Evolutionary and demographic history of the Californian scrub white oak species complex: An integrative approach

ARTICLE *in* MOLECULAR ECOLOGY · NOVEMBER 2015

Impact Factor: 6.49 · DOI: 10.1111/mec.13457

4 AUTHORS:

Joaquín Ortego

Estación Biológica de Doñana

71 PUBLICATIONS **885** CITATIONS

SEE PROFILE

Víctor Noguerales

Spanish National Research Council

12 PUBLICATIONS **7** CITATIONS

SEE PROFILE

Paul F. Gugger

University of Maryland Center for Environm…

26 PUBLICATIONS **307** CITATIONS

SEE PROFILE

Victoria L Sork

University of California, Los Angeles

118 PUBLICATIONS **5,169** CITATIONS

SEE PROFILE

MOLECULAR ECOLOGY

Molecular Ecology (2015) **24**, 6188–6208 doi: 10.1111/mec.13457

Evolutionary and demographic history of the Californian scrub white oak species complex: an integrative approach

JOAQUÍN ORTEGO,* VÍCTOR NOGUERALES,* PAUL F. GUGGER†[1] and VICTORIA L. SORK† ‡

*Department of Integrative Ecology, Estación Biológica de Doñana, EBD-CSIC, Avda. Américo Vespucio s/n, E-41092 Seville, Spain, †Department of Ecology and Evolutionary Biology, University of California, Box 957239, Los Angeles, CA 90095, USA, ‡Institute of the Environment and Sustainability, University of California, Box 951496, Los Angeles, CA 90095-1496, USA

Abstract

Understanding the factors promoting species formation is a major task in evolutionary research. Here, we employ an integrative approach to study the evolutionary history of the Californian scrub white oak species complex (genus *Quercus*). To infer the relative importance of geographical isolation and ecological divergence in driving the speciation process, we (i) analysed inter- and intraspecific patterns of genetic differentiation and employed an approximate Bayesian computation (ABC) framework to evaluate different plausible scenarios of species divergence. In a second step, we (ii) linked the inferred divergence pathways with current and past species distribution models (SDMs) and (iii) tested for niche differentiation and phylogenetic niche conservatism across taxa. ABC analyses showed that the most plausible scenario is the one considering the divergence of two main lineages followed by a more recent pulse of speciation. Genotypic data in conjunction with SDMs and niche differentiation analyses support that different factors (geography vs. environment) and modes of speciation (parapatry, allopatry and maybe sympatry) have played a role in the divergence process within this complex. We found no significant relationship between genetic differentiation and niche overlap, which probably reflects niche lability and/or that multiple factors, have contributed to speciation. Our study shows that different mechanisms can drive divergence even among closely related taxa representing early stages of species formation and exemplifies the importance of adopting integrative approaches to get a better understanding of the speciation process.

Keywords: Bayesian inference, genetic structure, hybridization, niche divergence, palaeo-distribution modelling, *Quercus*, species distribution modelling

Received 1 September 2015; revision received 30 October 2015; accepted 3 November 2015

Introduction

Understanding the factors promoting species divergence is a major task in ecological and evolutionary research (Dobzhansky 1937; Mayr 1942; Simpson 1953; Papadopulos *et al.* 2011; Butlin *et al.* 2012). Opportunities for speciation have been linked to geographic isolation (Mayr 1942), environmental heterogeneity (i.e. ecological speciation; Orr & Smith 1998; Schluter 2001; Nosil 2012) and a combination of both, which can result in different mechanisms of speciation with or without complete reproductive isolation among diverging lineages (Orr & Smith 1998; Schluter 2001, 2009; Graham *et al.* 2004). Accordingly, two major modes of speciation have been proposed: (i) allopatric/parapatric speciation, which leads to population divergence in total or partial geographical isolation. Under this model, taxa diverged in vicariance can conserve similar environmental niches or experience ecological specialization after divergence due to either genetic drift or environmentally mediated divergent selection (Graham *et al.* 2004; Warren *et al.* 2008); and (ii) sympatric speciation, which results from ecological and reproductive isolation due to divergent

Correspondence: Joaquín Ortego, Fax: (+34) 954 621 125; E-mail: joaquin.ortego@csic.es

[1]Present address: Appalachian Laboratory, University of Maryland Center for Environmental Science, 301 Braddock Road, Frostburg, MD 21532, USA

selection in the absence of physical barriers to gene flow (Graham *et al.* 2004; Papadopulos *et al.* 2011). Disentangling the relative roles of geography and ecological factors in promoting species/lineage diversification requires analyses of both the historical geographical separation of taxa that could have led to speciation as well as the present-day environmental conditions that shape the current distribution and ecological requirements of species (Rundell & Price 2009; Ribera *et al.* 2011). For hybridizing species, identifying the underlying mechanisms behind divergence is even more challenging, because interspecific gene flow can confound the genetic and phenotypic signals left during the diversification process (e.g. Papadopulos *et al.* 2011, 2013; Eaton & Ree 2013). In the presence of interspecific genetic exchange, the maintenance of species boundaries requires selection against hybrids due to their intrinsic low performance, maladaptation to both parental species environments or habitat-induced variation in assortative mating (e.g. Anderson 1953; Papadopulos *et al.* 2011; Singhal & Moritz 2012), so that hybrid/introgressed individuals constitute an ephemeral state that only persist under some particular circumstances (e.g. in areas with intermediate environments; Papadopulos *et al.* 2013; Ortego *et al.* 2014a).

California has a complex biogeographic history, with deep topographic and climatic gradients that have contributed to shape one of the richest floras and faunas on the planet (Raven & Axelrod 1978; Calsbeek *et al.* 2003; Davis *et al.* 2008; Lancaster & Kay 2013). The high biodiversity and local endemism of this region have been linked to the important role of topographic and climatic complexity in creating multiple ecological niches, favouring high rates of intra- and interspecific diversification (Calsbeek *et al.* 2003; Davis *et al.* 2008). In parallel, climate buffering associated with complex topography seems to have contributed to a particularly low extinction rate of species/lineages adapted to contrasting environmental conditions since the Tertiary (Lancaster & Kay 2013). Comparative studies have often revealed common phylogeographic breaks and genetic discontinuities across multiple taxa, suggesting a similar history of diversification by vicariance associated with geological events and climatic fluctuations (Calsbeek *et al.* 2003; Lapointe & Rissler 2005; Rissler *et al.* 2006; Vandergast *et al.* 2008; Gugger *et al.* 2013; Sork & Werth 2014). However, other studies have also pointed to ecological adaptation as an important mode of diversification of California biota (e.g. Ortego *et al.* 2012; Langin *et al.* 2015). This environmental complexity and high diversity makes California Floristic Province an ideal scenario to study the proximate factors underlying the speciation process, which can ultimately contribute to understand the origin of its high rates of endemism and guide conservation policies aimed to preserve this biologically rich but increasingly threatened ecoregion (Myers *et al.* 2000; Vandergast *et al.* 2008).

Here, we employ an integrative approach to analyse the evolutionary and demographic history of the Californian scrub white oak species complex (genus and section *Quercus*), infer the most plausible modes of speciation underlying the diversification process within the complex and understand the relative importance of geographical isolation vs. ecological-mediated selection in driving taxonomic divergence and maintaining species genetic identity. There are six recognized taxa within this complex, namely California interior (*Quercus berberidifolia*, Liebmann 1854), leather (*Quercus durata*, Jepson 1909), coastal sage (*Quercus dumosa*, Nuttall 1842), island (*Quercus pacifica*, Nixon & Muller 1994), Muller (*Quercus cornelius-mulleri*, Nixon & Steele 1981) and Tucker (*Quercus john-tuckeri*, Nixon & Muller 1994) scrub oaks (Nixon 2002). Most of these species are found in Southern California, where they live in close geographical proximity but not usually in mixed stands (Roberts 1995). Many of these taxa have been described in the last few decades (Nixon & Steele 1981; Nixon & Muller 1994) and, with the exception of *Q. durata*, all of them were generally treated before as part of a broad complex of *Q. dumosa* (Nixon 2002; e.g. Forde & Faris 1962). These taxa are edaphic specialist (*Q. durata*), island endemics (*Q. pacifica*) or encountered in different habitats such as coastal sage scrub (*Q. dumosa*), desert margins (*Q. cornelius-mulleri*), and chaparral and mixed forest (*Q. berberidifolia* and *Q. john-tuckeri*), suggesting that different modes of speciation (allopatric speciation in island endemics vs. sympatric/parapatric speciation among nearby ecological specialist) may have contributed to the diversity of the complex (Nixon 2002). Thus, this system offers an interesting case study to understand how potentially different mechanisms may have contributed to taxonomic divergence in a complex of dominant keystone species with a high importance in ecosystem functioning (Sawyer & Keeler-Wolf 1995).

To study the evolutionary and demographic history of the Californian scrub white oak species complex, we collected over 800 individuals from all putative species covering their entire distribution ranges across California (Fig. 1). Specimens were genotyped at 16 nuclear and five chloroplast simple sequence repeat (SSR) markers in order to infer the demographic history of the complex and the patterns of genetic structure and hybridization within and among species, respectively. We also built species distribution models (SDMs) for the different species that were projected into the last glacial maximum (LGM; *c.* 21 000 years BP) using different climate models to estimate range shifts and characterize the environmental envelope of each taxon. In

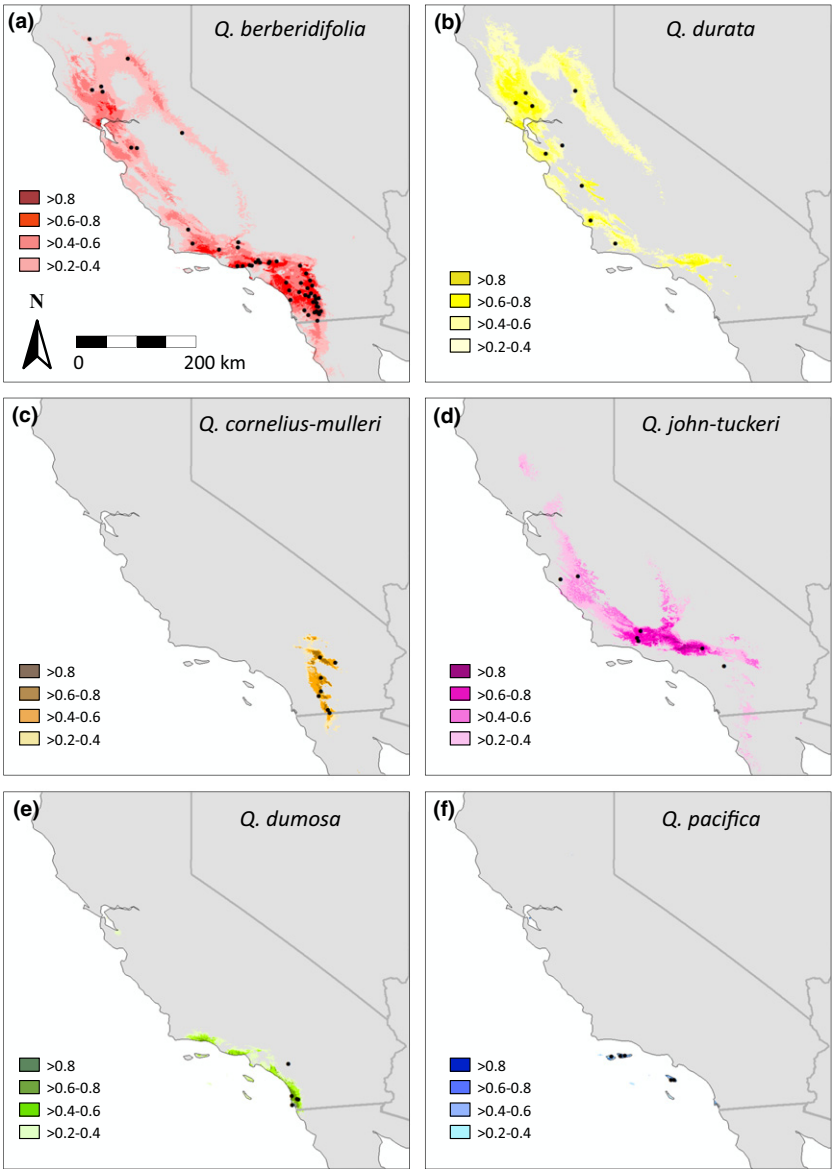

Fig. 1 Distribution modelling for the six scrub white oak study species from California. Colour scales refer to logistic probability of occurrence, and black dots indicate sampling sites for each species.

particular, (i) we first analysed patterns of inter- and intraspecific genetic differentiation among all putative species and employed an approximate Bayesian computation (ABC) framework (Beaumont *et al.* 2002; Beaumont & Rannala 2004) to compare different plausible scenarios of divergence. This allowed us to infer the evolutionary and demographic history of the different taxa within the complex. In a second step, to elucidate the potential mechanism of speciation (allopatric/parapatric speciation vs. ecological speciation in sympatry), we (ii) linked the inferred divergence pathways with current and past species distribution ranges and (iii) tested for environmental niche differentiation among taxa (Graham *et al.* 2004; Mao & Wang 2011; Papadopulos *et al.* 2013). Finally, we (iv) analysed whether pairwise species genetic differentiation is associated with niche overlap. If niches are highly conserved or if allopatric divergence is the dominant process, it would be expected that closely related species share a similar niche space (Warren *et al.* 2008; Nakazato *et al.* 2010). The opposite pattern would arise if phylogenetically distant taxa tend to share more similar environments than those representing earlier stages of species formation and would point to niche divergence as an important mechanism promoting divergence (Cavender-Bares *et al.* 2004; Warren *et al.* 2008). A lack of association between niche overlap and genetic differentiation is expected if environmental niches are highly labile (Knouft *et al.* 2006; Warren *et al.* 2008), which has been previously documented in trees and can result from phenotypic plasticity and/or some level of interspecific gene exchange (Cornuault *et al.* 2015).

Methods

Sampling

Between 2010 and 2014, we sampled 812 scrub white oaks from a total of 87 localities in California (Table S1, Supporting information; Fig. 1). Sampling was designed using occurrence records available for each taxon at the species level in the Calflora database (http://www.calflora.org/). We aimed to collect samples from populations located across the entire distribution range of all the species and make a more exhaustive sampling in southern California, where more taxa are present and distribution boundaries of each putative species are not well defined (Roberts 1995; eFloras 2015). Our sampling only excluded a small portion of the range of three species (*Q. berberidifolia*, *Q. dumosa* and *Q. cornelius-mulleri*) at their southernmost distribution limit in northern Baja California (Mexico) (Fig. 1). Most populations from the endangered coastal scrub oak (*Q. dumosa*) have been extirpated due to human development in southern California and only a few stands of this species could be found during our surveys across its former distribution range (IUCN 2010; Fig. 1).

SSR genotyping

We ground about 50 mg of frozen leaf tissue in tubes with a tungsten bead using a mixer mill and DNA extraction and purification was performed with the cetyl trimethyl ammonium bromide protocol (Doyle & Doyle 1990). To genotype scrub white oaks, we used 16 nuclear (nuSSR) and five chloroplast (cpSSR) SSR markers previously developed for other species (Table S2, Supporting information). Analyses of neutrality based on the F_{ST}-outliers tests implemented in LOSITAN (100 000 simulations assuming a stepwise-mutation model; Beaumont & Nichols 1996; Antao *et al.* 2008) indicated that no locus consistently deviated from neutral expectations in the studied species (data not shown). Amplifications were carried out in 10 μL reaction mixtures containing 2.5 μL of DNA (~20 ng) and 7.5 μL of PCR mix, which contained the following reagents: 3.5 μL of Type-it Multiplex PCR Master Mix (QIAGEN), 0.2 μL of BSA 10×, 2.5 μL of dye-labelled primer mix (see Table S2, Supporting information for specific primer concentrations), and 1.3 μL of ddH_2O. The PCR profile consisted of an initial denaturing of 15 min at 95 °C, followed by 44 cycles of 30-s denaturing at 95 °C, 90-s annealing at 56 °C, and 30-s extension at 72 °C, followed by a final extension step of 56 °C for 30 min. Amplification products were electrophoresed using an ABI PRISM 3700 capillary sequencer (Applied Biosystems) and genotypes were scored using GENEMAPPER 3.7 (Applied Biosystems).

Nuclear SSR genotypes were tested for departure from Hardy–Weinberg equilibrium within each sampling population/species at each locus using an exact test (Guo & Thompson 1992) based on 900 000 Markov chain iterations as implemented in the program ARLEQUIN 3.1 (Excoffier *et al.* 2005). We also used ARLEQUIN 3.1 to test for linkage equilibrium between each pair of loci for each sampling population/species using a likelihood-ratio statistic, whose distribution was obtained by a permutation procedure (Excoffier *et al.* 2005). Sequential Bonferroni corrections were applied to account for multiple comparisons (Rice 1989).

Genetic structure

For nuSSR, we analysed patterns of genetic structure using the Bayesian Markov chain Monte Carlo clustering analysis implemented in the program STRUCTURE 2.3.3 (Pritchard *et al.* 2000; Falush *et al.* 2003; Hubisz *et al.* 2009). We ran STRUCTURE assuming correlated allele frequencies and admixture and without using prior population information (Hubisz *et al.* 2009). We conducted 10 independent runs for each value of K = 1–10 to estimate the 'true' number of clusters with 200 000 MCMC cycles, following a burn-in step of 100 000 iterations. The number of populations best fitting the data set was defined using log probabilities [$\Pr(X|K)$] (Pritchard *et al.* 2000) and the ΔK method (Evanno *et al.* 2005), as implemented in STRUCTURE HARVESTER (Earl & vonHoldt 2012). We used CLUMPP 1.1.2 and the Greedy algorithm to align multiple runs of STRUCTURE for the same K value (Jakobsson & Rosenberg 2007) and DISTRUCT 1.1 (Rosenberg 2004) to visualize as bar plots the individual's probabilities of population membership. Clustering analyses for nuSSR were complemented with an individual-based principal component analysis (PCA), which does not rely on Hardy–Weinberg or linkage equilibrium and can be useful to detect genetic clines or complex patterns of genetic structure. We employed a MANOVA to compare simultaneously the PCA scores obtained for the first two principal component (PC) axes across all the species. Post hoc Tukey tests were used to examine differences between each species pair. PCA was performed using the R 3.0.2 package ADEGENET (Jombart 2008) and MANOVA in SPSS 22.0.

We examined population genetic structure for cpSSR markers using the 'clustering with linked loci' analysis in BAPS 6.0 (Corander *et al.* 2008). We conducted four independent runs for K = 1–40 to determine the K value with the highest likelihood. Results from BAPS were also visualized as bar plots using DISTRUCT 1.1. Complementarily, we constructed a median-joining haplotype network coloured by species using NETWORK 4.6 (Bandelt *et al.* 1999).

The distribution of genetic variation in both nuclear and chloroplast loci was assessed using AMOVAs as implemented in ARLEQUIN 3.1 (Excoffier *et al.* 2005). Genetic variation was hierarchically partitioned into among species, among populations within species and among individuals within populations. Significance was tested using 10 000 permutations of the original data.

Phylogenetic analyses

To visualize the phylogenetic relationship among all taxa and populations, we reconstructed a population-based neighbour-joining (NJ) tree in POPULATIONS 1.2.30 (Langella 2007) using Cavalli-Sforza & Edwards (1967) chord distances based on nuSSR and cpSSR allele frequency data. This genetic distance has been suggested to be the most effective in recovering the correct tree topology for SSR markers under a variety of evolutionary scenarios without making assumptions regarding constant population size or mutation rates among loci (Takezaki & Nei 1996). Levels of confidence of phylogenetic groupings were estimated using bootstrapping methods, resampling across loci 1000 times (Langella 2007). Only populations with five or more genotyped individuals were included in NJ analyses. These analyses were performed both considering all sampled individuals and excluding hybrids/introgressed individuals from the data set ($q < 0.90$; e.g. Ortego & Bonal 2010).

Demographic history: ABC

We used an ABC (Beaumont 2010; Bertorelle *et al.* 2010) statistical framework to compare three plausible scenarios of divergence among Californian scrub white oaks and estimate timing of divergence and effective population sizes of each taxa (Fig. 2a). Scenarios were built considering previous morphology-based knowledge on the taxonomy of the group (Nixon 2002) and STRUCTURE analyses (see Results section; e.g. Liu *et al.* 2014; Tsuda *et al.* 2015), which point to an older divergence between *Q. berberidifolia*/*Q. durata* and the rest of the species than among taxa within these two groups. The first scenario ('null model') consisted of a simultaneous divergence of all the species at t5. The second scenario predicts that *Q. durata* and *Q. berberidifolia* diverged from the rest of the species at t5. Afterwards, *Q. durata* and *Q. berberidifolia* diverged at t4 and the other four species diverged simultaneously at t3. The third scenario, the most likely according with hierarchical STRUCTURE analyses, is similar to the previous one but predicts that *Q. dumosa* and *Q. pacifica* diverged from *Q. corneliusmulleri* and *Q. john-tuckeri* at t3 and these pairs of taxa subsequently split at t1 and t2, respectively (Fig. 2a). We also performed further ABC analyses exclusively focusing on *Q. dumosa* and the island endemic *Q. pacifica* to test in more detail different scenarios of divergence among the populations of these two closely related taxa (Fig. 2b) for which we have found a complex pattern of genetic structure and admixture (see results section). In this case, we considered three populations, one including all sampling localities for *Q. dumosa*, another for the populations of *Q. pacifica* from Santa Catalina island, and a third one including the populations of *Q. pacifica* from both Santa Cruz and Santa Rosa Islands (Fig. 2b).

Simulations for each scenario and ABC analyses were conducted using DIYABC 2.0 (Cornuet *et al.* 2014). We generated 1 million simulated data sets per scenario considering a 1:1 female to male sex ratio, a generalized stepwise-mutation model and uniform priors with default values for all parameters. We considered the following constraints on temporal parameters: t5 > t4, t5 > t3, t5 > t2, t5 > t1, t3 > t2, t3 > t1 for all species (Fig. 2a) and t2 > t1 for analyses focused on *Q. dumosa* and *Q. pacifica* (Fig. 2b). We used five summary statistics (SS): the mean values of expected heterozygosity (H_E) and number of alleles (A) for each population (i.e. species) and H_E, A, classification index and F_{ST} for each pair of taxa. To reduce the computational demands and avoid potentially confounding effects of contemporary hybridization (e.g. Ortego & Bonal 2010; Ortego *et al.* 2014a), we excluded hybrids/introgressed individuals from the data sets ($q < 0.90$; e.g. Besnard *et al.* 2014; Tsuda *et al.* 2015) and for analyses including all species we selected up to a maximum of 50 individuals per taxon representative of all sampled populations ($n = 281$ individuals in total). We ran ABC analyses considering only nuSSR and both nuSSR and cpSSR markers to evaluate the impact of using different kind of loci on scenario choice and posterior parameter estimation. The posterior probability of scenarios was assessed using a weighted polychotomous logistic regression on the 1% of simulated data sets closest to the observed data (Cornuet *et al.* 2008, 2014; Fontaine *et al.* 2013). We simulated 500 pseudo-observed data sets (PODs) under each scenario to estimate type I and type II error rates and assess confidence in scenario choice (Robert *et al.* 2011). For the best supported scenario, we estimated the posterior distribution of all parameters using local linear regressions on the 1% of the simulations closest to the observed data after a logit transformation of parameter values (Beaumont *et al.* 2002; Cornuet *et al.* 2008, 2014). The performance of parameter estimation was determined by calculating the median of the relative median of absolute errors (RMAE) from 500 PODs (Cornuet *et al.* 2010). Finally, we evaluated the goodness of fit of the best supported

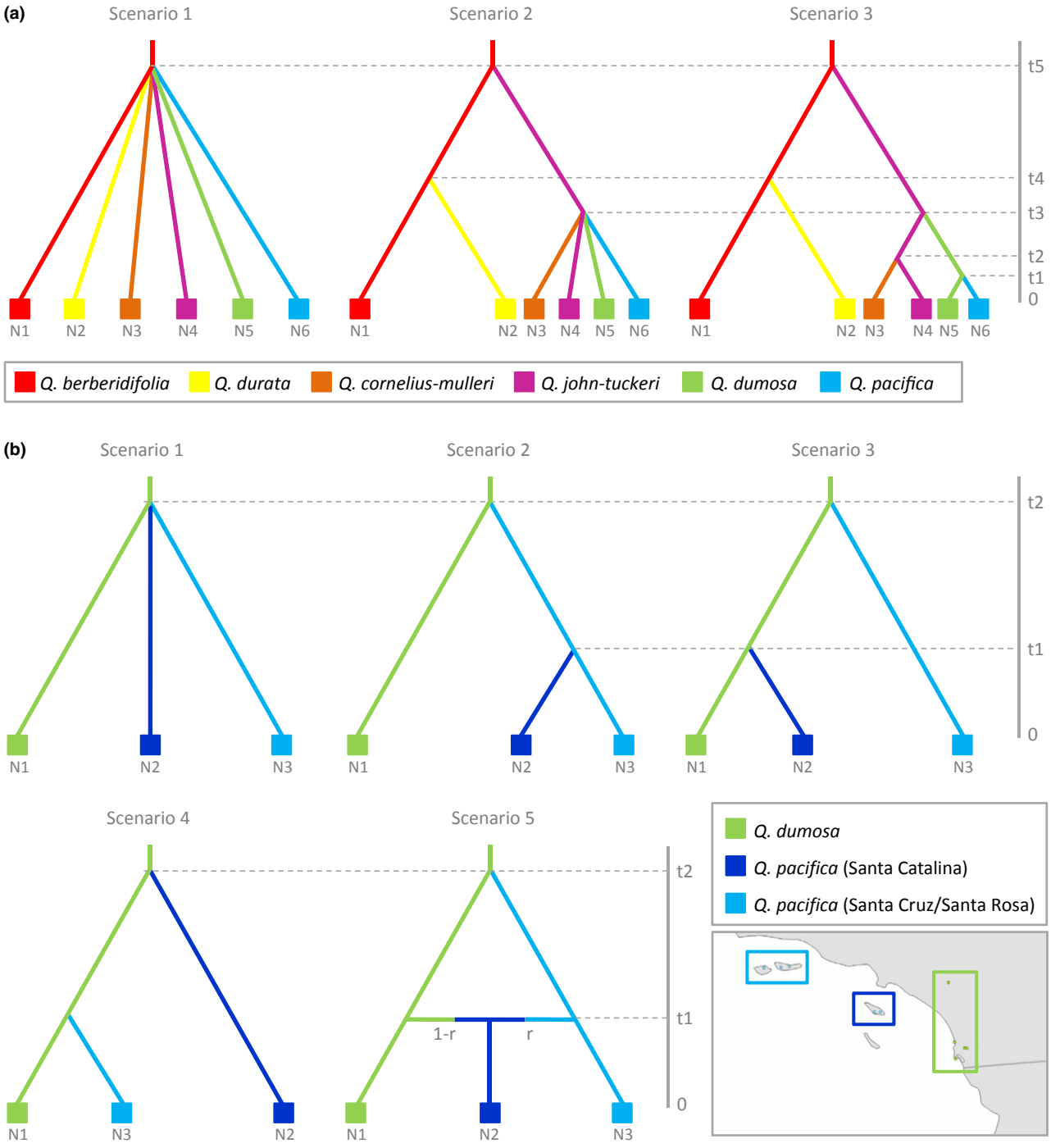

Fig. 2 Scenarios tested using approximate Bayesian computation (ABC) considering (a) all studied species and (b) only populations of coastal (*Quercus dumosa*) and island (*Quercus pacifica*) scrub oaks. In these scenarios, t# represents time in generations, N# effective population sizes of the different populations and r# immigration rates from donor populations.

scenario simulating 10 000 PODs from the posterior distribution of all parameters. As recommended by Cornuet *et al.* (2010), we used SS that had not been used for model selection or parameter estimation in previous ABC treatments. We applied a PCA on test statistics vectors to visualize the fit between simulated and observed data sets. Finally, we used ranking of SS (i.e. % of simulated data < observed data) to evaluate whether the best supported model can successfully reproduce observed data.

Species distribution modelling

We used SDM to predict the geographic distribution of suitable habitats for each studied species both in the present and during the LGM. We modelled current distributions using MAXENT 3.3.3 (Phillips *et al.* 2006; Phillips & Dudik 2008). Species occurrence data were obtained from sampling points as well as from herbarium records available at the Global Biodiversity Information Facility (http://www.gbif.org/), Consortium of California Herbaria (http://ucjeps.berkeley.edu/consortium/), Consortium of Pacific Northwest Herbaria (http://www.pnwherbaria.org/) and University of Arizona Herbarium (http://ag.arizona.edu/herbarium/). Prior to modelling, all records were mapped and examined to identify and exclude those having obvious georeferencing errors and misidentifications. For models, we only retained a single record among those falling within the same grid cell (see final sample sizes in Table S3, Supporting information). Initially we selected 26 variables that may potentially determine the current distribution of our species. These included growing degree days, soil organic carbon, soil pH, and soil moisture obtained from the Center for Sustainability and the Global Environment http://nelson.wisc.edu/sage/data-and-models/atlas/; ground frost frequency, vapour pressure, and wet day frequency obtained from the Intergovernmental Panel on Climate Change (http://www.ipcc-data.org/observ/clim/cru_ts2_1.html); and the 19 bioclimatic layers from the WorldClim data set (http://www.worldclim.org/) (Hijmans *et al.* 2005). The 19 bioclimatic layers were downloaded from WorldClim at 30-arcsec (~1-km) resolution. The other layers were downloaded at a spatial resolution of 0.5° and resampled to 30-arcsec resolution via bilinear interpolation using ARCMAP 10.2.1 (ESRI, Redlands, CA, USA). We assessed the correlation between all the layers using ENMTOOLS (Warren *et al.* 2010). When two layers were highly correlated ($r > 0.7$), we discarded the layer with the highest number of correlations with other layers. Accordingly, we selected a final set of seven layers to construct the models: mean diurnal range (Bio2), isothermality (Bio3), mean temperature of the wettest quarter (Bio8), precipitation of the wettest month (Bio13), precipitation of the driest month (Bio14), precipitation seasonality (Bio15), and precipitation of the warmest quarter (Bio18). Model evaluation statistics were produced from ten cross-validation replicate model runs and overall model performance was assessed using the area under the receiving operator characteristics curve (AUC).

We estimated the distribution of each scrub white oak species at the LGM (*c.* 21 000 years BP) projecting contemporary species-climate relationships to this period. We used the same seven bioclimatic layers available at WorldClim for two palaeoclimate models of the last glacial period: the Model for Interdisciplinary Research on Climate (Hasumi & Emori 2004) and the Community Climate System Model (CCSM3; Collins *et al.* 2006). The predicted probabilities of occurrence obtained for both palaeoclimate models were averaged using the R 3.0.2 package RASTER (R Core Team 2015).

Niche divergence/conservatism

We used identity and background tests in ENMTOOLS to analyse whether ecological niches are differentiated or conserved between each pair of species (Warren *et al.* 2008; e.g. Nakazato *et al.* 2010; Mao & Wang 2011). For this purpose, we first calculated actual niche overlap between each pair of species using two alternative statistics: *I* (Warren *et al.* 2008) and *D* (Schoener 1968), which are compared and described in detail in Warren *et al.* (2008). Then, we performed niche identity tests to compare the overlap of a species pair's actual niches to a distribution of niche overlaps obtained from pairs of pseudoniches ($n = 100$ pseudoreplicates) constructed based on randomly reshuffled occurrence points of the two species (Warren *et al.* 2008). Thus, this test examines the null hypothesis that a given pair of species is distributed in an identical environmental space. We also used background tests to analyse whether the ecological niches of a given pair of species overlap more or less than would be expected from the differences in the environmental backgrounds of the regions where they occur (Warren *et al.* 2008). This test compares the observed niche overlap of a given pair of species to a null distribution ($n = 100$ random samplings) of overlap values generated by comparing the ecological niche model of one taxon to an ecological niche model created from random points drawn from the geographic range of the other taxon (Warren *et al.* 2008). This process is repeated for both taxa in the comparison so that two null distributions are generated. The background area should include accessible areas for the species, not just the observed niche or an area tightly delimited by species occurrence (McCormack *et al.* 2010; Nakazato *et al.* 2010; Mao & Wang 2011). Given that the delimitation of the background area can influence the results of the analyses, we considered different background areas defined by buffer zones of 1, 5, and 10 km around the actual distribution delimited by occurrence points. These background areas should include all the accessible habitats consistent with restricted scale of acorn dispersal in oaks (Grivet *et al.* 2005, 2006, 2009; Gugger *et al.* 2013). Background areas were obtained using ARCMAP 10.2.1.

The environmental space of the studied species was also characterized using a PCA applied to scaled data of the same environmental variables employed for SDM and extracted from occurrence points. We employed a MANOVA to compare simultaneously the PCA scores obtained for the first two PC axes (68% of variance explained) across all the species and determine whether their sites of occurrence differ in bioclimatic conditions. Post hoc Tukey tests were used to examine differences between each species pair. PCA was performed in R 3.0.2 package PSYCH and MANOVA in SPSS 22.0.

Niche overlap and genetic differentiation

We analysed the correlation between genetic differentiation and the two estimates of niche overlap (*I* and *D*) using Mantel tests (e.g. Warren *et al.* 2008; Nakazato *et al.* 2010). We calculated different estimates of genetic differentiation for both nuSSR and cpSSR markers, including F_{ST}, Reynold's distance, G'_{ST} and Jost's *D* (Meirmans & Hedrick 2011). Pairwise F_{ST} and Reynold's distances were calculated using ARLEQUIN 3.1 (Excoffier *et al.* 2005) and G'_{ST} and Jost's *D* were calculated using the R 3.0.2 package MMOD (R Core Team 2015). Significance of F_{ST} values was tested using Fisher's exact tests after 10 000 permutations and a sequential Bonferroni adjustment. Mantel tests were performed using ZT software with 10 000 permutations (Bonnet & Van de Peer 2002).

Results

SSR data

All nuSSR markers were highly polymorphic and observed heterozygosity at each locus ranged from 0.21 to 0.91, with 10–35 alleles per locus (Table S2, Supporting information). Chloroplast SSR markers were less polymorphic than nuclear loci and had 4–6 alleles per locus (Table S2, Supporting information). After applying sequential Bonferroni corrections to compensate for multiple statistical tests, no nuSSR locus consistently deviated from HWE in all the studied species/populations. We did not find any evidence of genotypic linkage disequilibrium at any pair of nuSSR loci in any species/population (exact tests; all *P*-values > 0.05).

Genetic structure

STRUCTURE analyses and the statistic Δ*K* indicated an 'optimal' value of $K = 2$ when populations from all the species were included in the same analysis (Fig. S1a, Supporting information), grouping *Quercus berberidifolia*-*Quercus durata* in one cluster and the rest of the species

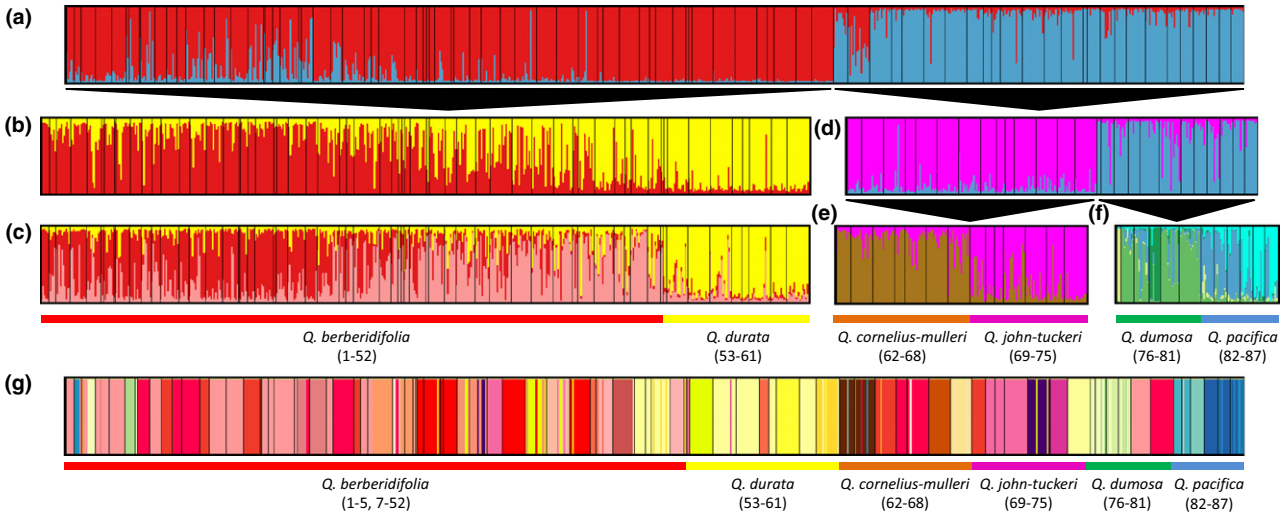

Fig. 3 Results of genetic assignments based on the Bayesian methods implemented in the programs (a–f) STRUCTURE (nuSSR markers) and (g) BAPS (cpSSR markers; $K = 34$). Panels a-f show hierarchical analyses in STRUCTURE considering (a) the six scrub white oak species ($K = 2$), (b, c) *Quercus berberidifolia* and *Quercus durata* ($K = 2$ and $K = 3$), (d) *Quercus cornelius-mulleri*, *Quercus john-tuckeri*, *Quercus dumosa* and *Quercus pacifica* ($K = 2$), (e) *Q. cornelius-mulleri* and *Q. john-tuckeri* ($K = 2$) and (f) *Q. dumosa* and *Q. pacifica* ($K = 5$). Each individual is represented by a vertical bar, which is partitioned into *K* coloured segments showing the individual's probability of belonging to the cluster with that colour. Thin vertical black lines separate individuals from different sampling localities, which are arranged according to their geographical location from southeast (left) to northwest (right) as presented in Table S1 (Supporting information). The only two individuals collected from locality six failed to amplify for most chloroplast markers and were not included in BAPS analyses. SSR, simple sequence repeat.

in another cluster (Fig. 3a). Subsequent hierarchical analyses only including samples from *Q. berberidifolia* and *Q. durata* showed that these two species split in two different clusters (Figs 3b and S1b, Supporting information). The rest of the species also split in two other clusters, one containing the populations of *Quercus cornelius-mulleri* and *Quercus john-tuckeri* and another including the populations of *Quercus dumosa* and *Quercus pacifica* (Figs 3d and S1c, Supporting information). Further, hierarchical analyses within these clusters showed that *Q. cornelius-mulleri* and *Q. john-tuckeri* split in two clusters (Figs 3e and S1d, Supporting information) whereas *Q. dumosa* and *Q. pacifica* split in five clusters (Figs 3f and S1e, Supporting information). All analyses revealed considerable rates of hybridization, particularly between *Q. berberidifolia* and *Q. cornelius-mulleri* in southern California and between *Q. berberidifolia* and *Q. durata* in northern California (Fig. 3). Further analyses within each of the species did not show any further genetic subdivision, with the exception of *Q. berberidifolia*, which showed two genetic clusters with a latitudinal cline (Fig. 3c). However, STRUCTURE analyses for *Q. berberidifolia* excluding individuals with admixed ancestry with other species ($q < 0.9$; e.g. Ortego & Bonal 2010) indicated an optimal value of $K = 1$ (data not shown), suggesting that the latitudinal cline of genetic differentiation within this species may be caused by hybridization with the different taxa with which it interbreeds in different parts of its distribution range. Further analyses excluding hybrid individuals for the groups *Q. cornelius-mulleri* and *Q. john-tuckeri* (Fig. 4a) and *Q. dumosa* and *Q. pacifica* (Fig. 4b) showed that optimal *K* values were the same than those yielded by global analyses including all individuals (Fig. 3). Genetic admixture within these two groups probably reflects both contemporary hybridization among nearby populations and shared ancestry due to recent species divergence (Fig. 4) (e.g. Tsuda *et al.* 2015).

The PCA on nuSSR data (PC1: 1.91% inertia; PC2: 0.90% inertia; Fig. 5a) and MANOVA analysis indicated that the obtained scores for each axis significantly differed among the studied species (PC1: $F_{5,805} = 582.36$, $P < 0.01$; PC2: $F_{5,805} = 115.63$, $P < 0.01$). Post hoc Tukey tests showed that the only nonsignificant pairwise comparisons were those between *Q. cornelius-mulleri* and *Q. john-tuckeri* (for PC1), *Q. dumosa* and *Q. pacifica* (for PC1), and *Q. durata* and *Q. cornelius-mulleri* (for PC2). Thus, all species were genetically differentiated for at least one of the two PCs.

BAPS analyses for linked chloroplast markers showed the highest likelihood (log maximum likelihood = −788.1) for $K = 34$ (Fig. 3g). This high number of genetic clusters indicates strong local genetic structure both among and within species, a pattern that has been previously reported in other Californian oaks (Grivet *et al.* 2006; Gugger *et al.* 2013). Twenty clusters were exclusively represented in a single species (i.e. not shared with other species) and the rest were represented in populations from more than one taxon (Fig. 3g; see Table S4, Supporting information). Accordingly, network analyses representing the relationships among the 45 haplotypes found in our study populations indicated that they are not clustered by species and nine of them were shared by two or three taxa (Fig. 6).

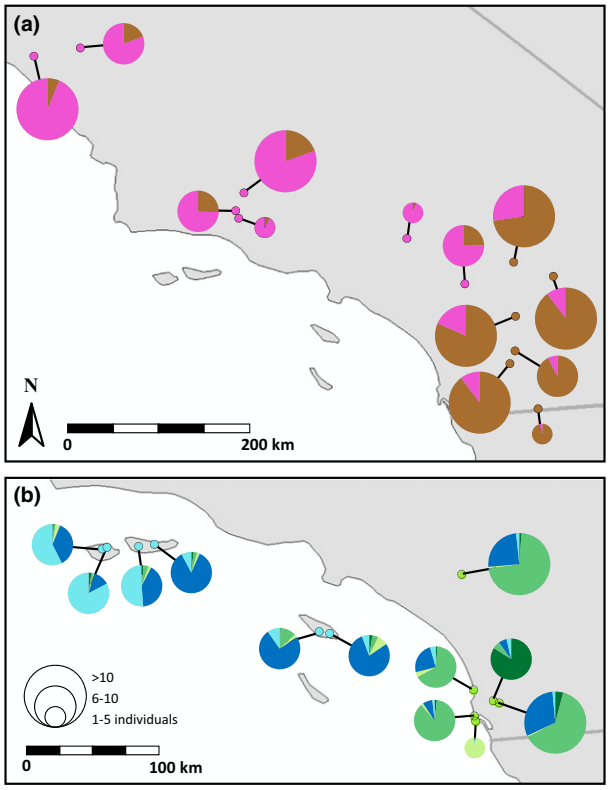

Fig. 4 Genetic assignment of the focal populations of (a) *Quercus cornelius-mulleri* and *Quercus john-tuckeri* ($K = 2$), and (b) *Quercus dumosa* and *Quercus pacifica* ($K = 5$) according to the Bayesian method implemented in the program STRUCTURE after excluding individuals with some degree of admixed ancestry ($q < 0.9$) with other species. Admixture proportions generated by STRUCTURE were represented using pie charts, with each colour indicating a different genotypic cluster. Pie chart size is proportional to the number of genotyped individuals at each location.

AMOVA analyses showed considerable differences between nuclear and chloroplast genomes in their respective distribution of genetic variation (Table 1). For both nuclear and chloroplast markers, the proportion of total variation among species was low (<10%; Table 1). However, the percentage of total variation attributed to differences among populations within

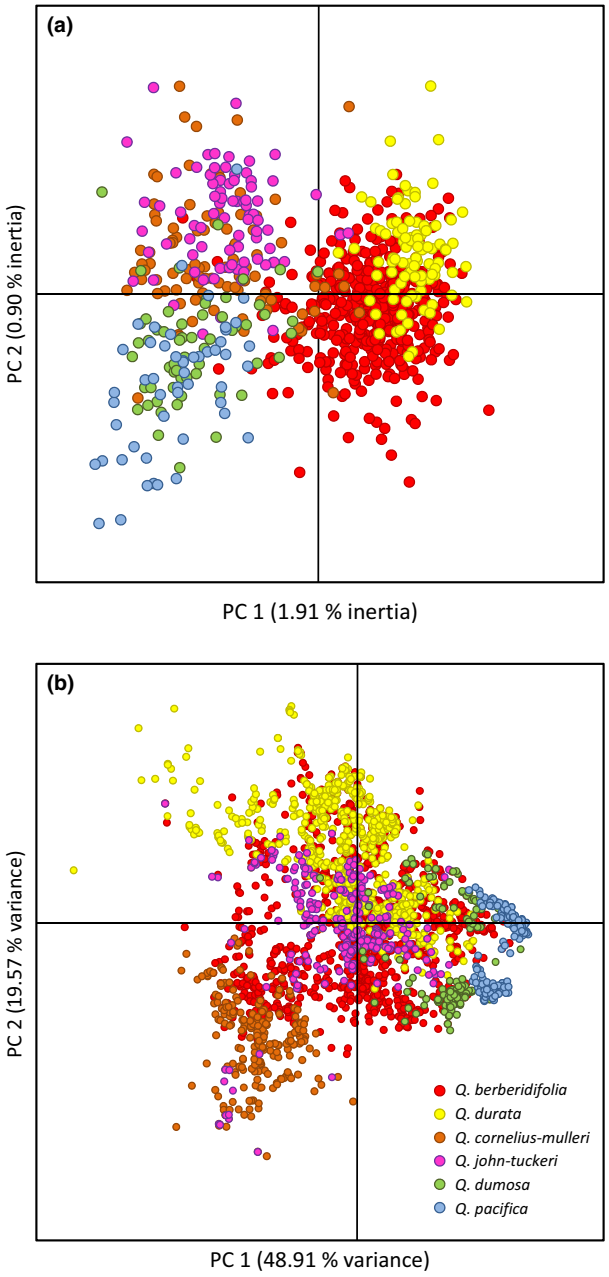

Fig. 5 Principal component analyses (PCA) for (a) genetic (16 nuSSR loci) and (b) environmental (seven bioclimatic variables) data for the six studied scrub white oaks from California. PCA on genetic data includes all the genotyped individuals (n = 812) and PCA on environmental variables considered all occurrence points used for species distribution modelling (n = 2610). Different species are represented with different colours. SSR, simple sequence repeat.

species was very low in nuclear markers (~3%) and very high in chloroplast markers (~81%). Conversely, variation among individuals within populations was low for cpSSR markers (~9%) and very high for nuSSR markers (~91%) (Table 1).

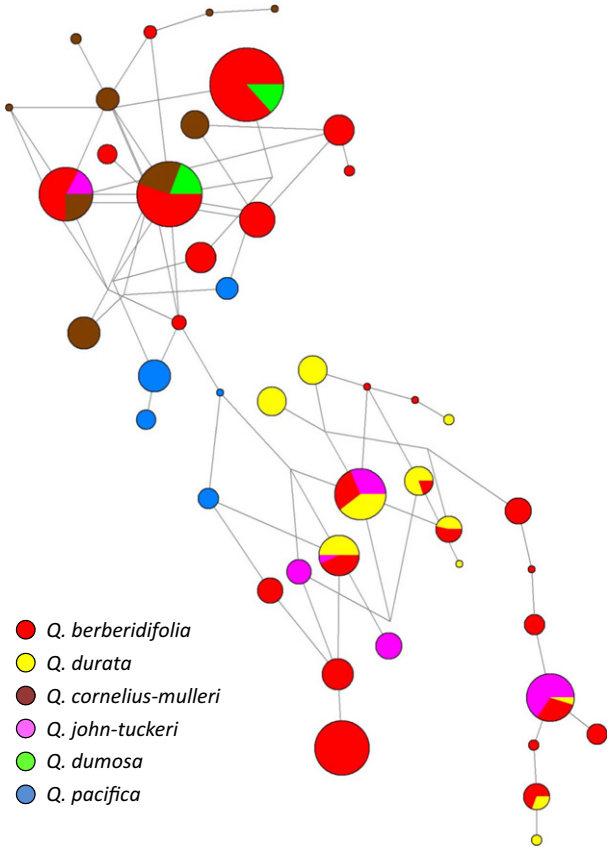

Fig. 6 Haplotype network for cpSSR markers. Pie chart size is proportional to the number of individuals and colours indicate the proportion of individuals of each species with each haplotype. SSR, simple sequence repeat.

Phylogenetic analyses

Neighbour-joining trees based on nuSSR markers showed the same patterns of genetic differentiation revealed by STRUCTURE analyses. *Quercus berberidifolia* and *Q. durata* are closely related and more distant from the rest of the species, which in turn form discrete clades but with a low degree of genetic differentiation among them (Fig. 7a). According with STRUCTURE analyses, neighbour-joining trees also suggested that the pairs of species *Q. cornelius-mulleri*/*Q. john-tuckeri* and *Q. dumosa*/*Q. pacifica* are more closely related with each other than with any of the two other species within the clade (Fig. 7a). Hybridization did not have a major impact on phylogenetic analyses and only population 63, a hybrid swarm between *Q. berberidifolia* and *Q. cornelius-mulleri* identified by STRUCTURE analyses (Fig. 3a), was originally classified as *Q. cornelius-mulleri* but placed within the *Q. berberidifolia* clade (see arrow in Fig. 7a). Neighbour-joining trees based on cpSSR markers showed that populations from the same species tend not to be phylogenetically clustered (Fig. 7b), suggesting

Table 1 Results of AMOVAs for nuclear (nuSSR) and chloroplast (cpSSR) markers. All comparisons were highly significant ($P < 0.001$ in all cases)

	nuSSR			cpSSR		
	Sum of squares	Variance components	Percentage of variation	Sum of squares	Variance components	Percentage of variation
Among species	464.83	0.38	5.88	155.47	0.15	9.98
Among populations within species	768.98	0.20	3.13	910.04	1.23	81.41
Among individuals within populations	8954.31	5.83	90.99	93.08	0.13	8.61

SSR, simple sequence repeat.

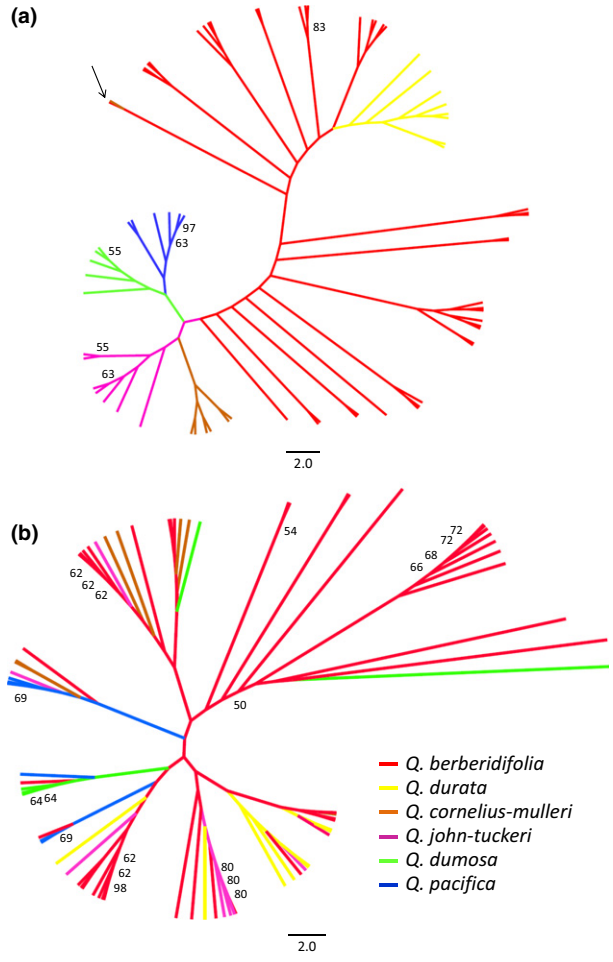

Fig. 7 Unrooted neighbour-joining trees based on Cavalli-Sforza's distance for (a) nuclear and (b) chloroplast simple sequence repeat (SSR) markers. Only populations with five or more genotyped individuals have been included in the analyses. The arrow indicates a hybrid swarm population between *Quercus berberidifolia* and *Quercus cornelius-mulleri*, and the scale bars show Cavalli-Sforza's genetic distance between samples. Bootstrap values below 0.50 are omitted.

considerable interspecific gene exchange and little taxonomic value of our set of chloroplast markers (see Petit & Excoffier 2009 and references therein). We obtained similar results for analyses considering five randomly picked individuals per population, indicating that differences in sample sizes among populations (range = 5–18 genotyped individuals/population; Table S1, Supporting information) had no considerable impact on tree topology (data not shown). Analyses performed excluding hybrid individuals ($q < 0.90$) from the data set also resulted in a similar tree topology (data not shown).

Demographic history: ABC

Considering the scenarios tested for all species and any subset of loci (nuSSR or nuSSR + cpSSR), scenario 2 had the highest posterior probability and 95% confidence intervals for this model did not overlap with those obtained for the other scenarios (Table 2). Analyses focused on populations of *Q. dumosa* and *Q. pacifica* indicated that a simultaneous split (scenario 1) is the best supported scenario and 95% confidence intervals for this model did not overlap with those obtained for the other scenarios (Table 2). Analyses to estimate confidence in scenario choice based on 500 PODs indicate that type I and type II errors for the best supported scenario were moderately low for analyses based on all species (Table 2). However, power analyses showed that type I errors (false positives) for ABC tests focused on *Q. dumosa* and *Q. pacifica* were high, indicating that confidence in scenario choice is poor. Principal component analyses showed that SS calculated for the posterior simulated data sets of the best supported scenario explained the observed data well. Accordingly, none of the SS used as test statistics significantly differed between observed and simulated data sets after sequential Bonferroni corrections (all *P*s > 0.1). RMAE values

Table 2 Posterior probability of each scenario and 95% confidence intervals (CI) based on the logistic regression approach for approximate Bayesian computation analyses (ABC) considering (a) all species and (b) only populations of coastal (*Quercus dumosa*) and island (*Quercus pacifica*) scrub oaks. Type I and type II errors for the best supported scenario (in bold) are indicated. Simulations and ABC analyses were performed only including nuclear markers (nuSSR) and considering both nuclear and chloroplast markers (nuSSR + cpSSR)

	nuSSR				nuSSR + cpSSR			
Scenario	Posterior probability	95% CI	Type I error	Type II error	Posterior probability	95% CI	Type I error	Type II error
(a) All species								
1	0.001	<0.001–0.300			0.008	<0.001–0.213		
2	**0.884**	**0.854–0.914**	**0.194**	**0.046**	**0.888**	**0.868–0.909**	**0.192**	**0.088**
3	0.115	<0.001–0.347			0.104	<0.001–0.268		
(b) *Q. dumosa* and *Q. pacifica*								
1	**0.563**	**0.555–0.571**	**0.450**	**0.150**	**0.429**	**0.421–0.438**	**0.510**	**0.122**
2	0.050	0.046–0.054			0.125	0.119–0.131		
3	0.089	0.084–0.094			0.173	0.166–0.180		
4	0.006	0.003–0.010			0.076	0.071–0.080		
5	0.291	0.284–0.299			0.197	0.190–0.204		

SSR, simple sequence repeat.

were moderately low (<0.2 in most cases), indicating that estimates of posterior parameters are reliable (Table 3; see also Figs S2 and S3, Supporting information). Analyses based on either nuSSR or nuSSR + cpSSR yielded very similar parameter estimates (Table 3). Assuming an average generation time of 50 years for scrub oaks (i.e. half of the generation time generally assumed for large oaks; e.g. Gugger *et al.* 2013), suggest that divergence between *Q. berberidifolia*/*Q. durata* and the rest of the species probably predates the Wisconsinan glaciation (*c.* 79–91 k years BP; Table 3). Divergence between *Q. berberidifolia* and *Q. durata* (*c.* 23–26 k years BP) and among the other four species (*c.* 22 k years BP) probably happened around the LGM. Parameter estimates indicate that the divergence among the populations of *Q. dumosa* and *Q. pacifica* probably also took place around the LGM (*c.* 19–26 k years BP) (Table 3). ABC analyses exclusively based on cpSSR loci resulted in very low support for any tested scenario, probably as a result of the scarce resolution provided by only five markers with low variability from a single locus (Table S2, Supporting information) and the considerable haplotype sharing among species (Figs 6 and 7b and Table S4, Supporting information; see also Petit *et al.* 2002, 2004). These two factors strongly limit the power of our chloroplast markers to reconstruct the evolutionary history of the studied species in comparison with nuclear loci (e.g. see Petit & Excoffier 2009).

Species distribution modelling

All SDM had high AUC values (range: 0.89–0.99; Table S3, Supporting information) and predicted well current species distributions (Fig. 1). Projected distributions into the past suggest that the ranges of some species contracted during the LGM (*Q. berberidifolia*, *Q. cornelius-mulleri*) whereas others experienced considerable expansions (*Q. john-tuckeri*, *Q. pacifica*) (Fig. S4, Supporting information). The expansion predicted for *Q. pacifica* during the LGM suggests that the geographic range of this species probably overlapped with the inland populations of *Q. berberidifolia* and *Q. dumosa* during this period. It is also remarkable the isolated distribution of *Q. cornelius-mulleri* in both time periods, with very little range overlap with *Q. berberidifolia* and *Q. john-tuckeri* even though they live in close geographical proximity. Finally, it should be noted that the predicted distribution of *Q. durata* in both the present and the LGM must be interpreted with caution given that this species is associated with serpentine soils that were not considered in our SDMs (Whittaker 1954; Forde & Faris 1962; Harrison 1999). For instance, the predicted southward expansion of this species into southern California during the LGM is highly unlikely due to the lack of serpentine soils in the region (Kruckeberg 1984). Serpentine soils are very patchily distributed in California, represent less than ~1% of land surface of this state (Kruckeberg 1984) and no detailed maps are available for SDM.

Niche divergence/conservatism

All niche identity tests were significant for the two estimates of niche overlap (*I* and *D*), indicating that the studied species are not distributed in an identical environmental space (all Ps < 0.01). Background tests

Table 3 Posterior parameter estimates (median and 95% confidence intervals) for the best supported scenarios considering (a) all species (scenario 2) and (b) only populations of coastal (*Quercus dumosa*) and island (*Quercus pacifica*) scrub oaks (scenario 1). Estimates are based on 1% of simulated data sets closest to the observed values. Simulations and approximate Bayesian computation analyses were performed only including nuclear markers (nuSSR) and considering both nuclear and chloroplast markers (nuSSR + cpSSR). Relative median absolute errors (RMAE) based on 500 pseudo-observed data sets are also indicated for each parameter

	nuSSR				nuSSR + cpSSR			
Parameter	Median	*q* (2.5)	*q* (97.5)	RMAE	Median	*q* (2.5)	*q* (97.5)	RMAE
(a) All species								
N1 (*Quercus berberidifolia*)	7950	4980	9800	0.151	8010	5110	9790	0.148
N2 (*Quercus durata*)	3610	1200	8740	0.175	3980	1300	8900	0.195
N3 (*Quercus cornelius-mulleri*)	5980	2660	9530	0.147	6450	3210	9550	0.157
N4 (*Quercus john-tuckeri*)	6370	2750	9640	0.171	6480	2990	9600	0.167
N5 (*Q. dumosa*)	5440	2200	9410	0.164	5070	2120	9180	0.165
N6 (*Q. pacifica*)	6990	3440	9680	0.175	7580	4460	9770	0.159
t3	423	146	914	0.199	416	157	943	0.187
t4	412	133	1140	0.248	460	136	1250	0.234
t5	1440	512	6020	0.118	1230	406	5570	0.122
μ (nuSSR)	7.8×10^{-04}	4.7×10^{-04}	9.9×10^{-04}	0.165	7.7×10^{-04}	4.6×10^{-04}	9.9×10^{-04}	0.190
μ (cpSSR)	—	—	—	—	1.4×10^{-04}	1.0×10^{-04}	3.5×10^{-04}	0.204
(b) *Q. dumosa* and *Q. pacifica*								
N1	4310	1630	9170	0.178	4690	1930	8920	0.179
N2	7690	3830	9850	0.182	6210	2930	9450	0.197
N3	6980	3100	9780	0.176	7000	3540	9660	0.191
t2	376	136	769	0.207	519	179	1910	0.206
μ (nuSSR)	7.1×10^{-04}	3.8×10^{-04}	9.8×10^{-04}	0.203	8.2×10^{-04}	4.8×10^{-04}	1.0×10^{-03}	0.214
μ (cpSSR)	—	—	—	—	1.6×10^{-04}	1.0×10^{-04}	3.5×10^{-04}	0.243

SSR, simple sequence repeat.
N = effective population sizes for each taxa; *t* = time in generations; μ = mutation rate.

showed mixed results depending on the species' pair and the direction of the test (Table 4). Niche overlap between most pairs of species was lower than expected based on the differences in the environmental background in which they occur, particularly when the larger background area was considered (Table 4). Exceptions generally involved comparisons with the island endemic *Q. pacifica* and the widespread *Q. berberidifolia*, which often showed either nonsignificant differences or higher niche overlap with other species than expected from differences in their environmental backgrounds. Contrarily, the comparisons involving *Q. cornelius-mulleri* and *Q. john-tuckeri* generally revealed that niche overlap with the other species is lower than expected from their environmental backgrounds (Table 4). As indicated above for SDM, comparisons involving *Q. durata* must be interpreted with caution given that this species is associated with serpentine soils, a trait that suggest a remarkable ecological niche differentiation with respect to the other studied taxa.

The first two axes of the PCA jointly explained 68.47% of the variance (PC1: 48.91%; PC2: 19.57%; Fig. 5b) and MANOVA analyses indicated that the obtained scores for each axis significantly differed among the studied species (PC1: $F_{5,2607} = 835.46$, $P < 0.01$; PC2: $F_{5,2607} = 308.44$, $P < 0.01$). Post hoc Tukey tests showed that the only nonsignificant pairwise comparisons were those between *Q. john-tuckeri* and *Q. durata* (for PC1), and *Q. berberidifolia* and *Q. pacifica* (for PC2). Thus, all species occupied a different environmental space for at least one of the two PCs.

Niche overlap and genetic differentiation

All pairwise F_{ST} values estimated for nuSSR markers between species were highly significant (all *P*s < 0.01) and ranged between 0.03 for *Q. dumosa*-*Q. pacifica* and 0.16 for *Q. durata*-*Q. john-tuckeri* (Table S5, Supporting information). For cpSSR markers, all pairwise F_{ST} values were also highly significant (all *P*s < 0.01) and ranged between 0.10 for *Q. berberidifolia*-*Q. cornelius-mulleri* and

Table 4 Results of background tests considering different indexes of niche overlap (Warren's *I*: left; Schoener's *D*: right) and background areas obtained using different distance buffers around occurrence points. Table indicates whether actual values of niche overlap of two species are more or less similar than expected based on the differences in the environmental background in which they occur

Species for observed distribution	Species for background	*I*	*D*	1-km buffer	5-km buffer	10-km buffer
Quercus berberidifolia	*Quercus durata*	0.849	0.618	More**/More**	NS/NS	Less*/Less***
Q. berberidifolia	*Quercus cornelius-mulleri*	0.280	0.088	NS/NS	NS/NS	Less*/Less***
Q. berberidifolia	*Quercus john-tuckeri*	0.717	0.437	NS/Less**	Less*/Less***	Less**/Less***
Q. berberidifolia	*Quercus dumosa*	0.607	0.318	NS/NS	NS/NS	Less***/Less**
Q. berberidifolia	*Quercus pacifica*	0.126	0.042	NS/NS	NS/NS	NS/NS
Q. durata	*Q. berberidifolia*	0.849	0.618	Less***/Less***	Less***/Less***	Less***/Less**
Q. durata	*Q. cornelius-mulleri*	0.081	0.020	Less*/NS	Less***/Less**	Less***/Less***
Q. durata	*Q. john-tuckeri*	0.538	0.316	Less***/Less**	Less***/Less**	Less***/Less***
Q. durata	*Q. dumosa*	0.439	0.202	NS/NS	NS/NS	NS/NS
Q. durata	*Q. pacifica*	0.079	0.022	NS/NS	NS/NS	NS/NS
Q. cornelius-mulleri	*Q. berberidifolia*	0.280	0.088	More*/More**	NS/More**	NS/NS
Q. cornelius-mulleri	*Q. durata*	0.081	0.020	Less***/Less***	Less***/Less***	Less***/Less***
Q. cornelius-mulleri	*Q. john-tuckeri*	0.424	0.152	Less**/Less***	Less***/Less***	Less***/Less***
Q. cornelius-mulleri	*Q. dumosa*	0.041	0.010	Less***/Less***	Less***/Less***	Less***/Less***
Q. cornelius-mulleri	*Q. pacifica*	0.002	0.001	Less***/Less***	Less***/Less***	Less***/Less***
Q. john-tuckeri	*Q. berberidifolia*	0.717	0.437	Less***/Less***	Less***/Less***	Less***/Less***
Q. john-tuckeri	*Q. durata*	0.538	0.316	Less***/Less***	Less***/Less***	Less***/Less***
Q. john-tuckeri	*Q. cornelius-mulleri*	0.424	0.152	More**/NS	NS/NS	NS/Less**
Q. john-tuckeri	*Q. dumosa*	0.357	0.148	NS/NS	Less**/NS	Less**/Less**
Q. john-tuckeri	*Q. pacifica*	0.059	0.014	Less***/Less***	Less***/Less***	Less***/Less***
Q. dumosa	*Q. berberidifolia*	0.607	0.318	More***/More***	More***/More***	More***/More***
Q. dumosa	*Q. durata*	0.439	0.202	More***/More***	Less*/Less*	Less***/Less***
Q. dumosa	*Q. cornelius-mulleri*	0.041	0.010	Less***/Less***	Less***/Less***	Less***/Less***
Q. dumosa	*Q. john-tuckeri*	0.357	0.148	Less***/Less***	Less***/Less***	Less***/Less***
Q. dumosa	*Q. pacifica*	0.384	0.162	Less***/NS	NS/NS	NS/NS
Q. pacifica	*Q. berberidifolia*	0.126	0.042	More***/More***	NS/NS	Less**/NS
Q. pacifica	*Q. durata*	0.079	0.022	More***/More**	NS/NS	Less***/Less***
Q. pacifica	*Q. cornelius-mulleri*	0.002	0.001	NS/NS	NS/NS	NS/NS
Q. pacifica	*Q. john-tuckeri*	0.059	0.014	Less**/Less**	Less***/Less***	Less***/Less***
Q. pacifica	*Q. dumosa*	0.384	0.162	More**/NS	More*/NS	More**/NS

NS, not significant.
*$P < 0.1$; **$P < 0.05$; ***$P < 0.01$.

0.48 for *Q. durata*-*Q. dumosa* (Table S5, Supporting information). Genetic differentiation estimated at nuSSR markers was not significantly correlated with genetic differentiation estimated at cpSSR markers (Mantel tests, F_{ST}: $r = 0.30$; $P = 0.16$; Reynold's distance: $r = 0.37$; $P = 0.10$; G'_{ST}: $r = -0.01$; $P = 0.54$; Jost's *D*: $r = 0.09$; $P = 0.33$). For nuSSR markers, Mantel tests showed that niche overlap was not significantly correlated with pairwise genetic differentiation estimated with F_{ST} (*I*: $r = -0.17$; $P = 0.28$; *D*: $r = -0.20$; $P = 0.25$), Reynold's distance (*I*: $r = -0.17$; $P = 0.28$; *D*: $r = -0.20$; $P = 0.26$), G'_{ST} (*I*: $r = -0.05$; $P = 0.46$; *D*: $r = -0.08$; $P = 0.43$) or Jost's *D* (*I*: $r = -0.04$; $P = 0.47$; *D*: $r = -0.07$; $P = 0.44$). For cpSSR markers, Mantel tests did not show either a significant correlation between niche overlap and pairwise genetic differentiation estimated with F_{ST} (*I*: $r = -0.26$; $P = 0.18$; *D*: $r = -0.27$; $P = 0.18$), Reynold's distance (*I*: $r = -0.25$; $P = 0.18$; *D*: $r = -0.27$; $P = 0.19$), G'_{ST} (*I*: $r = -0.28$; $P = 0.15$; *D*: $r = -0.29$; $P = 0.14$) or Jost's *D* (*I*: $r = -0.21$; $P = 0.22$; *D*: $r = -0.23$; $P = 0.22$). We obtained qualitatively similar results excluding from the analyses the soil-specialist *Q. durata* or when genetic distances were calculated excluding from the data set individuals with different degrees of admixed ancestry with other species (i.e. hybrids or introgressed individuals; see next section) (data not shown).

Discussion

Species evolutionary and demographic history

Demographic, environmental niche and SDM analyses suggest that different factors have probably driven the speciation process within the Californian scrub white oak complex. ABC analyses indicate that the most

plausible demographic scenario is the one considering the divergence between the *Quercus berberidifolia*/*Quercus durata* clade and the rest of the species during the early Wisconsinan glaciation (*c.* 79–91 k years BP) followed by a split *Q. berberidifolia* and *Q. durata* around the LGM (*c.* 23–26 k years BP). Hierarchical Bayesian clustering analyses, PCA, and SSR-based phylogenetic analyses suggest that *Quercus cornelius-mulleri* and *Quercus john-tuckeri*, on one hand, and *Quercus dumosa* and *Quercus pacifica*, on the other hand, diverged more recently than did both species pairs. The geographical ranges of these four species also suggest such a pattern of sequential divergence: species within both pairs of taxa have nonoverlapping or only partly overlapping distribution ranges but they are distributed immediately adjacent to each other (Fig. 1). However, ABC analyses suggest a simultaneous speciation among these four taxa around the LGM (*c.* 22 k years BP), a scenario also supported by pilot runs exclusively focused on these four species (data not shown). These discrepancies could have resulted from the low power of ABC analyses and our set of nuclear and chloroplast SSR markers to detect subtle differences in divergence times among taxa that have very recently split (Papadopoulou & Knowles 2015).

The ABC analyses suggest a recent split among the studied taxa, but our estimates of divergence time must be interpreted with caution due to the wide 95% CIs obtained for the inferred demographic parameters (Table 3), long overlaps between generations in tree species, and considerable uncertainty in average generation time (see Tsuda *et al.* 2015). Regarding the latter, we considered a generation time of 50 years for scrub oaks, which is half of the generation time generally assumed for large trees (e.g. oaks: Gugger *et al.* 2013; birches: Tsuda *et al.* 2015), but our confidence around this value is low due to the lack of direct estimates and the general controversy around its calculation in long-lived trees (Petit & Hampe 2006; Tsuda *et al.* 2015). Generation time may also potentially vary over time (e.g. in cold–dry vs. warm–wet periods) and among taxa inhabiting different environments (e.g. desert margins from southern California vs. mixed forest from northern California). Moreover, interspecific hybridization (see Discussion below) is expected to have resulted in our DIYABC analyses, which do not accommodate gene flow after divergence (Cornuet *et al.* 2014), have underestimated divergence times even though we excluded from the data set introgressed individuals (e.g. Tsuda *et al.* 2015).

Different lines of evidence suggest that the divergence between *Q. berberidifolia* and *Q. durata* probably took place in peripatry or sympatry. These two species share a similar distribution range and are the only two taxa within the complex that often form mixed stands or grow in very close geographical proximity, with populations separated by a few hundreds of metres (Forde & Faris 1962). *Quercus durata* is an edaphic specialist of serpentine soils, where *Q. berberidifolia* is unable to form stable populations (Forde & Faris 1962). These soils represent less than ~1% of California's land surface and are patchily distributed across the landscape (Whittaker 1954; Kruckeberg 1984; Harrison 1999; Harrison & Najakaruna 2011), which is likely to have favored the presence of populations of both species at a dispersal distance of each other and resulted in divergent selection in peripatry or sympatry (*sensu lato*, i.e. absence of physical barriers to gene flow; Papadopulos *et al.* 2011, 2013; Anacker 2014). Our data support a monophyletic origin of populations of *Q. durata*, indicating that populations of this species do not represent soil-adapted ecotypes evolved repeatedly from adjacent nonserpentine types as has been found in other Californian endemics (Rajakaruna *et al.* 2003; Brady *et al.* 2005). Thus, it is possible that a single divergence event associated with the adaptation to serpentine soils took place in allopatry and subsequent range expansion after divergence has resulted in current sympatry/peripatry with *Q. berberidifolia* (Harrison & Najakaruna 2011; Moyle *et al.* 2012; Anacker 2014).

More intricate is the evolutionary history of the four putative species within the *Q. dumosa* group (Figs 3–6). Results of SDMs and analyses of niche overlap suggest different modes of speciation within the *Q. dumosa* group after the divergence of its common ancestor from the *Q. berberidifolia*/*Q. durata* clade. Parapatric divergence followed by significant ecological differentiation has probably shaped the split between the closely related *Q. cornelius-mulleri* and *Q. john-tuckeri*. The distributions of these two species are adjacent to each other but scarcely overlapping (Fig. 1), not even during the LGM when *Q. john-tuckeri* seemed to experience a considerable range expansion (Fig. S4, Supporting information). Thus, their divergence is likely to have been favored by the Transverse Ranges, which separate most populations of both taxa and have been previously identified as an important geographical barrier contributing to genetic splits in oaks (Gugger *et al.* 2013; Ortego *et al.* 2015) and many other taxa (Calsbeek *et al.* 2003; Chatzimanolis & Caterino 2007; Vandergast *et al.* 2008). *Quercus dumosa* and *Q. cornelius-mulleri* have probably diverged in allopatry, with no range overlap and showing a remarkable niche differentiation. *Quercus dumosa* and *Q. john-tuckeri* show significant niche differentiation and distinct distributions in the present but SDM predict a considerable overlap during the LGM when the split was estimated to have taken place, suggesting that ecological divergence in parapatry is the

most plausible scenario under the simultaneous divergence suggested by ABC analyses.

Speciation in allopatry is a priori the most likely scenario of divergence for the island endemic *Q. pacifica*, but the relationship between this taxa and its most likely sibling species *Q. dumosa* was fairly complex. STRUCTURE analyses showed the presence of five genetic clusters, three mostly represented in *Q. dumosa* and two in *Q. pacifica*, with considerable degree of admixture (Figs 4g and 5b). ABC analyses focused on these two species revealed that a simultaneous split of *Q. dumosa* and *Q. pacifica* populations is the most supported scenario, suggesting a polyphyletic origin of *Q. pacifica* (Table 2b; Fig. 2b). SDMs indicate that the predicted distributions of both species considerably overlapped during the LGMs, which may have been favored by the fact that lower sea levels during the last glaciation almost connected all Northern Channel Islands with the mainland (Johnson 1978). Accordingly, ABC analyses indicate that divergence among populations of *Q. dumosa* and *Q. pacifica* may have taken place around the LGM (*c.* 19–26 k years BP), which could reflect population fragmentation as consequence of the progressively reduced connectivity among California Channel islands and the mainland due to the sea level rising during the early Holocene (Johnson 1978). Thus, the complicated demographic history and patterns of genetic structure observed in these two putative species may be related with the multiple episodes of population reconnections and admixture characteristics of species inhabiting aggregated island complexes linked to the sea level oscillations that took place during the Pleistocene (Esselstyn & Brown 2009; Papadopoulou & Knowles 2015).

The different modes of speciation evidenced by demographic analysis in conjunction with SDM and niche divergence tests are also supported by analyses of the association between genetic differentiation and niche overlap. We found no significant relationship between pairwise species genetic differentiation and environmental niche overlap (i.e. a lack of phylogenetic niche conservatism *sensu* Losos 2008), which indicates that related species do not show a tendency to occur sympatrically, that environmental niches are highly labile (e.g. Cornuault *et al.* 2015) and, most likely, that different mechanisms (e.g. geographic isolation in vicariance vs. ecological divergence in sympatry or peripatry) have driven speciation within the Californian scrub white oak complex (Losos 2008; Warren *et al.* 2008; Nakazato *et al.* 2010). However, given that these analyses cannot tease apart the geographic signals of evolutionary and ecological processes, our results must be interpreted with caution in terms of the underlying factors that may have driven the observed biogeographic patterns (Warren *et al.* 2014).

Distribution of genetic variation in nuclear and chloroplast genomes

Our comparative analyses on chloroplast and nuclear genomes revealed different demographic histories and highly contrasting patterns in the spatial distribution of genetic variation for each, which is expected given that they have different forms of inheritance, mutation rates, and dispersal vectors (Petit *et al.* 2005). Interspecific genetic differentiation was not significantly correlated between both genomes (Table S5, Supporting information) and AMOVA analyses showed that most variation for cpSSR markers was explained by differences among populations whereas the largest proportion of total genetic variation for nuSSR markers was attributed to differences among individuals within populations (Table 1). AMOVA analyses also showed that only a low proportion of variance was explained by differences among species (<10% for both genomes; Table 1), which contrasts with previous studies comparing other closely related tree species (e.g. Du *et al.* 2015). The low degree of genetic differentiation among populations and the high levels of genetic diversity within populations for nuSSR loci relative to cpSSR loci probably reflect high rates of pollen-mediated gene flow via wind vs. the more restricted gene flow through seed dispersal by animals, a pattern well documented in previous studies analysing contemporary processes of pollen and seed dispersal (e.g. Dow & Ashley 1998; Ortego *et al.* 2014b; Sork *et al.* 2015). Conversely, the high genetic differentiation among populations for cpSSR markers is probably reflecting the fact that most populations have been monopolized by one or a few chloroplast haplotypes (see haplotype network and BAPS analyses; Figs 3a and 6), which has been reported in previous studies on the Californian Valley oak (*Quercus lobata*) and is congruent with low rates of acorn dispersal (Grivet *et al.* 2005, 2006, 2009). Thus, our results are in agreement with previous studies showing strong local genetic structure for maternally inherited chloroplast genomes (e.g. Grivet *et al.* 2005, 2006; Gugger *et al.* 2013) and pervasive haplotype sharing among sympatric populations of different species that yet maintain their taxonomic, morphological, and nuclear genetic identity in presence of frequent interspecific hybridization (e.g. Whittemore & Schaal 1991; Petit *et al.* 2002, 2004; Lexer *et al.* 2006).

Interspecific hybridization

Oaks represent a paradigmatic example of the complexity around the biological species concept due to their high levels of interspecific gene flow (Coyne & Orr 2004; e.g. Howard *et al.* 1997; Curtu *et al.* 2007; Cavender-Bares & Pahlich 2009; Ortego & Bonal 2010), which

in turn has made the systematics of the genus challenging (Van Valen 1976; Manos *et al.* 1999; Hipp *et al.* 2014). Our studied species are not an exception and Bayesian clustering analyses have shown considerable rates of hybridization among different taxa within the complex (Fig. 3). Introgression was mostly detected in zones of sympatry/parapatry between two or more taxa, as is well exemplified in the codistributed populations of *Q. berberidifolia* and *Q. cornelius-mulleri* in southern California (Riordan *et al.* 2015) and *Q. berberidifolia* and *Q. durata* in central-northern California. Although we found some populations that constitute hybrid swarms (e.g. site 63; Figs 3b and 6), all the putative species maintain their taxonomic and genetic identity across their distribution ranges (Muir *et al.* 2000; Ortego *et al.* 2014a). Differences in flowering times among taxa or selection against hybrids due to their lower viability or performance in either parental environment may be responsible of maintaining the distinctness of the studied species in the presence of gene flow (Forde & Faris 1962; Muir *et al.* 2000; Ortego *et al.* 2014a). With the exception of *Q. dumosa* and *Q. pacifica*, our analyses also indicate a lack of intraspecific genetic structure across the entire ranges of the different taxa, even when some species (*Q. berberidifolia*, *Q. durata*; Fig. 1) are distributed over a large geographic area for which previous studies have found genetic subdivision in other codistributed oaks (Ortego *et al.* 2012; Gugger *et al.* 2013; Ortego *et al.* 2015; Fig. 1). This suggests that widespread gene flow in these wind-pollinated trees is a remarkable homogenizing factor across populations within each studied species, but selection against hybrids or assortative mating is probably preventing all species from converging in a hybrid swarm (Muir *et al.* 2000).

Conclusions

Overall, our study exemplifies the complex processes that can underlie early stages of species formation and suggest that different processes leading to divergence are likely to be involved in speciation within the Californian scrub white oak complex. With the exception of the island endemic *Quercus pacifica* and the closely related *Quercus dumosa*, our analyses support the taxonomic distinctiveness of all the species from either a molecular and/or an ecological perspective. Further studies taking advantage of high-throughput sequencing technology can provide in the future a higher resolution to address in more detail the demographic history of the complex (Emerson *et al.* 2010; Papadopoulou & Knowles 2015) and understand if some loci under selection related with adaptation to particular environmental conditions (e.g. serpentine soils in *Quercus durata* and xeric conditions in *Quercus cornelius-mulleri*) are involved in species divergence, which may explain the subtle genetic differentiation observed among the different taxa when examined at neutral markers (Funk *et al.* 2012; Shafer & Wolf 2013; e.g. Hohenlohe *et al.* 2010; Gompert *et al.* 2013; Papadopulos *et al.* 2013).

Acknowledgements

We thank the staff of the University of California Natural Reserve System at Santa Cruz Island, The Conservancy at Santa Catalina Island and the US National Park Service for providing sampling permissions and valuable logistic support during field work. We also wish to thank Isabel Afán (LAST-EBD) for her valuable help in GIS analyses, Mario B. Pesendorfer for providing samples from Santa Rosa Island and Juan Miguel Arroyo and Néstor Pérez for their advice during genotyping. Prof. Rémy J. Petit and three anonymous referees provided useful discussion and valuable comments on an earlier draft of this manuscript. JO was supported by a Ramón y Cajal Fellowship (RYC-2013-12501) and a research contract funded by Severo Ochoa Program (SEV-2012-0262). VN was supported by a FPI predoctoral fellowship from Ministerio de Economía y Competitividad. This work was supported by an internal EBD 'Microproyectos' grant to JO, financed by the Spanish Ministry of Economy and Competitiveness through the Severo Ochoa Program for Centres of Excellence in R + D + I (SEV-2012-0262).

References

Anacker BL (2014) The nature of serpentine endemism. *American Journal of Botany*, **101**, 219–224.

Anderson E (1953) Introgressive hybridization. *Biological Reviews of the Cambridge Philosophical Society*, **28**, 280–307.

Antao T, Lopes A, Lopes RJ, Beja-Pereira A, Luikart G (2008) LOSITAN: a workbench to detect molecular adaptation based on a F_{ST}-outlier method. *BMC Bioinformatics*, **9**, 323.

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

Beaumont MA (2010) Approximate Bayesian computation in evolution and ecology. *Annual Review of Ecology, Evolution and Systematics*, **41**, 379–406.

Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society of London B: Biological Sciences*, **263**, 1619–1626.

Beaumont MA, Rannala B (2004) The Bayesian revolution in genetics. *Nature Reviews Genetics*, **5**, 251–261.

Beaumont MA, Zhang WY, Balding DJ (2002) Approximate Bayesian computation in population genetics. *Genetics*, **162**, 2025–2035.

Bertorelle G, Benazzo A, Mona S (2010) ABC as a flexible framework to estimate demography over space and time: some cons, many pros. *Molecular Ecology*, **19**, 2609–2625.

Besnard G, Dupuy J, Larter M *et al.* (2014) History of the invasive African olive tree in Australia and Hawaii: evidence for sequential bottlenecks and hybridization with the Mediterranean olive. *Evolutionary Applications*, **7**, 195–211.

Bonnet E, Van de Peer Y (2002) ZT: a software tool for simple and partial Mantel tests. *Journal of Statistical Software*, **10**, 1–12.

Brady KU, Kruckeberg AR, Bradshaw HD (2005) Evolutionary ecology of plant adaptation to serpentine soils. *Annual Review of Ecology Evolution and Systematics*, **36**, 243–266.

Butlin R, Debelle A, Kerth C *et al.* (2012) What do we need to know about speciation? *Trends in Ecology & Evolution*, **27**, 27–39.

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 LL, Edwards AWF (1967) Phylogenetic analysis models and estimation procedures. *American Journal of Human Genetics*, **19**, 233–257.

Cavender-Bares J, Pahlich A (2009) Molecular, morphological and ecological niche differentiation of sympatric sister oak species, *Quercus virginiana* and *Q. geminata* (Fagaceae). *American Journal of Botany*, **96**, 1690–1702.

Cavender-Bares J, Ackerly DD, Baum DA, Bazzaz FA (2004) Phylogenetic overdispersion in Floridian oak communities. *The American Naturalist*, **163**, 823–843.

Chatzimanolis S, Caterino MS (2007) Toward a better understanding of the "Transverse Range Break": lineage diversification in southern California. *Evolution*, **61**, 2127–2141.

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.

Cornuault J, Khimoun A, Cuneo P, Besnard G (2015) Spatial segregation and realized niche shift during the parallel invasion of two olive subspecies in south-eastern Australia. *Journal of Biogeography*, **42**, 1930–1941.

Cornuet JM, Santos F, Beaumont MA *et al.* (2008) Inferring population history with DIY ABC: a user-friendly approach to approximate Bayesian computation. *Bioinformatics*, **24**, 2713–2719.

Cornuet JM, Ravigne V, Estoup A (2010) Inference on population history and model checking using DNA sequence and microsatellite data with the software DIYABC (v1.0). *BMC Bioinformatics*, **11**, 401.

Cornuet JM, Pudlo P, Veyssier J *et al.* (2014) DIYABC v2.0: a software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism, DNA sequence and microsatellite data. *Bioinformatics*, **30**, 1187–1189.

Coyne JA, Orr HA (2004) *Speciation*. Sinauer Associates, Sunderland, Massachusetts.

Curtu AL, Gailing O, Finkeldey R (2007) Evidence for hybridization and introgression within a species-rich oak (*Quercus* spp.) community. *BMC Evolutionary Biology*, **7**, 284.

Davis EB, Koo MS, Conroy C, Patton JL, Moritz C (2008) The California Hotspots Project: identifying regions of rapid diversification of mammals. *Molecular Ecology*, **17**, 120–138.

Dobzhansky T (1937) *Genetics and the Origin of Species*. Columbia University Press, New York, New York.

Dow BD, Ashley MV (1998) High levels of gene flow in bur oak revealed by paternity analysis using microsatellites. *Journal of Heredity*, **89**, 62–70.

Doyle J, Doyle J (1990) Isolation of plant DNA from fresh tissue. *Focus*, **12**, 13–15.

Du S, Wang Z, Ingvarsson PK *et al.* (2015) Multilocus analysis of nucleotide variation and speciation in three closely related *Populus* (Salicaceae) species. *Molecular Ecology*, **24**, 4994–5005.

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.

Eaton DAR, Ree RH (2013) Inferring phylogeny and introgression using RADseq data: an example from flowering plants (*Pedicularis*: *Orobanchaceae*). *Systematic Biology*, **62**, 689–706.

Emerson KJ, Merz CR, Catchen JM *et al.* (2010) Resolving postglacial phylogeography using high-throughput sequencing. *Proceedings of the National Academy of Sciences of the USA*, **107**, 16196–16200.

Esselstyn JA, Brown RM (2009) The role of repeated sea-level fluctuations in the generation of shrew (Soricidae: Crocidura) diversity in the Philippine Archipelago. *Molecular Phylogenetics and Evolution*, **53**, 171–181.

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, Laval G, Schneider S (2005) ARLEQUIN ver. 3.0: an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*, **1**, 47–50.

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.

eFloras (2015) *Flora of North America*, Missouri Botanical Garden, St. Louis, Missouri, and Harvard University Herbaria, Cambridge, Massachusetts. Available from http://www.efloras.org.

Fontaine MC, Austerlitz F, Giraud T *et al.* (2013) Genetic signature of a range expansion and leap-frog event after the recent invasion of Europe by the grapevine downy mildew pathogen *Plasmopara viticola*. *Molecular Ecology*, **22**, 2771–2786.

Forde MB, Faris DG (1962) Effect of introgression on serpentine endemism of *Quercus durata*. *Evolution*, **16**, 338–347.

Funk WC, McKay JK, Hohenlohe PA, Allendorf FW (2012) Harnessing genomics for delineating conservation units. *Trends in Ecology & Evolution*, **27**, 489–496.

Gompert Z, Lucas LK, Nice CC, Buerkle CA (2013) Genome divergence and the genetic architecture of barriers to gene flow between *Lycaeides idas* and *L. melissa*. *Evolution*, **67**, 2498–2514.

Graham CH, Ron SR, Santos JC, Schneider CJ, Moritz C (2004) Integrating phylogenetics and environmental niche models to explore speciation mechanisms in dendrobatid frogs. *Evolution*, **58**, 1781–1793.

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.

Grivet D, Robledo-Arnuncio JJ, Smouse PE, Sork VL (2009) Relative contribution of contemporary pollen and seed dispersal to the effective parental size of seedling population of

California valley oak (*Quercus lobata*, Née). *Molecular Ecology*, **18**, 3967–3979.

Gugger PF, Ikegami M, Sork VL (2013) Influence of late Quaternary climate change on present patterns of genetic variation in valley oak, *Quercus lobata* Née. *Molecular Ecology*, **22**, 3598–3612.

Guo SW, Thompson EA (1992) A monte-carlo method for combined segregation and linkage analysis. *American Journal of Human Genetics*, **51**, 1111–1126.

Harrison S (1999) Local and regional diversity in a patchy landscape: native, alien, and endemic herbs on serpentine. *Ecology*, **80**, 70–80.

Harrison S, Najakaruna N (2011) *Serpentine: The Evolution and Ecology of a Model System*. University of California Press, Los Angeles, California.

Hasumi H, Emori S (eds) (2004) *K-1 Coupled GCM (MIROC) Description*. K-1 Technical Report No. 1. Center for Climate System Research, University of Tokyo, Tokyo, Japan.

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.

Hipp AL, Eaton DAR, Cavender-Bares J *et al.* (2014) A framework phylogeny of the American oak clade based on sequenced RAD data. *PLoS ONE*, **9**, e93975.

Hohenlohe PA, Bassham S, Etter PD *et al.* (2010) Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genetics*, **6**, e1000862.

Howard DJ, Preszler RW, Williams J, Fenchel S, Boecklen WJ (1997) How discrete are oak species? Insights from a hybrid zone between *Quercus grisea* and *Quercus gambelii*. *Evolution*, **51**, 747–755.

Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources*, **9**, 1322–1332.

IUCN (2010) *IUCN red list of threatened species*. Version 2010.4. Available from http://www.iucnredlist.org. Downloaded on 31 July 2015.

Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, **23**, 1801–1806.

Johnson DL (1978) Origin of island mammoths and quaternary land bridge history of Northern Channel Islands, California. *Quaternary Research*, **10**, 204–225.

Jombart T (2008) Adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, **24**, 1403–1405.

Knouft JH, Losos JB, Glor RE, Kolbe JJ (2006) Phylogenetic analysis of the evolution of the niche in lizards of the *Anolis sagrei* group. *Ecology*, **87**, S29–S38.

Kruckeberg AR (1984) *California Serpentines: Flora, Vegetation, Geology, Soils, and Management Problems*. University of California Publications in Botany, Los Angeles, California.

Lancaster LT, Kay KM (2013) Origin and diversification of the California flora: re-examining classic hypotheses with molecular phylogenies. *Evolution*, **67**, 1041–1054.

Langella O (2007) *POPULATIONS 1.2.30: population genetic software, individuals or populations distances, phylogenetic trees*. Available from http://bioinformatics.org/~tryphon/populations/.

Langin KM, Sillett TS, Funk WC *et al.* (2015) Islands within an island: repeated adaptive divergence in a single population. *Evolution*, **69**, 653–665.

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

Lexer C, Kremer A, Petit RJ (2006) Shared alleles in sympatric oaks: recurrent gene flow is a more parsimonious explanation than ancestral polymorphism. *Molecular Ecology*, **15**, 2007–2012.

Liu CP, Tsuda Y, Shen HL *et al.* (2014) Genetic structure and hierarchical population divergence history of *Acer mono* var. *mono* in South and Northeast China. *PLoS ONE*, **9**, e87187.

Losos JB (2008) Phylogenetic niche conservatism, phylogenetic signal and the relationship between phylogenetic relatedness and ecological similarity among species. *Ecology Letters*, **11**, 995–1003.

Manos PS, Doyle JJ, Nixon KC (1999) Phylogeny, biogeography, and processes of molecular differentiation in *Quercus* subgenus *Quercus* (Fagaceae). *Molecular Phylogenetics and Evolution*, **12**, 333–349.

Mao JF, Wang XR (2011) Distinct niche divergence characterizes the homoploid hybrid speciation of *Pinus densata* on the Tibetan Plateau. *The American Naturalist*, **177**, 424–439.

Mayr E (1942) *Systematics and the Origin of Species*. Columbia University Press, New York, New York.

McCormack JE, Zellmer AJ, Knowles LL (2010) Does niche divergence accompany allopatric divergence in *Aphelocoma* jays as predicted under ecological speciation?: insights from tests with niche models. *Evolution*, **64**, 1231–1244.

Meirmans PG, Hedrick PW (2011) Assessing population structure: F_{ST} and related measures. *Molecular Ecology Resources*, **11**, 5–18.

Moyle LC, Levine M, Stanton ML, Wright JW (2012) Hybrid sterility over tens of meters between ecotypes adapted to serpentine and non-serpentine soils. *Evolutionary Biology*, **39**, 207–218.

Muir G, Fleming CC, Schlotterer C (2000) Taxonomy—species status of hybridizing oaks. *Nature*, **405**, 1016.

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

Nakazato T, Warren DL, Moyle LC (2010) Ecological and geographic models of species divergence in wild tomatoes. *American Journal of Botany*, **97**, 680–693.

Nixon K (2002) The oak (*Quercus*) biodiversity of California and adjacent regions. In: *Proceedings of the Fifth Symposium on Oak Woodlands: Oaks in California's Changing Landscape, October 22–25, 2001, San Diego, California* (eds Standiford RB, McCreary D, Purcell KL), pp. 3–20. USDA Forest Service General Technical Report PSW-GTR-184, San Diego, California.

Nixon KC, Muller CH (1994) New names in California oaks. *Novon*, **4**, 391–393.

Nixon KC, Steele KP (1981) A new species of *Quercus* (Fagaceae) from southern California. *Madroña*, **28**, 210–219.

Nosil P (2012) *Ecological Speciation*. Oxford University Press, New York, New York.

Orr MR, Smith TB (1998) Ecology and speciation. *Trends in Ecology and Evolution*, **13**, 502–506.

Ortego J, Bonal R (2010) Natural hybridisation between kermes (*Quercus coccifera* L.) and holm oaks (*Q. ilex* L.) revealed by microsatellite markers. *Plant Biology*, **12**, 234–238.

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.

Ortego J, Gugger PF, Riordan EC, Sork VL (2014a) Influence of climatic niche suitability and geographical overlap on hybridization patterns among southern Californian oaks. *Journal of Biogeography*, **41**, 1895–1908.

Ortego J, Bonal R, Munoz A, Aparicio JM (2014b) Extensive pollen immigration and no evidence of disrupted mating patterns or reproduction in a highly fragmented holm oak stand. *Journal of Plant Ecology*, **7**, 384–395.

Ortego J, Gugger PF, Sork VL (2015) Climatically stable landscapes predict patterns of genetic structure and admixture in the Californian canyon live oak. *Journal of Biogeography*, **42**, 328–338.

Papadopoulou A, Knowles LL (2015) Genomic tests of the species-pump hypothesis: recent island connectivity cycles drive population divergence but not speciation in Caribbean crickets across the Virgin Islands. *Evolution*, **69**, 1501–1517.

Papadopulos AST, Baker WJ, Crayn D *et al.* (2011) Speciation with gene flow on Lord Howe Island. *Proceedings of the National Academy of Sciences of the USA*, **108**, 13188–13193.

Papadopulos AST, Price Z, Devaux C *et al.* (2013) A comparative analysis of the mechanisms underlying speciation on Lord Howe Island. *Journal of Evolutionary Biology*, **26**, 733–745.

Petit RJ, Excoffier L (2009) Gene flow and species delimitation. *Trends in Ecology & Evolution*, **24**, 386–393.

Petit RJ, Hampe A (2006) Some evolutionary consequences of being a tree. *Annual Review of Ecology Evolution and Systematics*, **37**, 187–214.

Petit RJ, Csaikl UM, Bordacs S *et al.* (2002) 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.

Petit RJ, Bodenes C, Ducousso A, Roussel G, Kremer A (2004) Hybridization as a mechanism of invasion in oaks. *New Phytologist*, **161**, 151–164.

Petit RJ, Duminil J, Fineschi S *et al.* (2005) Comparative organization of chloroplast, mitochondrial and nuclear diversity in plant populations. *Molecular Ecology*, **14**, 689–701.

Phillips SJ, Dudik M (2008) Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography*, **31**, 161–175.

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

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

R Core Team (2015) *R: A Language and Environment for Statistical Computing*. Version 3.0.2. R Foundation for Statistical Computing, Vienna, Austria. Available from http://www.r-project.org.

Rajakaruna N, Baldwin BG, Chan R *et al.* (2003) Edaphic races and phylogenetic taxa in the *Lasthenia californica* complex (Asteraceae: Heliantheae): an hypothesis of parallel evolution. *Molecular Ecology*, **12**, 1675–1679.

Raven PH, Axelrod DI (1978) Origin and relationships of the California flora. *University of California Publications in Botany*, **72**, 1–134.

Ribera I, Castro A, Diaz JA *et al.* (2011) The geography of speciation in narrow-range endemics of the 'Haenydra' lineage (Coleoptera, Hydraenidae, Hydraena). *Journal of Biogeography*, **38**, 502–516.

Rice WR (1989) Analyzing tables of statistical tests. *Evolution*, **43**, 223–225.

Riordan EC, Gugger PF, Ortego J *et al.* (2015) Association of genetic and phenotypic variability with geography and climate in three southern California oaks (Fagaceae). *American Journal of Botany*, in press.

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. *The American Naturalist*, **167**, 655–666.

Robert CP, Cornuet JM, Marin JM, Pillai NS (2011) Lack of confidence in approximate Bayesian computation model choice. *Proceedings of the National Academy of Sciences of the USA*, **108**, 15112–15117.

Roberts FM (1995) *Illustrated Guide to the Oaks of the Southern California Floristic Province*. F.M. Roberts Publications, Encinitas, California.

Rosenberg NA (2004) DISTRUCT: a program for the graphical display of population structure. *Molecular Ecology Notes*, **4**, 137–138.

Rundell RJ, Price TD (2009) Adaptive radiation, nonadaptive radiation, ecological speciation and nonecological speciation. *Trends in Ecology & Evolution*, **24**, 394–399.

Sawyer JO, Keeler-Wolf T (1995) *A Manual of California Vegetation*. California Native Plant Society, Sacramento, California.

Schluter D (2001) Ecology and the origin of species. *Trends in Ecology & Evolution*, **16**, 372–380.

Schluter D (2009) Evidence for ecological speciation and its alternative. *Science*, **323**, 737–741.

Schoener TW (1968) Anolis lizards of Bimini—resource partitioning in a complex fauna. *Ecology*, **49**, 704–726.

Shafer ABA, Wolf JBW (2013) Widespread evidence for incipient ecological speciation: a meta-analysis of isolation-by-ecology. *Ecology Letters*, **16**, 940–950.

Simpson GG (1953) *The Major Features of Evolution*. Columbia University Press, New York, New York.

Singhal S, Moritz C (2012) Strong selection against hybrids maintains a narrow contact zone between morphologically cryptic lineages in a rainforest lizard. *Evolution*, **66**, 1474–1489.

Sork VL, Werth S (2014) Phylogeography of *Ramalina menziesii*, a widely distributed lichen-forming fungus in western North America. *Molecular Ecology*, **23**, 2326–2339.

Sork VL, Scofield D, Grivet D, Smouse PE (2015) Consequences of dispersal for genetic diversity in seedling in a California savanna oak species, *Quercus lobata* Née. *Evolutionary Ecology*, in press.

Takezaki N, Nei M (1996) Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. *Genetics*, **144**, 389–399.

Tsuda Y, Nakao K, Ide Y, Tsumura Y (2015) The population demography of *Betula maximowicziana*, a cool-temperate tree species in Japan, in relation to the last glacial period: its admixture-like genetic structure is the result of simple population splitting not admixing. *Molecular Ecology*, **24**, 1403–1418.

Van Valen L (1976) Ecological species, multispecies, and oaks. *Taxon*, **25**, 233–239.

Vandergast AG, Bohonak AJ, Hathaway SA, Boys J, Fisher RN (2008) Are hotspots of evolutionary potential adequately protected in southern California? *Biological Conservation*, **141**, 1648–1664.

Warren DL, Glor RE, Turelli M (2008) Environmental niche equivalency versus conservatism: quantitative approaches to niche evolution. *Evolution*, **62**, 2868–2883.

Warren DL, Glor RE, Turelli M (2010) ENMTools: a toolbox for comparative studies of environmental niche models. *Ecography*, **33**, 607–611.

Warren DL, Cardillo M, Rosauer DF, Bolnick DI (2014) Mistaking geography for biology: inferring processes from species distributions. *Trends in Ecology & Evolution*, **29**, 572–580.

Whittaker RH (1954) The ecology of serpentine soils. *Ecology*, **35**, 258–259.

Whittemore AT, Schaal BA (1991) Interspecific gene flow in sympatric oaks. *Proceedings of the National Academy of Sciences of the USA*, **88**, 2540–2544.

J.O. and V.L.S. conceived the study. J.O., P.F.G. and V.L.S. collected the samples. J.O. and V.N. analysed the data. J.O. performed the genetic analyses. J.O. wrote the manuscript with inputs from all authors.

Data accessibility

Simple sequence repeat data, climate layers and occurrence data used for MAXENT and ENMTOOLS analyses, neighbour-joining phylogenetic trees, and input files used for PCAs, Mantel tests and analyses in POPULATIONS, STRUCTURE and BAPS are stored and accessible through the Dryad data repository: doi: 10.5061/dryad.52504.

Supporting information

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

Fig. S1 Results of Bayesian clustering analyses in STRUCTURE.

Fig. S2 Prior and posterior probability distribution of parameters calculated in DIYABC analyses based on nuclear SSR markers.

Fig. S3 Prior and posterior probability distribution of parameters calculated in DIYABC analyses based on both nuclear and chloroplast SSR markers.

Fig. S4 Predicted distribution for the six studied scrub white oaks from California during the Last Glacial Maximum (LGM; *c.* 21 000 BP).

Table S1 Geographical location of scrub white oak sampling sites in California.

Table S2 Nuclear and chloroplast SSR loci used to genotype Californian scrub white oaks.

Table S3 Summary of species distribution modeling (SDM) results.

Table S4 Number of individuals of each species assigned to each genetic cluster identified by BAPS analyses on chloroplast SSR markers.

Table S5 Pairwise F_{ST}, Reynold's distances, G'_{ST} and Jost's D between the studied Californian scrub white oak species for nuclear and chloroplast SSR markers.