New Phytologist

Genomic data reveal cryptic lineage diversification and introgression in Californian golden cup oaks (section *Protobalanus*)

Joaquín Ortego[1], Paul F. Gugger[2] and Victoria L. Sork[3,4]

[1]Department of Integrative Ecology, Estación Biológica de Doñana, EBD-CSIC, Avda. Américo Vespucio 26, Seville E-41092, Spain; [2]Appalachian Laboratory, University of Maryland Center for Environmental Science, 301 Braddock Road, Frostburg, MD 21532, USA; [3]Department of Ecology and Evolutionary Biology, University of California, Box 957239, Los Angeles, CA 90095, USA; [4]Institute of the Environment and Sustainability, University of California, Box 951496, Los Angeles, CA 90095-1496, USA

Author for correspondence:
Joaquín Ortego
Tel: +34 954 232 340
Email: joaquin.ortego@csic.es

Received: *17 October 2017*
Accepted: *14 November 2017*

New Phytologist (2018) **218:** 804–818
doi: 10.1111/nph.14951

Key words: California, ddRAD-seq, genetic admixture, hybridization, introgression, oak, *Quercus*.

Summary

• Here we study hybridization, introgression and lineage diversification in the widely distributed canyon live oak (*Quercus chrysolepis*) and the relict island oak (*Q. tomentella*), two Californian golden cup oaks with an intriguing biogeographical history.

• We employed restriction-site-associated DNA sequencing and integrated phylogenomic and population genomic analyses to study hybridization and reconstruct the evolutionary past of these taxa.

• Our analyses revealed the presence of two cryptic lineages within *Q. chrysolepis*. One of these lineages shares its most recent common ancestor with *Q. tomentella*, supporting the paraphyly of *Q. chrysolepis*. The split of these lineages was estimated to take place during the late Pliocene or the early Pleistocene, a time corresponding well with the common presence of *Q. tomentella* in the fossil records of continental California. Analyses also revealed historical hybridization among lineages, high introgression from *Q. tomentella* into *Q. chrysolepis* in their current area of sympatry, and widespread admixture between the two lineages of *Q. chrysolepis* in contact zones.

• Our results support that the two lineages of *Q. chrysolepis* behave as a single functional species phenotypically and ecologically well differentiated from *Q. tomentella*, a situation that can be only accommodated considering hybridization and speciation as a continuum with diffuse limits.

Introduction

The hybridization continuum (*sensu* Hochkirch, 2013) ranges from sporadic interspecific gene flow, with little or no impact on the evolutionary and demographic trajectories of the taxa involved, to the formation of new species reproductively isolated from parental forms (Barton, 2001; Abbott *et al.*, 2013). At one end of the range, hybrids can constitute an ephemeral state through which species are able to colonize new habitats (Bacilieri *et al.*, 1996; Petit *et al.*, 2004) and exchange alleles involved in the adaptation to novel environmental conditions (Lewontin & Birch, 1966; Baskett & Gomulkiewicz, 2011) without compromising the genetic and ecological distinctiveness of parental taxa (i.e. collective evolution; Morjan & Rieseberg, 2004). At the opposite end, hybridization has been recognized to be an important engine of diversification in many organism groups (Dowling & Secor, 1997; Rieseberg, 1997), and involved in the functional extinction of rare species through demographic swamping and genetic assimilation by more abundant species (Levin *et al.*, 1996; Rhymer & Simberloff, 1996). Hybrid speciation mechanisms include allopolyploid speciation, which generally results in the formation of new hybrid species with strong reproductive barriers that prevent interbreeding between the hybrid and its parents, and homoploid hybrid speciation, which leads to the formation of stable hybrid species that are often reproductively compatible with the parental forms (Rieseberg, 1997; Wood *et al.*, 2009; Abbott *et al.*, 2013). The hybridization/divergence continuum also comprises a broad range of intermediate situations (Hochkirch, 2013; Edwards *et al.*, 2016), including asymmetrical contributions of parental taxa to the genetic composition of hybrid species/lineages (e.g. Sun *et al.*, 2014), low levels of historical introgression followed by reproductive isolation (e.g. Eaton & Ree, 2013; Escudero *et al.*, 2014; Eaton *et al.*, 2015) or strong introgression or resistance to introgression limited to specific genomic regions with important consequences on fitness (Fitzpatrick *et al.*, 2010; Poelstra *et al.*, 2014; Liu *et al.*, 2015).

Disentangling reticulate evolutionary histories of either stable hybrid species or introgressed lineages has been challenging for a long time due to the inherent limitations of phylogenetic tools to deal with nonstrictly bifurcating processes of species

diversification, the poor resolution of small subsets of genetic markers to discern introgression from ancestral polymorphism, and the scarce integration of population genetic approaches into phylogenetic reconstruction (Linder & Rieseberg, 2004; McBreen & Lockhart, 2006; Abbott *et al.*, 2016; Payseur & Rieseberg, 2016; McVay *et al.*, 2017). This task is even more challenging when one of the parental taxa is extinct (e.g. Wang *et al.*, 1990; Currat & Excoffier, 2011) or locally extinct (Melo-Ferreira *et al.*, 2009), when hybridization only involves populations from a portion of the geographical distribution of the focal taxon (e.g. Nettel *et al.*, 2008; Melo-Ferreira *et al.*, 2009; Wall *et al.*, 2013; de Manuel *et al.*, 2016) or if introgression has induced phenotypic assimilation (Rheindt *et al.*, 2014; Huang, 2016).

Here, we employ genomic tools to study lineage diversification, introgression and hybridization between the canyon live oak (*Quercus chrysolepis* Liebmann) and the island oak (*Q. tomentella* Engelmann), two Californian golden cup oaks (section *Protobalanus*) with an intriguing biogeographical history (Muller, 1967; Nixon, 2002; eFloras, 2017). Canyon live oak is a widely distributed tree in California, with relictual populations in Arizona, Nevada, New Mexico, Oregon and northern Baja California (Thornburgh, 1990; eFloras, 2017). Previous microsatellite-based studies have shown that this species presents a deep genetic structure, with two genetic clusters separating populations located north and south of the Transverse ranges (Ortego *et al.*, 2015a; Bemmels *et al.*, 2016). This phylogeographical subdivision contrasts with the shallow patterns of genetic differentiation found among other Californian oak species with similar distribution ranges, including both red oaks (section *Lobatae*; Dodd & Kashani, 2003) and white oaks (section *Quercus*; Ashley *et al.*, 2015; Fitz-Gibbon *et al.*, 2017; Ortego *et al.*, 2017; but see Gugger *et al.*, 2013). Contrasting with the wide distribution of *Q. chrysolepis* across continental California, *Q. tomentella* is currently found in small populations confined to Guadalupe Island (Mexico) and the Channel Islands off the coast of southern California (USA) (Muller, 1967; Nixon, 2002; eFloras, 2017), which has motivated its inclusion in the IUCN Red List of Threatened Species with the status 'endangered' (Beckman & Jerome, 2017). Unlike many other endemic taxa from the Channel Islands that have speciated *in situ* (Thorne, 1969; Backs & Ashley, 2016; Riley & McGlaughlin, 2016), the widespread representation of *Q. tomentella* in late Tertiary fossil floras from continental California supports its origin from a previously more broadly distributed taxon that included the mainland (Axelrod, 1944a,b, 1967; Muller, 1967). Canyon live oak is also known to be present in some of the California Channel Islands, where it occurs as scattered individuals at the highest elevations and often showing strong phenotypic signs of introgression with *Q. tomentella* (Muller, 1967; Thorne, 1969; Nixon, 2002; Ashley *et al.*, 2010; eFloras, 2017).

In this study, we use restriction-site-associated DNA sequencing (ddRAD-seq; Peterson *et al.*, 2012) and a suite of population and phylogenomic analytical approaches to unravel the evolutionary and hybridization history of *Q. chrysolepis* and *Q. tomentella*. In particular, we (1) first use genome-wide single-nucleotide polymorphisms (SNP) data to test for hybridization between the two species in their overlapping distribution at some of the Californian Channel Islands and confirm the presence of interspecific gene flow suggested by previous anecdotal observations and studies (Muller, 1967; Nixon, 2002; Ashley *et al.*, 2010; eFloras, 2017). Second, we employ (2) a coalescent-based simulation framework to infer the timing and directionality of gene flow between *Q. tomentella* and continental populations *Q. chrysolepis*. We hypothesize that interspecific gene flow may have occurred due to increased geographical contact between the two species during the Miocene/Pliocene epochs (23.03–2.58 Myr ago (Ma)), when *Q. tomentella* was widely distributed in continental California (Axelrod, 1944a,b; Muller, 1967; eFloras, 2017), and/or during the Pleistocene glacial periods (2.59–0.01 Ma), when the lower sea levels almost connected the northern Channel Islands (Santa Rosa, Santa Cruz and Anacapa) to the mainland (Johnson, 1978). We (3) also explore whether the deep genetic structure of *Q. chrysolepis* reported in previous studies (Ortego *et al.*, 2015a; Bemmels *et al.*, 2016) represents a phylogeographical break driven by the past geological history of the region and/or hides a cryptic history of hybridization or introgression with *Q. tomentella* (Calsbeek *et al.*, 2003; Chatzimanolis & Caterino, 2007; Vandergast *et al.*, 2008). Finally, we (4) test whether lineages/populations involved in hybrid gene flow interactions show higher levels of genetic diversity than those that have not experienced genetic admixture (Nettel *et al.*, 2008; Streicher *et al.*, 2014).

Materials and Methods

Population sampling

Between 2010 and 2014, we sampled 12 continental populations of canyon live oak (*Quercus chrysolepis* Liebmann) ($n = 88$ individuals) and three populations of island oak (*Q. tomentella* Engelmann) ($n = 36$ individuals) from Santa Cruz, Santa Catalina and San Clemente Islands (Supporting Information Table S1). We also collected samples from *Q. chrysolepis* from Santa Cruz ($n = 9$) and San Clemente Islands ($n = 1$), most of them tentatively identified as hybrids with *Q. tomentella* based on their morphological appearance (eFloras, 2017). We designed sampling using occurrence records available in the Calflora database (http://www.calflora.org/) and aimed to collect samples representing populations located across the entire distribution of the two species in California (Fig. 1a; Table S1). Samples from Palmer oak (*Q. palmeri* Engelmann) ($n = 5$) and huckleberry oak (*Q. vaccinifolia* Kellogg) ($n = 5$), species also belonging to the section *Protobalanus* (eFloras, 2017), were used in phylogenomic analyses (see below). Spatial coordinates were registered using a Global Positioning System (GPS) unit and leaf samples were stored frozen (–20°C) until needed for genomic analyses (Table S1).

DNA extraction, library preparation and sequencing

We used a mixer mill to grind *c.* 50 mg of frozen leaf tissue in tubes with a tungsten bead and performed DNA extraction and

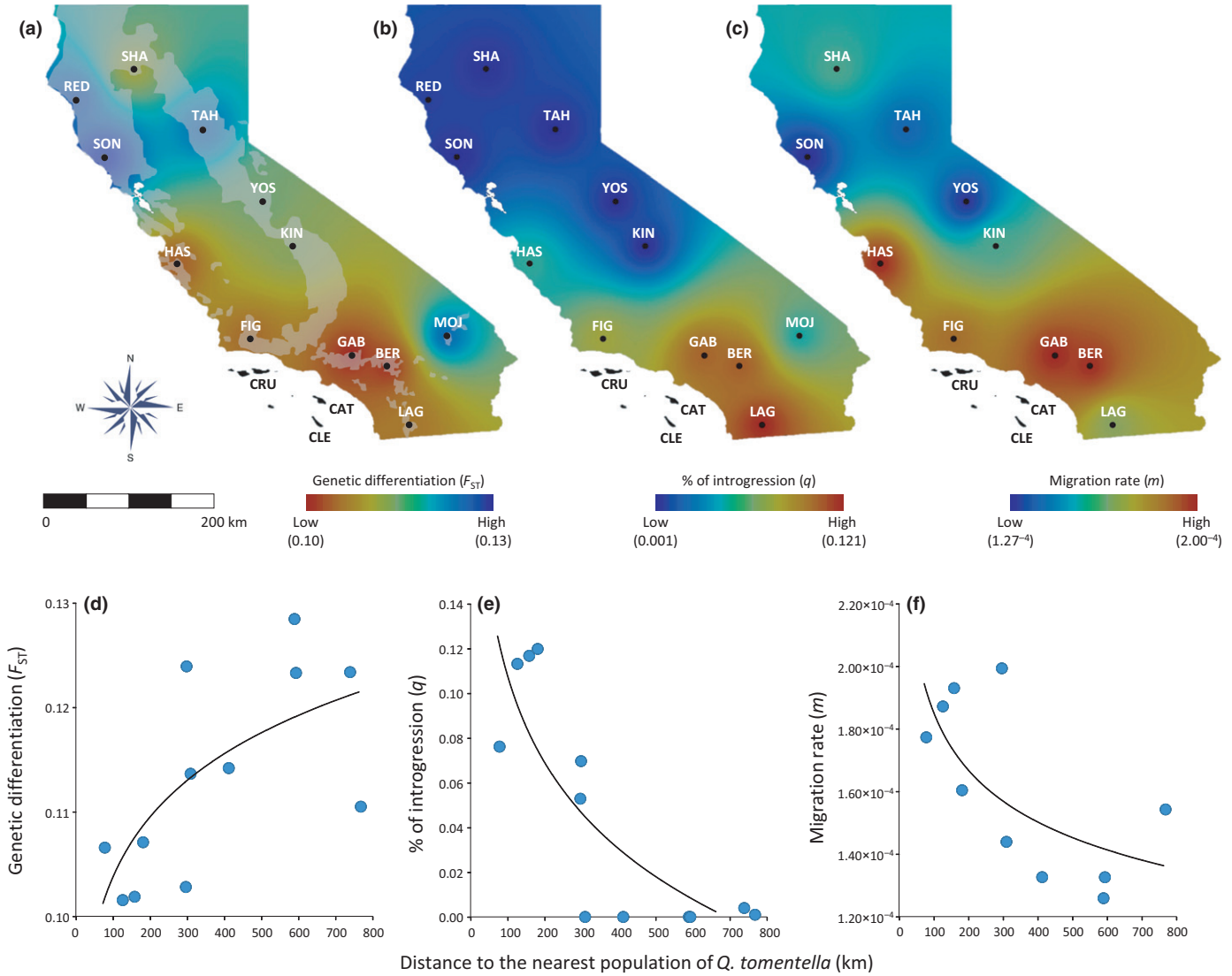

Fig. 1 Maps from California representing the studied populations of canyon live oak (*Quercus chrysolepis*) and island oak (*Q. tomentella*). Color gradients represent spatial interpolations for (a) pairwise genetic differentiation (F_{ST}) between *Q. tomentella* and each continental population of *Q. chrysolepis*, (b) probability of membership (*q*) of each continental population of *Q. chrysolepis* to the *Q. tomentella* genetic cluster as inferred by STRUCTURE analyses (i.e. % of introgression), and (c) pairwise migration rates (*m*) from *Q. tomentella* to each continental population of *Q. chrysolepis* as inferred by FASTSIMCOAL2 analyses. (d, e and f) These panels represent the same genetic parameters in relation to the geographical distance separating each continental population of *Q. chrysolepis* to the nearest population of *Q. tomentella*. Regression lines represent the function yielding the highest fit to observed data (d and f, power; e, logarithmic; see Supporting Information Table S2). Shadowed area in panel (a) shows the distribution of *Q. chrysolepis* in California based on occurrence data for the species obtained from sampling points and herbarium record databases (Ortego *et al.*, 2015a). Only populations with eight sampled individuals ($n = 10$ populations) were considered in analyses of migration rates with FASTSIMCOAL2 and, for this reason, localities MOJ ($n = 5$) and RED ($n = 3$) are not represented in panel (c). Population codes on maps are described in Table S1.

purification with NucleoSpin Plant II kits (Machery-Nagel, Düren, Germany). Genomic DNA was individually barcoded and processed into two libraries using the double-digestion restriction-fragment-based procedure (ddRAD-seq) described in Peterson *et al.* (2012). Briefly, DNA was double-digested using *Eco*RI and *Mse*I restriction enzymes (New England Biolabs, Ipswich, MA, USA), followed by the ligation of Illumina adaptors and unique 7-bp barcodes. Ligation products were pooled, size-selected between 350 and 450 bp using a Pippin Prep (Sage Science, Beverly, MA, USA) machine, and amplified by iProof™ High-Fidelity DNA Polymerase (Bio-Rad) with 12 cycles. Single-read 150-bp sequencing was performed on an Illumina HiSeq2500 platform at The Centre for Applied Genomics (Toronto, ON, Canada).

Bioinformatics and data filtering

We used both STACKS v.1.35 (Hohenlohe *et al.*, 2010; Catchen *et al.*, 2011, 2013) and PYRAD v.3.0.66 (Eaton, 2014) to assemble our sequences into *de novo* loci and call genotypes. This allowed us to examine the robustness of our analyses based on single nucleotide polymorphism (SNP) datasets obtained using

two of the most popular programs currently available to assemble RAD-seq data (Catchen *et al.*, 2011; Eaton, 2014). Phylogenetic, demographic and population structure analyses assume that the employed loci are selectively neutral (Luikart *et al.*, 2003; e.g. Guichoux *et al.*, 2013; Benestan *et al.*, 2016). For this reason, we identified loci putatively under selection and created SNP datasets only containing neutral loci (e.g. Brauer *et al.*, 2016) (see details in Methods S1). The choice of different filtering thresholds using either STACKS or PYRAD had little impact on the obtained inferences (see the Results section) (e.g. Eaton *et al.*, 2015). For this reason, all downstream analyses were performed using a SNP dataset obtained with STACKS and including only selectively neutral loci represented in at least five populations and half of the individuals of each population. SNP datasets are available at the Mendeley Data Repository (doi: 10.17632/g49jk9rwk9.2). See Methods S1 for additional details on sequence assembling and data filtering.

Genetic structure and hybrid identification

We identified hybrids and analysed population genetic structure and admixture using the Bayesian Markov chain Monte Carlo (MCMC) clustering method implemented in the program STRUCTURE v.2.3.3 (Pritchard *et al.*, 2000; Falush *et al.*, 2003; Hubisz *et al.*, 2009). We ran STRUCTURE using a random subset of 10 000 SNPs from six different datasets obtained with STACKS and PYRAD and considering different parameters ($P = 5$ and $P = 10$ for STACKS and $c = 0.85$ and $c = 0.90$ for PYRAD; see Methods S1 for further details). For each dataset, we ran STRUCTURE assuming correlated allele frequencies and admixture and without using prior population information (Hubisz *et al.*, 2009). We conducted 15 independent runs for each value of $K = 1$–8 to estimate the 'true' number of clusters with 200 000 MCMC cycles, following a burn-in step of 100 000 iterations. We retained the ten runs having the highest likelihood for each value of K and defined the number of populations best fitting the dataset using log probabilities of $X|K$ (Pritchard *et al.*, 2000) and the ΔK method (Evanno *et al.*, 2005), as implemented in STRUCTURE HARVESTER (Earl & vonHoldt, 2012). We used CLUMPP v.1.1.2 and the Greedy algorithm to align multiple runs of STRUCTURE for the same K value (Jakobsson & Rosenberg, 2007) and DISTRUCT v.1.1 (Rosenberg, 2004) to visualize as bar plots the individual's probabilities of population membership. Complementary to Bayesian clustering analyses, we performed a principal components analysis (PCA) using the R 3.3.2 (R Core Team, 2017) package ADEGENET (Jombart, 2008).

Coalescent analyses and model comparison

We compared different demographic models that considered three populations defined on the basis of STRUCTURE and PCA analyses (e.g. Eaton *et al.*, 2015; Lanier *et al.*, 2015). These analyses (1) revealed the presence of two lineages that separate populations of *Q. chrysolepis* located north and south of the Transverse Ranges and the Mohave Desert (hereafter referred to as northern lineage and southern lineage, respectively) and (2) suggested a certain degree of introgression (*c.* 7–10%) from *Q. tomentella* into the southern lineage of *Q. chrysolepis* (see the Results section for further details). In total, we tested six models that considered a combination of two migration matrices (total absence of post-divergence gene flow vs a full migration matrix of asymmetric gene flow) and three alternative scenarios of lineage divergence/formation (see Fig. 2).

We estimated the composite likelihood of the observed data given a specified model using the site frequency spectrum (SFS) and the simulation-based approach implemented in FASTSIMCOAL2 (Excoffier & Foll, 2011; Excoffier *et al.*, 2013). For the simulations, we selected 24 individuals from the three localities of *Q. tomentella* (CLE, CRU and CAT), 24 individuals from three populations representing the southern lineage of *Q. chrysolepis* (GAB, LAG and BER) and 24 individuals from three populations representing the northern lineage of *Q. chrysolepis* (SHA, TAH and SON). In the case of *Q. chrysolepis*, we selected for each lineage the three populations showing the highest probability of assignment to their specific genetic cluster (i.e. we excluded highly admixed populations located in contact zones; see the Results section). Note that we found weak evidence for introgression from *Q. vaccinifolia*/*Q. palmeri* into the populations of *Q. chrysolepis*/*Q. tomentella* considered for FASTSIMCOAL2 analyses (see the Results section). Thus, it is unlikely that contemporary or historical hybridization with the two Californian taxa belonging to the section *Protobalanus* and used as outgroups in phylogenomic analyses can bias the inferences obtained from our coalescent-based demographic reconstructions in FASTSIMCOAL2. A folded joint SFS was calculated considering a single SNP per locus to avoid the effects of linkage disequilibrium. Because we did not include invariable sites in the SFS, we fixed the effective population size for one population group (*Q. tomentella*; θ_T) to enable the estimation of other parameters in FASTSIMCOAL2 (e.g. Lanier *et al.*, 2015; Papadopoulou & Knowles, 2015). The effective population size fixed in the model was calculated from the level of nucleotide diversity (π) and estimates of mutation rate per site per generation (μ), because $N_e = (\pi/4\mu)$. Nucleotide diversity (π) for *Q. tomentella* was estimated from polymorphic and nonpolymorphic loci using STACKS ($\pi = 0.0005$; Table S1). The average mutation rate per site per generation (5.92×10^{-8}) was inferred from *Populus* (Tuskan *et al.*, 2006) following Gossmann *et al.* (2012) and considering that the average generation time of *Q. tomentella*/*Q. chrysolepis* is *c.* 50 yr (Bemmels *et al.*, 2016). To remove all missing data for the calculation of the joint SFS and minimize errors with allele frequency estimates, each population group was downsampled to 15 individuals using a custom Python script written by Qixin He and available on Dryad (Papadopoulou & Knowles, 2015). The final SFS contained 5668 variable SNPs.

Each model was run 100 replicated times considering 100 000–250 000 simulations for the calculation of the composite likelihood, 10–40 expectation-conditional maximization (ECM) cycles, and a stopping criterion of 0.001 (Papadopoulou & Knowles, 2015). We used an information-theoretic model selection approach based on the Akaike's information criterion (AIC) to determine the probability of each model given the

observed data (Burnham & Anderson, 1998; e.g. Abascal *et al.*, 2016; Thome & Carstens, 2016). After the maximum likelihood was estimated for each model in every replicate, we calculated the AIC scores as detailed in Thome & Carstens (2016). AIC values for each model were rescaled (ΔAIC) calculating the difference between the AIC value of each model and the minimum AIC obtained among all competing models (i.e. the best model has ΔAIC = 0). Point estimates of the different demographic parameters for the best-supported model were selected from the run with the highest maximum composite likelihood. Finally, we calculated confidence intervals of parameter estimates from 100 parametric bootstrap replicates by simulating SFS from the maximum composite likelihood estimates and re-estimating parameters each time (Excoffier *et al.*, 2013; e.g. Lanier *et al.*, 2015; Papadopoulou & Knowles, 2015). These analyses took *c.* 3 months of run-time using 30 processors on a HP ProLiant XL230a Gen9 with two Intel Haswell E5-2680 processors.

Phylogenomic inference

In order to explore the evolutionary and demographic history of *Q. tomentella* and the two lineages of *Q. chrysolepis* from a phylogenetic perspective, we reconstructed species trees using the multispecies coalescent model implemented in the SNAPP (Bryant *et al.*, 2012) plugin in BEAST v.2.4.1 (Bouckaert *et al.*, 2014). We considered the same genetic groups and populations used for FASTSIMCOAL2 analyses (*Q. tomentella*, and southern and northern lineages of *Q. chrysolepis*) plus five individuals of Palmer oak (*Q. palmeri*) and five individuals of huckleberry oak (*Q. vaccinifolia*). Due to large computational demands of the program, we analyzed five individuals per genetic group and performed independent analyses considering three different subsets of populations to determine the consistence of the obtained inferences (e.g. Kolar *et al.*, 2016). Initially, we ran analyses with different theta priors to allow for different current and ancestral population sizes (alpha = 2, beta = 200; alpha = 2, beta = 2000; alpha = 2, beta = 20 000) and leaving default settings for all other parameters. These analyses yielded the same topology (not shown) and only results for the intermediate prior for theta are presented (alpha = 2, beta = 2000). For each subset of populations, we used different starting seeds to run three independent replicate runs for *c.* 1 million generations sampled every 1000 steps. We used TRACER v.1.4 to examine log files, check stationarity and convergence of the chains, and confirm that effective sampling sizes (ESS) for all parameters were ≫ 200. We removed 10% of trees as burn-in and combined tree and log files for replicated runs using LOGCOMBINER v.2.4.1. We used TREEANNOTATOR v.1.8.3 to obtain maximum credibility trees. The full set of likely species trees was displayed with DENSITREE v.2.2.1 (Bouckaert, 2010), which is expected to show fuzziness in parts of the tree due to gene flow or other causes of phylogenetic conflict.

Genetic introgression and geographical distance

We calculated different estimates of introgression/gene flow from *Q. tomentella* into continental populations of *Q. chrysolepis* and analyzed their association with the geographical distance separating each population of *Q. chrysolepis* from the nearest population of *Q. tomentella*. In particular, we analysed: (1) pairwise genetic

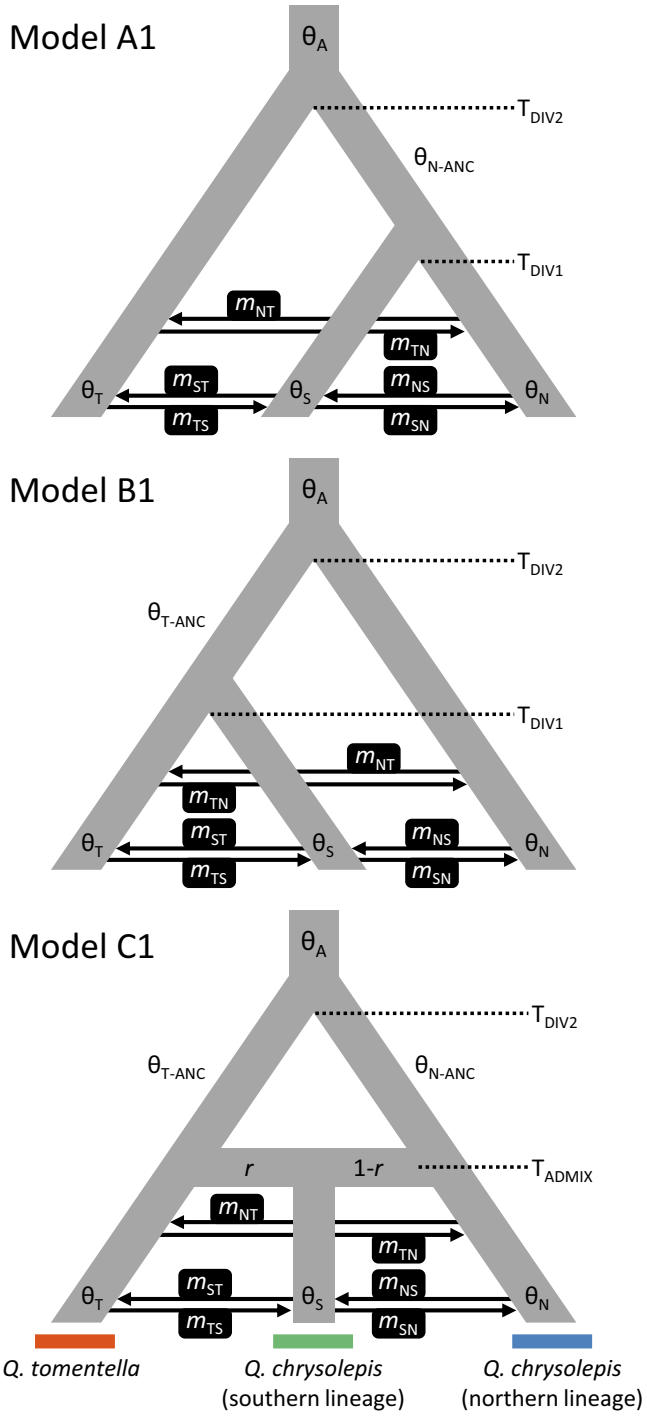

Fig. 2 Alternative demographic models tested using FASTSIMCOAL2. The three models consider different origins of the southern lineage of canyon live oak (*Quercus chrysolepis*) with respect to the island oak (*Q. tomentella*) and the northern lineage of canyon live oak. Parameter abbreviations include mutation-scaled effective population sizes (θ), migration rates per generation (*m*), timing of population divergence (T_{DIV}) and, in the case of the admixture model (Model C1), timing of admixture (T_{ADMIX}) and proportion of lineages transferred (*r*) from source to sink populations. Identical models not considering gene flow were also tested (A2, B2 and C2).

differentiation (F_{ST}) between *Q. tomentella* and each continental population of *Q. chrysolepis*; (2) the probability of membership (q) of each continental population of *Q. chrysolepis* to the *Q. tomentella* genetic cluster as inferred by STRUCTURE analyses (i.e. % of introgression); and (3) migration rates (m) from *Q. tomentella* to each continental population of *Q. chrysolepis*. We pooled all individuals of *Q. tomentella* and calculated their degree of genetic differentiation (F_{ST}) with each population of *Q. chrysolepis* using STACKS. Estimates of F_{ST} obtained with STACKS and ARLEQUIN v.3.5 (Excoffier & Lischer, 2010) were highly correlated ($r = 0.97$) and provided analogous results (data not shown). Pairwise migration rates (point estimates ± 95% CI) between *Q. tomentella* ($n = 24$ individuals from CRU, CAT and CLE) and each continental population of *Q. chrysolepis* were inferred with FASTSIMCOAL2 as described above (for a similar approach, see Barley *et al.*, 2015; Papadopoulou & Knowles, 2015). Only continental populations of *Q. chrysolepis* with eight genotyped individuals ($n = 10$ populations) were considered for these analyses (Table S1). The SFS was calculated downsampling to 15 and five individuals for *Q. tomentella* and each population of *Q. chrysolepis*, respectively. We tested four different gene flow models considering all possible migration matrices, including total absence of post-divergence gene flow. A full migration model considering asymmetrical gene flow was the most supported in all pairwise comparisons ($\Delta AIC > 2$ in all cases) (not shown) and subsequent analyses were based on estimates of migration rates per generation obtained using this model. The relationship between the different estimates of introgression/gene flow and the distance to the nearest population of *Q. tomentella* was analysed using linear and nonlinear (logarithmic, inverse, power and exponential) regressions in SPSS v.23. We considered nonlinear functions because these may potentially explain spatial patterns of introgression/gene flow better than a standard linear regression (e.g. if gene flow is mediated by long-distance pollen dispersal; Pluess *et al.*, 2009).

Results

Genomic data

A total of 151 285 222 (mean ± SD = 1719 150 ± 271 257 reads per individual) and 57 818 009 (mean ± SD = 1 606 055 ± 251 996 reads per individual) reads were obtained for 88 and 36 individuals of *Q. chrysolepis* and *Q. tomentella*, respectively. The number of reads retained after data processing and assembly averaged 86% per individual (Fig. S1). The datasets obtained with STACKS and parameters $P = 5$ and $P = 10$ contained 30 022 and 17 947 SNPs, respectively. The datasets obtained with PYRAD and parameters $c = 0.85$ and $c = 0.90$ contained 12 971 and 16 009 SNPs, respectively. See Methods S1 for further details.

Genetic structure and hybrid identification

STRUCTURE analyses and the statistic ΔK indicated an 'optimal' value of $K = 2$ (Fig. S2), splitting populations of *Q. chrysolepis* and *Q. tomentella* into two distinct genetic clusters (Figs 3, S3). STRUCTURE analyses for $K = 3$ divided populations of *Q. chrysolepis* located north and south of the Transverse Ranges and the Mohave Desert, with some genetic admixture in contact zones (FIG, HAS and KIN; Figs 3, S3). Bayesian clustering analyses confirmed that individuals of *Q. chrysolepis* collected from Santa Cruz and San Clemente Islands are hybrids with *Q. tomentella* and all of them showed a high degree of admixed ancestry (*c.* 50%) typical of first generation (F_1) hybrids or backcrosses between them (Figs 3, S3). Finally, STRUCTURE analyses revealed that southern populations of *Q. chrysolepis* (MOJ, LAG, BER, GAB and FIG) present some degree of introgression (*c.* 10%) from *Q. tomentella*, a pattern that was particularly clear for $K = 2$ (Figs 3, S3). Results obtained with STRUCTURE were supported by PCAs: PC1 separated populations of *Q. tomentella* and *Q. chrysolepis*, and PC2 split populations of *Q. chrysolepis* located north and south of the Transverse Ranges and the Mohave Desert (Fig. 4). Hybrid individuals from Santa Cruz and San Clemente Islands presented intermediate values along PC1 and continental populations of *Q. chrysolepis* from contact zones (FIG, HAS and KIN) had intermediate scores along PC2 (Fig. 4).

Coalescent analyses and model comparison

Results of FASTSIMCOAL2 showed that the most supported scenario (Model C1) was the one considering an admixed ancestry for the southern lineage of *Q. chrysolepis* (Table 1; Fig. 2). Parameter estimates obtained under this model indicate that this lineage shares a higher proportion of ancestry with *Q. tomentella* (*c.* 93%) than with the northern lineage of *Q. chrysolepis* (*c.* 7%) (Table 2). In fact, this model was statistically equivalent ($\Delta AIC = 2.3$; Burnham & Anderson, 1998) to Model B1, which considered that the southern lineage of *Q. chrysolepis* and *Q. tomentella* share their most recent common ancestor (Table 1; Fig. 2). Due to the similarity in parameters estimated by these two models, our inferences are based on model-averaged parameter values (e.g. Thome & Carstens, 2016; Table 2). Assuming an average generation time of 50 yr for these species (e.g. Bemmels *et al.*, 2016), the split between the northern lineage of *Q. chrysolepis* and the two other lineages (T_{DIV2}) was estimated to happen during the Pliocene (*c.* 3.7 Ma). The event leading to the split of *Q. tomentella* and the southern lineage of *Q. chrysolepis* (T_{ADMIX} in model C1 and T_{DIV1} in model B1) was estimated to take place during the early Pleistocene or the late Pliocene (*c.* 1.8–1.0 Ma) (Table 2). Note, however, that our estimates of divergence time must be interpreted with extreme caution due to the wide confidence intervals around our point estimates and considerable uncertainty in mutation rates and generation time for long-lived tree species (Table 2) (see Ortego *et al.*, 2015b; Tsuda *et al.*, 2015). Migration rates per generation (m) were estimated to be significantly higher (i.e. 95% CI do not overlap) from *Q. tomentella* to the two lineages of *Q. chrysolepis* than in the opposite direction (Table 2). Also, the estimate of migration rate from *Q. tomentella* to the southern lineage of *Q. chrysolepis* was significantly higher than that inferred from *Q. tomentella* to the northern lineage of *Q. chrysolepis* (Table 2). Pilot runs for more complex models

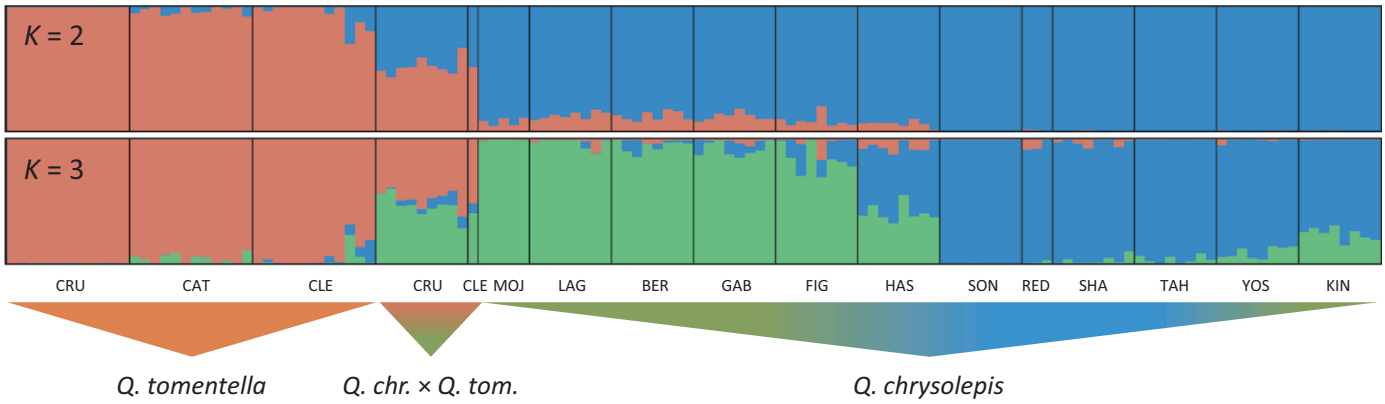

Fig. 3 Results of genetic assignments for canyon live oak (*Quercus chrysolepis*), island oak (*Q. tomentella*) and their hybrids (*Q. chrysolepis* × *Q. tomentella*) based on the Bayesian method implemented in the program STRUCTURE ($K = 2$ and $K = 3$). Analyses are based on a random subset of 10 000 SNPs data. Each individual is represented by a vertical bar, which is partitioned into K colored segments showing the individual's probability of belonging to the cluster with that color. Thin vertical black lines separate individuals from different sampling localities/species. Population codes are described in Supporting Information Table S1.

considering different sets of incomplete migration matrices (e.g. lack of migration between *Q. tomentella* and the northern lineage of *Q. chrysolepis*), gene flow after the first genetic split (T_{DIV2}) and/or migration limited to a certain period of time consistently provided a lower model support (e.g. see Fig. 2 in Filatov *et al.*, 2016) (data not shown).

Phylogenomic inference

Analyses with SNAPP considering three different combinations of populations showed that *Q. tomentella* and the southern lineage of *Q. chrysolepis* share their most recent common ancestor (Fig. 5), supporting the results obtained with FASTSIMCOAL2. Phylogenetic analyses also suggested the presence of gene flow between *Q. vaccinifolia* and the northern lineage of *Q. chrysolepis* (Fig. 5). Accordingly, STRUCTURE analyses performed considering the same three subsets of populations and individuals used for SNAPP analyses (i.e. balanced in terms of sample size for all taxa/lineages) revealed asymmetric hybridization from *Q. chrysolepis* into *Q. vaccinifolia* (Figs S4, S5). Three out of the five genotyped individuals of *Q. vaccinifolia* showed a high degree (*c.* 50%) of admixed ancestry (Fig. S5). In agreement with SNAPP analyses, we found no signature of hybridization or introgression between *Q. palmeri* and the other taxa/lineages (Fig. S5).

Genetic introgression and geographical distance

All estimates of introgression/gene flow from *Q. tomentella* to continental populations of *Q. chrysolepis* were significantly associated with the geographical distance to the nearest population of *Q. tomentella* (Table S2; Fig. 1). Genetic differentiation (F_{ST}) was positively associated with geographical distance following a power function (Table S2). The proportion of genetic introgression (q) estimated with STRUCTURE and pairwise migration rates per generation (m) inferred with FASTSIMCOAL2 (Table S3) were negatively associated with geographical distance following a logarithmic and a power function, respectively (Table S2; Fig. 1). Note, however, that simpler linear functions provided similar fits to the data than nonlinear functions for all estimates of introgression/gene flow (Table S2).

Genetic diversity and effective population sizes

The southern lineage of *Q. chrysolepis* presented higher estimates of genetic diversity and effective population sizes (N_e) than *Q. tomentella* or the northern lineage of *Q. chrysolepis* (Fig. 6; Table S1; see also Table S4 for genetic drift). Also, the southern lineage of *Q. chrysolepis* presented higher estimates of N_e than *Q. vaccinifolia* and *Q. palmeri*, the two other Californian golden cup oaks (Fig. 6). Estimates of genetic diversity and N_e were remarkably low for all populations of *Q. tomentella*, suggesting that they have experienced strong genetic drift due to their small size and geographical isolation (Fig. 6). Accordingly, STRUCTURE analyses showed that genetic drift after divergence (F-value) for the cluster of *Q. tomentella* was double the estimates obtained for the two lineages of *Q. chrysolepis* (Table S4).

Discussion

Our analyses showed that *Quercus tomentella* and *Q. chrysolepis* comprise three lineages, one corresponding with *Q. tomentella* and two separating populations of *Q. chrysolepis* located north and south of the Transverse Ranges and the Mohave Desert (Figs 3–5). These lineages present levels of genetic differentiation (F_{ST} *c.* 0.12) similar to that typically reported for other sister or closely related oak taxa (e.g. Muir & Schlotterer, 2005; Ortego *et al.*, 2015b, 2017). Both phylogenomic and population genomic coalescent-based analyses supported that the southern lineage of *Q. chrysolepis* shares its most recent common ancestor with *Q. tomentella* rather than with the northern lineage *Q. chrysolepis*. The northern lineage of *Q. chrysolepis* was estimated to have originated during the Pliocene, whereas the split between *Q. tomentella* and the southern lineage of *Q. chrysolepis*

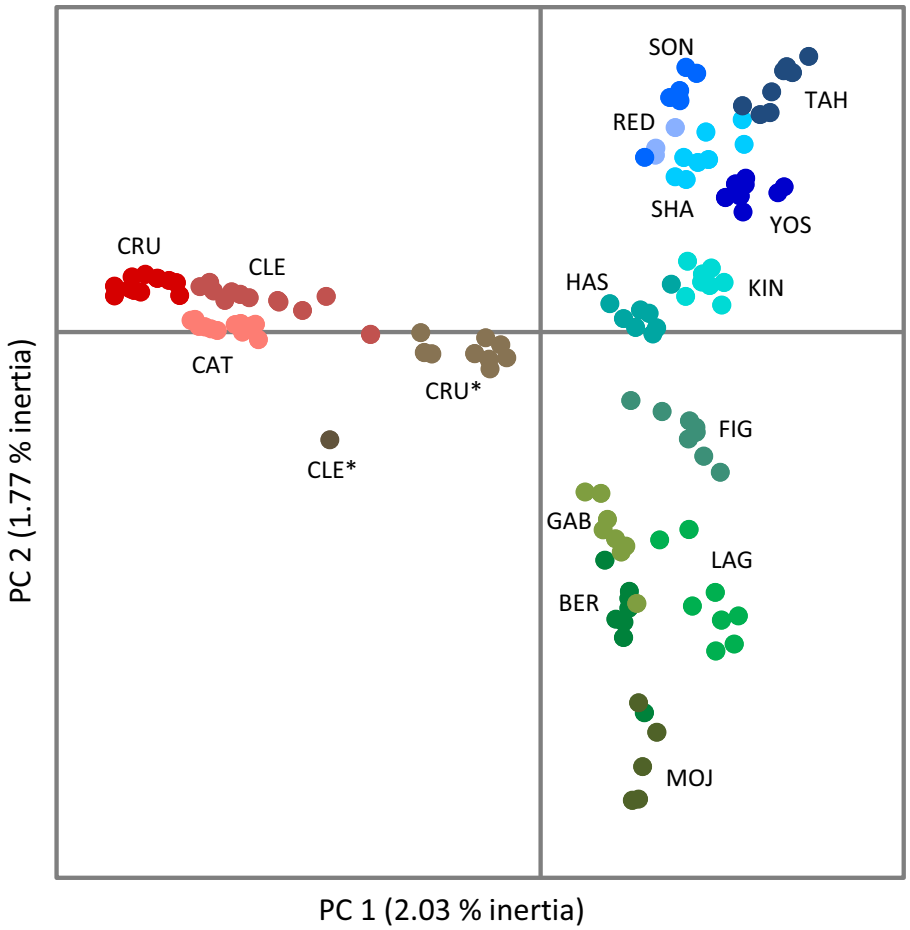

Fig. 4 Principal component analysis (PCA) of genetic variation (29 811 single nucleotide polymorphisms) for canyon live oak (*Quercus chrysolepis*) and island oak (*Q. tomentella*). Asterisks denote interspecific hybrids from Santa Cruz and San Clemente Islands. Population codes are described in Supporting Information Table S1.

Table 1 Comparison of demographic models analyzed with FASTSIMCOAL2 for the island oak (*Quercus tomentella*) and the two lineages of canyon live oak (*Quercus chrysolepis*)

Model	$\log_{10}$L	*k*	AIC	ΔAIC	ω_i
A1	−11 579.74	12	23 183.49	36.48	0.00
A2	−17 391.03	6	34 794.06	11 647.06	0.00
B1	**−11 562.81**	**12**	**23 149.62**	**2.62**	**0.21**
B2	−11 824.79	6	23 661.58	514.57	0.00
C1	**−11 559.50**	**14**	**23 147.01**	**0.00**	**0.79**
C2	−11 811.82	8	23 639.64	492.63	0.00

Best supported models are indicated in bold. Model description is detailed in Fig. 2. $\log_{10}$L, maximum likelihood estimate of the model; *k*, number of parameters in the model; AIC, Akaike's information criterion value; ΔAIC, difference in AIC value from that of the strongest model; ω_i, AIC weight.

probably occurred during the late Pliocene or the early Pleistocene. Although our estimated dates for lineage formation must be interpreted with extreme caution, they correspond well with the common presence of *Q. chrysolepis* and *Q. tomentella* (or their extinct forms, *Q. hannibali* and *Q. declinata*, respectively) in the Miocene and Pliocene fossil records of continental California (Hannibal, 1911; Axelrod, 1944a,b; see also Axelrod, 1938, 1958). Our analyses also revealed signatures of historical introgression from *Q. tomentella* into the continental populations of the southern lineage of *Q. chrysolepis*. Thus, the past coexistence of these two taxa or their extinct ancestors in the continent probably offered an ideal scenario for hybridization until the progressively drier and cooler climate during the Miocene and Pliocene pushed *Q. tomentella* and many other temperate species toward a shrinking belt of mild climate restricted to the California coastal strip and the Channel Islands (Hannibal, 1911; Axelrod, 1958, 1967, 1983; Muller, 1967).

A complex evolutionary history

Our coalescent analyses showed that the southern lineage of *Q. chrysolepis* shares a higher proportion of ancestry with *Q. tomentella* (*c.* 93%) than with its conspecific northern lineage, supporting the paraphyly of *Q. chrysolepis* (Fig. 5). In contrast with other oak complexes including taxa with very similar phenotypes (e.g. the Californian scrub white oak species complex; Roberts, 1995; Nixon, 2002; Ortego *et al.*, 2015b), *Q. tomentella* can be easily distinguished from *Q. chrysolepis* by its larger and thicker leaves with more prominent regular teeth and a characteristic corrugated leaf blade (Nixon, 2002; see also Manos, 1993 for differences in foliar trichome variation). In fact, *Q. chrysolepis* shows a higher phenotypic affinity with *Q. vaccinifolia* and *Q. palmeri*, the most divergent taxa within the section (Fig. 5), than with *Q. tomentella* (Tucker, 1980; Axelrod, 1983; see also Manos, 1993). Thus, the most

Table 2 Parameters inferred from coalescent simulations with FASTSIMCOAL2 under the two most supported demographic models for the island oak (*Quercus tomentella*) and the two lineages of canyon live oak (*Q. chrysolepis*)

Parameter	Model B1	Model C1	Model average	Lower bound	Upper bound
θ_A	105 451	84 741	89 090	26 879	167 286
θ_S	726 282	515 891	560 073	353 715	1118 399
θ_N	249 080	175 747	191 147	106 267	415 700
$\theta_{N\text{-}ANC}$	308 920	136 204	172 474	74 596	374 292
$\theta_{T\text{-}ANC}$		314 952	248 812	95 586	430 452
T_{DIV2}	97 732	67 405	73 774	45 861	313 678
T_{DIV1}	36 392		36 392	17 060	108 348
T_{ADMIX}		19 792	19 792	7285	106 647
r		0.93	0.93	0.63	0.97
m_{TS}	2.18×10^{-04}	1.98×10^{-04}	2.02×10^{-04}	1.68×10^{-04}	2.33×10^{-04}
m_{TN}	9.77×10^{-05}	1.16×10^{-04}	1.12×10^{-04}	8.54×10^{-05}	1.40×10^{-04}
m_{ST}	3.95×10^{-07}	1.56×10^{-09}	8.43×10^{-08}	3.95×10^{-11}	4.00×10^{-06}
m_{SN}	8.92×10^{-06}	9.58×10^{-06}	9.44×10^{-06}	8.61×10^{-07}	1.08×10^{-05}
m_{NT}	3.03×10^{-07}	4.56×10^{-09}	6.73×10^{-08}	2.78×10^{-11}	6.11×10^{-06}
m_{NS}	1.34×10^{-05}	2.48×10^{-05}	2.24×10^{-05}	7.63×10^{-06}	5.10×10^{-05}

Table shows point estimates under models B1 and C1 (illustrated in Fig. 2), model averaged estimates, and lower and upper 95% confidence intervals. Note that the effective population size of *Quercus tomentella* (θ_T) is not presented in this table because it was fixed in FASTSIMCOAL2 analyses to enable the estimation of other parameters (see the Materials and Methods section for further details). θ, mutation-scaled effective population sizes; T, timing of population divergence or admixture (given in number of generations); *r*, proportion of lineages transferred from source to sink populations; *m*, migration rates per generation. Each specific parameter is illustrated in Fig. 2.

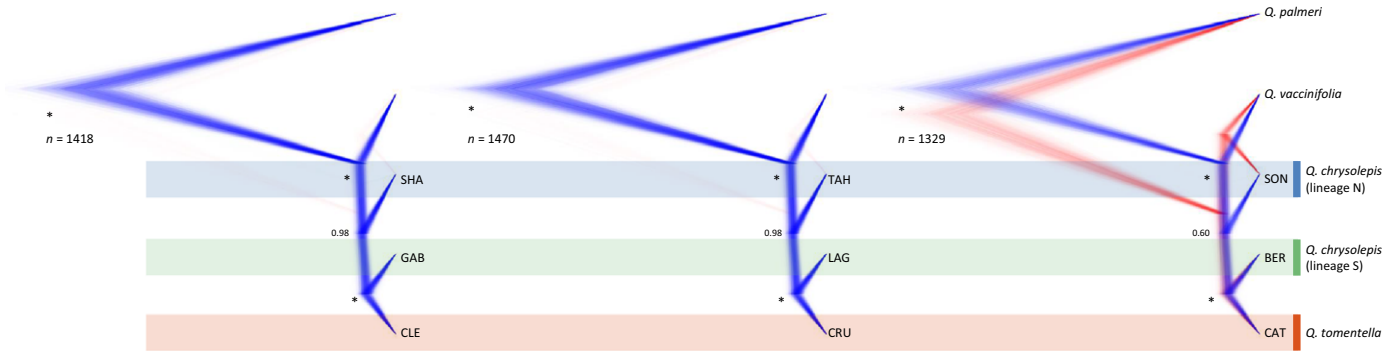

Fig. 5 SNAPP trees reconstructed considering different subsets of populations from the three genetic clusters inferred by STRUCTURE and principal component analysis (PCA) analyses for canyon live oak (*Quercus chrysolepis*) and island oak (*Q. tomentella*). Huckleberry oak (*Q. vaccinifolia*) and Palmer oak (*Q. palmeri*), species also belonging to the section *Protobalanus*, were also included. The first (blue) and second (red) most supported topologies are shown. The number of loci (*n*) shared across all clades and retained for SNAPP analyses are presented with each tree and posterior probabilities for the most supported topology are indicated on the nodes (* = 1). Population codes are described in Supporting Information Table S1.

parsimonious explanation for the morphological and ecological integrity of *Q. chrysolepis* across its entire distribution range is that this species represents the ancestral phenotypic state and that *Q. tomentella* has phenotypically and ecologically diverged as a separate species after it split from the southern lineage of *Q. chrysolepis*. Fossil records indicate that *Q. tomentella* gradually evolved larger leaves in response to milder oceanic climate when it migrated coastward, supporting a phenotypic transition in this taxon that has not been documented in the fossil record of *Q. chrysolepis* (Axelrod, 1941, 1944b, 1958).

An alternative explanation for the phenotypic and ecological integrity of *Q. chrysolepis* across its entire distribution range is that the two lineages of this taxon have converged in response to similar environmental conditions. Accordingly, previous studies have shown that oaks present considerable evolutionary lability in important physiological and morphological traits and communities composed by phylogenetically distant species tend to converge to similar phenotypes (Cavender-Bares *et al.*, 2004; Hipp *et al.*, 2018). Another possibility to explain the phenotypic integrity of *Q. chrysolepis* may be related with processes of phenotypic assimilation (Rheindt *et al.*, 2014; Huang, 2016) resulted from gene flow between its two lineages (Fig. 3) and introgression at a few genomic islands or specific genes involved in ecological adaptation and trait expression (Poelstra *et al.*, 2014; Yeaman *et al.*, 2016). Previous studies on oaks and other organisms have shown that convergence toward one parental phenotype may occur even when the genetic background of the other parental species is high (Hercus & Hoffmann, 1999; Masta *et al.*, 2002; Ortego & Bonal, 2010). Anecdotal evidence supporting this hypothesis in our system comes from the single individual *Q. chrysolepis* collected on San Clemente Island. This individual, the only one found on the island, did not present phenotypic

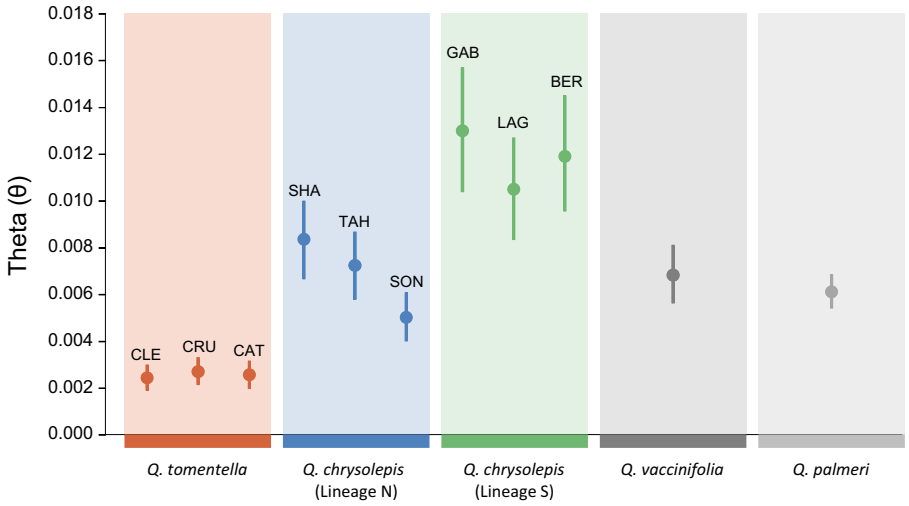

Fig. 6 Estimates of theta (θ) (mean ± 95% CIs) inferred by SNAPP for different populations of canyon live oak (*Quercus chrysolepis*), island oak (*Q. tomentella*), Palmer oak (*Q. palmeri*) and huckleberry oak (*Q. vaccinifolia*). Estimates are presented for those populations used in the three subsets of SNAPP analyses presented in Fig. 5. Population codes are described in Supporting Information Table S1.

signs of hybridization in spite of having a high degree of admixed ancestry with *Q. tomentella* (*c.* 54%; Figs 3, 4).

Historical vs contemporary introgression

Estimated migration rates per generation inferred with FASTSIMCOAL2 indicate higher rates of gene flow from *Q. tomentella* to continental populations of *Q. chrysolepis* than in the opposite direction (Table 2). In agreement with the results of Bayesian clustering analyses, the migration rate from *Q. tomentella* to the southern lineage of *Q. chrysolepis* was estimated to be significantly higher than that inferred from *Q. tomentella* to the northern lineage of *Q. chrysolepis* (Table 2; Fig. 3). Accordingly, different estimates of genetic introgression were negatively associated with the distance from each continental population of *Q. chrysolepis* to the nearest population of *Q. tomentella*, which suggests a pattern of isolation-by-distance probably resulted from a diffusion cline of neutral gene flow (Rieux *et al.*, 2013). Wind currents around the California Channel Islands have been largely stable for the past 10 000 years and are dominated by northwesterly winds and the periodic Santa Ana easterly winds (Riley & McGlaughlin, 2016 and references therein). Thus, although oaks have a high potential for long-distance pollen dispersal (Buschbom *et al.*, 2011; Hampe *et al.*, 2013), it is expected that the amount of pollen from *Q. tomentella* reaching nowadays continental populations from *Q. chrysolepis* is very limited to explain observed patterns of introgression. STRUCTURE analyses also showed that introgressed individuals of *Q. chrysolepis* from the continent present a small and homogeneous background of genetic admixture characteristic of populations at equilibrium, which indicates that observed patterns of introgression have not resulted from contemporary gene flow but reflect a past history of hybridization (Barton & Gale, 1993; Brelsford & Irwin, 2009). These lines of evidence, together with the high admixture between the two lineages of *Q. chrysolepis*, support that past hybridization between *Q. tomentella* and *Q. chrysolepis* in southern California and subsequent gene flow among continental populations of the latter is the most likely explanation for the observed correlation between genetic introgression and the distance to the nearest stand populations of *Q. tomentella*.

Exclusively focusing on the populations from the Channel Islands, STRUCTURE analyses revealed that the rate of introgression from *Q. tomentella* into *Q. chrysolepis* (*c.* 54%) is much higher than in the opposite direction (*c.* 3%; Fig. 3). The fact that all individuals of *Q. chrysolepis* from San Clemente and Santa Cruz Islands have a high degree of admixed ancestry (typical of F_1 hybrids or backcrosses between them) suggests that nowadays they constitute a hybrid swarm and confirms previous descriptive studies noticing that the two species have a strong history of hybridization on the Channel Islands (Muller, 1967; Thorne, 1969; Nixon, 2002; Ashley *et al.*, 2010; eFloras, 2017). Although *Q. tomentella* sustain higher census numbers than *Q. chrysolepis* on the Channel Islands, the populations of the former are also generally limited to small groves (e.g. Santa Cruz or Santa Catalina Islands) or scattered individuals (e.g. San Clemente Island) (Junak *et al.*, 1995; McCune, 2005; Ashley *et al.*, 2010). Thus, the highly asymmetrical introgression of *Q. chrysolepis* by *Q. tomentella* at different temporal scales suggests that the latter may present stronger pre- or post-zygotic barriers to interspecific gene flow, show a higher invasive capacity (e.g. Bacilieri *et al.*, 1996; Ortego *et al.*, 2017) and/or experience processes of parental genotype reconstruction (*sensu* Cannon & Scher, 2017; see also Petit *et al.*, 2004). Our analyses also revealed asymmetric introgression from the northern lineage of *Q. chrysolepis* into *Q. vaccinifolia* (Fig. 5 and Fig. S5), which is in agreement with previous studies describing hybridization between these two taxa (Myatt, 1980; Tucker, 1980; Hickman, 1993; Nixon, 2002; eFloras, 2017). However, we found no signature of introgression between *Q. palmeri* and the other taxa/lineages (Fig. S5). This supports previous studies suggesting that *Q. palmeri* hybridizes with relictual populations of *Q. chrysolepis* in Arizona (Tucker, 1980; Nixon, 2002), whereas ecological isolation prevents interspecific gene flow between these two taxa in California (Tucker, 1980).

Genetic diversity of parental and admixed lineages

Estimates of genetic variation and effective population sizes (N_e) are consistent with the inferred history of hybridization (Nettel *et al.*, 2008; Streicher *et al.*, 2014). The introgressed lineage of *Q. chrysolepis* from southern California presented the highest levels of genetic diversity and N_e (Fig. 6; Table S1). This supports the notion that hybridization can boost levels of genetic diversity of natural populations (Nettel *et al.*, 2008; Ortego *et al.*, 2014; Streicher *et al.*, 2014), particularly when the resulting hybrid population/lineages are not reproductively isolated from parental taxa and have not passed through severe demographic bottlenecks during the speciation/divergence process (Templeton, 1981; Rieseberg, 1997). Estimates of genetic diversity and N_e were remarkably low for *Q. tomentella*, suggesting that their small-size and isolated populations have experienced strong bottlenecks (Beckman & Jerome, 2017; Ashley *et al.*, 2010). This is also supported by STRUCTURE analyses, which showed that genetic drift after divergence (*F*-value) for the cluster of *Q. tomentella* was twice as high as the estimates obtained for the two lineages of *Q. chrysolepis* (Table S4; e.g. Papadopoulou & Knowles, 2015). Such considerable genetic drift can also explain why the two lineages of *Q. chrysolepis* clustered together for $K = 2$ when both demographic and phylogenomic analyses strongly support that the southern lineage shares its most recent common ancestor with *Q. tomentella* (Fig. 3).

Conclusions

Overall, this study highlights the importance of integrating demographic and phylogenomic analyses to understand lineage diversification in taxa with complex biogeographic histories. Our results support that the two lineages of *Q. chrysolepis* can be considered as part of a single paraphyletic taxon ecologically and phenotypically well differentiated from *Q. tomentella* (Thornburgh, 1990; Fralish & Franklin, 2002; Nixon, 2002). Populations resulted from admixture between the two lineages *Q. chrysolepis* are not restricted to narrow contact zones and extend across large geographical areas (Fig. 3; see also Ortego *et al.*, 2015a; Bemmels *et al.*, 2016). This represents a very different situation to that frequently described for most hybridizing oak species that, although can occasionally form hybrid swarms, maintain their genetic and phenotypic distinctiveness across overlapping portions of their respective distribution ranges and when they co-occur in mixed stands (Ortego *et al.*, 2014). Collectively, these results broaden the debate about species definition toward more complicated circumstances that can be probably only accommodated considering hybridization and speciation processes as a continuum with diffuse limits (Abbott *et al.*, 2013; Hochkirch, 2013; Edwards *et al.*, 2016). Future studies considering detailed phenotypic information and genome scans to detect potential loci under selection implicated in phenotypic trait expression and ecological adaptation would be of great help to get a better understanding of the processes underlying the evolutionary history of this and other intriguing species complexes (Hohenlohe *et al.*, 2013; Stolting *et al.*, 2013; Poelstra *et al.*, 2014; Suarez-Gonzalez *et al.*, 2016).

Acknowledgements

We wish to thank to Anna Papadopoulou for her valuable help in data analysis, Julie Lambert and Emily Howe (San Diego State University) for collecting samples from San Clemente Island, Amparo Hidalgo-Galiana for library preparation, and Sergio Pereira (The Centre for Applied Genomics) for Illumina sequencing. Dylan Burge and two anonymous referees provided valuable comments on an earlier draft of this manuscript. 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 logistic support during field work. Logistical support was provided by the Laboratorio de Ecología Molecular from Estación Biológica de Doñana (LEM-EBD). We also wish to thank to Centro de Supercomputación de Galicia for access to computer resources. J.O. was supported by a Ramón y Cajal (RYC-2013-12501) research fellowship. This work was supported by an internal EBD 'Microproyectos' grant to J.O., financed through the Severo Ochoa Program for Centres of Excellence (SEV-2012-0262). Support for fieldwork was provided by UCLA research funds of V.L.S.

Author contributions

J.O. conceived and designed the study, analysed the data and wrote the manuscript; J.O. and P.F.G. collected the samples; and J.O., P.F.G. and V.L.S. interpreted and discussed the results and edited the manuscript.

ORCID

Joaquín Ortego http://orcid.org/0000-0003-2709-429X

References

Abascal F, Corvelo A, Cruz F, Villanueva-Canas JL, Vlasova A, Marcet-Houben M, Martinez-Cruz B, Cheng JY, Prieto P, Quesada V *et al.* 2016. Extreme genomic erosion after recurrent demographic bottlenecks in the highly endangered Iberian lynx. *Genome Biology* **17**: 251.

Abbott R, Albach D, Ansell S, Arntzen JW, Baird SJE, Bierne N, Boughman JW, Brelsford A, Buerkle CA, Buggs R *et al.* 2013. Hybridization and speciation. *Journal of Evolutionary Biology* **26**: 229–246.

Abbott RJ, Barton NH, Good JM. 2016. Genomics of hybridization and its evolutionary consequences. *Molecular Ecology* **25**: 2325–2332.

Ashley MV, Abraham ST, Backs JR, Koenig WD. 2015. Landscape genetics and population structure in Valley oak (*Quercus lobata* Nee). *American Journal of Botany* **102**: 2124–2131.

Ashley MV, Abraham S, Kindsvater LC, Knapp D, Craft K. 2010. Population structure and genetic variation of Island oak, *Quercus tomentella* Engelmann on Santa Catalina Island. In: Knapp DA, ed. *Oak ecosystem restoration research on Santa Catalina Island, California. Proceeding of an on-island workshop, February 2–4, 2007.* Avalon, CA, USA: Catalina Island Conservancy, 125–135.

Axelrod DI. 1941. The concept of ecospecies in Tertiary paleobotany. *Proceedings of the National Academy of Sciences, USA* **27**: 545–551.

Axelrod DI. 1944a. The Sonoma flora (California). *Carnegie Institution of Washington Publication* **553**: 166–206.

Axelrod DI. 1944b. The Mulholland flora (California). *Carnegie Institution of Washington Publication* **553**: 102–146.

Axelrod DI. 1958. Evolution of the Madro-Tertiary geoflora. *Botanical Review* **24**: 433–509.

Axelrod DI. 1967. Geologic history of the California insular flora. In: Philbrick RN, ed. *Proceedings of the symposium on the biology of the California Islands.* Santa Barbara, CA, USA: Santa Barbara Botanical Gardens, 267–316.

Axelrod DI. 1983. Biogeography of oaks in the Arcto-Tertiary Province. *Annals of the Missouri Botanical Garden* **70**: 629–657.

Axelrod DJ. 1938. The stratigraphic significance of a southern element in later Tertiary floras of western America. *Journal of the Washington Academy of Science* **28**: 314–322.

Bacilieri R, Ducousso A, Petit RJ, Kremer A. 1996. Mating system and asymmetric hybridization in a mixed stand of European oaks. *Evolution* **50**: 900–908.

Backs JR, Ashley MV. 2016. Evolutionary history and gene flow of an endemic island oak: *Quercus pacifica*. *American Journal of Botany* **103**: 2115–2125.

Barley AJ, Monnahan PJ, Thomson RC, Grismer LL, Brown RM. 2015. Sun skink landscape genomics: assessing the roles of micro-evolutionary processes in shaping genetic and phenotypic diversity across a heterogeneous and fragmented landscape. *Molecular Ecology* **24**: 1696–1712.

Barton NH. 2001. The role of hybridization in evolution. *Molecular Ecology* **10**: 551–568.

Barton NH, Gale KS. 1993. Genetic analysis of hybrid zones. In: Harrison RG, ed. *Hybrid zones and the evolutionary process.* New York, NY, USA: Oxford University Press, 13–45.

Baskett ML, Gomulkiewicz R. 2011. Introgressive hybridization as a mechanism for species rescue. *Theoretical Ecology* **4**: 223–239.

Beckman E, Jerome D. 2017. *Quercus tomentella.* In: *The IUCN red list of threatened species 2017: e.T30959A2799049.* [WWW document] URL http://dx.doi.org/10.2305/IUCN.UK.2017-2.RLTS.T30959A2799049.en [accessed 11 December 2017].

Bemmels JB, Title PO, Ortego J, Knowles LL. 2016. Tests of species-specific models reveal the importance of drought in postglacial range shifts of a Mediterranean-climate tree: insights from integrative distributional, demographic and coalescent modelling and ABC model selection. *Molecular Ecology* **25**: 4889–4906.

Benestan L, Quinn BK, Maaroufi H, Laporte M, Clark FK, Greenwood SJ, Rochette R, Bernatchez L. 2016. Seascape genomics provides evidence for thermal adaptation and current-mediated population structure in American lobster (*Homarus americanus*). *Molecular Ecology* **25**: 5073–5092.

Bouckaert R, Heled J, Kuhnert D, Vaughan T, Wu CH, Xie D, Suchard MA, Rambaut A, Drummond AJ. 2014. BEAST2: a software platform for Bayesian evolutionary analysis. *PLoS Computational Biology* **10**: e1003537.

Bouckaert RR. 2010. DensiTree: making sense of sets of phylogenetic trees. *Bioinformatics* **26**: 1372–1373.

Brauer CJ, Hammer MP, Beheregaray LB. 2016. Riverscape genomics of a threatened fish across a hydroclimatically heterogeneous river basin. *Molecular Ecology* **25**: 5093–5113.

Brelsford A, Irwin DE. 2009. Incipient speciation despite little assortative mating: the yellow-rumped warbler hybrid zone. *Evolution* **63**: 3050–3060.

Bryant D, Bouckaert R, Felsenstein J, Rosenberg NA, RoyChoudhury A. 2012. Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. *Molecular Biology and Evolution* **29**: 1917–1932.

Burnham KP, Anderson DR. 1998. *Model selection and inference: a practical information-theoretic approach.* New York, NY, USA: Springer.

Buschbom J, Yanbaev Y, Degen B. 2011. Efficient long-distance gene flow into an isolated relict oak stand. *Journal of Heredity* **102**: 464–472.

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.

Cannon CH, Scher CL. 2017. Exploring the potential of gametic reconstruction of parental genotypes by F-1 hybrids as a bridge for rapid introgression. *Genome* **60**: 713–719.

Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA. 2013. STACKS: an analysis tool set for population genomics. *Molecular Ecology* **22**: 3124–3140.

Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH. 2011. Stacks: building and genotyping loci *de novo* from short-read sequences. *G3: Genes – Genomes – Genetics* **1**: 171–182.

Cavender-Bares J, Ackerly DD, Baum DA, Bazzaz FA. 2004. Phylogenetic overdispersion in Floridian oak communities. *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.

Currat M, Excoffier L. 2011. Strong reproductive isolation between humans and Neanderthals inferred from observed patterns of introgression. *Proceedings of the National Academy of Sciences, USA* **108**: 15 129–15 134.

Dodd RS, Kashani N. 2003. Molecular differentiation and diversity among the California red oaks (Fagaceae; *Quercus* section *Lobatae*). *Theoretical and Applied Genetics* **107**: 884–892.

Dowling TE, Secor CL. 1997. The role of hybridization and introgression in the diversification of animals. *Annual Review of Ecology and Systematics* **28**: 593–619.

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. 2014. PyRAD: assembly of *de novo* RADseq loci for phylogenetic analyses. *Bioinformatics* **30**: 1844–1849.

Eaton DAR, Hipp AL, Gonzalez-Rodriguez A, Cavender-Bares J. 2015. Historical introgression among the American live oaks and the comparative nature of tests for introgression. *Evolution* **69**: 2587–2601.

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

Edwards SV, Potter S, Schmitt CJ, Bragg JG, Moritz C. 2016. Reticulation, divergence, and the phylogeography-phylogenetics continuum. *Proceedings of the National Academy of Sciences, USA* **113**: 8025–8032.

Escudero M, Eaton DAR, Hahn M, Hipp AL. 2014. Genotyping-by-sequencing as a tool to infer phylogeny and ancestral hybridization: a case study in *Carex* (Cyperaceae). *Molecular Phylogenetics and Evolution* **79**: 359–367.

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, Dupanloup I, Huerta-Sanchez E, Sousa VC, Foll M. 2013. Robust demographic inference from genomic and SNP data. *PLoS Genetics* **9**: e1003905.

Excoffier L, Foll M. 2011. Fastsimcoal: a continuous-time coalescent simulator of genomic diversity under arbitrarily complex evolutionary scenarios. *Bioinformatics* **27**: 1332–1334.

Excoffier L, Lischer HEL. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* **10**: 564–567.

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.

Filatov DA, Osborne OG, Papadopulos AST. 2016. Demographic history of speciation in a *Senecio* altitudinal hybrid zone on Mt. Etna. *Molecular Ecology* **25**: 2467–2481.

Fitz-Gibbon S, Hipp AL, Pham KK, Manos PS, Sork VL. 2017. Phylogenomic inferences from reference-mapped and *de novo* assembled short-read sequence data using RADseq sequencing of California white oaks (*Quercus* section *Quercus*). *Genome* **60**: 743–755.

Fitzpatrick BM, Johnson JR, Kump DK, Smith JJ, Voss SR, Shaffer HB. 2010. Rapid spread of invasive genes into a threatened native species. *Proceedings of the National Academy of Sciences, USA* **107**: 3606–3610.

eFloras. 2017. *Flora of North America.* St Louis, MO, USA & Cambridge, MA, USA: Missouri Botanic Garden & Harvard University Herbaria. [WWW document] URL http://www.efloras.org [accessed 12 November 2017].

Fralish JS, Franklin SB. 2002. *Taxonomy and ecology of woody plants in North American forests (excluding Mexico and Subtropical Florida).* New York, NY, USA: John Wiley & Sons.

Gossmann TI, Keightley PD, Eyre-Walker A. 2012. The effect of variation in the effective population size on the rate of adaptive molecular evolution in Eukaryotes. *Genome Biology and Evolution* **4**: 658–667.

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

Guichoux E, Garnier-Gere P, Lagache L, Lang T, Boury C, Petit RJ. 2013. Outlier loci highlight the direction of introgression in oaks. *Molecular Ecology* **22**: 450–462.

Hampe A, Pemonge MH, Petit RJ. 2013. Efficient mitigation of founder effects during the establishment of a leading-edge oak population. *Proceedings of the Royal Society B* **280**: 20131070.

Hannibal H. 1911. A Pliocene flora from the Coast Ranges of California. *Bulletin of the Torrey Botanical Club* **38**: 329–342.

Hercus MJ, Hoffmann AA. 1999. Does interspecific hybridization influence evolutionary rates? An experimental study of laboratory adaptation in hybrids between *Drosophila serrata* and *Drosophila birchii*. *Proceedings of the Royal Society B* **266**: 2195–2200.

Hickman JC, ed. 1993. *The Jepson manual: higher plants of California*. Berkeley, CA, USA: University of California Press.

Hipp AL, Manos PS, González-Rodríguez A, Hahn M, Kaproth M, McVay JD, Avalos SV, Cavender-Bares J. 2018. Sympatric parallel diversification of major oak clades in the Americas and the origins of Mexican species diversity. *New Phytologist* **217**: 439–452.

Hochkirch A. 2013. Hybridization and the origin of species. *Journal of Evolutionary Biology* **26**: 247–251.

Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, Cresko WA. 2010. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genetics* **6**: e1000862.

Hohenlohe PA, Day MD, Amish SJ, Miller MR, Kamps-Hughes N, Boyer MC, Muhlfeld CC, Allendorf FW, Johnson EA, Luikart G. 2013. Genomic patterns of introgression in rainbow and westslope cutthroat trout illuminated by overlapping paired-end RAD sequencing. *Molecular Ecology* **22**: 3002–3013.

Huang JP. 2016. Parapatric genetic introgression and phenotypic assimilation: testing conditions for introgression between Hercules beetles (Dynastes, Dynastinae). *Molecular Ecology* **25**: 5513–5526.

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.

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.

Junak S, Ayers T, Scott R, Wilken D, Young D. 1995. *A flora of Santa Cruz Island*. Santa Barbara, CA, USA: Santa Barbara Botanic Garden.

Kolar F, Fuxova C, Zaveska E, Nagano AJ, Hyklova L, Lucanova M, Kudoh H, Marhold K. 2016. Northern glacial refugia and altitudinal niche divergence shape genome-wide differentiation in the emerging plant model *Arabidopsis arenosa*. *Molecular Ecology* **25**: 3929–3949.

Lanier HC, Massatti R, He Q, Olson LE, Knowles LL. 2015. Colonization from divergent ancestors: glaciation signatures on contemporary patterns of genomic variation in Collared Pikas (*Ochotona collaris*). *Molecular Ecology* **24**: 3688–3705.

Levin DA, Francisco-Ortega J, Jansen RK. 1996. Hybridization and the extinction of rare plant species. *Conservation Biology* **10**: 10–16.

Lewontin RC, Birch LC. 1966. Hybridization as a source of variation for adaptation to new environments. *Evolution* **20**: 315–336.

Linder CR, Rieseberg LH. 2004. Reconstructing patterns of reticulate evolution UN plants. *American Journal of Botany* **91**: 1700–1708.

Liu KJ, Steinberg E, Yozzo A, Song Y, Kohn MH, Nakhleh L. 2015. Interspecific introgressive origin of genomic diversity in the house mouse. *Proceedings of the National Academy of Sciences, USA* **112**: 196–201.

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

Manos PS. 1993. Foliar trichome variation in *Quercus* section *Protobalanus* (Fagaceae). *SIDA* **117**: 391–403.

de Manuel M, Kuhlwilm M, Frandsen P, Sousa VC, Desai T, Prado-Martinez J, Hernandez-Rodriguez J, Dupanloup I, Lao O, Hallast P *et al.* 2016. Chimpanzee genomic diversity reveals ancient admixture with bonobos. *Science* **354**: 477–481.

Masta SE, Sullivan BK, Lamb T, Routman EJ. 2002. Molecular systematics, hybridization, and phylogeography of the *Bufo americanus* complex in Eastern North America. *Molecular Phylogenetics and Evolution* **24**: 302–314.

McBreen K, Lockhart PJ. 2006. Reconstructing reticulate evolutionary histories of plants. *Trends in Plant Science* **11**: 398–404.

McCune J. 2005. *Report on the census and survey of Island oak (Quercus tomentella Engelm.) and canyon live oak (Quercus chrysolepis Liebm.) groves on Catalina Island, 2004 and 2005*. Avalon, CA, USA: Catalina Island Conservancy.

McVay JD, Hipp AL, Manos PS. 2017. A genetic legacy of introgression confounds phylogeny and biogeography in oaks. *Proceedings of the Royal Society B* **284**: 20170300.

Melo-Ferreira J, Alves PC, Freitas H, Ferrand N, Boursot P. 2009. The genomic legacy from the extinct *Lepus timidus* to the three hare species of Iberia: contrast between mtDNA, sex chromosomes and autosomes. *Molecular Ecology* **18**: 2643–2658.

Morjan CL, Rieseberg LH. 2004. How species evolve collectively: implications of gene flow and selection for the spread of advantageous alleles. *Molecular Ecology* **13**: 1341–1356.

Muir G, Schlotterer C. 2005. Evidence for shared ancestral polymorphism rather than recurrent gene flow at microsatellite loci differentiating two hybridizing oaks (*Quercus* spp.). *Molecular Ecology* **14**: 549–561.

Muller CH. 1967. Relictual origins of insular endemics in *Quercus*. In: Philbrick RN, ed. *Proceedings of the symposium on the biology of the California Islands*. Santa Barbara, CA, USA: Santa Barbara Botanical Gardens, 73–77.

Myatt RG. 1980. Canyon live oak vegetation in the Sierra Nevada. In: Plumb TR, technical coordinator. *Proceedings of the Symposium on the Ecology, Management and Utilization of California Oaks, June 26–28, 1979, Claremont, California*. Berkeley, CA, USA: US Department of Agriculture Forest Service, Pacific Southwest Forest and Range Experiment Station, General Technical Report PSW-144, 86–91.

Nettel A, Dood RS, Afzal-Rafii Z, Tovilla-Hernandez C. 2008. Genetic diversity enhanced by ancient introgression and secondary contact in East Pacific black mangroves. *Molecular Ecology* **17**: 2680–2690.

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

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

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

Ortego J, Gugger PF, Sork VL. 2017. Impacts of human-induced environmental disturbances on hybridization between two ecologically differentiated Californian oak species. *New Phytologist* **213**: 930–943.

Ortego J, Noguerales V, Gugger PF, Sork VL. 2015b. Evolutionary and demographic history of the Californian scrub white oak species complex: an integrative approach. *Molecular Ecology* **24**: 6188–6208.

Papadopoulou A, Knowles LL. 2015. Species-specific responses to island connectivity cycles: refined models for testing phylogeographic concordance

across a Mediterranean Pleistocene Aggregate Island Complex. *Molecular Ecology* **24**: 4252–4268.

Payseur BA, Rieseberg LH. 2016. A genomic perspective on hybridization and speciation. *Molecular Ecology* **25**: 2337–2360.

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE. 2012. Double digest RADseq: an inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. *PLoS ONE* **7**: e37135.

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

Pluess AR, Sork VL, Dolan B, Davis FW, Grivet D, Merg K, Papp J, Smouse PE. 2009. Short distance pollen movement in a wind-pollinated tree, *Quercus lobata* (Fagaceae). *Forest Ecology and Management* **258**: 735–744.

Poelstra JW, Vijay N, Bossu CM, Lantz H, Ryll B, Muller I, Baglione V, Unneberg P, Wikelski M, Grabherr MG *et al.* 2014. The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. *Science* **344**: 1410–1414.

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

R Core Team. 2017. *R: a language and environment for statistical computing, v.3.3.2*. Vienna, Austria: R Foundation for Statistical Computing. [WWW document] URL http://www.rproject.org. [accessed 12 November 2017]

Rheindt FE, Fujita MK, Wilton PR, Edwards SV. 2014. Introgression and phenotypic assimilation in *Zimmerius* flycatchers (Tyrannidae): population genetic and phylogenetic inferences from genome-wide SNPs. *Systematic Biology* **63**: 134–152.

Rhymer JM, Simberloff D. 1996. Extinction by hybridization and introgression. *Annual Review of Ecology and Systematics* **27**: 83–109.

Rieseberg LH. 1997. Hybrid origins of plant species. *Annual Review of Ecology and Systematics* **28**: 359–389.

Rieux A, Lenormand T, Carlier J, de Bellaire LD, Ravigne V. 2013. Using neutral cline decay to estimate contemporary dispersal: a generic tool and its application to a major crop pathogen. *Ecology Letters* **16**: 721–730.

Riley L, McGlaughlin ME. 2016. Endemism in native floras of California's Channel Islands correlated with seasonal patterns of aeolian processes. *Botany-Botanique* **94**: 65–72.

Roberts FM. 1995. *The oaks of the Southern California Floristic Province*. F. M. Encinitas, CA, USA: Roberts Publications.

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

Stolting KN, Nipper R, Lindtke D, Caseys C, Waeber S, Castiglione S, Lexer C. 2013. Genomic scan for single nucleotide polymorphisms reveals patterns of divergence and gene flow between ecologically divergent species. *Molecular Ecology* **22**: 842–855.

Streicher JW, Devitt TJ, Goldberg CS, Malone JH, Blackmon H, Fujita MK. 2014. Diversification and asymmetrical gene flow across time and space: lineage sorting and hybridization in polytypic barking frogs. *Molecular Ecology* **23**: 3273–3291.

Suarez-Gonzalez A, Hefer CA, Christe C, Corea O, Lexer C, Cronk QCB, Douglas CJ. 2016. Genomic and functional approaches reveal a case of adaptive introgression from *Populus balsamifera* (balsam poplar) in *P. trichocarpa* (black cottonwood). *Molecular Ecology* **25**: 2427–2442.

Sun YS, Abbott RJ, Li LL, Li L, Zou JB, Liu JQ. 2014. Evolutionary history of Purple cone spruce (*Picea purpurea*) in the Qinghai-Tibet Plateau: homoploid hybrid origin and Pleistocene expansion. *Molecular Ecology* **23**: 343–359.

Templeton AR. 1981. Mechanisms of speciation – a population genetic approach. *Annual Review of Ecology and Systematics* **12**: 23–48.

Thome MTC, Carstens BC. 2016. Phylogeographic model selection leads to insight into the evolutionary history of four-eyed frogs. *Proceedings of the National Academy of Sciences, USA* **113**: 8010–8017.

Thornburgh DA. 1990. *Quercus chrysolepis* Liebm. canyon live oak. In: Burns RM, Honkala BH, eds. *Silvics of North America, vol. 2, Hardwoods, Agricultural Handbook 654*. Washington, DC, USA: Forest Service USDA, 618–624.

Thorne RF. 1969. California Islands. *Annals of the Missouri Botanical Garden* **56**: 391–408.

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.

Tucker JM. 1980. Taxonomy of California oaks. In: Plumb TR, technical coordinator. *Proceedings of the Symposium on the Ecology, Management and Utilization of California Oaks, June 26–28, 1979, Claremont, California*. Berkeley, CA, USA: US Department of Agriculture Forest Service, Pacific Southwest Forest and Range Experiment Station, General Technical Report PSW-144, 19–29.

Tuskan GA, DiFazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A *et al.* 2006. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* **313**: 1596–1604.

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.

Wall JD, Yang MA, Jay F, Kim SK, Durand EY, Stevison LS, Gignoux C, Woerner A, Hammer MF, Slatkin M. 2013. Higher levels of Neanderthal ancestry in East Asians than in Europeans. *Genetics* **194**: 199–209.

Wang XR, Szmidt AE, Lewandowski A, Wang ZR. 1990. Evolutionary analysis of *Pinus densata* Masters, a putative tertiary hybrid. 1. Allozyme variation. *Theoretical and Applied Genetics* **80**: 635–640.

Wood TE, Takebayashi N, Barker MS, Mayrose I, Greenspoon PB, Rieseberg LH. 2009. The frequency of polyploid speciation in vascular plants. *Proceedings of the National Academy of Sciences, USA* **106**: 13875–13879.

Yeaman S, Hodgins KA, Lotterhos KE, Suren H, Nadeau S, Degner JC, Nurkowski KA, Smets P, Wang TL, Gray LK *et al.* 2016. Convergent local adaptation to climate in distantly related conifers. *Science* **353**: 1431–1433.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information tab for this article:

Fig. S1 Number of reads per individual before and after different quality filtering steps by STACKS.

Fig. S2 Mean (± SD) log probability of the data (Ln Pr($X|K$)) for each value of K and the magnitude of ΔK over 10 runs of STRUCTURE for canyon live oak (*Quercus chrysolepis*) and island oak (*Q. tomentella*).

Fig. S3 Results of genetic assignments for canyon live oak (*Quercus chrysolepis*) and island oak (*Q. tomentella*) based on the Bayesian method implemented in the program STRUCTURE.

Fig. S4 Mean (± SD) log probability of the data (Ln Pr($X|K$)) for each value of K and the magnitude of ΔK over 10 runs of STRUCTURE for huckleberry oak (*Quercus vaccinifolia*), Palmer oak (*Q. palmeri*), canyon live oak (*Q. chrysolepis*) and island oak (*Q. tomentella*).

Fig. S5 Results of genetic assignments for huckleberry oak (*Quercus vaccinifolia*), Palmer oak (*Q. palmeri*), canyon live oak (*Q. chrysolepis*) and island oak (*Q. tomentella*) based on the Bayesian method implemented in the program STRUCTURE.

Table S1 Geographical location and genetic statistics (P, H_O, H_E and π) for the studied populations of canyon live oak (*Quercus chrysolepis*) and island oak (*Q. tomentella*)

Table S2 Linear and nonlinear regressions analyzing the relationship between different genetic parameters (F_{ST}, q and m) and the geographical distance from each continental population of canyon live oak (*Quercus chrysolepis*) to the nearest population of island oak (*Q. tomentella*)

Table S3 Pairwise estimates of migration rates per generation (m) between island oak (*Quercus tomentella*) and continental populations of canyon live oak (*Q. chrysolepis*) inferred from coalescent simulations with FASTSIMCOAL2

Table S4 Amount of genetic drift after divergence (F-values) estimated by STRUCTURE for each inferred genetic cluster

Methods S1 Supplemental methods.

Please note: Wiley Blackwell are not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.