MOLECULAR ECOLOGY

Molecular Ecology (2013) **22**, 3598–3612 doi: 10.1111/mec.12317

Influence of late Quaternary climate change on present patterns of genetic variation in valley oak, *Quercus lobata* Née

PAUL F. GUGGER,* MAKIHIKO IKEGAMI†[1] and VICTORIA L. SORK*‡

*Department of Ecology and Evolutionary Biology, University of California, Box 957239, Los Angeles, CA 90095, USA, †Donald Bren School of Environmental Science and Management, University of California, Santa Barbara, CA 93106-5131, USA, ‡Institute of the Environment and Sustainability, University of California, Box 951496, Los Angeles, CA 90095, USA

Abstract

Phylogeography and ecological niche models (ENMs) suggest that late Quaternary glacial cycles have played a prominent role in shaping present population genetic structure and diversity, but have not applied quantitative methods to dissect the relative contribution of past and present climate vs. other forces. We integrate multilocus phylogeography, climate-based ENMs and multivariate statistical approaches to infer the effects of late Quaternary climate change on contemporary genetic variation of valley oak (*Quercus lobata* Née). ENMs indicated that valley oak maintained a stable distribution with local migration from the last interglacial period (~120 ka) to the Last Glacial Maximum (~21 ka, LGM) to the present compared with large-scale range shifts for an eastern North American white oak (*Quercus alba* L.). Coast Range and Sierra Nevada foothill populations diverged in the late Pleistocene before the LGM [104 ka (28–1622)] and have occupied somewhat distinct climate niches, according to ENMs and coalescent analyses of divergence time. In accordance with neutral expectations for stable populations, nuclear microsatellite diversity positively correlated with niche stability from the LGM to present. Most strikingly, nuclear and chloroplast microsatellite variation significantly correlated with LGM climate, even after controlling for associations with geographic location and present climate using partial redundancy analyses. Variance partitioning showed that LGM climate uniquely explains a similar proportion of genetic variance as present climate (16% vs. 11–18%), and together, past and present climate explains more than geography (19%). Climate can influence local expansion–contraction dynamics, flowering phenology and thus gene flow, and/or impose selective pressures. These results highlight the lingering effect of past climate on genetic variation in species with stable distributions.

Keywords: ecological niche modelling, genetic diversity, phylogeography, Pleistocene, *Quercus*, redundancy analysis

Received 10 May 2012; revision received 28 February 2013; accepted 6 March 2013

Introduction

Historical processes leave imprints on the genetic structure of present populations, especially of long-lived, sessile organisms. Thus, the present genetic structure of a variety of taxa has been used to infer historical vicariance associated with geological change (Avise 2000), dispersal history (Petit *et al.* 1997) and episodes of expansion and contraction associated with global climate change (Hewitt 2000). Recent research has integrated ecological niche models (ENM) based on the climate niche to independently infer historical demographic processes (Carstens & Richards 2007; Waltari *et al.* 2007; Knowles & Alvarado-Serrano 2010). Major insights include independent support for postgla-

Correspondence: Paul F. Gugger,
E-mail: pfg@ucla.edu
[1]Present address: The Bio-Protection Research Centre, Lincoln University, PO Box 84, Lincoln 7647, Canterbury, New Zealand

cial expansion from glacial refugia (Waltari *et al.* 2007; Svenning *et al.* 2008), the role of past and present climate in population subdivision (Carstens & Richards 2007; Pease *et al.* 2009) and that genetic diversity is related to niche stability through major periods of climate change (Carnaval *et al.* 2009; Ortego *et al.* 2012). Together, phylogeographic and ENM studies reveal a prominent role for the combined effects of drift, gene flow, isolation and other neutral, historical demographic processes in shaping present genetic structure. However, associated with all of these processes are temporal changes in climate. Thus, an important question to address is the extent to which past and present climate themselves have shaped allele frequencies relative to other forces.

Climate can influence genetic variation in two complementary ways. First, climate can shape genetic variation through purely neutral processes, such as affecting gene flow by affecting flowering phenology among populations (Knight *et al.* 2005; Sork *et al.* 2010; Ortego *et al.* 2012) or by controlling population expansion–contraction dynamics (James *et al.* 2011). Climate can also shape genetic variation through natural selection. In the absence of strong genome-wide linkage, this effect is expected to be primarily locus-specific in functionally important genomic regions. However, divergence by local adaptation and subsequent selection against nonlocal migrants can cause neutral loci to be in linkage disequilibrium with loci under selection that underlie adaptive differentiation, leading to apparent genome-wide signatures of adaptive divergence (Via *et al.* 2000; Nosil *et al.* 2008; Andrew *et al.* 2012). Under changing climate, both demographic and selective responses are expected (Davis & Shaw 2001) and could reinforce each other.

The reassortment of genetic variation is expected to lag behind climate change. For example, patterns of genetic variation might reach equilibrium under relatively stable climate during long glacial periods (~100 000 years) if relatively stable populations are maintained. As climate abruptly shifts to interglacial periods, genotype-climate disequilibrium would develop and might persist for many generations, depending on the rate of environmental change, demographic factors and levels of genetic variation (Pease *et al.* 1989; Bürger & Lynch 1995; Davis *et al.* 2005). In species with long generation times, this disequilibrium could last well into the subsequent, short interglacial (~10 000 years) such as the present. Thus, one question is how much genetic variation is associated with current vs. past climates or environments.

Multivariate statistical methods, especially constrained ordinations, are a particularly sensitive method for quantifying complex associations of genetic variation with climate. They distil the small associations of individual climatic and genetic variables to the principal relationships and can be used to partition variation among multiple classes of explanatory variables, for example, by partialing out other purely neutral forces associated with geographic structure (e.g. phylogeographic structure and spatial autocorrelation). In concert with other approaches to studying the influence of past climate on genetic variation, such as the combination of ecological niche modelling with phylogeographic analysis, we can begin to more precisely understand how past and present climates and climate change shape genetic variation.

To investigate the extent to which current and past climate and late Quaternary climate change affect contemporary genetic variation in tree populations, we examined present genetic variation in relation to the past and present climate niche of a declining, endemic California oak species, valley oak (*Quercus lobata* Née). Valley oak is found scattered throughout lowland parts of the California Floristic Province, which is the most biodiverse region of North America (Myers *et al.* 2000). One hypothesis for the exceptional biodiversity of California suggests that topographic and local climate heterogeneity (topoclimate, Thornthwaite 1953; Dobrowski 2011) combined with regional climate stability generate opportunity for genetic diversity to accumulate within populations and for genetic divergence to arise among regions through drift and local adaptation (Stebbins & Major 1965; Calsbeek *et al.* 2003). Previous studies in valley oak find patterns consistent with this idea. High diversity and strong population differentiation at maternally inherited chloroplast microsatellites (cpSSRs) suggest historically persistent local seed dispersal (Grivet *et al.* 2006), rather than dramatic range expansions found for oaks in other temperate regions (Petit *et al.* 2002a; Svenning *et al.* 2008). Nuclear microsatellites (nSSRs) reveal genetic differentiation among populations in the inland foothills of the Coast Range and the western foothills of the Sierra Nevada, with higher gene flow north–south along climatically similar mountain corridors than east–west across the Central Valley (Grivet *et al.* 2008; Sork *et al.* 2010), suggesting a role for environmentally mediated historical vicariance. Those studies identified centres of genetic diversity in the San Francisco Bay area and Transverse Ranges of southern California, which coincide with some glacial refugia in other taxa (Gugger & Sugita 2010). Given the inferred stability of local populations and expectations for stable populations without major bottlenecks under genetic drift, we hypothesize that niche stability leads to higher levels of genetic diversity (Carnaval *et al.* 2009; Ortego *et al.* 2012). Moreover, the predicted stability of valley oak populations offers a unique opportunity test for and quantify the association of present genetic variation with past climate, beyond present climate (Sork *et al.* 2010), consistent with

a lingering influence of past climate adaptation and/or past climate-mediated demographic responses.

Using phylogeographic inference, climate-based ENMs and multivariate statistical approaches, we test the following specific hypotheses for how present patterns of genetic structure and diversity arose: (i) the distribution of valley oak was stable through late Quaternary climate cycles and migration was local; (ii) population genetic diversity is higher in sites with more stable climate niches; (iii) past climate-induced isolation explains regional genetic structure and (iv) past climate shaped present genetic variation.

Materials and methods

DNA preparation

We compiled previously published range-wide data for six cpSSR markers from 190 individuals from 64 sites and six nSSR loci from 267 individuals from 65 sites that were generated using standardized methods across studies (Grivet *et al.* 2006, 2008; Sork *et al.* 2010) (Dryad doi:10.5061/dryad.g645d/3). We added nuclear DNA sequences (nDNA) for two commonly sequenced low-copy genes: *elongation factor 1-α* (*ef1α*; 21 individuals from 11 sites, 736 bp) and *glyceraldehyde 3-phosphate dehydrogenase* (*g3pdh*; 29 individuals from 29 sites, 606 bp). DNA extractions and PCR conditions followed Werth & Sork (2008) using the following primers with 58 °C annealing temperature: efa-pF.deg = 5′-ACCCCCAARTACTCCAAGGC-3′ and efa-pR.deg = 5′-GCAGCCTTGGTRACCTTG-3′ for *ef1α*, and gpd-cgF = 5′-TGGAATTGTTGAGGGTCTCAT-3′ and gpd-cgR = 5′-TGCTGTCACCAATGAAGTCG-3′ for *g3pdh*. Only the forward primer was used in sequencing reactions.

Sequences were aligned and edited using SEQUENCHER 4.8 (Gene Codes, Ann Arbor, MI). Haplotypes were resolved using the PHASE 2.1 algorithm (Stephens *et al.* 2001) implemented in DNASP 5.0 (Librado & Rozas 2009) allowing for recombination (Li & Stephens 2003). Haplotype networks were constructed in TCS 1.21 (Clement *et al.* 2000). We estimated standard diversity statistics: number of haplotypes, n_h; number segregating sites, S; gene diversity, H; nucleotide diversity, π. We tested for selective neutrality (and constant population size) using D (Tajima 1989) and for recombination with a four-gamete test (Hudson & Kaplan 1985) in DNASP to ensure these new molecular data met the assumptions of the IMA2 model (below).

Population structure

To describe regional and local population genetic structure for subsequent analyses, we used Bayesian clustering methods. For nSSRs, we estimated population structure with admixture and correlated allele frequencies using a Bayesian Markov Chain Monte Carlo (MCMC) clustering analysis in STRUCTURE 2.3.4 (Pritchard *et al.* 2000; Falush *et al.* 2003). We ran 10 replicates of 200 000 steps after 100 000 burn-in steps for 1–10 population clusters (K). We used the ΔK method (Evanno *et al.* 2005) to select the best-supported number of clusters, as implemented in STRUCTURE HARVESTER (Earl & vonHoldt 2012). For cpSSRs, we estimated population structure using the 'clustering with linked loci' analysis in BAPS 5.2 (Corander *et al.* 2008). We ran the model three times for $K = 1$–40 to ensure convergence and chose the K with the highest log likelihood. We also generated a median-joining haplotype network coloured by geographic region using NETWORK 4.6 (Bandelt *et al.* 1999). Results were visualized using ARCGIS 9.3 (ESRI, Redlands, CA, USA).

To quantify variation in nSSRs and cpSSRs among populations and genetic clusters, we performed analyses of molecular variance (AMOVA) (Excoffier *et al.* 1992) using Φ-statistics in GENALEX 6.1 (Peakall & Smouse 2006).

Divergence time, gene flow and population size changes

To set regional patterns of population structure in an historical context, divergence time among coastal and Sierran population clusters (see Results), effective sizes of ancestral and extant populations and bidirectional gene flow rates (as population migration rate, $2N_em$) were estimated using an isolation-with-migration model in IMA2 (Hey 2010). We compared three replicate runs (ESS > 50) of 40 geometrically heated chains ($ha = 0.96$, $hb = 0.9$) with a 10 000 000-step burn-in followed by >20 000 000 steps sampled each 100 steps to ensure convergence. We chose the maximum values for priors drawn from a truncated uniform distribution on the basis of short preliminary runs (Won & Hey 2005): $t = 5$, $q_1 = q_2 = q_A = 80$, and $m_1 = m_2 = 9$. We assumed an infinite sites model (Kimura 1969) of evolution for DNA sequences and a stepwise mutation model for nSSRs (Ohta & Kimura 1973). We excluded cpSSRs from the analysis because long runs (>3 months) with 200 heated chains did not reach stationarity when cpSSRs were included.

To convert parameter estimates to demographic units, we considered a range of mutation rates and generation times. For nDNA, we applied rates calculated for other nuclear loci in Fagaceae of 1.6×10^{-10} to 2.3×10^{-9} per bp per year (Cavender-Bares *et al.* 2011). For nSSRs, we considered rates ranging from 1.76×10^{-4} per gen in tomatoes (*Solanum lycopersicum*; Azaiez *et al.* 2006) to

8.8×10^{-4} per gen in *Arabidopsis thaliana* (Marriage *et al.* 2009). To convert these estimates to the per-locus, per-year rates required in IMa2, per-generation nSSR mutation rates were multiplied by a mean generation time in oaks of 100 years (Cavender-Bares *et al.* 2011; Gugger & Cavender-Bares 2013), and per-base pair nDNA rates were multiplied by the length of the locus in base pairs.

Ecological niche modelling

To test the hypothesis that the distribution of valley oak was stable through late Quaternary climate cycles, we compared ENMs for valley oak to ENMs of an eastern North American white oak (*Quercus alba* L.) expected to have experienced dramatic range shifts in response to glacial–interglacial cycles (a 'positive control') (Davis 1981). Specifically, we used MAXENT 3.3.3e (Phillips *et al.* 2004, 2006) to estimate their distributions during the present, Last Glacial Maximum (LGM, ~21 ka) and last interglacial (LIG, ~120–140 ka). MAXENT is a machine learning-based ENM that uses presence points and performs well with few occurrence points (Elith *et al.* 2006). The 'climatic distribution' of a species represents the modelled statistical association between its known occurrences and current climatic variables. This model was transformed from climate space into geographic space to generate maps showing probability of occurrence. Past distributions were then estimated by projecting this relationship onto scenarios of past climate, assuming that current relationships between climate and distribution are retained. This assumption may be a reasonable approximation for most purposes, but we also built models separately for coastal and Sierran subpopulations to control for observed divergence within valley oak (more below). Final models were constructed using 10 replicate runs. We evaluated model performance by running 10 replicate models divided into 70%/30% train/test partitions of occurrence points and then estimating the area under the receiver operating curve (AUC), where 1 is the maximum prediction and 0.5 suggests a random prediction.

We obtained 247 *Quercus lobata* presence points from Sork Lab sample sites, others sample sites (Sage *et al.* 2011) and personal observations representative of the entire presettlement distribution (Griffin & Critchfield 1972; Davis *et al.* 1998; Whipple *et al.* 2011) (Dryad doi:10.5061/dryad.g645d/3). We obtained 1947 occurrence points of *Quercus alba* from the Global Biodiversity Information Facility (GBIF) portal (http://data.gbif.org) and retained 1920 points after removing those outside the known current distribution (http://gec.cr.usgs.gov/data/atlas/little/) (Little 1971). The distribution of valley oak is thought to be primarily limited by growing-season precipitation or water availability, seasonality and temperature extremes (Sork *et al.* 2010; McLaughlin & Zavaleta 2012). Thus, the following eight climate variables were extracted from the WorldClim database (http://www.worldclim.org, Hijmans *et al.* 2005) at 2.5-arcmin resolution for present (1950–2000), LGM [the average climate of two commonly used LGM models: MIROC 3.2 (Hasumi & Emori 2004) and CCSM3 (Collins *et al.* 2006; Braconnot *et al.* 2007)] and LIG (Otto-Bliesner *et al.* 2006) climates: mean annual precipitation, precipitation in the warmest quarter, precipitation in coldest quarter, precipitation seasonality, mean annual temperature, maximum temperature of warmest month and minimum temperature of coldest month, temperature seasonality.

To further test for patterns consistent with local migration in valley oak, we estimated migration distance and direction of valley oak and white oak populations. From predicted climate suitability of ENMs, presence/absence maps were created from MAXENT's logistic probability of occurrence output using 'maximum sensitivity plus specificity' threshold (Jiménez-Valverde & Lobo 2007). We calculated the geographic centroids of the distributions for every 60×60 km^2 (i.e. 15×15 grid cells in 2.5 arcmin) in each time step. Then, to evaluate the potential source of populations from one time period to the next, we calculated the distance and direction between the centroid of the first time step to the nearest centroid in the second time step (i.e. from LIG to LGM and from LGM to present). The resulting maps display vectors expressing inferred migration distance and direction for each 60×60 km^2 area.

Finally, we built additional ENMs for each nSSR genetic cluster (coastal and Sierran; see Population structure in Results) to determine whether the climate niches were more distinct during the LIG or the LGM than the present. We calculated niche overlap among clusters for each time period by first determining presence/absence with a logistic threshold of equal sensitivity and specificity from test data in MAXENT and overlaying the niche distributions based on each nSSR cluster to calculate the proportion of niche overlap at each time step. A clear separation of projected past distributions among each cluster throughout the past would suggest that past climate-induced isolation explains genetic structure.

Genetic diversity and niche stability

To test the hypothesis that niche stability explains genetic diversity, we regressed LIG, LGM and present niche suitabilities (i.e. probabilities of occurrence) (N_{LIG}, N_{LGM}, N_{Pres}) and climatic niche stability from ENMs on nSSR genetic diversity in general linear models using R 2.14.1 (http://www.r-project.org/). Niche suitability

was included as a covariate because it could be considered a proxy for population size if more suitable habitat maintains larger populations. We defined climatic niche stability as two variables: $S_{\text{LIG-LGM}} = 1 - |N_{\text{LIG}} - N_{\text{LGM}}|$ and $S_{\text{LGM-Pres}} = 1 - |N_{\text{LGM}} - N_{\text{Pres}}|$. Because the independent variables are related to one another, we present a correlation matrix, which shows low to moderate correlations ($0.07 < |r| < 0.61$, $0 < P < 0.54$) among predictors (Table S1, Supporting information), and we measured variance inflation factors (VIF), a measure of how much collinearity increases variance in a model. We observed VIF < 2; VIF < 10 is evidence that multicollinearity does not confound the interpretation of individual predictors in our models. For measures of genetic diversity, we used expected heterozygosity (H_E) and allelic richness (A_r) and after standardizing with rarefaction in HP-RARE 1.0 (Kalinowski 2005). We tested three nested models and chose the one with the lowest Akaike information criterion (AIC) score. The first model contained all five explanatory variables relating to the LIG, LGM and present; the second contained only variables relating to the LGM and present (N_{LGM}, N_{Pres}, $S_{\text{LGM-Pres}}$); and the third contained only N_{Pres}. By comparing these three models, we evaluated how far back in time we could detect the influence of niche suitability and stability on present genetic diversity.

Genetic variation and past climate

To test the hypothesis that past climate shape patterns of present genetic variation, beyond the role of present climate and geography (Sork *et al.* 2010), we performed a series of partial redundancy analyses using the VEGAN 2.0-2 package (Oksanen *et al.* 2009) in R. Redundancy analysis is a multivariate analogue of linear regression similar to canonical correspondence analysis (ter Braak 1986; Legendre & Legendre 1998). We first tested our hypothesis by testing the association of LGM climate (CCSM, Braconnot *et al.* 2007) with cpSSR and nSSR variation, in each case partialing out the variation due to present climate and geographic position (i.e. latitude, longitude, squared terms, cubed terms and cross products). Significance was assessed with a permutation test. Following established procedures (Borcard *et al.* 1992; Liu 1997; Økland 1999), we then partitioned the explainable genetic variance (i.e. inertia, or in this case mean squared contingency coefficients) into percentages of variance explained by LGM climate, present climate, geography and their joint effects. For these analyses, genetic data were transformed to allele frequencies for each individual, removing one allele from each locus (Smouse & Williams 1982; Westfall & Conkle 1992), and climate data for the same eight variables used in ENMs were extracted from the WorldClim database for each sample.

Results

Nuclear DNA sequence diversity

All *g3pdh* and *ef1α* alleles were phased with probability $P > 0.9$, although a few specific bases at each locus were assigned with probability $0.5 \leq P < 0.9$ (unphased data in GenBank: JQ421340–JQ421389; aligned, phased data in Dryad: doi:10.5061/dryad.g645d/1 and doi:10.5061/dryad.g645d/2). Both nDNA loci showed moderate amounts of variation (*g3pdh*: $n_h = 10$, $S = 16$, $H = 0.75$, $\pi = 0.0067$; *ef1α*: $n_h = 7$; $S = 7$; $H = 0.66$; $\pi = 0.0013$) (Fig. S1, Supporting information) (Derory *et al.* 2010; Gugger *et al.* 2010). Tajima's D was not significantly different from zero for *ef1α* ($D = -1.09$, $P > 0.1$) nor for *g3pdh* ($D = 0.50$, $P > 0.1$), and neither showed evidence of intralocus recombination.

Population structure

ΔK based on STRUCTURE analysis of nSSRs supported two geographically meaningful population clusters ($\ln P(K) = -6066.6$, $\Delta K = 269.9$) (Table S2, Supporting information), a western one mostly contained to the inland foothills of the Coast Range and an eastern one mostly on the western Sierra Nevada foothills (hereforth referred to as 'coastal' and 'Sierran' clusters) (Fig. 1a). The northernmost part of the distribution shows evidence of interdigitation among population clusters, but not much admixture; most sample sites retain their signature of coastal or Sierran origin despite the potential for long-distance pollen dispersal in highly outcrossing wind-dispersed trees. East of the San Francisco Bay, the coastal population extends eastward into the Sierra Nevada foothills. Seventeen percent of nSSR variation is explained among the coastal and Sierran genetic clusters ($\Phi_{RT} = 0.17$, $P = 0.01$) and 10% among sample sites ($\Phi_{PT} = 0.27$, $P = 0.01$) (Table S3, Supporting information).

As supported by linked-locus analysis in BAPS ($K = 25$, log likelihood = −391.2), cpSSRs displayed strong local structure (Grivet *et al.* 2006) with some intermixing in the northern part of the distribution parallel to that of nSSRs (Fig. 1b). Although the high K raises suspicions of model overfitting, this result is consistent with the fact that most haplotypes are geographically restricted to a single sample sites or a few neighbouring ones. More important than its specific value, K is probably large, and structure is local. Among the 25 BAPS clusters $\Phi_{PT} = 0.92$, and among the 64 sample sites $\Phi_{PT} = 0.88$ ($P = 0.01$) (Table S3, Supporting information). Hierarchical AMOVA was not performed on cpSSR data because the 64 sample sites were not strictly nested within the 25 BAPS clusters.

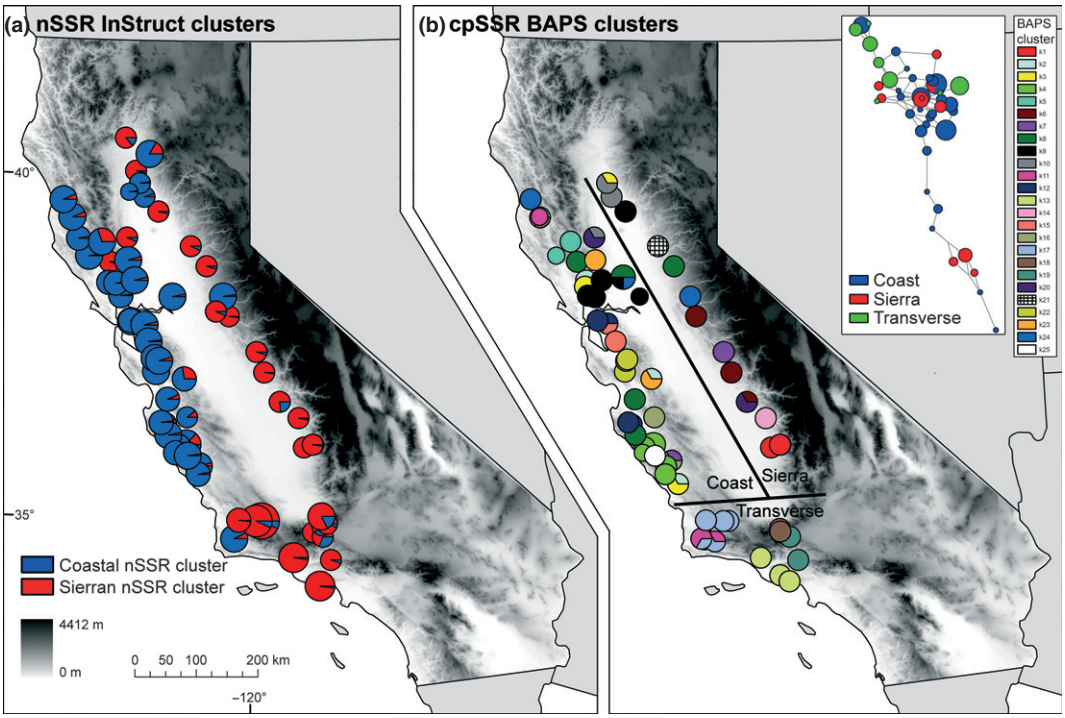

Fig. 1 Maps of genetic variation in valley oak showing (a) genetic clusters inferred from STRUCTURE based on nuclear microsatellites (nSSRs) and (b) genetic clusters inferred from BAPS based on chloroplast microsatellites (cpSSRs). In (b), the haplotype network is coloured according to the three regions delineated on the map (Coast, Sierra, Transverse). The size of points representing sample sites on maps and haplotypes in networks are proportional to number of individuals.

Divergence time, gene flow and population size changes

Assuming the mean of published mutation rates for our loci, divergence time (t) among coastal and Sierran population clusters was 104 ka (28–1622) and gene flow was more pronounced from the west to the east [$2N_em_{CS} = 6.96$ (3.3–10.5)] than east to west [$2N_em_{SC} = 0.02$ (0–2.9)] (Table 1). The present effective sizes of the coastal and Sierran population clusters were estimated as 18% and 19% of the ancestral effective population size, respectively. Considering only high-end or only low-end mutation rates led to considerable variation in divergence time estimates (Table 1; Fig. S2, Supporting information).

Ecological niche models

The valley oak ENM performed well (AUC = 0.975), whereas the white oak model performed moderately (AUC = 0.876), which can occur when large numbers of presence points are used (Franklin *et al.* 2009). The ENMs for all valley oak sites show a stable California distribution relative to white oak, which drastically contracted during the LGM and expanded in the Holocene (Fig. 2). Valley oak ENMs show modest local to regional changes in distribution. For example, during current and LIG climate conditions coastal and Sierran populations appear more isolated from each other, whereas they appear more interconnected during the LGM. Nonetheless, some connectivity is apparent in northern California through all time periods (Figs 2 and 3). The Transverse Range populations of southern California appear more connected to coastal rather than Sierran populations.

Migration vector analysis suggested the local movement of valley oak between periods of major climate transition, especially when compared with the dramatic northward expansion of white oak from the LGM to present (Fig. 3). Along the eastern range margin of valley oak, there is some evidence for elevational shift; from the LIG to LGM, vectors point from the Sierras towards the Central Valley and from the LGM to present, the process reverses.

In the southernmost and northernmost parts of valley oak's distribution, the migration vector analysis implies moderate latitudinal shifts, although those shifts are not nearly as large as those observed in white oak (Fig. 2). This is likely an artefact of the general tendency of the ENMs to overestimate the distribution of valley oak (e.g. compare Fig. 2a–d), suggesting nonclimatic environmental (McLaughlin & Zavaleta 2012), biotic or anthropogenic variables factors not included in our models are important in defining the realized niche of

Table 1 IMa2 results for effective population sizes of the coastal (N_{eC}) and Sierran (N_{eS}) population clusters and their ancestor (N_{eA}), population migration rates among east and west clusters ($2N_em$), and divergence time (t) in thousands of years (ka) with 95% highest posterior density intervals in parentheses considering mean, low, and high estimates of published mutation rates (μ) for run with the highest effective sample size (ESS > 50)

μ	N_{eC}	N_{eS}	N_{eA}	$2N_em_{CS}$	$2N_em_{SC}$	t
Mean	5158 (2996–9480)	4961 (2210–8989)	27 164 (14 294–51 822)	6.96 (3.3–10.5)	0.024 (0–2.9)	104 (28–1622)
Low	12 124 (7044–22 286)	11 663 (5196–21 131)	63 855 (33 602–121 821)	—	—	245 (66–3814)
High	3324 (1931–6109)	3197 (1424–5793)	17 505 (9212–33 396)	—	—	67 (18–1046)

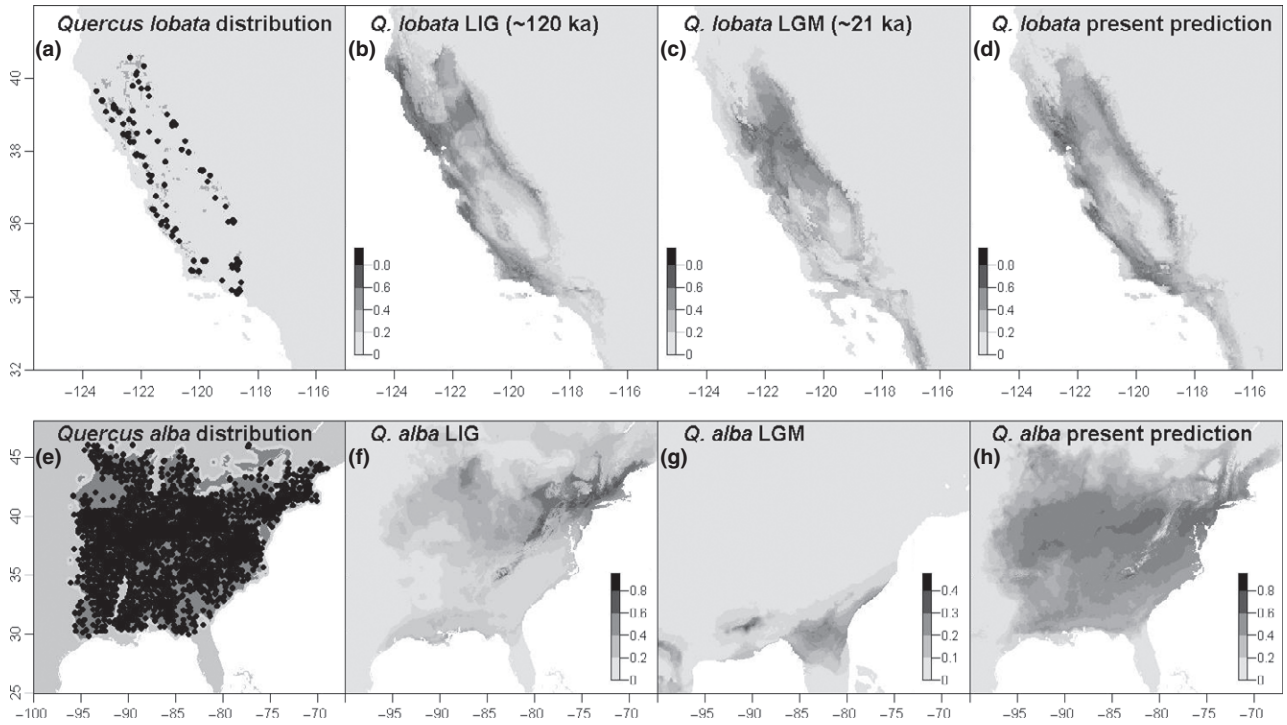

Fig. 2 Ecological niche models for (a–d) valley oak (*Quercus lobata*) and (e–h) white oak (*Quercus alba*) from eastern North America. In each, the first panel shows the points used to model distributions set against the present distribution, followed by model-predicted distributions for the last interglacial period (~120–140 ka), Last Glacial Maximum (~21 ka), and the present. Greyscale refers to probability of occurrence (habitat suitability) from MAXENT output.

valley oak. Nonetheless, the climate variables we included have been shown to be important to valley oak (Sork *et al.* 2010; McLaughlin & Zavaleta 2012), especially precipitation, which is consistent with the fact that precipitation of the warmest quarter (i.e. during the growing season) had the highest percent contribution and permutation importance among climate variables used to construct the ENMs (Table S4, Supporting information).

Models based on each nSSR genetic cluster show periodic, partial climate niche separation (Fig. 4). For example, the distribution of the coastal genetic cluster was not predicted in the Central Valley nor Sierras during the LIG (Fig. 4b), whereas the Sierran genetic cluster was not predicted on the coast during the LGM (Fig. 4g). Climate niches were most distinct during the LIG (niche overlap = 19.1%) and similarly distinct during the LGM (46.7%) and present (51.1%). The principal area of overlap during the LGM was in the northern Central Valley, similar to today.

Genetic diversity and niche stability

Nuclear SSR heterozygosity ranged from 0.43 to 0.74 and allelic richness averaged across loci after rarefaction to 6 ranged from 2.3 to 4.1 (Table S5, Supporting information). The model that best explained H_E and A_r contained N_{LGM}, N_{Pres} and $S_{LGM\text{-}Pres}$, but not terms relating to the LIG (Table 2). Both genetic diversity measures were significantly higher as recent niche stability ($S_{LGM\text{–}Pres}$) increased ($P < 0.037$). This result was only true when including N_{LGM} and N_{Pres} as covariates (Table 2);

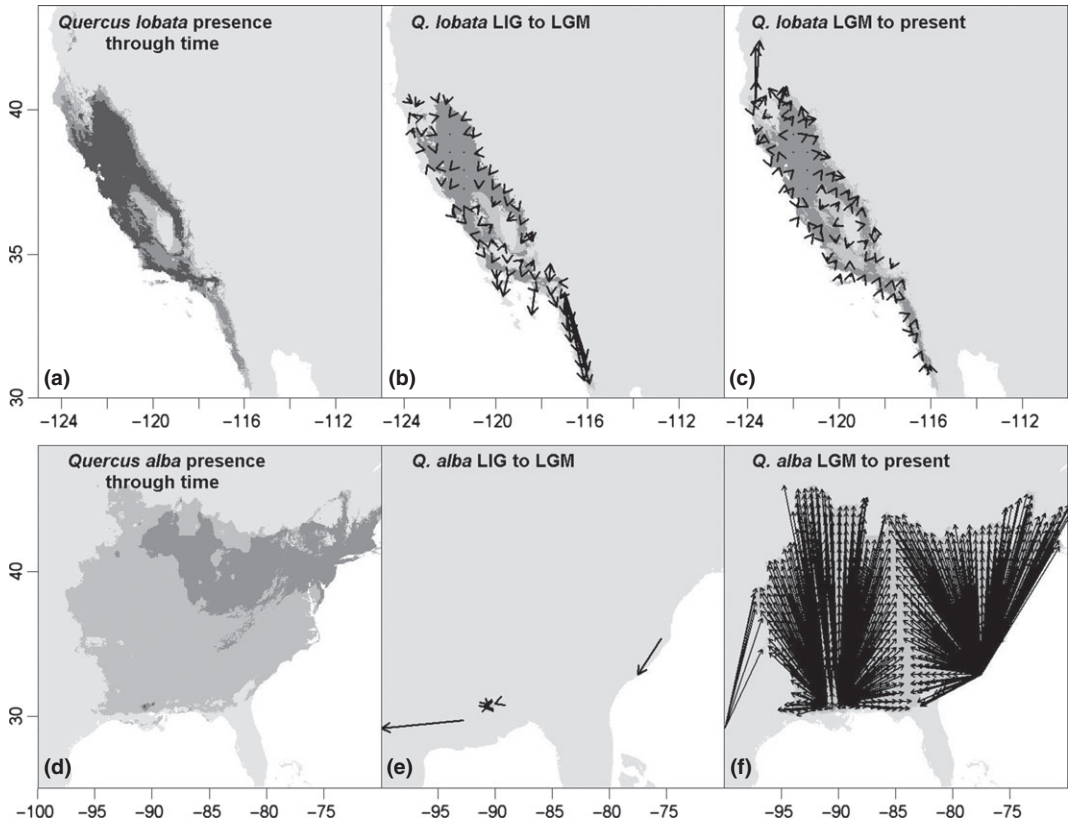

Fig. 3 Changes in *Quercus lobata* and *Quercus alba* distribution through time as summarized by (a, d) areas of overlap among valley oak distribution models from different time periods (darkest shade overlaps in all three time periods, medium in two, and lightest in one) and migration vector analysis of local changes in distribution between the (b, e) last interglacial (LIG) period and Last Glacial Maximum (LIG to LGM) and (c, f) LGM to present (see Methods for details).

in contrast, simple linear regression of $S_{LGM–Pres}$ on genetic diversity indicates an insignificant positive correlation ($0.12 < r < 0.18$, $P > 0.15$). This apparent discrepancy can occur when other predictors included in the model are important and thus reduce the residual error variance. Another factor may be that we have low power to detect the pattern in isolation, given the noisy estimates of genetic diversity expected when samples sizes are 3–4 individuals (6–8 alleles) per site. The significant relationship in the general linear model likely reflects a real relationship because we did not detect problematic multicollinearity in tests of variance inflation factors (VIF < 2). The covariates, N_{LGM} and N_{Pres}, had negative and positive relationships, respectively, with both measures of genetic diversity. Among those, only the relationship of N_{Pres} with A_r was significant ($P = 0.04$) and all others were marginally significant ($0.05 < P < 0.08$).

Genetic variation and past climate

After controlling for associations with present climate and geographic position, partial redundancy analysis suggested that present cpSSR (pseudo-$F = 4.82$, $P < 0.005$) and nSSR variation (pseudo-$F = 1.63$, $P < 0.005$) retain some signature of LGM climate. Minimum temperature during the LGM had the highest loading on the first redundancy analysis axis for both cpSSRs and nSSRs (Fig. S3, Supporting information). For nSSRs, LGM climate explained 16.5% of the present genetic variance, present climate explained 17.8%, geography explained 18.9% and their collinearities explained the rest (Fig. 5). For cpSSRs, LGM climate explained 15.7% of the present genetic variance, present climate explained 11.5%, geography explained 19.4% and their collinearities explained the rest (Fig. 5). Results were similar when using the MIROC instead of CCSM LGM climate data and when summarizing genetic data as allele frequencies per site rather than per individual (although the permutation test for nSSRs analysed by site was not significant, $P = 0.14$) (not shown).

Discussion

We found evidence that late Quaternary climate has shaped present genetic variation in valley oak in several ways consistent with the effects of drift under relatively

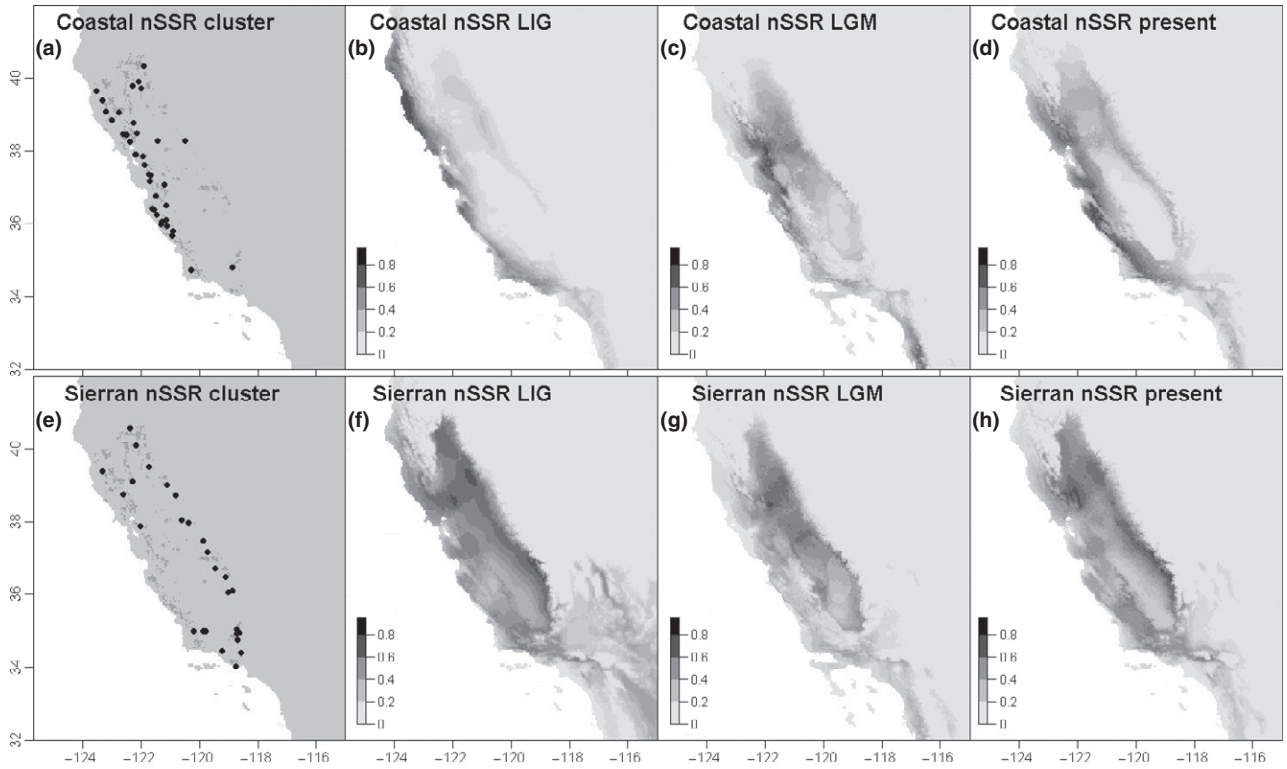

Fig. 4 Ecological niche models for the (a–d) coastal and (e–h) Sierran nuclear microsatellite (nSSR) genetic clusters of valley oak. In each, the first panel shows the points used to model distributions set against the present distribution, followed by model-predicted distributions for the last interglacial period (~120–140 ka), Last Glacial Maximum (~21 ka), and the present. Greyscale refers to probability of occurrence (habitat suitability) from MAXENT output.

stable climate, the influence of climate on demographic processes, and perhaps the influence of selection by climate. Specifically, ENMs and genetic data support a long history of relative stability in the distribution of valley oak with local, elevational expansion and contraction in response to glacial cycles (Figs 1, 2 and 3). At the site scale, niche stability from the LGM to present could contribute to higher levels of nuclear microsatellite diversity (Table 2). Within the broadly constant distribution, we also found evidence of a late Pleistocene (Table 1) subdivision of coastal and Sierran population clusters (Fig. 4) with subsequent intermixing in regions of more persistent niche overlap (Fig. 3a). Most strikingly, LGM climate explains a substantial portion of nuclear and chloroplast genetic variation (Fig. 5 and Fig. S2, Supporting information), suggesting the residual influence of LGM climate-mediated demographic responses or possibly of LGM climate adaptation on present genetic variation.

Stable distribution, local migration

The most prominent feature of the ENMs is that the conditions favourable to valley oak remain relatively constant through major global climate cycles from the LIG to LGM to present (Fig. 2a–d). This pattern is especially clear when compared with models for the eastern North American white oak (Fig. 2e–h) and European white oaks (Petit *et al.* 2002a; Svenning *et al.* 2008) that experienced major latitudinal contractions and expansions. General relative climate stability in California (Barry 1983; Bartlein *et al.* 1998) and local topographic heterogeneity throughout the state probably promoted the regional stability of valley oak's climate niche apparent in ENMs (Tzedakis *et al.* 2002). Migration vector analyses (Fig. 3) suggest some local migration along elevational gradients at the eastern range margin. For example, from the LIG to LGM, local migration proceeds from the Sierras down slope towards the Central Valley and from the LGM to present the process reverses. These shifts may have been associated with moderate regional population size changes according to coalescent analyses of molecular data (Table 1). Elevational migration and long-term stability have been proposed for California oaks (Byrne *et al.* 1991) and other taxa (Gugger & Sugita 2010) based on fossil records. These processes are thought to promote high regional genetic diversity and local population structure through drift and local adaptation (Hewitt 1996; Hampe & Petit 2005; Gugger *et al.* 2011) and help explain the higher

Table 2 General linear models including variables derived from ecological niche modelling that could explain population genetic diversity (A_r and H_E). Akaike information criterion (AIC) scores suggest the best model is the LGM/Pres model

	A_r				H_E			
	Estimate	SE*	t	P†	Estimate	SE*	t	P†
	Full model	AIC = 86.21			*Full model*	AIC = −145.19		
(Intercept)	1.89	0.57	3.30	**0.002**	0.41	0.10	4.34	**0.000**
N_{LIG}	−0.27	0.49	−0.54	0.589	0.02	0.08	0.25	0.806
N_{LGM}	−0.66	0.43	−1.53	0.133	−0.12	0.07	−1.69	*0.096*
N_{Pres}	1.16	0.58	1.99	*0.051*	0.14	0.10	1.42	0.160
$S_{LIG–LGM}$	−0.03	0.58	−0.06	0.956	−0.05	0.10	−0.48	0.634
$S_{LGM–Pres}$	1.58	0.71	2.22	**0.030**	0.25	0.12	2.09	**0.041**
	LGM/Pres model	AIC = 82.69			*LGM/Pres model*	AIC = −148.39		
(Intercept)	1.87	0.53	3.52	**0.001**	0.40	0.09	4.50	**0.000**
N_{LGM}	−0.75	0.39	−1.94	*0.057*	−0.12	0.06	−1.79	*0.078*
N_{Pres}	0.97	0.47	2.09	**0.041**	0.15	0.08	1.96	*0.054*
$S_{LGM–Pres}$	1.58	0.61	2.60	**0.012**	0.22	0.10	2.14	**0.037**
	Present model	AIC = 85.57			*Present model*	AIC = 147.50		
(Intercept)	3.12	0.21	14.53	**0.000**	0.57	0.04	16.12	**0.000**
N_{Pres}	0.41	0.42	0.98	0.334	0.07	0.07	1.03	0.307

LGM, Last Glacial Maximum.
*Standard error.
†Significant P-values in bold ($P < 0.05$) and marginally significant ($0.05 < P < 0.10$) italicized.

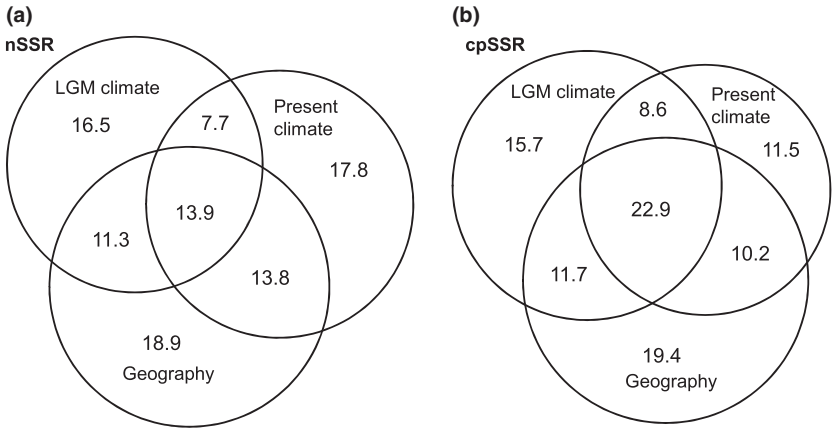

Fig. 5 Venn diagrams showing percentage of genetic 'variance' (i.e. inertia or mean squared contingency coefficient) explained by present climate, climate during the Last Glacial Maximum, geography and their joint effects for (a) nuclear microsatellites (nSSR) and (b) chloroplast microsatellites (cpSSR).

genetic diversity of valley oak (Fig. 1) (Grivet *et al.* 2006) compared with European oaks (Petit *et al.* 2002b) and eastern North American trees (Magni *et al.* 2005; McLachlan *et al.* 2005) that have undergone large-scale range shifts.

Given the long-term stability of valley oak's predicted distribution, we further hypothesized that niche stability would promote local genetic diversity. Our general linear models including niche stability and niche suitability support the idea that niche stability from the LGM to present predicts present local genetic diversity (Table 2). The role of LIG to LGM changes remains unclear as these terms were not included in the best model and not significant in the full model. In the best model, present niche suitability is also significantly or marginally significantly positively associated with genetic diversity, which makes sense if sites with more favourable habitat maintain larger population sizes. These results are consistent with the expectation that genetic diversity would be lost more slowly in stable populations.

Late Pleistocene divergence

ENMs suggest that the coastal and Sierran valley oak populations may have maintained partial isolation since at least the LIG. Both niche models and genetic data suggest isolation among coastal and Sierran populations in the southern part of valley oak's distribution since the LIG. Conversely, the northern part of the distribution

has been an area of periodic niche overlap among populations (Figs 2 and 3a), consistent with the interdigitation of nSSR and cpSSR genetic clusters in the region (Fig. 1). The inland penetration of coastal climate and the presence of riparian corridors may facilitate dispersal across Central Valley east of San Francisco Bay, especially eastward, which is supported by the higher migration rate ($2N_em$) estimate from coastal to Sierran populations (Table 1). This phenomenon has been termed the 'transvalley leak' (Peabody & Savage 1958) and has been observed in salamanders (Stebbins 1949; Martínez–Solano *et al.* 2007; Kuchta *et al.* 2009), spiders (Satler *et al.* 2011), lichens (Werth S, and Sork VL, unpublished data) and possibly annual plants (Baldwin *et al.* 2011).

The formation of the east–west split in valley oak may have coincided with the LIG when niche overlap among nSSR clusters was lowest (Fig. 4), but could have occurred at other times in the Middle to Late Pleistocene before the LGM (Table 1, Fig. S2, Supporting information). Although the divergence time cannot be pinpointed, we can exclude recent LGM or post-LGM divergence as contexts for the major subdivision within valley oak. Increased isolation among coastal and Sierran populations during warm interglacial periods (Fig. 4), causing migration of valley oak woodlands to somewhat higher elevations and perhaps out of parts of the Central Valley (Fig. 3) and reduced effective population sizes since the LGM (Table 1), could have facilitated divergence. In contrast, fossil pollen evidence suggests that oaks, although not necessarily valley oak specifically, were more abundant during interglacial periods across a variety of habitats: in the Coast Range of northern California (Clear Lake, Adam *et al.* 1981), the Central Valley (Tulare Lake, Davis 1999) and southern California (Santa Barbara Basin, Heusser 1995).

Regardless of the uncertainty in associating the divergence with a particular climate regime or transition, ENMs provide preliminary evidence of niche specialization among coastal and Sierran populations despite the interdigitation of genetic clusters. For example, there is some niche differentiation among genetic groups during interglacial periods, especially during the LIG when the coastal populations appear highly restricted to the coast and the Sierran populations appear restricted to inland areas (Fig. 4). However, niche models indicate substantial areas of overlap in the northern Central Valley during the LGM and present. In addition to geographic separation, niche differences may reinforce gene flow within each genetic cluster, consistent with earlier findings of gene flow along mountain corridors and correlations with climate (Grivet *et al.* 2008; Sork *et al.* 2010) (Figs 1 and 5).

Many phylogeographic studies of Californian taxa show the split observed among coastal and Sierran populations in nSSR data and ENMs (Lapointe & Rissler 2005). Transverse Range populations of other species tend to be more closely related to coastal populations (Calsbeek *et al.* 2003), but in some species, Transverse Range populations are more similar to Sierran populations (e.g. Macey *et al.* 2001). In valley oak, ENMs suggest that Transverse Range populations share closer ties to coastal populations, whereas Bayesian clustering analysis of nSSRs (Fig. 1a) indicates a closer relationship to Sierran populations. This relationship could reflect ongoing gene flow by pollen dispersal or insufficient time for lineage sorting at nuclear loci. The unique cpSSR composition of the Transverse Range suggests isolation, at least through seed-mediated gene dispersal (Fig. 1b).

Correlations with LGM climate

The correlation of LGM climate with present genetic variation after controlling for correlations with geography and present climate suggests that the effects of past climate on genetic variation can persist for many generations. Moreover, this remaining relationship from the LGM is about as strong as that with present climate or geography for both molecular marker types (Fig. 5). Some of this relationship could be explained by current environmental variables not included in the model that happen to correlate with past climate variables, but it seems unlikely that this explains the entire effect. Moreover, minimum LGM temperature had the highest loading on the first redundancy analysis axis for both cpSSRs and nSSRs (Fig. S3, Supporting information), which could result from post-LGM colonization of habitat that was uninhabitable during the LGM due to low minimum temperatures (i.e. climate-mediated demographic shifts) or selection during the cooler LGM. We hypothesize that genetic variation in tree populations is not in equilibrium with current climate, or that the reassortment of genetic variation in response to changing climate lags, due to long generation time and thus fewer opportunities per year for recombination, drift and/or selection. Consistent with this hypothesis, present climate explains proportionately more variation in nSSRs compared with cpSSRs, the latter of which represent a single nonrecombining locus. Furthermore, under rapidly changing environments, theory predicts a lag time for adaptive genetic variation (Pease *et al.* 1989; Bürger & Lynch 1995) that approximately corresponds to temporal scale in generations observed here.

Purely demographic forces could cause the correlation of past and present climate with allele frequency variation if climate acts as a demographic filter, permitting growth under some conditions and contraction under others, or if climate affects gene flow by affecting flowering phenology. Demographic forces have been proposed for similar correlations observed in pine bee-

tles near an expanding range margin (James *et al.* 2011). Because we factored out large-scale geographic patterns and their collinearity with climate (Fig. 5) and because no simple pattern of expansion and contraction occurred in valley oak (Figs 2 and 3), such demographic processes would have to be invoked on a local level or along climate gradients without linear, quadratic or cubic associations to latitude or longitude.

It is also possible that both demographic and selective forces account for associations with climate, as both are expected under changing conditions. The ENMs for each nSSR cluster suggest the long-term persistence of somewhat distinct niches for coastal and Sierran clusters and thus the potential for locally adapted populations (Fig. 4). This history could generate linkage disequilibrium between the neutral nSSR loci and genetic variation underlying local adaptation. Selection against nonlocal genotypes could continually reinforce climate-genotype associations in neutral markers, especially if flowering phenology differences or other potential barriers lead to effective, nascent reproductive isolation among the locally adapted clusters (Via *et al.* 2000; Nosil *et al.* 2008; Sork *et al.* 2010; Wang & Summers 2010; Andrew *et al.* 2012; Ortego *et al.* 2012). Although we cannot disentangle the effects of selection from climate-mediated demographic responses using these methods (Geffen *et al.* 2004), such correlations create conditions that would facilitate selection by climate (Endler 1986).

Conclusions

In contrast to the numerous studies of tree populations that experienced major bottlenecks due to glaciation, this investigation explores the impact of climate on the geographic genetic structure in a species that maintained a more stable distribution through recent glacial–interglacial cycles. We found evidence that present genetic variation is shaped by this history of a long-term stable distribution with local migration, late Pleistocene vicariance, climate-mediated demographic responses, and possibly past and present selection by climate. Together, these results support the hypothesis that long-term stability and variable topoclimate explain high biodiversity in the California Floristic Province. By investigating climate associations with present genetic variation at multiple timescales, we provide preliminary evidence for and quantify the endurance of past climate changes on genetic variation and establish an historical context for future genomic studies explicitly examining geographic patterns of adaptive molecular variation. To our knowledge, our study is the first to demonstrate that LGM climate is associated with present genetic variation, raising the question of whether this imprint will shape future responses of tree populations to climate change.

Acknowledgements

We thank K. Lundy and S. Werth for laboratory assistance; E.C. Riordan for help extracting data from WorldClim databases; J. Ortego for helpful conversations; and K. Gaddis, J. Ortego, A. Platt, E.C. Riordan, P. Thompson and four anonymous reviewers for constructive comments on this manuscript. Funding was provided by NSF-DEB-0516529 to V.L. Sork.

References

Adam DP, Sims JD, Throckmorton CK (1981) 130,000-yr continuous pollen record from Clear Lake, Lake County, California. *Geology*, **9**, 373–377.

Andrew RL, Ostevik KL, Ebert DP, Rieseberg LH (2012) Adaptation with gene flow across the landscape in a dune sunflower. *Molecular Ecology*, **21**, 2078–2091.

Avise JC (2000) *Phylogeography: The History and Formation of Species*. Harvard University Press, Cambridge, Massachusetts.

Azaiez A, Bouchard EF, Jean M, Belzile FJ (2006) Length, orientation, and plant host influence the mutation frequency in microsatellites. *Genome*, **49**, 1366–1373.

Baldwin BG, Kalisz S, Armbruster WS (2011) Phylogenetic perspectives on diversification, biogeography, and floral evolution of *Collinsia* and *Tonella* (Plantaginaceae). *American Journal of Botany*, **98**, 731–753.

Bandelt HJ, Forster P, Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, **16**, 37–48.

Barry RG (1983) Late-Pleistocene climatology. In: *Late-Quaternary Environments of the United States: The Late Pleistocene* (eds Wright HE, Porter SC), pp. 390–407. University of Minnesota Press, Minneapolis, Minnesota.

Bartlein PJ, Anderson KH, Anderson PM *et al.* (1998) Paleoclimate simulations for North America over the past 21,000 years: features of the simulated climate and comparisons with paleoenvironmental data. *Quaternary Science Reviews*, **17**, 549–585.

Borcard D, Legendre P, Drapeau P (1992) Partialling out the spatial component of ecological variation. *Ecology*, **73**, 1045–1055.

ter Braak CJF (1986) Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. *Ecology*, **67**, 1167–1179.

Braconnot P, Otto-Bliesner B, Harrison S *et al.* (2007) Results of PMIP2 coupled simulations of the Mid-Holocene and Last Glacial Maximum – Part 1: experiments and large-scale features. *Climate of the Past*, **3**, 261–277.

Bürger R, Lynch M (1995) Evolution and extinction in a changing environment: a quantitative genetic analysis. *Evolution*, **49**, 151–163.

Byrne R, Edlund E, Mensing S (1991) Holocene changes in the distribution and abundance of oaks in California. *USDA Forest Service Pacific Southwest Research Station. General Technical Report*, **126**, 182–191.

Calsbeek R, Thompson JN, Richardson JE (2003) Patterns of molecular evolution and diversification in a biodiversity hotspot: the California Floristic Province. *Molecular Ecology*, **12**, 1021–1029.

Carnaval AC, Hickerson MJ, Haddad CFB, Rodrigues MT, Moritz C (2009) Stability predicts genetic diversity in the Brazilian Atlantic forest hotspot. *Science*, **323**, 785–789.

Carstens BC, Richards CL (2007) Integrating coalescent and ecological niche modeling in comparative phylogeography. *Evolution*, **61**, 1439–1454.

Cavender-Bares J, González-Rodríguez A, Pahlich A, Koehler K, Deacon N (2011) Phylogeography and climatic niche evolution in live oaks (*Quercus* series *Virentes*) from the tropics to the temperate zone. *Journal of Biogeography*, **38**, 962–981.

Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, **9**, 1657–1660.

Collins WD, Bitz CM, Blackmon ML *et al.* (2006) The community climate system model version 3 (CCSM3). *Journal of Climate*, **19**, 2122–2143.

Corander J, Marttinen P, Siren J, Tang J (2008) Enhanced Bayesian modelling in BAPS software for learning genetic structures of populations. *BMC Bioinformatics*, **9**, 539.

Davis MB (1981) Quaternary history and the stability of forest communities. In: *Forest Succession: Concepts and Application* (eds West DC, Shugart HH, Botkin DB), pp. 132–153. Springer-Verlag, New York, New York.

Davis OK (1999) Pollen analysis of Tulare Lake, California: Great Basin-like vegetation in Central California during the full-glacial and early Holocene. *Review of Palaeobotany and Palynology*, **107**, 249–257.

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, California.

Davis MB, Shaw RG, Etterson JR (2005) Evolutionary responses to changing climate. *Ecology*, **86**, 1704–1714.

Derory J, Scotti-Saintagne C, Bertocchi E *et al.* (2010) Contrasting relationships between the diversity of candidate genes and variation of bud burst in natural and segregating populations of European oaks. *Heredity*, **104**, 438–448.

Dobrowski SZ (2011) A climatic basis for microrefugia: the influence of terrain on climate. *Global Change Biology*, **17**, 1022–1035.

Earl DA, vonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, **4**, 359–361.

Elith J, Graham CH, Anderson RP *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, New Jersey.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**, 2611–2620.

Excoffier L, Smouse P, Quattro J (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, **131**, 479–491.

Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, **164**, 1567–1587.

Franklin J, Wejnert KE, Hathaway SA, Rochester CJ, Fisher RN (2009) Effect of species rarity on the accuracy of species distribution models for reptiles and amphibians in southern California. *Diversity and Distributions*, **15**, 167–177.

Geffen ELI, Anderson MJ, Wayne RK (2004) Climate and habitat barriers to dispersal in the highly mobile grey wolf. *Molecular Ecology*, **13**, 2481–2490.

Griffin JR, Critchfield WB (1972) *The Distribution of the Forest Trees in California*. Pacific Southwest Forest and Range Experiment Station, U.S. Department of Agriculture Forest Service, Berkeley, California.

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.

Gugger PF, Cavender-Bares J (2013) Molecular and morphological support for a Florida origin of the Cuban oak. *Journal of Biogeography*, **40**, 632–645.

Gugger PF, Sugita S (2010) Glacial populations and postglacial migration of Douglas-fir based on fossil pollen and macrofossil evidence. *Quaternary Science Reviews*, **29**, 2052–2070.

Gugger PF, Sugita S, Cavender-Bares J (2010) Phylogeography of Douglas-fir based on mitochondrial and chloroplast DNA sequences: testing hypotheses from the fossil record. *Molecular Ecology*, **19**, 1877–1897.

Gugger PF, González-Rodríguez A, Rodríguez-Correa H, Sugita S, Cavender-Bares J (2011) Southward Pleistocene migration of Douglas-fir into Mexico: phylogeography, ecological niche modeling, and conservation of 'rear edge' populations. *New Phytologist*, **189**, 1185–1199.

Hampe A, Petit RJ (2005) Conserving biodiversity under climate change: the rear edge matters. *Ecology Letters*, **8**, 461–467.

Hasumi H, Emori S (2004) *K-1 coupled GCM (MIROC) Description*. Center for Climate System Research, University of Tokyo, Tokyo.

Heusser LE (1995) Pollen stratigraphy and paleoecologic interpretation of the 160-k.y. record from Santa Barbara Basin, Hole 893A. In: *Proceedings of the Ocean Drilling Program, Scientific Results* (eds Kennet JP, Baldauf JG & Lyle M), pp. 265–279. Texas A&M University, College Station, Texas.

Hewitt GM (1996) Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society*, **58**, 247–276.

Hewitt G (2000) The genetic legacy of the Quaternary ice ages. *Nature*, **405**, 907–913.

Hey J (2010) Isolation with migration models for more than two populations. *Molecular Biology and Evolution*, **27**, 905–920.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, **25**, 1965–1978.

Hudson RR, Kaplan NL (1985) Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics*, **111**, 147–164.

James PMA, Coltman DW, Murray BW, Hamelin RC, Sperling FAH (2011) Spatial genetic structure of a symbiotic beetle-fungal system: toward multi-taxa integrated landscape genetics. *PLoS ONE*, **6**, e25359.

Jiménez-Valverde A, Lobo JM (2007) Threshold criteria for conversion of probability of species presence to either–or presence–absence. *Acta Oecologica*, **31**, 361–369.

Kalinowski ST (2005) HP-Rare 1.0: a computer program for performing rarefaction on measures of allelic richness. *Molecular Ecology Notes*, **5**, 187–189.

Kimura M (1969) The number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutations. *Genetics*, **61**, 893–903.

Knight TM, Steets JA, Vamosi JC *et al.* (2005) Pollen limitation of plant reproduction: pattern and process. *Annual Review of Ecology Evolution and Systematics*, **36**, 467–497.

Knowles LL, Alvarado-Serrano DF (2010) Exploring the population genetic consequences of the colonization process with spatio-temporally explicit models: insights from coupled ecological, demographic and genetic models in montane grasshoppers. *Molecular Ecology*, **19**, 3727–3745.

Kuchta SR, Parks DS, Mueller RL, Wake DB (2009) Closing the ring: historical biogeography of the salamander ring species *Ensatina eschscholtzii*. *Journal of Biogeography*, **36**, 982–995.

Lapointe F-J, Rissler LJ (2005) Congruence, consensus, and the comparative phylogeography of codistributed species in California. *The American Naturalist*, **166**, 290–299.

Legendre P, Legendre L (1998) *Numerical Ecology*, 2nd English edn. Elsevier, Amsterdam.

Li N, Stephens M (2003) Modeling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data. *Genetics*, **165**, 2213–2233.

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, **25**, 1451–1452.

Little EL Jr (1971) *Atlas of United States Trees, Volume 1: Conifers and Important Hardwoods*. U.S. Department of Agriculture, Washington, District of Columbia.

Liu Q (1997) Variation partitioning by partial redundancy analysis (RDA). *Environmetrics*, **8**, 75–85.

Macey JR, Strasburg JL, Brisson JA, Vrendenburg VT, Jennings M, Larson A (2001) Molecular phylogenetics of western North American frogs of the *Rana boylii* species group. *Molecular Phylogenetics and Evolution*, **19**, 131–143.

Magni CR, Ducousso A, Caron H, Petit RJ, Kremer A (2005) Chloroplast DNA variation of *Quercus rubra* L. in North America and comparison with other Fagaceae. *Molecular Ecology*, **14**, 513–524.

Marriage TN, Hudman S, Mort ME, Orive ME, Shaw RG, Kelly JK (2009) Direct estimation of the mutation rate at dinucleotide microsatellite loci in *Arabidopsis thaliana* (Brassicaceae). *Heredity*, **103**, 310–317.

Martínez–Solano I, Jockusch EL, Wake DB (2007) Extreme population subdivision throughout a continuous range: phylogeography of *Batrachoseps attenuatus* (Caudata: Plethodontidae) in western North America. *Molecular Ecology*, **16**, 4335–4355.

McLachlan JS, Clark JS, Manos PS (2005) Molecular indicators of tree migration capacity under rapid climate change. *Ecology*, **86**, 2088–2098.

McLaughlin BC, Zavaleta ES (2012) Predicting species responses to climate change: demography and climate microrefugia in California valley oak (*Quercus lobata*). *Global Change Biology*, **18**, 2301–2312.

Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GAB, Kent J (2000) Biodiversity hotspots for conservation priorities. *Nature*, **403**, 853–858.

Nosil P, Egan SP, Funk DJ (2008) Heterogeneous genomic differentiation between walking-stick ecotypes: "isolation by adaptation" and multiple roles for divergent selection. *Evolution*, **62**, 316–336.

Ohta T, Kimura M (1973) A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. *Genetical Research*, **22**, 201–204.

Økland RH (1999) On the variation explained by ordination and constrained ordination axes. *Journal of Vegetation Science*, **10**, 131–136.

Oksanen J, Kindt R, Legendre P *et al.* (2009) *vegan: Community Ecology Package*. Available from http://cran.r-project.org/web/packages/vegan/index.html

Ortego J, Riordan EC, Gugger PF, Sork VL (2012) Influence of environmental heterogeneity on genetic diversity and structure in an endemic southern Californian oak. *Molecular Ecology*, **21**, 3210–3223.

Otto-Bliesner BL, Marshall SJ, Overpeck JT *et al.* (2006) Simulating Arctic climate warmth and icefield retreat in the last interglaciation. *Science*, **311**, 1751–1753.

Peabody FE, Savage JM (1958) Evolution of the coast range corridor in California and its effect on the origin and dispersal of living amphibians and reptiles. In: *Zoogeography* (ed. Hubbs CL), pp. 159–186. American Association for the Advancement of Science, Washington, District of Columbia.

Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel: population genetic software for teaching and research. *Molecular Ecology Notes*, **6**, 288–295.

Pease CM, Lande R, Bull JJ (1989) A model of population growth, dispersal and evolution in a changing environment. *Ecology*, **70**, 1657–1664.

Pease KM, Freedman AH, Pollinger JP *et al.* (2009) Landscape genetics of California mule deer (*Odocoileus hemionus*): the roles of ecological and historical factors in generating differentiation. *Molecular Ecology*, **18**, 1848–1862.

Petit RJ, Pineau E, Demesure B, Bacilieri R, Ducousso A, Kremer A (1997) Chloroplast DNA footprints of postglacial recolonization by oaks. *Proceedings of the National Academy of Sciences*, **94**, 9996–10001.

Petit RJ, Brewer S, Bordács S *et al.* (2002a) Identification of refugia and post-glacial colonisation routes of European white oaks based on chloroplast DNA and fossil pollen evidence. *Forest Ecology and Management*, **156**, 49–74.

Petit RJ, Csaikl UM, Bordács S *et al.* (2002b) Chloroplast DNA variation in European white oaks: phylogeography and patterns of diversity based on data from over 2600 populations. *Forest Ecology and Management*, **156**, 5–26.

Phillips SJ, Dudik M, Schapire RE (2004) A maximum entropy approach to species distribution modeling. *ACM International Proceedings Series*, **69**, 655.

Phillips SJ, Anderson RP, Schapire RE (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, **190**, 231–259.

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

Sage R, Koenig W, McLaughlin B (2011) Fitness consequences of seed size in the valley oak *Quercus lobata* Née (Fagaceae). *Annals of Forest Science*, **68**, 477–484.

Satler JD, Starrett J, Hayashi CY, Hedin M (2011) Inferring species trees from gene trees in a radiation of California trapdoor spiders (Araneae, Antrodiaetidae, *Aliatypus*). *PLoS ONE*, **6**, e25355.

Smouse PE, Williams RC (1982) Multivariate analysis of HLA-disease associations. *Biometrics*, **38**, 757–768.

Sork VL, Davis FW, Westfall R *et al.* (2010) Gene movement and genetic association with regional climate gradients in California valley oak (*Quercus lobata* Nee) in the face of climate change. *Molecular Ecology*, **19**, 3806–3823.

Stebbins RC (1949) Speciation in salamanders of the plethodontid genus *Ensatina*. In: *University of California Publications in Zoölogy* (eds Daniel JF, Eakin RM, Hall ER, Kirby H, Light SF, Miller AH), pp. 377–532. University of California, Berkeley, California.

Stebbins GL, Major J (1965) Endemism and speciation in the California flora. *Ecological Monographs*, **35**, 2–35.

Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics*, **68**, 978–989.

Svenning J-C, Normand S, Kageyama M (2008) Glacial refugia of temperate trees in Europe: insights from species distribution modelling. *Journal of Ecology*, **96**, 1117–1127.

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585–595.

Thornthwaite C (1953) A charter for climatology. *World Meteorological Organization Bulletin*, **2**, 40–46.

Tzedakis PC, Lawson IT, Frogley MR, Hewitt GM, Preece RC (2002) Buffered tree population changes in a Quaternary refugium: evolutionary implications. *Science*, **297**, 2044–2047.

Via S, Bouck AC, Skillman S (2000) Reproductive isolation between divergent races of pea aphids on two hosts. II. Selection against migrants and hybrids in the parental environments. *Evolution*, **54**, 1626–1637.

Waltari E, Hijmans RJ, Peterson AT, Nyári ÁS, Perkins SL, Guralnick RP (2007) Locating Pleistocene refugia: comparing phylogeographic and ecological niche model predictions. *PLoS ONE*, **2**, e563.

Wang IJ, Summers K (2010) Genetic structure is correlated with phenotypic divergence rather than geographic isolation in the highly polymorphic strawberry poison-dart frog. *Molecular Ecology*, **19**, 447–458.

Werth S, Sork VL (2008) Local genetic structure in a North American epiphytic lichen, *Ramalina menziesii* (Ramalinaceae). *American Journal of Botany*, **95**, 568–576.

Westfall RD, Conkle MT (1992) Allozyme markers in breeding zone designation. *Population Genetics of Forest Trees*, **42**, 279–309.

Whipple AA, Grossinger RM, Davis FW (2011) Shifting baselines in a California oak savanna: nineteenth Century data to inform restoration scenarios. *Restoration Ecology*, **19**, 88–101.

Won YJ, Hey J (2005) Divergence population genetics of chimpanzees. *Molecular Biology and Evolution*, **22**, 297–307.

P.F.G. is a Postdoctoral Scholar interested in combining fossil evidence, molecular data, and models to understand the ecological and evolutionary history of trees. 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 work. V.S. integrates population genomic, landscape genetic, and molecular ecology approaches to understand how gene flow and natural selection shape the evolutionary potential of natural tree populations.

Data accessibility

DNA sequences: GenBank accessions JQ421340–JQ421389 for unphased sequences, and Dryad doi:10.5061/dryad.g645d for aligned, phased sequences.

Sample locations, microsatellite data and occurrence points for ecological niche models: Dryad doi:10.5061/dryad.g645d.

Supporting information

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

Table S1 Correlation among independent variables used in general linear models to test the role of past niche suitability and stability on present genetic diversity.

Table S2 Mean log likelihood (ln$P(K)$), its standard deviation (SD), and ΔK calculated from 10 replicates for each of 1–10 clusters (K) using a Bayesian clustering analysis implemented in STRUCTURE.

Table S3 Analyses of molecular variance based on Φ-statistics for nuclear microsatellite (nSSR) sample sites within STRUCTURE clusters, chloroplast microsatellite (cpSSR) sites, and cpSSR BAPS clusters.

Table S4 Percent contribution and permutation importance of each variable used to construct valley oak (*Quercus lobata*) ecological niche models in MAXENT.

Table S5 Nuclear microsatellite (nSSR) allelic richness and heterozygosity, niche suitabilities, and niche stabilities by site for samples used in genetic analyses (additional sites were used to construct niche models).

Fig. S1 Maps of genetic variation in valley oak showing haplotype distributions for (a) *glyceraldehyde 3-phosphate dehydrogenase* (*g3pdh*) and (b) *elongation factor 1-α* (*ef1α*).

Fig. S2 Posterior distributions for per-locus mutation rate-scaled parameter estimates from the IMa2 run with the highest effective sample size.

Fig. S3 Biplots of partial redundancy analyses of (a) nuclear microsatellite and (b) chloroplast microsatellite variation on Last Glacial Maximum (~21 ka) climate after controlling for present climate and geographic position.