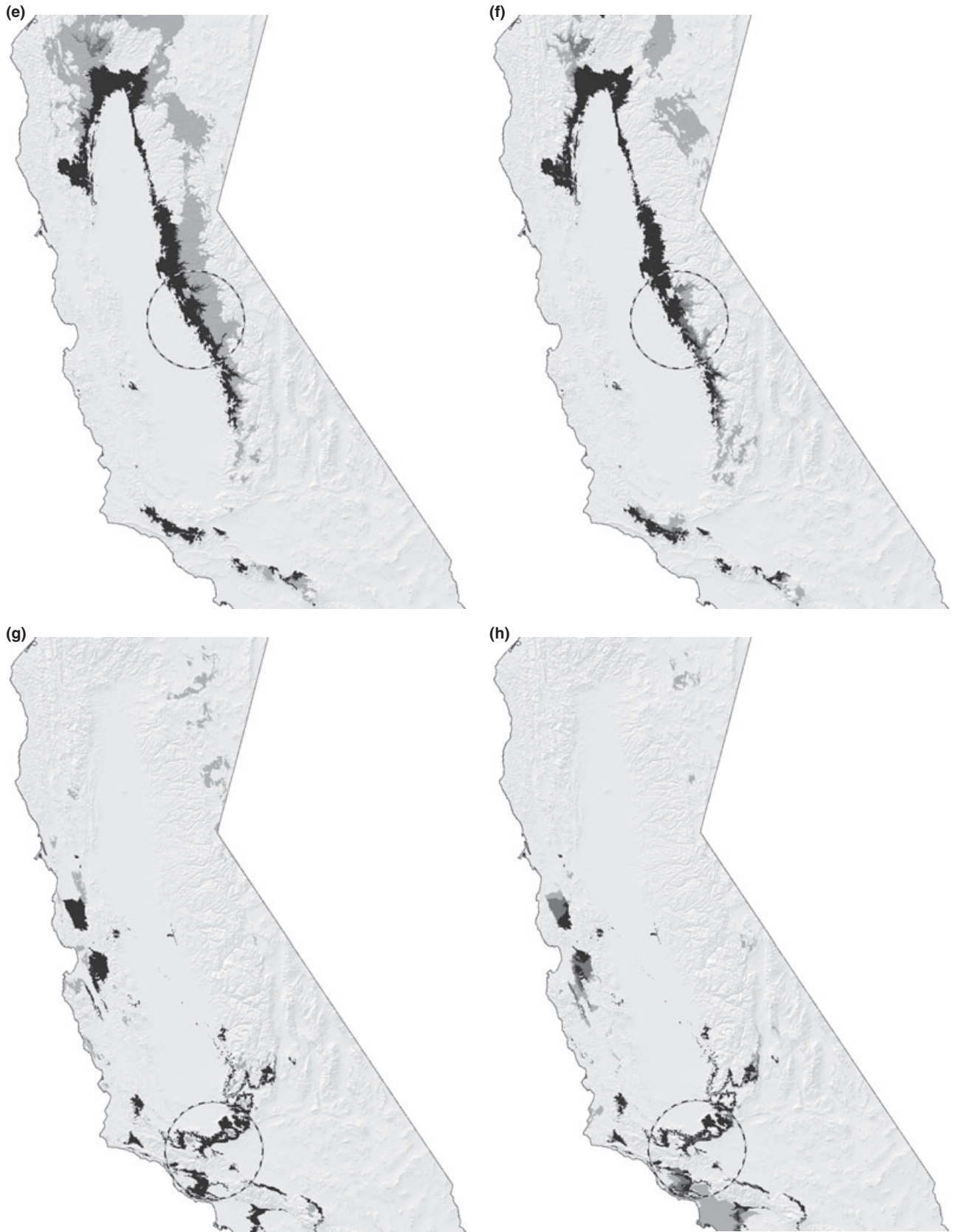


Fig. 4 (Continued)

The strong spatial structure of genetic variation in valley oak and associated inter-regional variation in climate-based distribution models suggest that valley oak populations will experience region-specific climate change impacts. The differences in the extent of climate displacement reflect both regional differences in the magnitude of projected climate change and the steepness of local topographically induced temperature and precipitation gradients. Our results indicate that central coastal populations are especially vulnerable because increases in temperature and seasonality are proportionally greater here and shallower spatial climate gradients equate to longer distances for range adjustments than for populations elsewhere in the range (Loarie *et al.* 2009).

We recognize that interpolated climate grids are imperfect models of actual local spatial variation and that such bioclimatic analyses would benefit from denser climate station networks. Comparison to other downscaling methods, such as nonparametric splines (Rehfeldt *et al.* 2006) and dynamic regional climate models (e.g. Kueppers *et al.* 2005) is also worth pursuing. Furthermore, other environmental factors could affect these results, notably local soil and hydrologic conditions. Valley oaks are associated with deep loamy soils as well as valley floor and riparian sites with relatively high water tables. We are currently examining the interaction of soil factors and climate factors and comparing different downscaling methods in regional distribution models for the species.

Over a long evolutionary period, tree populations have tremendous potential for gene flow through seeds and pollen (Aitken *et al.* 2008). In our estimates of contemporary pollen dispersal, the size of a genetic neighbourhood is on the scale of 100 m but with potential for rare long distant dispersal (Austerlitz *et al.* 2004; Pluess *et al.* 2009) that could spread novel alleles for several to tens of kilometres. Our studies of seed dispersal indicate very restricted dispersal, but here too occasional long distance dispersal events over a kilometre take place (Sork *et al.* manuscript in preparation). For example, birds such as jays can regularly transport acorns several kilometres and plant them in the soil (Gomez 2003). Palaeoecological studies of eastern North American oaks, based on recolonization after glaciations, estimate that migration rates are about 100 m per year (McLachlan & Clark 2004), or 6–8 km in 60–80 years, which is less than the climate zone shifts for many parts of the valley oak range predicted by our models. Thus, tracking climate change is probably not a likely scenario for many valley oak populations under the current rapid rates of climate change.

The future of valley oak has many challenges such as regeneration difficulties and landscape transforma-

tion (Tyler *et al.* 2006). The survival of extant populations will require either an ability to adapt to new climates or the presence of individuals that can tolerate the new conditions. As a species with relatively long generation time, adaptation is unlikely to be an effective response. However, it is possible that local populations might include individuals that can tolerate new conditions, which is especially likely for those populations in regions with variable climate conditions (e.g. Sierra foothills). Our regional climate envelopes indicate that the three climate variables we examined contribute differently across regions. For example, mean annual precipitation contributes most to the distribution of valley oak in the east, north, and south regions (42–53%), but its contribution is small in the west region (19%). In contrast, temperature seasonality is important in the west and south regions, but less so in the North and East regions, where seasonality is actually very high (see Fig. 3B). As future climates in the west become warmer and more seasonal, one can speculate whether existing populations have the plasticity to tolerate this shift in climate. Our analysis would suggest that western regional populations may be more vulnerable to climate change than the eastern ones and that tolerance to climate change, in general, is going to be regionally dependent on the interaction between local genetic variation and climate conditions.

Implications for adaptive genetic variation

As we consider the vulnerability of populations to respond to climate change, patterns of adaptive genetic variation will be more critical than those of neutral genetic variation (Holderegger *et al.* 2006). As a first step, it is useful to simply understand how geographic genetic structure is associated with climate variables so that we can make some first order approximations about potential responses (Manel *et al.* 2010). Forest geneticists have used genetic markers in this way to construct seed zones to select genotypes for plantations (see Westfall & Conkle 1992). The microsatellite markers examined in this study have described geographic genetic structure, and revealed genetic clines associated with geographical and climatic gradients. These clines provide the opportunity for selection on phenotypic traits but whether these clines are correlated with adaptive evolution has to be carefully demonstrated. Other processes, such as migration or range expansion, may generate such clines in loci across environmental gradients. After portioning out the effect of spatial location, we still find significant variation due to the genetic association with climate. The next step would be to identify candidate genes associated with these traits and assess whether geographic variation in SNPs

follows environmental gradients in a similar geographic pattern.

Population genomics is facilitating the identification of adaptive molecular variation by examining numerous loci or genome regions, and recent advances in genome-wide scans have allowed examining footprint of selection in numerous species (Luikart *et al.* 2003; Nielsen 2005; Biwas & Akey 2006). One could take two approaches. The first one would be to examine variation within candidate genes coding for known biological functions, potentially correlated with adaptive traits, such as phenology (Heuertz *et al.* 2006; Pyhajarvi *et al.* 2007), drought tolerance (González-Martínez *et al.* 2006; Pyhajarvi *et al.* 2007; Eveno *et al.* 2008; Grivet *et al.* 2009b) or cold resistance (Wachowiak *et al.* 2009). The second would use a random scan of the genome and look for loci or SNP potentially under selection (e.g. Namroud *et al.* 2008; Eckert *et al.* 2010). The next step is to link SNP's under selection with phenotypes measured in common gardens through association studies (e.g. González-Martínez *et al.* 2007). The geographic mapping of those SNP's could then provide a direct indicator of how various populations have responded to past climate change and thus may respond to future climate change. Currently, work is underway with valley oak to identify SNP variation at candidate genes linked to traits such as drought tolerance and phenology, which would allow us to map adaptive variation on the landscape and assess its association with climate variables.

Ultimately, intense genome scans linked with common garden measurements of phenotypes will allow us to assess the extent to which the mapping of candidate genetic variation provides insight about the ability of local populations to tolerate climate change. It is encouraging that studies have found associations between molecular markers and phenotype (Savolainen *et al.* 2007). If we can understand the linkages, future studies may be able use maps of adaptive genes to identify regions of concern with respect to climate change.

Conclusions

Valley oak populations show a great deal of historical genetic connectivity that indicates tremendous potential for gene movement throughout its range. Yet, the rate of rapid climate change that is now being modelled is likely to cause shifts in climate zones that are beyond the rate of migration for most local populations. The geographic analysis of genetic structure shows a strong association with climate variables that indicates that regional populations are likely to be adapted to local climate conditions. Because some regions will experi-

ence particularly dramatic shifts in climate zones, the underlying climatically based genetic structure may constrain the ability of populations to tolerate that rapid climate change. Future studies using candidate genes and common garden experiments are needed to evaluate the extent to which spatially distributed local populations can tolerate or adapt to future climate conditions.

Acknowledgements

We thank the anonymous reviewers who provide valuable suggestions on how to improve the manuscript. We are grateful to: Rodney Dyer and Doug Scofield for discussions of data analysis; Keith Gaddis, Karen Lundy, Stephanie Steele, and Pam Thompson for comments on manuscript; Charles Winder, Edith Martinez, and Karen Lundy for laboratory work; and Uma Dandekar and other staff at the UCLA Sequencing and Genotyping Core Facility. We thank all those people who were drafted into collecting oak leaves, including Brian Alfaro, Maria Valbuena-Carabaña, Kurt Merg, Andrea Pluess, Andrea Sork, Danielle Sork, Silke Werth, and Charles Winder. Over the years of data collection and analysis, VLS received funding from UCLA Division of Life Sciences, UCLA Senate Faculty Research Programme, NSF DEB-0089445, NSF DEB-0516529, and the National Geographic Society. FWD and MI were supported by California Energy Commission research project CEC 500-08-020. HW was supported by the China National Scientific and Technical Foundation Project (Grant No. 2006FY210100).

References

- Aitken SN, Yeaman S, Holliday JA, Wang TL, Curtis-McLane S (2008) Adaptation, migration or extirpation: climate change outcomes for tree populations. *Evolutionary Applications*, **1**, 95–111.
- Austerlitz F, Dick CW, Dutech C *et al.* (2004) Using genetic markers to estimate the pollen dispersal curve. *Molecular Ecology*, **13**, 937–954.
- Barton N (2000) Genetic hitchhiking. *Philosophical Transactions of the Royal Society of London B Biological Sciences*, **355**, 553–1562.
- Bawa KS, Dayanandan S (1998) Global climate change and tropical forest genetic resources. *Climatic Change*, **39**, 473–485.
- Bennett KD (1997) *Evolution and Ecology: The Pace of Life*. Cambridge University Press, Cambridge, UK.
- Biwas S, Akey JM (2006) Genomic insights into positive selection. *Trends in Genetics*, **22**, 437–446.
- Borcard D, Legendre P, Drapeau P (1992) Partialling out the spatial component of ecological variation. *Ecology*, **73**, 1045–1055.
- Box GEP, Draper NR (1987) *Empirical Model-Building and Response Surfaces* John Wiley & Sons, New York, NY.
- Cayan DR, Maurer EP, Dettinger MD, Tyree M, Hayhoe K (2008) Climate change scenarios for the California region. *Climatic Change*, **87**, S21–S42.
- Daly C, Halbleib M, Smith JI *et al.* (2008) Physiographically sensitive mapping of climatological temperature and

- precipitation across the conterminous United States. *International Journal of Climatology*, **28**, 2031–2064.
- Davis MB, Shaw RG (2001) Range shifts and adaptive responses to Quaternary climate change. *Science*, **292**, 673–679.
- Davis FW, Stoms DM, Hollander AD *et al.* (1998) The California Gap Analysis Project—Final Report. University of California, Santa Barbara, CA. Available from http://www.biogeog.ucsb.edu/projects/gap/gap_rep.html.
- Davis MB, Shaw RG, Etterson JR (2005) Evolutionary responses to changing climate. *Ecology*, **86**, 1704–1714.
- Deguilloux M-F, Dumolin-Lapègue S, Gielly L, Grivet D, Petit RJ (2003) A set of primers for the amplification of chloroplast microsatellites in *Quercus*. *Molecular Ecology Notes*, **3**, 24–27.
- Dutech C, Sork VL, Irwin AJ, Smouse PE, Davis FW (2005) Gene flow and fine-scale genetic structure in a wind pollinated tree species, *Quercus lobata* (Fagaceae). *American Journal of Botany*, **92**, 252–261.
- Dyer RJ (2009) GENETICSTUDIO. A suite of programs for the spatial analysis of genetic marker data. *Molecular Ecology Resources*, **9**, 110–113.
- Dyer RJ, Nason JD (2004) Population Graphs: the graph theoretic shape of genetic structure. *Molecular Ecology*, **13**, 1713–1727.
- Dyer RJ, Nason JD, Garrick RC (2010) Landscape modelling of gene flow: Improved power using conditional genetic distance derived from the topology of population networks. *Molecular Ecology*, **19**, 3746–3759.
- Eckert AJ, Bower AD, González-Martínez SC, Wegrzyn JL, Coop G, Neale DB (2010) Back to nature: ecological genomics of loblolly pine (*Pinus taeda*, Pinaceae). *Molecular Ecology*, **19**, 3789–3805.
- Elith J, Graham C, Anderson R *et al.* (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, **29**, 129–151.
- Endler JA (1986) *Natural Selection in the Wild*. Princeton University Press, Princeton, NJ.
- Eveno E, Collada C, Guevara MA *et al.* (2008) Contrasting patterns of selection at *Pinus pinaster* Ait. drought stress candidate genes as revealed by genetic differentiation analyses. *Molecular Biology and Evolution*, **25**, 417–437.
- Flint AL, Flint LE (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. U.S. Geological Survey Scientific Investigations Report 2007-5099.
- Flint AL, Flint LE (2010) Downscaling future climate scenarios to fine scales for hydrologic and ecologic modeling and analysis. In Review, U.S. Geological Survey, California Water Science Center, Sacramento, CA.
- Foll M, Gaggiotti O (2006) Identifying the environmental factors that determine the genetic structure of populations. *Genetics*, **174**, 875–891.
- Freedman AH, Thomassen HA, Buermann W, Smith TB (2010) Genomic signals of diversification along ecological gradients in a tropical lizard. *Molecular Ecology*, **19**, 3773–3788.
- Garrick RC, Nason JD, Meadows CA, Dyer RJ (2009) Not just vicariance: phylogeography of a Sonoran Desert euphorb indicates a major role of range expansion along the Baja peninsula. *Molecular Ecology*, **18**, 1916–1931.
- Gomez J (2003) Spatial patterns in long-distance dispersal of *Quercus ilex* acorns by jays in a heterogeneous landscape. *Ecography*, **26**, 573–584.
- González-Martínez SC, Ersoz E, Brown GR, Wheeler NC, Neale DB (2006) DNA sequence variation and selection of tag single-nucleotide polymorphisms at candidate genes for drought-stress response in *Pinus taeda* L. *Genetics*, **172**, 1915–1926.
- González-Martínez SC, Wheeler NC, Ersoz E, Nelson CD, Neale DB (2007) Association genetics in *Pinus taeda* L. I. Wood property traits. *Genetics*, **175**, 399–409.
- Griffin JR, Critchfield WB (1972) *The Distribution of the Forest Trees in California*. Pacific SW Forest and Range Experiment Station, U.S. Department of Agriculture Forest Service, Berkeley.
- Grivet D, Smouse PE, Sork VL (2005) A novel approach to an old problem: tracking dispersed seeds. *Molecular Ecology*, **14**, 3585–3595.
- Grivet D, Deguilloux M-F, Petit RJ, Sork VL (2006) Contrasting patterns of historical colonization in white oaks (*Quercus* spp.) in California and Europe. *Molecular Ecology*, **15**, 4085–4093.
- Grivet D, Sork VL, Westfall RD, Davis FW (2008) Conserving the evolutionary potential of California valley oak (*Quercus lobata* Nee): a multivariate genetic approach to conservation planning. *Molecular Ecology*, **17**, 139–156.
- Grivet D, Robledo-Arnuncio JJ, Smouse PE, Sork VL (2009a) Relative contribution of contemporary pollen and seed dispersal to the effective parental size of seedling population of California valley oak (*Quercus lobata* Nee). *Molecular Ecology*, **18**, 3967–3979.
- Grivet D, Sebastiani F, Gonzalez-Martinez SC, Vendramin GG (2009b) Patterns of polymorphism resulting from long-range colonization in the Mediterranean conifer Aleppo pine. *New Phytologist*, **184**, 1016–1028.
- Hamrick JL (2004) Response of forest trees to global environmental changes. *Forest Ecology and Management*, **197**, 323–335.
- Heuertz M, De Paoli E, Kallman T *et al.* (2006) Multilocus patterns of nucleotide diversity, linkage disequilibrium and demographic history of Norway spruce [*Picea abies* (L.) Karst]. *Genetics*, **174**, 2095–2105.
- Hidalgo HG, Dettlinger MD, Cayan DR (2008) Downscaling with constructed analogues: Daily precipitation and temperature fields over the United States. PIER Energy-Related Environmental Research Report CEC-500-2007-123. California Energy Commission, Sacramento.
- Holderegger R, Kamm U, Gugerli F (2006) Adaptive vs. neutral genetic diversity: implications for landscape genetics. *Landscape Ecology*, **21**, 797–807.
- Holsinger KE (1993) The evolutionary dynamics of fragmented plant populations. In: *Biotic Interactions and Global Change*, (eds Kingsolver JG, Kareiva PM, Huey RB), pp. 198–216. Sinauer Associates, Sunderland, MA.
- Jackson ST, Overpeck JT (2000) Responses of plant populations and communities to environmental changes of the late Quaternary. *Paleobiology*, **26**, 194–220.
- Joost S, Kalbermatten M, Bonin A (2008) Spatial analysis method (SAM): a software tool combining molecular and

- environmental data to identify candidate loci for selection. *Molecular Ecology Resources*, **8**, 957–960.
- Jump AS, Hunt JM, Martínez-Izquierdo JA, Peñuelas J (2006) Natural selection and climate change: temperature-linked spatial and temporal trends in gene frequency in *Fagus sylvatica*. *Molecular Ecology*, **15**, 3469–3480.
- Kelly PA, Phillips SE, Williams DF (2005) Documenting ecological change in time and space: the San Joaquin Valley of California. In: *Mammalian Diversification: From Chromosomes to Phylogeography (A Celebration of the Career of James L. Patton)* (eds Lacey EA, Myers P), University of California Press, Berkeley.
- Kueppers LM, Snyder MA, Sloan LC, Zavaleta ES, Fulfrost B (2005) Modeled regional climate change and California enaemic oak ranges. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 16281–16286.
- Ledig FT (1992) Human impacts on genetic diversity in forest ecosystems. *Oikos*, **63**, 87–108.
- Legendre P, Legendre L (1998) *Numerical Ecology*, 2nd English edn. Elsevier, Amsterdam.
- Loarie SR, Duffy PB, Hamilton H *et al.* (2009) The velocity of climate change. *Nature*, **462**, 1052–1111.
- Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2003) The power and promise of population genomics: From genotyping to genome typing. *Nature Review Genetics*, **4**, 981–994.
- Manel S, Schwartz MK, Luikart G, Taberlet P (2003) Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology & Evolution*, **18**, 189–197.
- Manel S, Joost S, Epperson BK *et al.* (2010) Perspectives on the use of landscape genetics to detect genetic adaptive variation in the field. *Molecular Ecology*, **19**, 3760–3772.
- McLachlan JS, Clark JS (2004) Reconstructing historical ranges with fossil data at continental scales. *Forest Ecology and Management Holsinger*, **197**, 139–147.
- Mitton JB (1997) *Selection in Natural Populations*. Oxford University Press, New York.
- Namroud MC, Beaulieu J, Juge N, Laroche J, Bousquet J (2008) Scanning the genome for gene single nucleotide polymorphisms involved in adaptive population differentiation in white spruce. *Molecular Ecology*, **17**, 3599–3613.
- Nielsen R (2005) Molecular signatures of natural selection. *Annual Review of Genetics*, **39**, 197–218.
- Parmesan C (2006) Ecological and evolutionary responses to recent climate change. *Annual Review of Ecology Evolution and Systematics*, **37**, 637–669.
- Parry ML, Canziani OF, Palutikof JP, van der Linden PJ, Hanson CE (2007) Working Group II Report. Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, UK.
- Phillips SJ, Dudi'k M (2008) Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography*, **31**, 161–175.
- Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modeling*, **190**, 231–259.
- Pluess AR, Sork VL, Dolan B *et al.* (2009) Short distance pollen movement in a wind-pollinated tree, *Quercus lobata* (Fagaceae). *Forest Ecology and Management*, **258**, 735–744.
- Pritchard JK, Pickrell JK, Coop G (2010) The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. *Current Biology*, **20**, R208–R215.
- Pyhajarvi T, Garcia-Gil MR, Knurr T *et al.* (2007) Demographic history has influenced nucleotide diversity in European *Pinus sylvestris* populations. *Genetics*, **177**, 1713–1724.
- Randin CF, Jaccard H, Vittoz P, Yoccoz NG, Guisan A (2009) Land use improves spatial predictions of mountain plant abundance but not presence-absence. *Journal of Vegetation Science*, **20**, 996–1008.
- Rehfeldt GE, Tchebakova NM, Parfenova YI *et al.* (2002) Intraspecific responses to climate in *Pinus sylvestris*. *Global Change Biology*, **8**, 912–929.
- Rehfeldt GE, Crookston NL, Warwell MV, Evans JS (2006) Empirical analyses of plant-climate relationships for the western United States. *International Journal of Plant Sciences*, **167**, 1123–1150.
- SAS Institute (1989) *SAS/STAT User's guide, Version 6, Fourth Edition, Volume 1*. Cary, NC.
- Savolainen O, Pyhajarvi T, Knurr T (2007) Gene flow and local adaptation in trees. *Annual Review of Ecology Evolution and Systematics*, **38**, 595–619.
- Scofield D, Sork VL, Smouse P (2010) Influence of acorn woodpecker social behaviour on transport of coast live oak (*Quercus agrifolia*) acorns in a southern California oak savanna. *Journal of Ecology*, **98**, 561–572.
- Seo C, Thorne JH, Hannah L, Thuiller W (2009) Scale effects in species distribution models: implications for conservation planning under climate change. *Biology Letters*, **5**, 39–43.
- Smouse PE, Williams RC (1982) Multivariate analysis of Hla disease associations. *Biometrics*, **38**, 757–768.
- Smouse PE, Dyer RJ, Westfall RD, Sork VL (2001) Two-generation analysis of pollen flow across a landscape. I. Male gamete heterogeneity among females. *Evolution*, **55**, 260–271.
- Solomon S, Qin D, Manning M *et al.* (2007) Working Group I Report. The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge.
- Sork VL, Smouse PE (2006) Genetic analysis of landscape connectivity in tree populations. *Landscape Ecology*, **21**, 821–836.
- Sork VL, Dyer R, Davis F, Smouse P (2002a) Mating system in California Valley oak, *Quercus lobata* Neé. In: *Proceedings of the Fifth Symposium on Oak Woodlands: Oaks in California's Changing Landscape*. 2001 October 22–25; San Diego, CA. Gen. Tech. Rep. PSW-GTR-184 (eds Standiford RBMD, McCreary D, Purcell KL), pp. 427–444. Pacific Southwest Research Station, Forest Service, U.S. Department of Agriculture, Albany, CA.
- Sork VL, Davis FW, Smouse PE *et al.* (2002b) Pollen movement in declining populations of California Valley oak, *Quercus lobata*: where have all the fathers gone? *Molecular Ecology*, **11**, 1657–1668.
- Sork VL, Davis FW, Grivet D (2009) Incorporating genetic information into conservation planning for California valley oak. Gen. Tech. Rep. PSW-GTR-217 497–509.
- Storfer A, Murphy MA, Evans JS *et al.* (2007) Putting the 'landscape' in landscape genetics. *Heredity*, **98**, 128–142.
- Thomassen HA, Cheviron ZA, Freedman AH, Harrigan RJ, Wayne RK, Smith TB (2010) Spatial modelling and

- landscape-level approaches for visualizing intra-specific variation. *Molecular Ecology*, **19**, 3532–3548.
- Trivedi MR, Berry PM, Morecroft MD, Dawson TP (2008) Spatial scale affects bioclimate model projections of climate change impacts on mountain plants. *Global Change Biology*, **14**, 1089–1103.
- Tyler CM, Kuhn B, Davis FW (2006) Demography and recruitment limitations of three oak species in California. *Quarterly Review of Biology*, **81**, 127–152.
- Wachowiak W, Balk PA, Savolainen O (2009) Search for nucleotide diversity patterns of local adaptation in dehydrins and other cold-related candidate genes in Scots pine (*Pinus sylvestris* L.). *Tree Genetics & Genomes*, **5**, 117–132.
- Westfall RD, Conkle MT (1992) Allozyme markers in breeding zone designation. *Population Genetics of Forest Trees*, **42**, 279–309.
- Whipple A, Grossinger R, Davis FW (2010) Shifting baselines in a California oak savanna: Nineteenth Century data to inform restoration scenarios. *Restoration Ecology*, doi: 10.1111/j.1526-100X.2009.00633.x.
- Young A, Boyle T, Brown T (1996) The population genetic consequences of habitat fragmentation for plants. *Trends in Ecology & Evolution*, **11**, 413–418.
- Zanetto A, Roussel G, Kremer A (1994) Geographic variation of inter-specific differentiation between *Quercus robur* L, *Quercus petraea* (Matt.) Liebl. I. *Monolocus patterns of variation*. *Forest Genetics*, **1**, 111–123.
-
- V.L.S. studies the ecological and evolutionary processes of tree populations in temperate and tropical habitats with a landscape genetic and environmental genomics approach to study gene flow, adaptation, and conservation in California oaks. F.W.D. is a biogeographer and landscape ecologist broadly interested in the coupling of vegetation pattern and ecological processes, California shrubland and oak woodland ecology, impacts of climate change in Mediterranean-climate regions, and reserve network design. R.D.W. is quantitative forest geneticist who studies geographical genetics, responses to sub-alpine trees species to climate changes, rock-ice features as biotic refugia, and landscape patterns and the thermal environments of American pika. A.L.F. is a U.S. Geological Survey research hydrologist working in downscaling current and future climate scenarios to ecologically relevant scales then translating those climates into hydrologic response across the landscape including recharge, runoff, evapotranspiration, and climatic water deficits to evaluate changes in habitat and potential habitat refugia under global climate change. M.I. is a plant ecologist working in the fields of forestry, biogeography, and conservation biology focusing on evaluating species distributions and performances under different environments through modeling and field works. H.W. applies the methods of landscape genetics and phylogeography to understand the response of single species or communities to the changing world. D.G. is interested in population genetics and genomics, gene flow and conservation genetics of temperate forest trees.
-

Supporting information

Additional supporting information may be found in the online version of this article.

Table S1 Summary of genetic diversity six chloroplast microsatellite markers genotyped 190 individuals and 64 sites distributed across species range and six nuclear microsatellite markers genotyped for 270 individuals sampled in 65 sites throughout species range

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.