Molecular Ecology (2008) **17**, 139–156 doi: 10.1111/j.1365-294X.2007.03498.x

Conserving the evolutionary potential of California valley oak (*Quercus lobata* Née): a multivariate genetic approach to conservation planning

DELPHINE GRIVET,*† VICTORIA L. SORK,* ROBERT D. WESTFALL‡ and FRANK W. DAVIS§
*Department of Ecology and Evolutionary Biology and Institute of the Environment, University of California Los Angeles, Box 951606, Los Angeles, CA 90095-1606, USA, †Department of Forest Systems and Resources, Forest Research Institute, CIFOR-INIA, Carretera de la Coruña km 7.5, 28040 Madrid, Spain, ‡Sierra Nevada Research Center, USDA Forest, Service, Pacific Southwest Research Station, PO Box 245, Berkeley, CA 94701, USA, §Donald Bren School of Environmental Science and Management, University of California, Santa Barbara, CA 93106, USA

Abstract

California valley oak (*Quercus lobata* Née) is a seriously threatened endemic oak species in California and a keystone species for foothill oak ecosystems. Urban and agricultural development affects a significant fraction of the species' range and predicted climate change is likely to dislocate many current populations. Here, we explore spatial patterns of multivariate genotypes and genetic diversity throughout the range of valley oak to determine whether ongoing and future patterns of habitat loss could threaten the evolutionary potential of the species by eradicating populations of distinctive genetic composition. This manuscript will address three specific questions: (i) What is the spatial genetic structure of the chloroplast and nuclear genetic markers? (ii) What are the geographical trends in the distribution of chloroplast and nuclear genotypes? (iii) Is there any part of the species' range where allelic diversity in either the chloroplast or nuclear genomes is particularly high? We analysed six chloroplast and seven nuclear microsatellite genetic markers of individuals widespread across the valley oak range. We then used a multivariate approach correlating genetic markers and geographical variables through a canonical trend surface analysis, followed by GIS mapping of the significant axes. We visualized population allelic richness spatially with GIS tools to identify regions of high diversity. Our findings, based on the distribution of multivariate genotypes and allelic richness, identify areas with distinctive histories and genetic composition that should be given priority in reserve network design, especially because these areas also overlap with landscape change and little degree of protection. Thus, without a careful preservation plan, valuable evolutionary information will be lost for valley oak.

Keywords: allelic richness, canonical trend surface analysis, colonization, gene flow, Geographic Information System

Received 8 March 2007; revision accepted 10 July 2007

Introduction

Human-altered landscapes affect the distribution and abundance of many species by fragmenting populations and creating threats to genetic connectivity and by reducing local population sizes, which can diminish genetic diversity and evolutionary potential (Ledig & Kitzmiller 1992; Young *et al*. 1996; Couvet 2002; Frankham *et al*. 2002; Stockwell *et al*. 2003). When a species is widely distributed, its evolutionary history can result in a geographical mosaic of genotypes that reflects the movement of genes and the impact of genetic bottlenecks (Moritz 1995; cf. references in Moritz & Faith 1998; Avise 2000). Some parts of a species' range may be 'evolutionary hotspots' that reflect past evolutionary processes and provide essential

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

genetic variation for future evolutionary dynamics (Rayburn & Moritz 2006). In this study, we ask whether the regions of dramatic landscape alteration comprise areas of evolutionary interest to California valley oak (*Quercus lobata* Née), which is considered the 'monarch of California oaks by virtue of its size, age, and beauty' (Pavlik *et al*. 1991). Like other tree oak species of California, valley oak defines the ecosystem and plays a keystone resource role for wildlife, insects, fungi, and lichen.

Many widely distributed woody plant species have special conservation significance, and yet present a particular challenge to the development of conservation strategies because these species have high genetic diversity within populations and low genetic differences among populations because of gene flow (Hamrick *et al*. 1992; Hamrick & Godt 1996). Indeed, it might seem that preserving any subset of populations should be sufficient, since theoretically, each harbours a representative fraction of the total gene pool, and therefore should include sufficient genetic variation to protect the evolutionary potential of the species. Instead, range-wide geographical patterns of these tree species are likely to include gradual trends in variation (e.g. Kinloch *et al*. 1986; Kremer & Zanetto 1997; Le Corre *et al*. 1998). For example, directionality in migration rates could create genetic clines across the species range (Epperson 2003). Alternatively, these species could show sharp genetic gradients that might be indicative of 'evolutionary hotspots' created by suture zones, areas of high ecological complexity, historical refugia, or recent colonization and adaptive divergence (Rayburn & Moritz 2006). Thus, even species with widespread distributions and tremendous potential for gene flow are likely to possess complex geographical patterns of genetic variation that are at risk when part of their range is lost.

Since European settlement, the distribution and abundance of valley oak has been reduced through conversion of oak savannah and riparian oak forest to agricultural crops, vineyards, and urban development (Pavlik *et al*. 1991; Kelly *et al*. 2005). In the San Joaquin Valley, the extent of valley oak woodland and savanna may have been reduced by as much as 95% (Kelly *et al*. 2005). Remnant woodlands occur mainly in the foothills and valleys of the Central and Northern Coast Ranges, Tehachapi Mountains and Sierra Nevada. Housing density exceeds one house per 64 ha (160 ac) on more than two-thirds of mapped valley oak woodlands and exceeds one house per 16 ha (40 ac) on nearly 20% of remaining woodlands. Based on population projections, residential development will continue in many of these areas (Fig. 1) because more than 90% of the area is privately owned and less than 3% is in formally designated reserves (Davis *et al*. 2000). Within existing habitat, recruitment of reproductive-aged individuals is rare and not sufficient to offset adult mortality (Brown & Davis 1991; Tyler *et al*. 2006). Thus, valley oak plant communities are considered threatened and of high priority for inventory by both the California Department of Fish and Game and by The Nature Conservancy (Davis *et al*. 2000). Combined with this predicted loss of habitat and declining population size in remaining areas is potential habitat dislocation due to climate change (Kueppers *et al*. 2005). Collectively, the issues of landscape and climate change call for a strategy for a network of reserves that protect the regions of evolutionary and genetic uniqueness.

In this study, we analyse geographical patterns of genetic variation to identify regions of potential conservation significance. Our strategy in detecting regions of evolutionary interest has been to allocate our sampling effort to maximize the number of sampling locations rather than the number of individuals within populations. This investigation provides a first-round analysis that examines a broad range of the distribution to identify areas of further interest. The results here can both identify areas of conservation priority and then inform a second-round analysis where additional sampling would be beneficial. We employ multivariate statistics because they are sensitive to detecting small genetic differences across loci (Conkle & Westfall 1984; Kremer & Zanetto 1997; Rajora & Dancik 1999). The use of multivariate genetic variables may be novel for conservation planning, but the methods of quantitative genetics or community ecology invoke the same approach. Multivariate statistics have been used to assess genetic variation within and among regions (e.g. Li & Adams 1989; Kremer & Zanetto 1997). The geographical analysis of multivariate genotypes has been used by forest geneticists to identify breeding zones for trees (Westfall & Conkle 1992). It is possible to statistically test whether genetic variables and environmental variables are significantly correlated through canonical correspondence analysis (CCA) (e.g. Gram & Sork 1999, 2001). Canonical trend surface analysis (Wartenberg 1985), an extension of canonical analysis, provides a useful technique first to test statistically for the directionality in continuous geographical trends of genetic data, treating these data like quantitative traits, and then to form hypotheses about historical gene movement. Depending on the genetic variables analysed with these multivariate statistics, different information will be revealed that can be integrated into conservation strategies. Uniparentally inherited markers, such as those in the chloroplast, offer useful information on (re)colonization through seeds. Nuclear markers encompass information of both pollen and seed dispersal, and their comparison with strictly maternally inherited markers provides access to pollen movement (Petit *et al*. 1993a; Ennos 1994). Thus, analysing simultaneously both sets of markers can reveal areas of genetic distinctiveness and of common evolutionary history over the species' range.

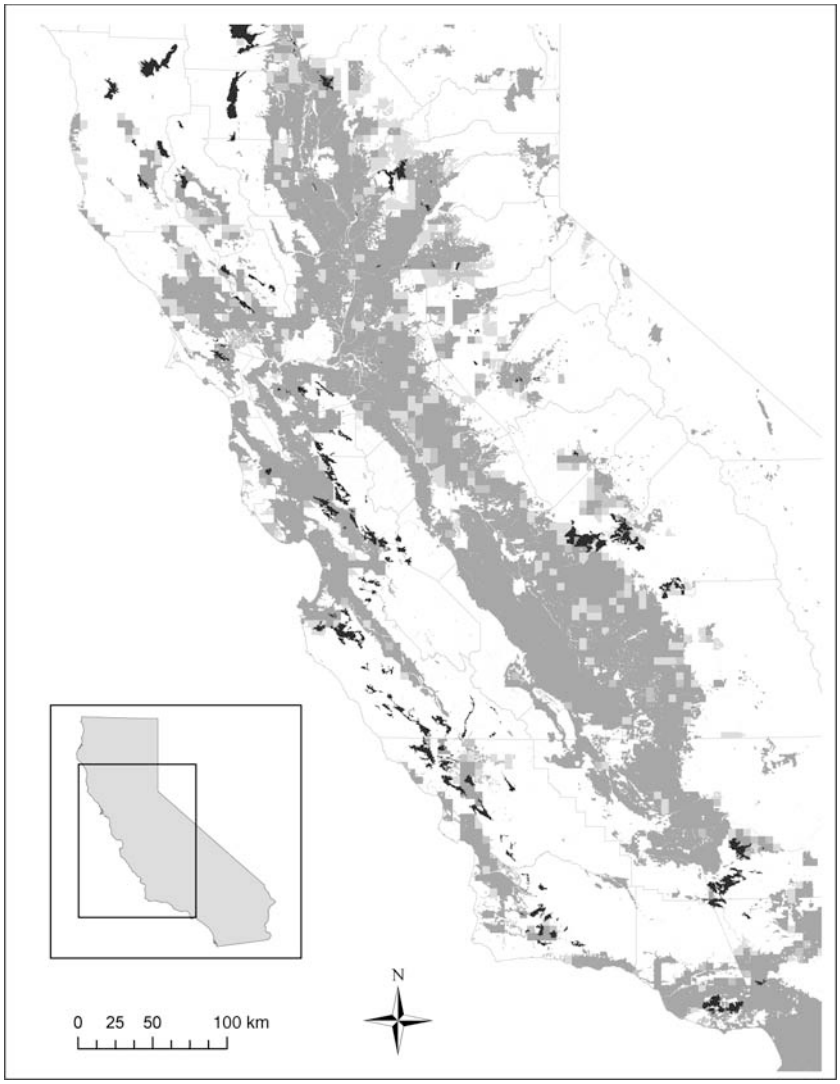

Fig. 1 General distribution of valley oak woodlands (black areas) in California showing their juxtaposition with urban/suburban (> 1 dwelling unit per 2 ha) and agricultural (excluding rangelands) areas (dark grey), as well as areas projected to be urbanized by 2040 common era (light grey). The maps was produced by combining the Gap Analysis map of extant valley oak woodland (Davis *et al*. 2000; http://frap.cdf.ca.gov/data.html) with the California Department of Forestry and Fire Protection's (CDF & FP) maps of land use and land cover and housing projections (http://frap.cdf.ca.go/data.html).

The overall questions of this study are whether California valley oak has regions of evolutionary interest that should be considered by conservation planners, and whether these regions are under threat because of agricultural and urban land development. Previously, we have found that chloroplast haplotypes of valley oak display high richness and are clustered in patches across the species range (Grivet *et al*. 2006). The comparison of the historical colonization of this California oak with that of the well-studied European oaks led us to suggest that valley oak populations have experienced a series of local contractions/expansions due the combined impacts of climatic and geological heterogeneities (Grivet *et al*. 2006). In this study, we will use chloroplast (cp) and nuclear (n) genetic markers to address three specific questions: (i) What is the spatial genetic structure of the chloroplast and nuclear genetic markers? For this question, we will describe the genetic structure of the sampled populations with analysis of molecular variance (AMOVA, Excoffier *et al*. 1992) and the pattern of spatial autocorrelation for multi-allelic loci (Smouse & Peakall 1999). (ii) What are the geographical trends in the distribution of chloroplast and nuclear genotypes? We will answer this question by conducting a canonical trend surface analysis of chloroplast and nuclear markers to identify geographical patterns across the species range and to detect the direction of correlations between genetic variables and spatial location. (iii) Is there any part of the species' range where allelic diversity in either the chloroplast or nuclear genomes is particularly high? Here, we will examine geographical trends in allelic richness for both sets of markers to detect potential centre(s) of genetic diversity. In the discussion, we will discuss these findings in terms of human threats to this species because of landscape alteration and predicted climate change.

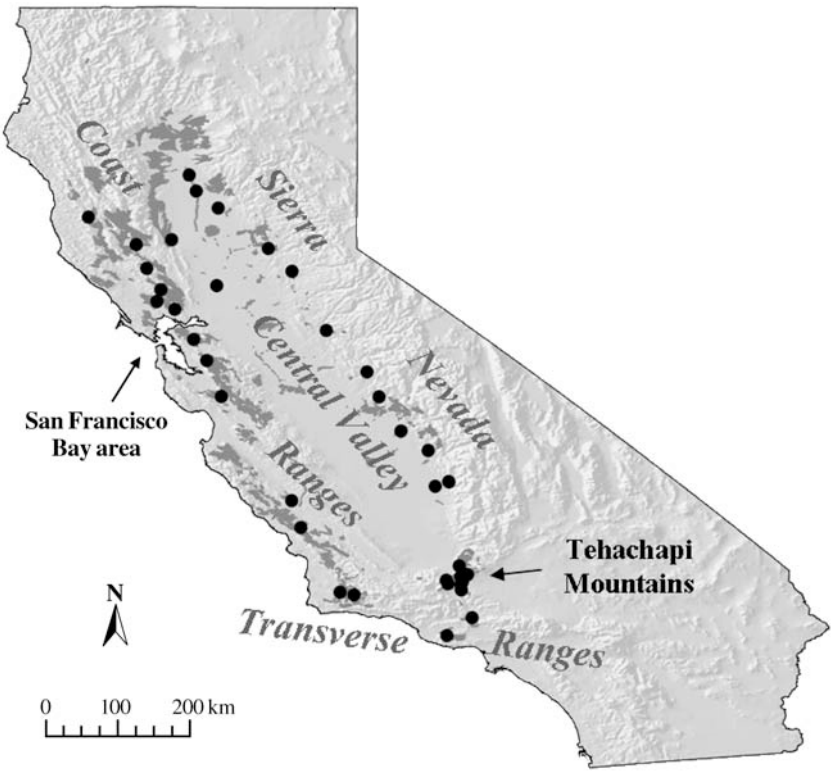

Fig. 2 Location of the 37 sampled populations (black dots), along with the distribution range of *Quercus lobata* (grey areas), and the main mountain ranges in California.

Materials and methods

Study species and study area

Quercus lobata (Née) is a diploid, wind-pollinated, monoecious, and predominantly outcrossing tree species, based on Ritland's (1990) MLTR mating system analysis (Sork *et al*. 2002b). Acorns mature in late September to early November of the year of pollination and germinate within 1–2 months after maturation. Acorns are dispersed by gravity, small mammals, acorn woodpeckers, and possibly jays. Valley oak populations occur at low densities in oak savanna habitat in the Central Valley, and more abundantly on the foothills of the Sierra Nevada, Coastal Ranges, and Transverse Ranges that surround the Central Valley (Fig. 2). Because of their extended latitudinal distribution (34–40° latitude) and altitudinal range (from sea level to 1700 m, mainly at elevations lower than 600 m, Pavlik *et al*. 1991), *Q. lobata* populations are spread across various climatic and geographical zones.

Valley oak has recently been the focus of genetic studies on several scales at the University of California Sedgwick Reserve, located in the Santa Ynez Valley, Santa Barbara County, California, USA. 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-distance pollen dispersal (Smouse *et al*. 2001; Sork *et al*. 2002a; Sork *et al*. 2002b; A. Pluess, V. Sork, B. Dolan, F. Davis, K. Merg, D. Grivet, J. Papp, P. Smouse, submitted), although long-distance dispersal occurs. Finally, acorn woodpecker-mediated movement is restricted, in the order of a 100 m (Grivet *et al*. 2005). Fine-scale genetic analysis of adults found within a ~230-ha area indicates isolation by distance because of historically restricted gene flow in valley oak estimated at ~350 m (σ) for seed and pollen together (Dutech *et al*. 2005). Thus, studies of contemporary and historical gene flow both indicate that the scale of most dispersal is on the order of 100 m.

Sampling

In 2003 and 2004, we sampled individuals from 37 populations across the contemporary valley oak distribution range (Fig. 2). The data set used for chloroplast markers includes 97 individuals spread across 32 populations (Grivet *et al*. 2006). A post-hoc analysis of the optimal allocation of resources (Sokal & Rohlf 1981), using the sampling variances of observed heterozygosities (Weir 1996) in previously sampled allozyme data (C.I. Millar, L.A. Riggs, D.L. Delany, R.D. Westfall, unpublished) indicated that a sampling design with many sites and two to three individuals per site would be sufficient. The nuclear marker data set includes 113 individuals sampled across all 37 populations because of the inclusion of additional samples from the Tehachapi Mountains (Tejon Ranch, Lebec, California). The characterization of the sampled valley oak populations (latitude, longitude, elevation) is provided in Sork *et al*. (in press).

DNA extraction and genotyping

We analysed microsatellite genetic markers for both the chloroplast and nuclear genomes. Microsatellite markers are useful for conservation studies because their relatively high level of polymorphism provides sensitive measures of genetic variation among populations. Microsatellite markers have been effective in describing genetic structure, levels of inbreeding, estimates of effective population size, and patterns of gene movement — all important in evaluation of the evolutionary history and future potential of a taxon (Schwartz *et al*. 2007).

Chloroplast data set. We used the genetic data obtained from our previous study (Grivet *et al*. 2006), where we describe the DNA extraction method and the polymerase chain reaction (PCR) conditions. These samples were genotyped for six chloroplast microsatellite primers: ccmp10 was designed on *Nicotiana tabacum* (Weising & Gardner 1999), while dt1, dt3, dt4, cd4, and cd5 were designed on *Quercus robur* (Deguilloux *et al*. 2003).

Nuclear data set. We ground about 50 mg of frozen leaf tissue in tubes with a tungsten ball using the Mixer Mill MM301 (Retsch), and DNA extraction was performed with the DNeasy Plant Mini Kit (QIAGEN). We genotyped all samples for seven nuclear microsatellites: MSQ4 (Dow *et al*. 1995), QpZAG1/5, QpZAG9, QpZAG36, QpZAG110 (Steinkellner *et al*. 1997), QrZAG11, and QrZAG20 (Kampfer *et al*. 1998). Amplifications were carried out in 10- L reaction mixture containing 2.5 L of DNA (≈ 20 ng) and 7.5 L of PCR mix, which contained the following reagents: 5 L of Multiplex Mix (QIAGEN), 0.4 L of BSA 10×, 1 L of primer mix (at 2 M for MSQ4, QpZAG9, QpZAG36, QpZAG110, QrZAG20, and at 1 M for QpZAG1/5 and QrZAG11), 1.1 L of water. The PCR touchdown profile consisted of an initial denaturing of 15 min at 95 °C, followed by 12 cycles of 30-s denaturing at 94 °C, 90-s annealing from 60 to 55 °C, and 60-s extension at 72 °C, followed by 33 cycles of 30-s denaturing at 89 °C, 90-s annealing at 55 °C, and 60-s extension at 72 °C, with a final extension step of 60 °C for 30 min. We measured the length of the amplified sequence by running an aliquot of each PCR product on an ABI PRISM 3700 capillary sequencer at the University of California, Los Angeles (UCLA) Sequencing & Genotyping Core Facility (http://www.genetics.ucla.edu/sequencing/index.php). To verify repeatability, each sample was re-genotyped, after repeating the PCRs.

Statistical analyses

Chloroplast and nuclear genetic structure. To estimate the differentiation levels among populations, we computed an analogue of *F*-statistics, G_{ST}, for which populations have equal weight, irrespective of sample size (Pons & Petit 1996), using the program SPAGEDI version 1.2 (Hardy & Vekemans 2002). We also used SPAGEDI to statistically compare our various *F*-statistics by computing jackknife standard errors for multilocus average statistics to calculate the 95% confidence intervals.

To determine whether populations have regional structure, we ran an AMOVA with the GENALEX version 6 software (Peakall & Smouse 2006) in a hierarchical model that included variation within populations, among populations within regions, and among three regions, using samples from the Coastal Ranges, from the Sierra Nevada and from the Transverse Ranges. These three mountain ranges were chosen as natural delimitations that can potentially affect valley oak population movements because they are characterized by different uplift timings and climatic/ecological conditions (cf. Materials and Methods section in Grivet *et al*. 2006). This analysis discarded one population in the Central Valley East of San Francisco, because it was not in spatial proximity to any of these three regions.

Finally, we examined the scale of historical gene flow among populations by performing a spatial autocorrelation analysis based on genetic distance methods for multiple allelic and multilocus loci (Smouse & Peakall 1999). This spatial autocorrelation method describes the genetic structure across the studied site by employing a multivariate approach combining alleles and loci that allow reducing stochastic (allele-to-allele and locus-to-locus) noise, strengthening therefore the spatial signal. The generated autocorrelation coefficient (r) is closely related to Moran's I, and provides a measure of the genetic similarity between pairs of individuals. The results for the chloroplast markers have been published in Grivet *et al*. (2006). In this study, we present the results for the nuclear data set.

Geographical trends in chloroplast and nuclear genotypes. For a multivariate analysis of the microsatellite markers, we first transformed the single locus genotypes into allelic variables (Westfall & Conkle 1992). Using the method of Smouse & Williams (1982), we converted each allele into a single variable based on presence or absence. For haploid markers, the score would be 1 if present and 0 if absent. For diploid markers, the score of a single allelic variable would be 1 if present as a homozygote, 0.5 if a heterozygote, and 0 if not present. The number of variables created at each locus would be the number of alleles minus one. As Westfall & Conkle (1992) point out, linear combinations of these variables usually result in normal distributions of scored data when sufficient numbers of loci are present.

For the chloroplast data, we used results from six primers, with two to five alleles each, resulting in a total of 16 transformed allelic variables, after removing one allele per locus. We discarded an additional allele from our

analysis that was distributed in limited areas (ccmp10_2), resulting in 15 cp allelic variables. The nuclear genotypes were derived from seven nuclear microsatellite markers and from 7 to 23 alleles per locus. These data yielded 78 transformed allelic variables, after removing at least one allele per locus. To reduce the high number of nuclear allelic variables into a smaller number of genetic variables appropriate for our sample size, we conducted principal component analysis (PCA) and then used the first 15 principal component (PC) axes to represent the genotypic data. These 15 axes accounted for 48% of the variation in the markers.

To test the relationship between multilocus genetic and geographical variables (longitude, latitude, elevation) and estimate continuous genetic patterns, we performed canonical trend surface analyses (Lee 1969; Wartenberg 1985) for the genetic markers vs. the geographical variables, using a second-order surface equation of latitude (projected as metres north in the Albers equal-area conic projection), longitude (metres east in the Albers projection), and elevation (metres above sea level). We used standard cartographic projection parameters for California, including standard parallels of 34°N and 40°30′N, a central meridian of –120° and a false northing of 4000 000 m. Canonical trend surface analysis, a variant of canonical correlation analysis, estimates the association between two sets of variables (Legendre & Legendre 1998), a multivariate equivalent of multiple regression. We conducted this analysis using the PROC CANCORR function of SAS (SAS Institute 1989). In our analyses, we tested the normality of residuals of the canonical models using SAS PROC UNIVARIATE. We then passed observed scores for the first two canonical axes of the cp and nuclear marker models to diagnostically describe the forms of the equations (Box & Draper 1987) 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.

Maps of canonical variates were produced using ESRI ARCMAP 9.1 geographical information system software (ESRI). We produced 500 grids of northing, easting, and elevation [re-sampled from a 100 m US Geological Survey digital elevation grid (http://edc.usgs.gov/products/elevation/dem.html)], and calculated first and second canonical axis scores for each genetic marker by applying the polynomial functions at every grid cell. The boundaries of the maps were defined by the ecological subregions (Hickman 1993) where *Q. lobata* occurs.

Geographical patterns of allelic richness. Both allelic richness and heterozygosity measure genetic variation, but studies have shown that the former is more sensitive to the effects of short, severe bottlenecks, and may also reflect more effectively a population's long-term evolutionary potential (references cited in Leberg 2002). We computed the allelic richness for multiple loci for each population for chloroplast and nuclear markers with the program FSTAT version 2.9.3.2 (Goudet 2001). We did not correct the measures of allelic richness for variation in sample size since we have a balanced design (three individuals per population except two that have four individuals). We then overlaid these values on the California map by running the Geographic Information System (GIS) software ARCMAP 9.1 (ESRI). We used the Geostatistical Analyst to derive a surface with the inverse-distance weighting method (determines cell values using a linearly weighted combination of a set of sample points; the weight being a function of inverse distance).

Results

Chloroplast and nuclear genetic structure

The chloroplast genome shows much greater genetic structure among populations than the nuclear genome as indicated by the *F*-statistics: G_{ST} = 0.805 (95% CI = 0.7661–0.8433) for the chloroplast markers and G_{ST} = 0.052 (95% CI = 0.0327–0.0715) for the nuclear loci. The hierarchical analysis of genetic structure indicates a significant amount of variation (19%) among the three mountain ranges for the chloroplast genome, but not for the nuclear genome (Table 1a). These results suggest that the Coast Ranges, Sierra Nevada, and Transverse Ranges are genetically

Table 1 Analyses of molecular variance for chloroplast and nuclear markers for three levels of structure (a), and for western and eastern samples (b) of California valley oak, *Quercus lobata*

(a) Summary of AMOVA results within and among major regions

Variance component	d.f.	SS	MS	Percentage of total
Chloroplast (N = 31)				
Among regions	2	27.3	13.7	19.2
Among populations/regions	28	106.2	3.8	63.4
Within populations	62	19.7	0.3	17.4
Nuclear (N = 36)				
Among regions	2	13.6	6.8	0
Among populations/regions	33	213.1	6.5	12.3
Within populations	73	330.3	4.5	87.5

(b) Comparison of genetic structure (G_{ST}) between two major regions of valley oak

Source of variation	Western samples	Eastern samples
Chloroplast (N_{West} = 14; N_{East} = 13)		
Among populations within regions	0.606	0.968
Nuclear (N_{West} = 14; N_{East} = 13)		
Among populations within regions	0.047	0.065

d.f., degree of freedom; SS, sums of squares; MS, mean square.

Table 2 Summary of canonical correlation analysis based on chloroplast genetic markers. (a) Statistical results of canonical correlation analysis for genetic variables versus the geographical variables of latitude, longitude, and elevation for the significant canonical axes. (b) Canonical correlations of the 15 chloroplast allelic variables with the predicted scores of the first two canonical axes. (c) Canonical correlations of the linear and higher order combination of the geographical variables for the first two canonical axes

(a) Statistical results

Canonical axis	Canonical correlation	Adjusted canonical correlation	Squared canonical correlation	Eigenvalue	Approximate *F* value	Numerator d.f.	Denominator d.f.	Pr > F
1	0.93	0.91	0.87	6.73	7.7	135	583.16	<0.0001
2	0.91	0.88	0.82	4.54	6.28	112	530.55	<0.0001
3	0.84	0.80	0.70	2.36	5.04	91	476.01	<0.0001
4	0.78	0.73	0.60	1.53	4.27	72	419.29	<0.0001
5	0.72	0.68	0.52	1.09	3.68	55	360.00	<0.0001
6	0.66	0.63	0.43	0.77	3.12	40	297.62	<0.0001
7	0.55	0.48	0.30	0.43	2.54	27	231.36	0.0001
8	0.48	0.47	0.23	0.30	2.23	16	160.00	0.0062

(b) Correlation of chloroplast haplotypes allelic variables with predicted scores for the two first canonical axes, W1 and W2

Allelic variables	W1	W2
dt1_1	–0.05	0.35
dt1_2	0.01	0.28
dt1_3	0.15	–0.34
cd4_1	0.51	0.02
cd4_2	0.20	0.46
cd4_3	–0.35	0.06
dt3_1	–0.15	–0.21
dt3_2	–0.09	0.07
dt3_3	–0.13	0.39
dt3_4	–0.44	–0.08
cd5_1	–0.15	–0.01
cd5_2	0.23	0.11
ccmp10_1	–0.22	0.03
ccmp10_3	–0.56	0.31
dt4_1	0.30	0.34

(c) Correlation of geographical variables with predicted scores for the first two canonical axes, W1 and W2

Longitude (X)	0.69	–0.37
Latitude (Y)	–0.69	0.60
Elevation (Z)	0.26	–0.63
X*X	–0.37	0.42
Y*Y	0.84	–0.27
Z*Z	0.35	–0.47
X*Y	–0.52	–0.44
X*Z	0.80	–0.14
Y*Z	–0.73	0.43

interconnected for the nuclear markers, and to a lesser extent for the chloroplast markers.

Our comparison of the genetic structure of populations in the western vs. eastern samples (Table 1b) shows that the chloroplast genome has significantly much greater structure in the eastern than western populations: G_{ST} = 0.968 (95% CI = 0.9018–1.0338) vs. G_{ST} = 0.606 (95% CI = 0.517–0.695), respectively. The nuclear markers show a similar, albeit nonsignificant trend between the eastern and western populations: G_{ST} = 0.0653 (95% CI = 0.0379–0.0927) vs. G_{ST} = 0.0473 (95% CI = 0.0109–0.0837), respectively (Table 1b). Thus, it appears that historical gene flow among populations was greater in the western region, especially for seed movement.

The spatial autocorrelation analyses are consistent with the higher structure in the chloroplast markers vs. the nuclear markers for oak populations: the uniparentally inherited genome displays a positive spatial autocorrelation for individual pairs distant from 250 km or less ($0 \le r \le 0.2$, Grivet *et al*. 2006), while the biparentally inherited genome shows a positive spatial autocorrelation for individual pairs separated by 100 km or less ($0 \le r \le 0.06$). This spatial pattern most likely reflects restricted gene flow for both genomes (Sokal & Wartenberg 1983; Epperson 1990; Smouse & Peakall 1999).

Geographical trends in chloroplast and nuclear genotypes

Chloroplast data. The canonical correlation model revealed significant correlation of the multivariate genetic data with latitude, longitude, and elevation. More specifically, the first eight of the nine canonical correlations are significant (Table 2a), the first two explaining most of the multilocus patterns (38% and 25%, respectively). With the eigenvalues and the trace (the sum of the eigenvalues), which are the multivariate equivalent of the ratio of the sums of squares due to the model(s) and the residual sums of squares, respectively, we can compute percent of the total allelic variation represented by the models. These values are roughly equivalent to a multivariate form of F_{ST} (Kremer *et al*. 1997). For these first two

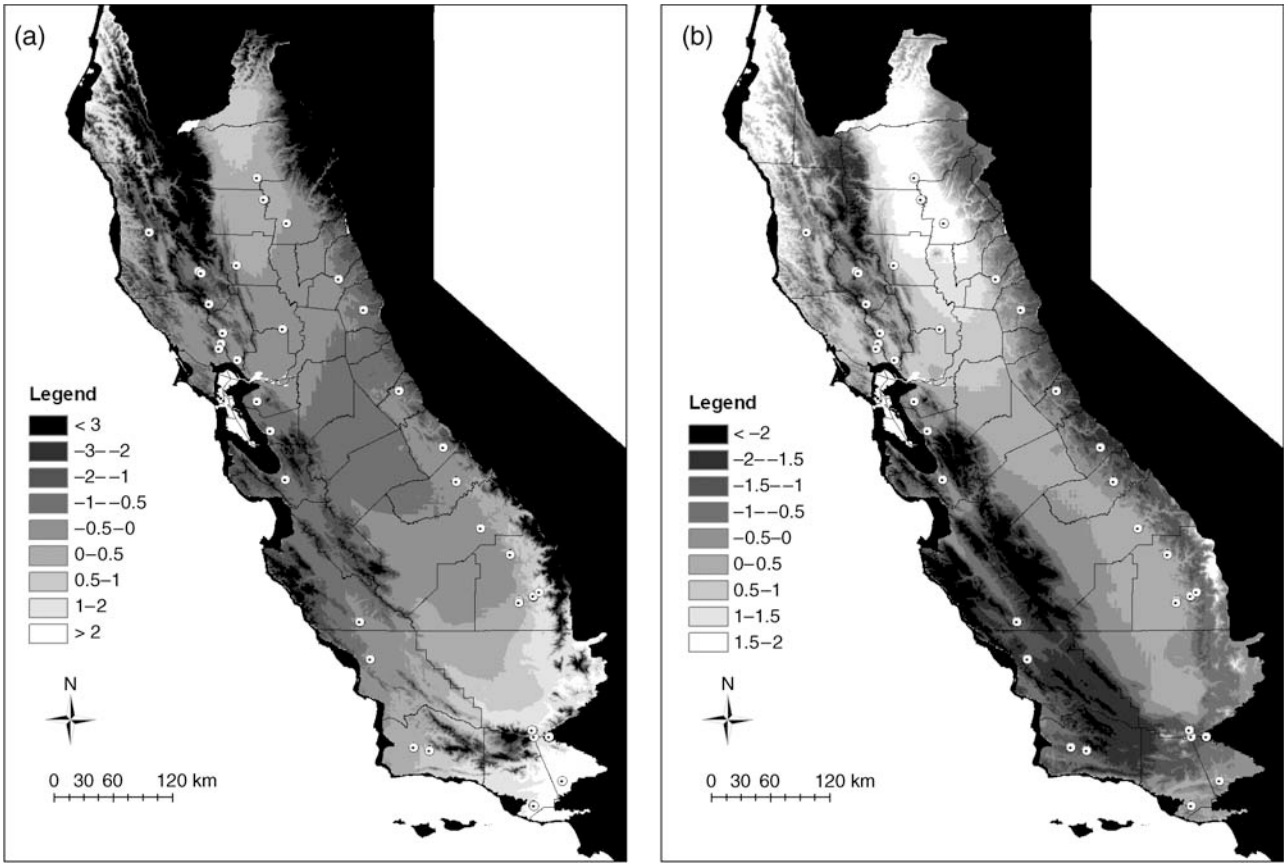

Fig. 3 Spatial trends in chloroplast genetic markers using the population mean standardized canonical scores based on the canonical correlation analysis of multivariate chloroplast genotypes vs. geographical variables. (a) Trend surface of the first canonical axis. (b) Trend surface of the second canonical axes. See Table 2 for statistical results. The shadings indicate the classes of canonical scores (see legends); dark colours represent negative values, and light colours positive values around the population mean, but no biological significance is attached to having positive or negative scores. The boundaries of the contour maps are determined by the Jepson Ecoregions where *Quercus lobata* occurs.

models, the estimates are 36% and 24%, respectively, which indicates a high amount of multivariate structure explained by the spatial variables. In each of these two canonical axes (W1 and W2), the predicted scores are correlated more or less evenly with the various haplotypic variables, with a few variables highly correlated (Table 2b). The correlations for the geographical variables show that latitude and longitude correlate heavily with the first geographical canonical variable W1. The second canonical axis (W2) reflects the contribution of the elevation and the latitude, but the longitude also contributes to this axis (Table 2c). Biologically, the analysis indicates that the chloroplast has a great deal of genetic structure and this structure is spatially organized mainly with north–south, but also with east–west, and elevational gradients (Fig. 3a, b).

The map of the trend surface for the first canonical axis (Fig. 3a) shows the score values, reflecting multilocus genotypes, rising to the north or south and falling to the west with rapid changes in gradients in the western portion of the distribution. We performed a diagnostic analysis in response surface analysis of the first axis surface (using SAS' JMP) that indicates a saddle-shaped surface for fixed elevation with the long axis of the saddle roughly to the northwest and with the centre of the saddle (the stationary point) near the geographical centre of the data. This surface is similar to two-dimensional, anisotropic isolation by distance illustrated in Epperson (2003, p. 25). In other words, the first canonical axis seems to capture population evolutionary history due to isolation by distance. This pattern is illustrated in Fig. 3(a) by the contour of the gradients that show that pairs of points are less likely to be similar as they become more distant. Also, note that contours of the gradients are longer going north–south than east–west, which indicates that gene flow appears more restricted in the east–west south direction.

The second canonical axis for the chloroplast data again shows contours running north–south and east–west with a great deal of complexity in the west (Fig. 3b). The contour profiler in JMP shows a bowl-shaped surface

Table 3 Summary of canonical correlation analysis based on nuclear genetic markers. (a) Summary of the statistical results of canonical correlation analysis for genetic variables versus the spatial variables of latitude, longitude, and elevation for the significant canonical axes. (b) Canonical correlations of the 15 PC axes of the 97 nuclear variables with the predicted scores of the first two canonical axes. (c) Canonical correlations of the linear and higher order combination of the geographical variables for the first two canonical axes

(a) Statistical results

Canonical axis	Canonical correlation	Adjusted canonical correlation	Squared canonical correlation	Eigen value	Approximate *F* value	Numerator d.f.	Denominator d.f.	Pr > *F*
1	0.66	—	0.43	0.76	1.62	135	723.21	<0.0001
2	0.65	—	0.42	0.73	1.38	112	656.78	0.0091

(b) Correlation of allelic variables with predicted scores for the two first canonical axes, W1 and W2

Allelic variables	W1	W2
Prin1	–0.04	0.03
Prin2	–0.10	0.25
Prin3	–0.13	0.14
Prin4	–0.11	–0.10
Prin5	–0.08	–0.10
Prin6	–0.21	0.18
Prin7	–0.23	0.35
Prin8	0.04	0.03
Prin9	0.19	0.19
Prin10	0.24	–0.14
Prin11	0.18	0.25
Prin12	0.15	–0.04
Prin13	0.12	0.04
Prin14	–0.23	–0.17
Prin15	0.27	–0.08

(c) Correlation of geographical variables with predicted scores for the two first canonical axes, W1 and W2

Longitude (X)	0.27	0.67
Latitude (Y)	0.16	–0.76
Elevation (Z)	–0.24	0.37
X*X	–0.48	–0.66
Y*Y	–0.22	0.86
Z*Z	–0.18	0.33
X*Y	0.04	–0.20
X*Z	0.13	0.59
Y*Z	0.17	–0.59

for the second model, tipped so that values increase to the east and the north. However, values decrease with increasing elevation, so the surface appears more like a rising ridge, with the ridge top orientated along the Central Valley. For this model, the stationary point is to the west of the California Central Coast, 1400 m above the sea surface (for details on this method of analysis, see Box & Draper 1987, pp. 323–380). A simple interpretation of this pattern would be greater restriction in gene flow in the northeast direction and greater restriction relative to the pattern in the first canonical axis.

In sum, the trend surface maps for the first and second canonical axes illustrate that the distribution of chloroplast markers is highly structured across the species range (Fig. 3a, b). The western part of the species range shows more north–south heterogeneity than the eastern part along the Sierra Nevada foothills. The map also indicates steeper genetic heterogeneity in the southern end of the range through the San Francisco Bay area, and in the northwestern corner of the range. Finally, the maps illustrate that multivariate scores differ greatly between the west and east, suggesting restricted gene flow between those two regions.

Nuclear data. The canonical correlation model for the nuclear data is significant, but it accounts for less of the total variation than the chloroplast results. Only the first and second canonical correlations are significant (Table 3a); each explains about 29% of the fitted models and about 20% of the total allelic variation represented in the PCs. Because the 15 PCs account for about half of the total genetic variation, the two canonical models account for only ~10% of the total variation. The correlations for the genetic variables with the predicted canonical scores (W1, W2) show that the first and second canonical variables include contributions from almost all of the 15 contributing principal variables (Table 3b), which means that the axes truly represent multivariate genotypes.

In this model, the first geographical canonical axis (W1) is weakly correlated with the linear terms of the three dimensions and more strongly correlated with the nonlinear term for longitude (Table 3c). The mapped spatial trends illustrate steeper gradients east–west than north–south (Fig. 4a). The response surface analysis indicates a bowl-shaped surface, with the base of the valley following the California coast; the stationary point is near the Mexican border, 80 m below the earth's surface. These data are consistent with an isolation-by-distance process, but we must point out that sometimes a cline from one part of the plot to another is consistent with a

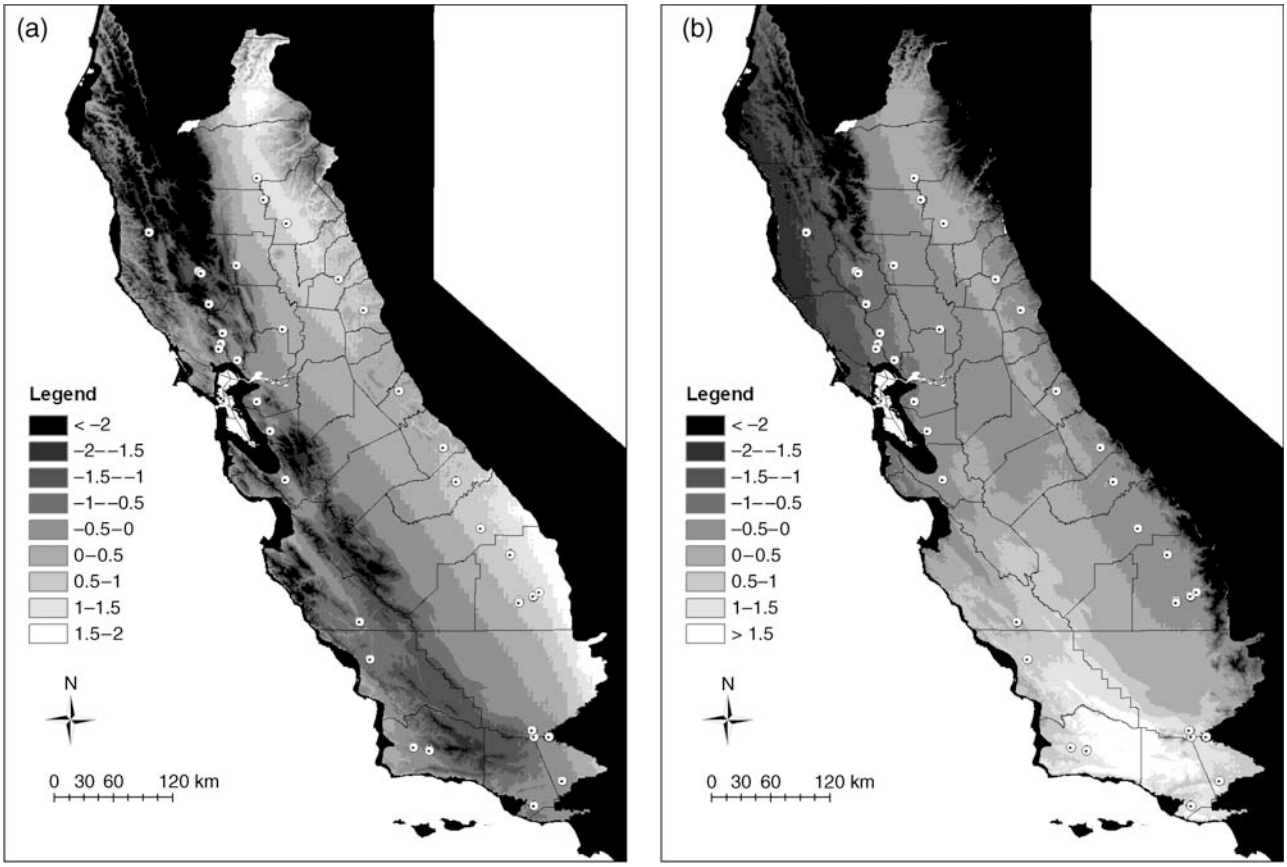

Fig. 4 Spatial trends in nuclear genetic markers using the population mean standardized canonical scores based on the canonical correlation analysis of multivariate chloroplast genotypes vs. geographical variables. (a) Trend surface of the first canonical axis. (b) Trend surface of the second canonical axes. See legend of Fig. 3 for more details.

selection gradient caused by an environmental gradient. This association is possible for putative neutral markers because we are examining the loci as multivariate genotypes that can accumulate small effects across loci because of selection on correlated loci associated with quantitative traits.

Latitude and longitude correlate strongly with the second geographical canonical variable W2 (Table 3c). The visualization of the standardized scores shows a saddle shape, centred in the Sierran foothills, northeast of San Francisco, with scores increasing to the north and south and decreasing to the east and west (Fig. 4b). Thus, this response surface for the nuclear data indicates anisotropic isolation by distance, although to a lesser extent compared to the chloroplast variables. Again changes in scores are greater west to east across California and the structure among populations appears lower for the nuclear than for the chloroplast genomes.

The genetic map of the first canonical vector based on the nuclear data shows that the western region shows much more genetic heterogeneity than the Sierran foothill populations (Fig. 4a). Both maps indicate that the western and eastern populations differ genetically at the north and south ends and that the populations in the south differ from those in the north (Fig. 4a, b).

Geographical patterns of allelic diversity

Total richness for chloroplast markers ranged from 2 to 5 haplotypes per locus and from 6 to 12 total alleles per population (mean = 7); and for nuclear makers, observed values of richness ranged from 7 to 23 alleles per locus and from 15 to 28 total alleles per population (mean = 22). The geographical trend for the chloroplast data shows two areas of high allelic richness, one in the central Coast Ranges around the San Francisco Bay area towards the Sacramento Valley and the other in the southern Coast Ranges (Fig. 5a). For nuclear markers, populations in the northern region around the Bay area and further east to the Sacramento Valley show high allelic richness. Populations in the entire southern part of the distribution also contain high allelic richness (Fig. 5b).

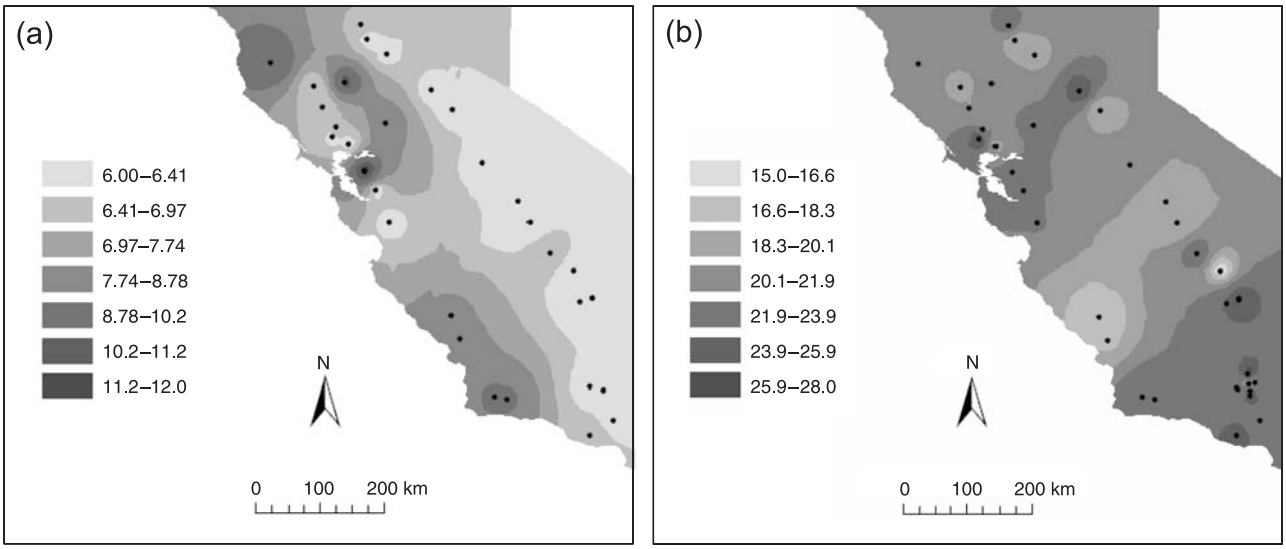

Fig. 5 Spatial locations and range of total allelic richness per population summarized across (a) six chloroplast and (b) seven nuclear microsatellite markers. The legend indicates the mean number of alleles per locus for the shadings, with light colours representing low values and dark colours high values. The boundaries of the map are set to enclose the range of *Quercus lobata*.

Discussion

California valley oak manifests pronounced genetic variation across the species' range. Some regions show sharp gradients in chloroplast and nuclear genetic data and these regions are also associated with high allelic richness. These same areas have experienced significant habitat loss through landscape change due to agricultural and urban development. The pattern of the genotypic gradients in both chloroplast and nuclear genome provides strong evidence that the biogeography of valley oak has been strongly shaped by isolation by distance in gene flow. The chloroplast and nuclear genomes display similar patterns, indicative of common evolutionary forces, but the higher genetic structure in the chloroplast genome suggests that the history of seed movement has established an underlying genetic structure that is present in contemporary nuclear genetic data. We will first comment below on the evolutionary and conservation significance of our findings and then discuss the implications for the human threats to the distribution of valley oak because of landscape and global climate changes.

Genetic structure and gene flow in valley oak

The analyses of molecular variance, autocorrelation, and the canonical trend surface analyses indicate greater genetic structure for the chloroplast genome than for the nuclear genome. This result is consistent with the maternal inheritance of chloroplast DNA vs. biparental inheritance of nuclear DNA. Gene flow is mediated by seed dispersal for the former as opposed to both pollen and seed for the latter (Crawford 1984; Petit *et al*. 1993a; Ennos 1994). The genetic variation of the two genomes can be used to compute the ratio of pollen to seed migration under an island model of migration (Petit 1992; Ennos 1994). We calculate a ratio of ~60 for *Quercus lobata*, which indicates significantly greater gene flow through pollen than seeds. Thus, pollen movement homogenizes the gene pool of valley oak to a large extent, but significant structure remains across the range, as shown by the spatial genetic trend detected in this study (cf. following section). It is a bit surprising that the pollen–seed ratio in valley oak is so much smaller than the values of ~200–500 reported for European oaks (Ennos 1994; El Mousadik & Petit 1996). This difference is due to the lower estimates of chloroplast genetic differentiation for valley oaks (G_{ST} = 0.8) vs. European oaks (G_{ST} = 0.9; Petit *et al*. 1993b), combined with higher nuclear genetic differentiation for valley oaks (G_{ST} = 0.05) vs. European oaks (G_{ST} = 0.02 for *Quercus petraea*; G_{ST} = 0.03 for *Quercus robur*; Zanetto *et al*. 1994). Biologically, this difference in chloroplast structure may be explained by a rapid Holocene expansion of European oaks through seed colonization that resulted in populations mainly fixed for one haplotype. In contrast, the California populations' expansion was more restricted, despite the higher chloroplast diversity within populations (Grivet *et al*. 2006). The higher chloroplast diversity could be due to the fact that California oak populations, especially in the west, have been present for a longer period of time than the European ones. Thus, the difference in the nuclear structure between Europe and California

might be due to the impact of gene flow on relatively young, expanding populations in Europe coming largely from the southern refugia vs. older, widely distributed populations in California that have been alternately contracting and coalescing over time.

We recognize that the differences in pollen to seed ratios between European and California populations may be confounded by differences in methodology, because of different genetic markers [in the European studies PCR-RFLP (restriction fragment length polymorphism) markers and isozymes were used to compute the chloroplast and nuclear estimates, respectively, rather than microsatellites) and to slightly different statistical approaches used to compute the genetic parameters. Although homoplasy in microsatellite markers can sometimes create bias, and could cause increased spatial autocorrelation with distance, we found the opposite, suggesting no such problem with our data. Without redoing the studies all with the same markers, we cannot remove the concern about different markers or formulas. Moreover, we are unable to statistically compare the European vs. California estimates of G_{ST}. However, the differences in the distribution of cp genotypes and diversity found in California and Europe that we report elsewhere (Grivet *et al*. 2006) suggests that the populations in the two regions have experienced different kinds of evolutionary history that would lead to the differences in pollen to seed ratios we have observed.

We have another potential sampling problem in this study, which is high genetic diversity of the microsatellite nuclear markers. To see whether we had an appropriate sample size per population, we conducted a post-hoc analysis of our sample design. This analysis uses the between- and within-population variances estimated from the AMOVA to estimate optimal sample allocations (cf. Materials and methods section, Sampling paragraph). By dividing the expectation of mean squares for populations in the AMOVA by the product of the number of populations and number of trees sampled per populations, we get the total mean square error. Then by iterating over populations and trees per population, we can find the combinations of population number and trees per population that give us the asymptotic minimum for this error. For the chloroplast data, one tree per population would have been sufficient for the same number of populations sampled in this study. In contrast, for the nuclear data, five trees per population would have been optimal for the number of populations sampled. The under-sampling of individuals for our analysis of nuclear genetic structure would increase the variance in our estimate and reduce our sensitivity to detection of rare alleles, but it should not bias the estimate of genetic structure. Nonetheless, future work with more individuals per site and more sites is the best way to validate the estimates of pollen–ovule ratios.

Geographical trends in chloroplast and nuclear genomes

Based on the geological history of California, we subdivided our populations into different subregions and then assessed the extent to which patterns of gene flow were similar or divergent among geographical regions. We tried two methods of geographical partitioning to identify alternative groupings: STRUCTURE (Pritchard *et al*. 2000), and BARRIER (Manni *et al*. 2004). Results for both markers and both analyses detected multiple genetic groups, but they did not identify any specific clusters according to major geographical regions. So, we did not use those statistical approaches, but instead used geographical groupings, and found contrasting patterns in populations in the west and the east. In the eastern samples from the Sierran foothills, data from the cpDNA markers suggest much more restricted seed movement (G_{ST} = 0.968) than found in the western samples from the coastal range foothills (G_{ST} = 0.606). The difference in the nuclear markers is far less dramatic, but again, the data indicate less gene flow among Sierran populations (G_{ST} = 0.065 for the eastern populations vs. G_{ST} = 0.047 for the western populations). We cannot be sure whether such differences are due to topographically related differences that have altered gene movement through pollen and seed differently in the east and west, or to climatological differences that have created more dramatic demographic contractions and genetic bottlenecks in the east than in the west.

The patterns of multivariate genotypes and allelic richness data identify pronounced latitudinal trends and, particularly highlight the southern part of the range and the central Coast Ranges as important centres of genetic diversity (i.e. sharp genetic gradients and high allelic richness). These may be regions where valley oak populations persisted through Pleistocene range expansions and contractions receiving cross currents of gene flow, particularly along the Coast Ranges, during expansions, in a manner similar to that in Petit *et al*. (2003). The genetic similarity of these two regions suggests dynamic movements of valley oak across its entire range, with the capacity to maintain local structure. An alternative explanation would be that the observed genetic spatial trend reflects the influence of selection pressures associated with environmental change. However, chloroplast DNA normally do not represent adaptive variation, and the nuclear trend is weaker as one would expect for pollen movement rather than stronger, which is what one might expect if selection were enhancing the structure due to colonization. It is very possible that adaptive variation might indicate a latitudinal gradient, but the detection of such a pattern would be more reliable with markers associated with adaptive traits.

The distribution of haplotypes and the canonical analyses provide valuable insight about the evolutionary history of valley oak in California. Using these chloroplast data,

Grivet *et al*. (2006) showed that the Sierran foothill populations are usually fixed for different haplotypes, while those in the coastal areas showed admixture of haplotypes, indicating that seed movement was more fluent among coastal populations. Two haplotypes in the northern Sierran populations are found in the San Francisco Bay Area populations indicating that seed colonization may have occurred through the Sacramento River watershed between the Sierran foothills and the Bay Area. The canonical surfaces show a similar, though more complex pattern. Latitudinal pairs of Sierran and Coast Range populations in the central portion of the range are rather similar, falling within the same contour interval in both canonical patterns, in contrast to steep gradients through the Transverse Ranges and between Coast Range and Sierran populations in the north. This process is part of the reason for the high genetic heterogeneity in the San Francisco Bay Area. Surprisingly, the two canonical patterns in the nuclear markers showed greater restriction to gene flow between the Sacramento and Bay area populations, which could indicate that ongoing pollen exchange subsequent to colonization has not been extensive. All four canonical patterns indicate isolation between the North Coast Range and the Sacramento River populations. Similarly, all four canonical patterns showed greater isolation across the Kern River/Kaweah Gap between the Tehachapi and southern Sierran populations than between the former and the southern Coast Range populations. Thus, valley oak populations may have arisen from these regions of discontinuities so that they now differ from each other dramatically in genetic composition.

The geological and climatic history of California should be taken into account when considering the evolutionary history of valley oak. From late Pliocene to the Pleistocene, California experienced consecutive changes in climate with cold and wet conditions during the glacial ages and warm and dry conditions during the interglacial (Thompson *et al*. 1993). The topography also changed during that period because of the uplift of the Coastal and Sierra Nevada ranges (Harden 1997). The vegetation assemblages fluctuated from low densities for oaks in glacial periods to expansion during the interglacial periods (Pisias *et al*. 2001), and these fluctuations led to the establishment of the current oak populations (Raven & Axelrod 1978). The upper elevation of the Sierra Nevada were subject to several glacial advances and retreats (Benson *et al*. 1996) and the presence of mountain glaciers was probably not favourable to oaks on the foothills of these mountain ranges. The arid flats of the Central Valley were not optimal during the interglacial periods for oaks either, whereas the middle elevations and proximity to the ocean of the Coastal Ranges constituted a more favourable environment for oak population to develop. These past conditions would have favoured the continued historical presence of oak populations along the coast, although the data also suggest persistence of Sierran foothill populations (e.g. Millar & Woolfenden 1999). The foothills would be connected to the west through the Californian riparian corridors, where the greater animal movements and suitable valley oak habitat would facilitate the migration of the oaks throughout the state. Populations in the northern areas would have been connected to each other through a rich river system (i.e. Sacramento and San Joaquin Rivers). This riparian network would have allowed the connection of coastal and Sierran populations, as shown by the shared chloroplast haplotypes found between eastern and western populations north of the Monterey Bay area (Grivet *et al*. 2006) and supported by the canonical trends. The southern populations would have been part of a distinct group, since they are located in one of the driest part of the species range, thus limiting oak movement. The propagation of oaks across the state would have occurred through successive localized waves of migration, illustrated by the rapid surge in oak pollen in Sierran lakes at 1600 m (Anderson & Smith 1994) and in Tulare Lake (Davis 1999), peaking during the Mid-Holocene Warm, 8000 years BP. These dynamics are suggested by the pattern of isolation by distance detected in our analyses. Thus, valley oak response to past climatic events, although complex in timing and space, would have resulted in a genetic footprint structured enough to be detected through our analyses.

Our interpretation is consistent with the patterns of other California plant and animal taxa that show north–south split near the Transverse Ranges and east–west phylogeographical breaks between the Sierra Nevada and the Coastal Ranges (Calsbeek *et al*. 2003; Lapointe & Rissler 2005; Rissler *et al*. 2006). Comparative phylogeographical studies revealed some correlation of phylogenetic groups with climatic variables, as well as correlation between molecular divergence dates and the uplifts of the principal mountain ranges in California (Calsbeek *et al*. 2003; Rissler *et al*. 2006). It appears therefore that some common events that affected the genetic pattern of other species in California may have also affected the oaks; however, additional analyses beyond the scope of this study will need to be conducted to assess the extent of this common impact.

The utility of canonical trend surface analysis (CTSA)

The application of multivariate analyses to genetic data is not new, and many people simply map geographical patterns of the principal component axes (Cavalli-Sforza *et al*. 1993; Le Corre *et al*. 1998; Heckel *et al*. 2005). However, the advantage of a CTSA approach is that we can test hypotheses about the relationship between spatial and genetic variables. The null hypothesis in CTSA is that populations are arrayed in a regular simplex (Dyer & Nason 2004). Another way of stating this hypothesis is that

the correlation between any linear function of allelic frequencies (as in quantitative genetic models; e.g. Kempthorne 1969) and any linear combination of geographical location is zero. We have shown that the null hypothesis can be rejected in our data and that a portion of the sampled genome has geographical structure. A relevant question is whether populations are instead clustered in discrete groups in the form of a series of stepped clines and the canonical model simply passes through these. In our data, though, examinations of residuals indicate this pattern is not the case. Indeed, CTSA also behaves well in rugged landscapes, in comparison to principal components (Wartenberg 1985). Our experience indicates that the genetic landscapes (or surfaces) described in the first one or two canonical vectors are fairly regular, whereas those in subsequent vectors are rugged, even if statistically significant. Finally, we should point out that codominant markers in CTSA behave well and according to statistical assumptions (Westfall & Conkle 1992). However, we caution that dominant markers where the alternative allele is a null, as with random amplified polymorphic DNAs (RAPD), are not well suited to CTSA analysis. Thus, for certain genetic markers found in populations sampled throughout the range of a species, CTSA offers a sensitive approach to the identification of genetic patterns that can be used to detect regions of evolutionary and conservation interest.

When should an evolutionary conservation approach be used?

The complex genetic structure of valley oak suggests a strategic approach for conserving genetic diversity in this widely distributed species. This species is an ecologically significant California endemic that has experienced extensive loss of populations throughout its range and diminished population recruitment in many local populations. Additional habitat loss and fragmentation are inevitable given ongoing development pressure and vineyard expansion, so range-wide conservation planning is both timely and appropriate.

In a separate study, we applied the data reported here to a maximum coverage (MC) approach that identifies a reserve network design with the locations of a minimum number of sites that would maximize the coverage of greatest number of chloroplast and nuclear alleles for individual loci across the fewest number of sites (Sork *et al*. in press). We conducted that analysis separately for nuclear and chloroplast loci and also with a model combining the markers. Interestingly, that approach yields similar recommendations for reserve locations to those we report here, but there are also some notable differences. On one hand, findings of both studies underscore the importance of preserving populations in the south. This conclusion is particularly important because the southern part of valley oak range is under significant population pressure for development. On the other, the MC approach identified the Bay region as being an area of interest only for the analysis based on chloroplast markers. In fact, the model that combines the cp and nuclear DNA markers overlooked that region. Thus, a maximum coverage model can provide a useful framework for selecting sites, but it might be necessary to include specific evolutionary criteria as weighting factors when deciding which local sites to preserve.

Given the cost and effort required to produce evolutionary information relevant to conservation planning, is the effort justified? Are there shortcuts? Will conservation plans based on geography (e.g. ecoregional representation) and ecological diversity (e.g. representation of the range of habitats where the species occurs) suffice to conserve the evolutionary potential of widespread, genetically diverse species like valley oak? Our study focuses on putatively neutral markers, and it would be of great interest to assess geographical trends in adaptive markers. A recent simulation study in European oaks compared the genetic variability at neutral markers and quantitative trait loci in a subdivided population under selection (Le Corre & Kremer 2003) and showed that, for different levels of gene flow and different types of selection, the geographical structure can differ markedly between neutral genes and those associated with quantitative trait loci (QTL) (see also Reed & Frankham 2001). This lack of association highlights the fact that, depending on the marker used, different phenomenon will be revealed. Theoretically, neutral loci across the genome should be affected similarly by demography and evolutionary effects of populations, while loci under selection will often behave differently because they are connected to traits important for fitness and adaptation (Luikart *et al*. 2003; Toro & Caballero 2005). If we can identify genes associated with adaptive traits that will be useful for future population survival, then evolutionary conservation strategies might benefit from identification of the spatial distribution of those genes. Additional studies comparing conservation planning outcomes with and without genetic information would be of immediate practical value to answer these questions. Given the current state on knowledge, identification of areas with high genotypic variation and unique genotypes provides a first step towards incorporation of genetic information into reserve network design.

Conserving the evolutionary potential of valley oak: policy recommendations

Based on the genetic evidence presented here, we infer that valley oak has a complex evolutionary history in California and that, using the criteria outlined by Rayburn and Moritz at a recent meeting of the Society for Conservation

Biology (2006), some regions within the species' range have especially high conservation value. Two areas that stand out in our study — the Bay area and the southern area near the Transverse Ranges and southern Coastal Ranges — have been identified through other studies as areas of interest (e.g. Calsbeek *et al*. 2003; Lapointe & Rissler 2005; Rissler *et al*. 2006). In addition, these two areas constitute important repositories of plant species endemism (Coastal Ranges, Stebbins & Major 1965) and plant species richness (central Coast Ranges and southwest California, Richerson & Lum 1980). Conversion to vineyards and rural residential development are significant threats to remaining valley oak woodlands (Fig. 1), and conservation nongovernmental organizations (NGO) such as The Nature Conservancy have already committed large resources to oak woodland protection in these regions. Local and state oak conservation policies are also helping to steer development away from or at least partially mitigate development in oak woodlands (see the California Oak Foundation's information at www.californiaoaks.org/index.html).

The preservation of evolutionary potential is especially important in the face of rapid climate change. Kueppers *et al*. (2005) analysed the implications of climate change for the distribution of valley oak. Based on a regional climate change model, they predict that the range of valley oak will shrink to about 54% of its current distribution. In southern California, their model shows that much of the suitable habitat will be altered by climatic conditions (aside from landscape change) and that populations will expand upward in elevation, if dispersal, colonization, and competition with conifer species permit. If valley oak is to persist either in its present range, or shift its distribution northward and upward, contemporary populations and their genetic composition will be needed to retain genetic variation that will allow an adaptive response to these changing environmental conditions.

The preliminary analyses presented here indicate that land-use change threatens to reduce the valley oak genetic diversity and evolutionary potential, and that genetic analyses will be important in developing a statewide strategy for preserving this potential. However, it is not possible to recommend a comprehensive reserve design at this time. We need to sample (i) populations in the geographical regions that are under represented in the current study (e.g. south of the Bay area); (ii) populations with steep genetic gradients that would benefit from further refinement; and (iii) populations in oak habitats with few samples (especially riparian corridors that are so important to oak movement and represent a major oak ecosystem that has been seriously damaged). Factors such as degree of threat, cost of acquisition, and availability of preserved sites that include comparable genetic composition should be addressed in any reserve design plan (e.g. Davis *et al*. 2006).

Conclusions

California valley oak has experienced a significant reduction in its abundance over the last 200+ years since European settlement. The landscape transformation due to agricultural land use has reduced the abundance of valley oak, sometimes leaving the riparian corridors intact and thus allowing oak movement. The more recent increase in land-use change poses a greater risk to the valley oak because corridors are also disappearing. The combined spatial analysis of chloroplast and nuclear genotypes has identified several regions of California of high evolutionary interest and these regions happen to be active regions of landscape alteration. Moreover, the populations in the southern part of the range are valuable sources of genetic variation for range shifts due to global climate change. A comprehensive reserve plan needs to be designed in the near future for this keystone tree species or California may lose this precious ecosystem in many parts of the state.

Acknowledgements

We are grateful to Jeanette Papp and Uma Dandekar (from the UCLA Sequencing and Genotyping Core Facility) for handling the nuclear genotypic data, David K. Jacobs for information on California history, Valérie Le Corre, Rodney J. Dyer, Thomas W. Gillespie, John Novembre, and Peter E. Smouse for comments on this manuscript as well as two anonymous reviewers for constructive comments on an earlier version of the manuscript. V.L.S. and D.G. were supported by funds from UCLA and from NSF DEB-0089445. F.W.D. was supported by NSF HSD-0527286.

References

Anderson RS, Smith SJ (1994) Paleoclimatic interpretations of meadow sediment and pollen stratigraphies from California. *Geology*, **22**, 723–726.

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

Benson LV, Burdett JW, Kashgarian M *et al*. (1996) Climatic and hydrologic oscillations in the Owens Lake basin and adjacent Sierra Nevada. *California. Science*, **274**, 746–749.

Box GEP, Draper NR (1987) *Empirical Model-Building and Response Surfaces*. John Wiley & Sons, New York.

Brown RW, Davis FW (1991) Historical mortality of valley oak in the Santa Ynez Valley, Santa Barbara County, CA. In: *Proceedings of the Symposium on Oak Woodlands and Hardwood Rangeland Management. General Technical Report PSW-126* (ed. Standiford R), pp. 202–207. Pacific Southwest Research Station, Forest Service. US Department of Agriculture, Albany, California.

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.

Cavalli-Sforza L, Menozzi P, Piazza A (1993) Demic expansions and human-evolution. *Science*, **259**, 639–646.

Conkle MT, Westfall RD (1984) Evaluating breeding zones for ponderosa pine in California. In: *Progeny testing. Proceedings of*

the Servicewide Genetics Workshop, pp. 89–98. USDA Forest Service, Timber Management, Charleston, South Carolina.

Couvet D (2002) Deleterious effects of restricted gene flow in fragmented populations. *Conservation Biology*, **16**, 369–376.

Crawford TJ (1984) The estimation of neighbourhood parameters for plant populations. *Heredity*, **52**, 273–283.

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 Paleobotany and Paleoecology*, **107**, 249–257.

Davis FW, Costello C, Strom D (2006) Efficient conservation in a utility-maximization framework. *Ecology and Society*, **11**, 33. [online] URL: http://www.ecologyandsociety.org/vol11/iss1/art33/.

Davis FW, Kuhn W, Alagona P, Campopiano M, Brown R (2000) *Santa Barbara County Oak Woodland Inventory and Monitoring Program. Final report to the County of Santa Barbara department of planning and development*. University of California, Santa Barbara, California.

Deguilloux MF, Dumolin-Lapegue 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.

Dow BD, Ashley MV, Howe HF (1995) Characterization of highly variable (GA/CT) (n) microsatellites in the Bur Oak, *Quercus macrocarpa. Theoretical and Applied Genetics*, **91**, 137–141.

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, Nason JD (2004) Population graphs: The graph-theoretic shape of genetics structure. *Molecular Ecology*, **13**, 1713–1728.

El Mousadik A, Petit RJ (1996) Chloroplast DNA phylogeography of the argan tree of Morocco. *Molecular Ecology*, **5**, 547–555.

Ennos RA (1994) Estimating the relative rates of pollen and seed migration among plant populations. *Heredity*, **72**, 250–259.

Epperson BK (1990) Spatial autocorrelation of genotypes under directional selection. *Genetics*, **124**, 757–771.

Epperson BK (2003) *Geographical Genetics*. Princeton University Press, Princeton, New Jersey.

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

Frankham R, Ballou JD, Briscoe DA (2002) *Introduction to Conservation Genetics*. Cambridge University Press, Cambridge, UK.

Goudet J (2001) *FSTAT, a Program to Estimate and Test Gene Diversities and Fixation Indices* (version 2.9.3). Available from http://www2.unil.ch/popgen/softwares/fstat.htm.

Gram WK, Sork VL (1999) Population density as a predictor of genetic variation for woody plant species. *Conservation Biology*, **13**, 1079–1087.

Gram WK, Sork VL (2001) Association between environmental and genetic heterogeneity in forest tree populations. *Ecology*, **82**, 2012–2021.

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 MF, Petit RJ, Sork VL (2006) Contrasting patterns of historical colonization in white oaks (*Quercus* spp.) in California and Europe. *Molecular Ecology*, **15**, 4085–4093.

Hamrick JL, Godt MJW (1996) Effects of life history traits on genetic diversity in plant species. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **351**, 1291–1298.

Hamrick JL, Godt MJW, Sherman-Broyles SL (1992) Factors influencing levels of genetic diversity in woody plant species. *New Forests*, **6**, 95–124.

Harden DR (1997) *California Geology*. Prentice Hall, Inc, Upper Saddle River, New Jersey.

Hardy OJ, Vekemans X (2002) SPAGEDI: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes*, **2**, 618–620.

Heckel G, Burri R, Fink S, Desmet JF, Excoffier L (2005) Genetic structure and colonization processes in European populations of the common vole, Microtus arvalis. *Evolution*, **59**, 2231–2242.

Hickman JC (1993) *The Jepson Manual of Higher Plants of California*. University of California Press, Berkeley, California.

Kampfer S, Lexer C, Glossl J, Steinkellner H (1998) Characterization of (GA) (n) microsatellite loci from *Quercus robur*. *Hereditas*, **129**, 183–186.

Kelly PA, Phillips SE, Williams DF (2005) Documenting ecological changes 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). UC Publication in Zoology, Berkeley, California.

Kempthorne O (1969) *An Introduction to Genetic Statistics*. The Iowa State University Press, Aimes, Iowa.

Kinloch BB, Westfall RD, Forrest GI (1986) Caledonian Scots pine: origins and genetic structure. *New Phytologist*, **104**, 729.

Kremer A, Zanetto A (1997) Geographical structure of gene diversity in *Quercus petraea* (Matt) Liebl. 2. Multilocus patterns of variation. *Heredity*, **78**, 476–489.

Kremer A, Zanetto A, Ducousso A (1997) Multilocus and multitrait measures of differentiation for gene markers and phenotypic traits. *Genetics*, **145**, 1229–1241.

Kueppers LM, Snyder MA, Sloan LC, Zavaleta ES, Fulfrost B (2005) Modeled regional climate change and California endemic oak ranges. *Proceedings of the National Academy of Sciences, USA*, **102**, 16281–16286.

Lapointe FJ, Rissler LJ (2005) Notes and comments — congruence, consensus, and the comparative phylogeography of codistributed species in California. *American Naturalist*, **166**, 290–299.

Le Corre V, Kremer A (2003) Genetic variability at neutral markers, quantitative trait loci and trait in a subdivided population under selection. *Genetics*, **164**, 1205–1219.

Le Corre V, Roussel G, Zanetto A, Kremer A (1998) Geographical structure of gene diversity in *Quercus petraea* (Matt.) Liebl. III. Patterns of variation identified by geostatistical analyses. *Heredity*, **80**, 464–473.

Leberg PL (2002) Estimating allelic richness: effects of sample size and bottlenecks. *Molecular Ecology*, **11**, 2445–2449.

Ledig FT, Kitzmiller JH (1992) Genetic strategies for reforestation in the face of global climate change. *Forest Ecology and Management*, **50**, 153–169.

Lee PJ (1969) The theory and application of canonical trend surfaces. *Journal of Geology*, **77**, 303–318.

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

Li P, Adams WT (1989) Range-wide patterns of allozyme variation in Douglas-fir (Pseudotsuga menziesii). *Canadian Journal of Forest Research*, **19**, 149–161.

Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2003) The power and promise of population genomics: from genotyping to genome typing. *Nature Reviews Genetics*, **4**, 981–994.

Manni F, Guérard E, Heyer E (2004) Geographic patterns of (genetic, morphologic, linguistic) variation: how barriers can be detected by 'Monmonier's algorithm'. *Human Biology*, **76**, 173–190.

Millar CI, Woolfenden WB (1999) Sierra Nevada forests: where did they come from? Where are they going?. In: *Natural Resources Management: Perceptions and Realities. Trans. 64th North American Wildlife and Natural Resources Conference, San Francisco, March 26–30, 1999* (eds McCabe R, Loos S), pp. 206–236. Wildlife Management Institute, Washington, DC.

Moritz C (1995) Uses of molecular phylogenies for conservation. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **349**, 113–118.

Moritz C, Faith DP (1998) Comparative phylogeography and the identification of genetically divergent areas for conservation. *Molecular Ecology*, **7**, 419–429.

Pavlik BM, Muick PC, Johnson SG, Popper M (1991) *Oaks of California*. California Press and the California Oak Foundation, Los Olivos, California.

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

Petit RJ (1992) *Polymorphisme de l'ADN chloroplastique dans un complexe d'espèces: les chênes blancs européens. Approche de génétique de populations*. PhD diss., Paris XI.

Petit RJ, Kremer A, Wagner DB (1993a) Finite island model for organelle and nuclear genes in plants. *Heredity*, **71**, 630–641.

Petit RJ, Kremer A, Wagner DB (1993b) Geographic structure of chloroplast DNA polymorphisms in European oaks. *Theoretical and Applied Genetics*, **87**, 122–128.

Petit RJ, Aguinagalde I, de Beaulieu JL *et al*. (2003) Glacial refugia: hotspots but not melting pots of genetic diversity. *Science*, **300**, 1563–1565.

Pisias NG, Mix AC, Heusser L (2001) Millennial scale climate variability of the northeast Pacific Ocean and northwest North America based on radiolaria and pollen. *Quaternary Science Reviews*, **20**, 1561–1576.

Pons O, Petit RJ (1996) Measuring and testing genetic differentiation with ordered versus unordered alleles. *Genetics*, **144**, 1237–1245.

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

Rajora OP, Dancik B (1999) Population genetic variation, structure, and evolution in Engelmann spruce, white spruce, and their natural hybrid complex in Alberta. *Canadian Journal of Botany*, **78**, 768–780.

Raven PH, Axelrod DI (1978) *Origin and Relationship of the California Flora*. University of California, Publications in Botany, Sacramento, California.

Rayburn R, Moritz C (2006) Using evolutionary hotspots to identify important areas for conservation in California. California State Department of Parks and Recreation, Sacramento, California.

Reed DH, Frankham R (2001) How closely correlated are molecular and quantitative measures of genetic variation? A meta-analysis. *Evolution*, **55**, 1095–1103.

Richerson PJ, Lum K-L (1980) Patterns of plant species diversity in California: relation to weather and topography. *The American Naturalist*, **116**, 504–536.

Rissler LJ, Hijmans RJ, Graham CH, Moritz C, Wake DB (2006) Phylogeographic lineages and species comparisons in conservation analyses: a case study of California herpetofauna. *American Naturalist*, **167**, 655–666.

Ritland K (1990) A series of FORTRAN computer programs for estimating plant mating systems. *Journal of Heredity*, **81**, 235–237.

SAS Institute (1989) SAS/STAT *User's Guide*, Version 6, 4th edn, Vol. 1. Cary, North Carolina.

Schwartz MK, Luikart G, Waples RS (2007) Genetic monitoring as a promising tool for conservation and management. *Trends in Ecology & Evolution*, **22**, 25–33.

Smouse PE, Peakall R (1999) Spatial autocorrelation analysis of individual multiallele and multilocus genetic structure. *Heredity*, **82**, 561–573.

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.

Sokal RR, Rohlf FJ (1981) *Biometry*, 2nd edn. Freeman & Co, San Francisco, California.

Sokal RR, Wartenberg DE (1983) A test of spatial auto-correlation analysis using an isolation-by-distance model. *Genetics*, **105**, 219–237.

Sork V, Davis F, Smouse P *et al*. (2002a) Pollen movement in declining populations of California Valley oak, *Quercus lobata*: where have the fathers gone? *Molecular Ecology*, **11**, 1657–1668.

Sork VL, Davis FW, Dyer RJ, Smouse PE (2002b) Mating patterns in a savanna population of valley oak (*Quercus Lobata* Née*). USDA Forest Service General Technical Report* PSW-GTR-184, 427–439.

Sork VL, Davis FW, Grivet D (in press) Incorporating genetic information into conservation planning for California valley oak.

Stebbins GL, Major J (1965) Endemism and speciation in the California flora. *Ecological Monographs*, **35**, 1–35.

Steinkellner H, Fluch S, Turetschek E *et al*. (1997) Identification and characterization of (GA/CT) (n)-microsatellite loci from *Quercus petraea*. *Plant Molecular Biology*, **33**, 1093–1096.

Stockwell CA, Hendry AP, Kinnison MT (2003) Contemporary evolution meets conservation biology. *Trends in Ecology & Evolution*, **18**, 94–101.

Thompson RS, Whitlock C, Bartlein PJ, Harrison SP, Spaulding WG (1993) Climate changes in the western United States since 18 000 years BP. In: *Global Climates Since the Last Glacial Maximum* (eds Wright HJ, Kutzbach J, Webb TI *et al*.), pp. 468–513. University of Minnesota Press, Minneapolis.

Toro MA, Caballero A (2005) Characterization and conservation of genetic diversity in subdivided populations. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **360**, 1367–1378.

Tyler CM, Kuhn B, Davis FW (2006) Demography and recruitment limitations of three oak species in California. *Quarterly Review of Biology*, **81**, 127–152.

Wartenberg D (1985) Canonical trend-surface analysis—a method for describing geographic patterns. *Systematic Zoology*, **34**, 259–279.

Weir BS (1996) *Genetic Data Analysis*, 2nd edn. Sinauer & Associates, Sunderland, Massachusetts.

Weising K, Gardner RC (1999) A set of conserved PCR primers for the analysis of simple sequence repeat polymorphisms in chloroplast genomes of dicotyledonous angiosperms. *Genome*, **42**, 9–19.

Westfall RD, Conkle MT (1992) Allozyme markers in breeding zone designation. *New Forests*, **6**, 279–309.

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.

All four authors have extensive research programs concerning the conservation biology and genetics of forest tree species. D Grivet is interested in the phylogeography, gene flow, genomics and adaptation of temperate forest tree species. VL Sork studies the ecology, evolution and the conservation of temperate and tropical trees especially oaks. The research program of FW Davis includes the ecology, management and conservation of California plant communities. RD Westfall, is a statistical forest geneticist within interests in the historical population dynamics in tree species, particularly conifers.