Received: 16 July 2017 | Revised: 25 July 2018 | Accepted: 29 August 2018

DOI: 10.1111/mec.14869

ORIGINAL ARTICLE

WILEY MOLECULAR ECOLOGY

RADseq data reveal ancient, but not pervasive, introgression between Californian tree and scrub oak species (*Quercus* sect. *Quercus*: Fagaceae)

Bernard Y. Kim[1] | **Xinzeng Wei**[1,2] | **Sorel Fitz-Gibbon**[1] | **Kirk E. Lohmueller**[1,3] | **Joaquín Ortego**[4] | **Paul F. Gugger**[1,5] | **Victoria L. Sork**[1,6]

[1]Department of Ecology and Evolutionary Biology, University of California at Los Angeles, Los Angeles, California

[2]Key Laboratory of Aquatic Botany and Watershed Ecology, Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan, Hubei, China

[3]Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, California

[4]Department of Integrative Ecology, Estación Biológica de Doñana, EBD-CSIC, Seville, Spain

[5]Appalachian Laboratory, University of Maryland Center for Environmental Science, Frostburg, Maryland

[6]Institute of the Environment and Sustainability, University of California, Los Angeles, California

Correspondence
Victoria L. Sork, Department of Ecology and Evolutionary Biology, University of California at Los Angeles, Los Angeles, CA.
Email: vlsork@ucla.edu

Funding information
National Science Foundation, Grant/Award Number: IOS-1444611

Abstract

A long-term debate in evolutionary biology is the extent to which reproductive isolation is a necessary element of speciation. Hybridizing plants in general are cited as evidence against this notion, and oaks specifically have been used as the classic example of species maintenance without reproductive isolation. Here, we use thousands of SNPs generated by RAD sequencing to describe the phylogeny of a set of sympatric white oak species in California and then test whether these species exhibit pervasive interspecific gene exchange. Using RAD sequencing, we first constructed a phylogeny of ten oak species found in California. Our phylogeny revealed that seven scrub oak taxa occur within one clade that diverged from a common ancestor with *Q. lobata*, that they comprise two subclades, and they are not monophyletic but include the widespread tree oak *Q. douglasii*. Next, we searched for genomic patterns of allele sharing consistent with gene flow between long-divergent tree oaks with scrub oaks. Specifically, we utilized the *D*-statistic as well as model-based inference to compare the signature of shared alleles between two focal tree species (*Q. lobata* and *Q. engelmannii*) with multiple scrub species within the two subclades. We found that introgression is not equally pervasive between sympatric tree and scrub oak species. Instead, gene flow commonly occurs from scrub oaks to recently sympatric *Q. engelmannii*, but less so from scrub oaks to long-sympatric *Q. lobata*. This case study illustrates the influence of ancient introgression and impact of reproductive isolating mechanisms in preventing indiscriminate interspecific gene exchange.

KEYWORDS
ancient introgression, *D*-statistic test, hybridization, phylogeny, systematics

1 | INTRODUCTION

The ability to detect ancient introgression with high-throughput genomic tools provides new ways to test classical hypotheses about the long-term role of gene flow in speciation and the maintenance of species boundaries. The unexpected finding that anatomically modern humans and archaic hominids interbred (Plagnol & Wall, 2006; Wall, Lohmueller, & Plagnol, 2009) and the sequencing of Neanderthals (Green et al., 2010) precipitated the development of new statistical tools to identify and quantify gene flow between evolutionarily distinct lineages (Green et al., 2010; Patterson et al., 2012; Sankararaman et al., 2014; Vernot & Akey, 2014). These tools, notably the *D*-statistic (Green et al., 2010) and model-based approaches (Gutenkunst, Hernandez, Williamson, & Bustamante, 2009), have been applied to many non-model systems and have shown that gene flow between species is common (Payseur & Rieseberg, 2016). Modern

genomic data create a way to examine the role of gene flow in the speciation process, which is typified by the classical biological species concept (BSC), which posits that species are distinct reproductive entities, or in other words, that complete reproductive isolation is a necessary condition for speciation (Dobzhansky & Dobzhansky, 1937). Moreover, gene flow among species undermines the assumption of bifurcating species trees, leading some to call for a reevaluation of the bifurcating species model (Cui et al., 2013; Ellstrand, 2014; Mallet, Besansky, & Hahn, 2016; Novikova et al., 2016). Now, the availability of new genomic data and statistical tools allows us to revisit the contribution of gene flow to the nature of species.

The idea that gene flow between species occurs frequently—especially in plants—dates to Darwin (1859), and the role of hybrids in facilitating this process was debated during the modern synthesis (Clausen, 1951; Mayr, 1942; Stebbins, 1950). Oaks (*Quercus* spp.) are a classic illustration of the limitations of the BSC (Burger, 1975; Van Valen, 1976) due to their widespread hybridization (Burger, 1975; Lexer, Kremer, & Petit, 2006; Ortego & Bonal, 2010; Ortego, Gugger, Riordan, & Sork, 2014; Palmer, 1948; Petit, Bodenes, Ducousso, Roussel, & Kremer, 2004; Tucker, 1990; Van Valen, 1976; Whittemore & Schaal, 1991). Particularly drawing attention to hybridization are the cases of introgression between sympatric oak species in eastern North America (Whittemore & Schaal, 1991) and Europe (Petit, Kremer, & Wagner, 1993; Petit et al., 1997), based on shared chloroplast markers or other evidence (Rushton, 1993). Nuclear genetic markers have also identified oak hybrids across the Northern Hemisphere, including Mexico (Penaloza-Ramirez et al., 2010; Tovar-Sanchez & Oyama, 2004), California (Dodd & Afzal-Rafii, 2004; Ortego, Gugger, & Sork, 2015a; Ortego et al., 2014; Riordan et al., 2016; Tucker, 1990) and China (Zeng, Liao, Petit, & Zhang, 2011). Nevertheless, the extent to which ancient hybridization has shaped the genome of contemporary oaks is unclear.

The focus of our study was to assess ancient interspecific gene flow between Californian tree and scrub white oak species (*Quercus* sect. *Quercus*)—species with evidence of contemporary hybridization. We decided to investigate gene flow between tree and scrub oak species because they are sufficiently different genetically and phenotypically to identify the species. The long divergence times between species also makes patterns of allele sharing between species more likely to be from introgression rather than from the long-term persistence of structure in the common ancestor. We sampled a focal California endemic tree oak species, *Q. lobata*, which is often sympatric with blue oak (*Q. douglasii)* and numerous scrub oak species in central and northern California, and have records of hybrids (Craft, Ashley, & Koenig, 2002; Hickman, 1993; Tucker, 1990). *Q. lobata* was previously considered to be closely related to *Q. douglasii* (Nixon, 2002), but recent evidence reported here and elsewhere (Fitz-Gibbon, Hipp, Pham, Manos, & Sork, 2017; Hipp et al., 2018; McVay, Hauser, Hipp, & Manos, 2017a; McVay, Hipp, & Manos, 2017b) indicates *Q. lobata* is in a separate group from the one they share.

We also sampled a non-endemic tree oak found in southern California, *Q. engelmannii*, which is related to a lineage found in northwest Mexico and southern Arizona (McVay et al., 2017b), which diverged from a common ancestor shared with *Q. lobata* about 34 MYA (Hipp et al., 2018). The scrub oak taxa occupy a variety of habitats, which are dispersed geographically across California, and many of them are often difficult to distinguish from each other due to recent divergence from a common ancestor and hybridization (Nixon, 1997, 2002; Tucker, 1974). Importantly, *Q. lobata* has been spatially proximate with several endemic scrub oaks as well as *Q. douglasii* for hundreds of thousands of years if not longer, and much longer than the distantly related *Q. engelmannii*. Many of these white oak species are sympatric with each other (see Figure 1a–c;

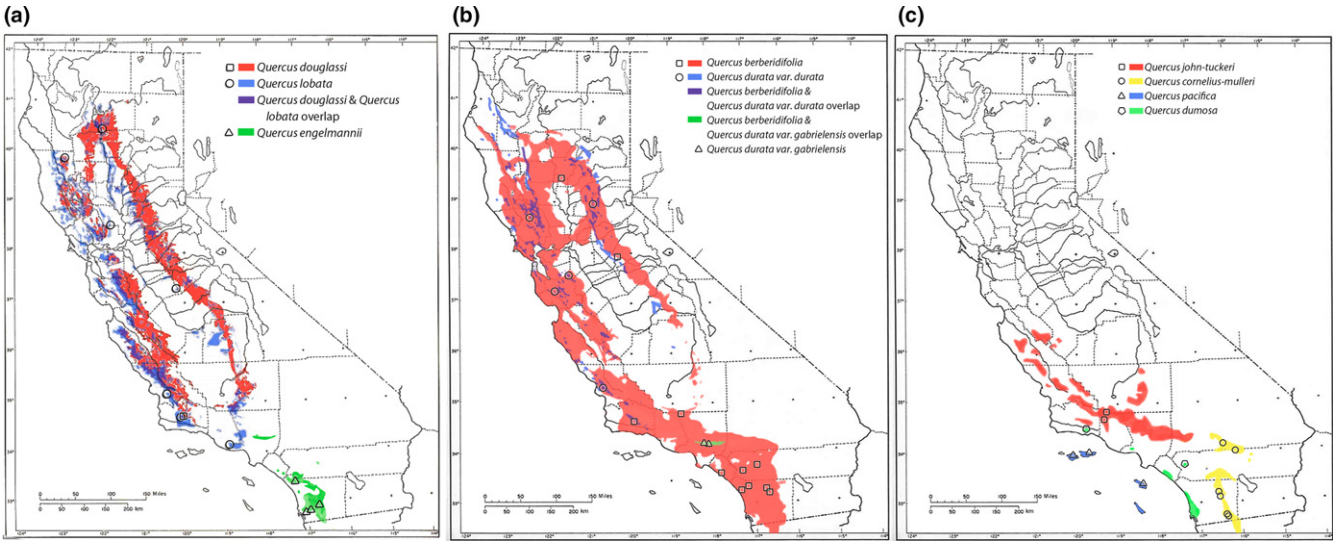

FIGURE 1 Map of the species distribution of study oak species of *Quercus* section *Quercus* occurring in California: (a) Tree white oak species: *Q. douglasii*, *Q. engelmannii* and *Q. lobata*. (b) Scrub white oak species: *Q. berberidifolia*, *Q. durata* var. *durata*, *Q. durata* var. *gabrielensis*. (c) Scrub oak species: *Q. cornelius-mulleri*, *Q. dumosa*, *Q. john-tuckeri* and *Q. pacifica*. Maps are based on a synthesis of several studies (Gugger et al., 2013), with modification based on personal observations (D. Burge, F. Roberts, J. Ortego and V. Sork)

Supporting information Table S1), providing ample opportunity for hybridization and introgression.

Because the two focal tree oaks, *Q. lobata* and *Q. engelmannii*, have very different temporal scales of contact with the scrub oaks, we compare patterns and directions of gene exchange between tree and scrub oaks to learn about biological barriers to gene flow. Both tree oaks are not only sympatric with the scrub oaks, but they are also sufficiently distantly related from each other and the scrub oaks that it is more feasible to distinguish introgression from shared ancestry than if we were to analyse introgression among the scrub oaks. If species are sympatric and pervasive hybridization results in introgression, we should observe evidence of introgression between all sympatric tree and scrub oaks. In particular, we should consistently observe introgression between *Q. lobata* and the scrub oaks due to their long history of proximity and evolution in the California Floristic Province. If long-term contact has instead led to selection against hybrids in the process of speciation (e.g., evolution of reinforcement, Matute, 2010), there should be more instances of introgression between *Q. engelmannii* and scrub oaks as compared to *Q. lobata* and the scrub oaks, since *Q. engelmannii* expanded northwards to become sympatric with the California endemic scrub oaks.

As background for our study of ancient introgression, we first generate a phylogeny of the ten white oak species using thousands of loci generated by RADseq found in California, including a few outgroup oak species. Second, we analyse the population genetic structure and admixture of the specimens to select samples for the analysis of introgression that are good representatives of species, rather than contemporary F1 hybrids or backcrosses. Third, we use the *D*-statistic (Green et al., 2010) to test whether the California scrub oak species complex showed equal evidence for gene flow with the endemic oak *Q. lobata*, compared to the more recently arrived *Q. engelmannii* of Mexican origin, and whether introgression is pervasive between all pairs of these tree and scrub oaks, which are or have been sympatric. Finally, we quantify the demographic history, including migration rates and asymmetry in direction of gene flow, by fitting a series of demographic models to the site frequency spectrum of scrub and tree oaks (Gutenkunst et al., 2009). Findings from this study demonstrate non-pervasive introgression, evidence suggesting the evolution of reproductive isolation between some oak species.

2 | MATERIALS AND METHODS

2.1 | Study species and sampling

The white oaks (*Quercus* sect. *Quercus*, Fagaceae) include a dozen California endemic tree and scrub oak species and subspecies (Nixon, 1997, 2002; Pavlik, Muick, Johnson, & Popp, 1995). The phylogenetic relationships of subsets of the California oaks have been recently explored using microsatellite data (Ortego, Noguerales, Gugger, & Sork, 2015b), AFLP markers (Pearse & Hipp, 2009) and RADseq (Eaton, Hipp, Gonzalez-Rodriguez, & Cavender-Bares, 2015; Fitz-Gibbon et al., 2017; Hipp et al., 2014; McVay et al., 2017b). The most abundant state-wide white oak species are *Q. lobata* and *Q. douglasii*, which occur below the boundary of glaciation during the last glacial maximum. The other California endemic species is *Q. garryana* with the tree form occurring in the northern part of the state and ranging north of the glacial boundary of the LGM and some shrub varieties occurring in southern Sierra and Tehachapi ranges (Griffin & Critchfield, 1972).

The first focal tree species is *Q. lobata*, which is a widespread endemic to most of California (see Figure 1) and has been present in California for millions of years (Axelrod, 1983). While the current southern boundary of its current distribution is located in Los Angeles County (Figure 1a), species distribution models predict that its southern boundary went as far south as northern Baja California during the last glacial maximum (Gugger, Ikegami, & Sork, 2013). This southern boundary is supported by records of hybrids between *Q. lobata* and *Q. cornelius-mulleri* in Joshua Tree National Park, a desert region of southern California (Tucker, 1968). Thus, it has probably had the opportunity to be sympatric with almost all of the scrub oak species in this study, as well as *Q. douglasii*, since that more recent clade evolved.

In contrast to *Q. lobata*, the second focal tree species is a non-California endemic tree species, *Q. engelmannii*, which is found in southern California and northern Baja California, Mexico, and is related to a Mexican clade with its nearest relatives being *Q. oblongifolia* and *Q. arizonica* (McVay et al., 2017b). It is not known how long *Q. engelmannii* has been sympatric with the scrub oak species, but molecular and morphological evidence documents contemporary hybridization with *Q. berberidifolia* and *Q. cornelius-mulleri* in several southern California localities (Ortego et al., 2014; Riordan et al., 2016; Roberts, 1995; Sork et al., 2016b). Herbaria records include several observations of hybrid individuals among California endemic oaks based on morphology, and some are sufficiently common to be named (Pavlik et al., 1995; Roberts, 1995), including *Q. x alvordiana* (*Q. douglasii* × *Q. john-tuckeri*), *Q. x macdonaldii* (*Q. pacifica* × *Q. lobata*) and *Q. x acutidens* (*Q. engelmannii* × *Q. cornelius-mulleri*).

This study analyses 55 specimens of ten white oak species (*Quercus* sect. *Quercus*) currently found in California (Table 1). Our samples were largely collected between 2010 and 2014 as part of a separate study of Californian scrub white oaks (Ortego et al., 2015b) with additional specimens of those species plus *Q. durata* var. *gabrielensis* collected during 2015 (see Supporting information Table S2). These samples were also included in a phylogenetic methods study using California oaks (Fitz-Gibbon et al., 2017). Samples represent the entire geographic ranges of all seven Californian scrub white oak taxa within California, but do not include southernmost distribution limits in northern Baja California, Mexico, of *Q. berberidifolia*, *Q. dumosa* and *Q. cornelius-mulleri*. Leaves were collected from wild plants in the field, kept fresh on ice during transport and stored at −80°C until DNA extraction. Species identification was based on leaf, bark, acorn and stem characters following several sources (Hickman, 1993; Roberts, 1995) and the Flora of North America (http://www.efloras.org/flora_page.aspx?flora_id=1).

We created maps of the geographic distributions for the ten California white oak taxa in this study (Figure 1a–c) by integrating

information from several sources (Griffin & Critchfield, 1972; Kruckeberg, 1984; Ortego et al., 2015b; Riordan et al., 2016; Roberts, 1995; Tyler, Kuhn, & Davis, 2006) as well as personal observations of VL Sork, J Ortego, D. Burge and F. Roberts. We also utilized information from herbarium specimens and maps of localities available from CalFlora (http://www.calflora.org), although cautiously, because herbarium specimens of California oaks, especially the scrub oaks, are often misidentified due to morphological intergradation and changes in taxonomic circumscription over time.

2.2 | Library preparation and high-throughput sequencing

2.2.1 | DNA extraction and purification

DNA was extracted using the Qiagen DNeasy Mini Plant Kit and purified with AMPure XP beads using the Agencourt AMPure XP Purification Protocol. Samples were loaded in a 96-well plate, with some in duplicate or triplicate and shipped to Floragenex, Inc. (Eugene, OR), for library prep, including barcode ligation, using same methods as other oak phylogenetic studies (Cavender-Bares et al., 2015; Hipp et al., 2014; McVay et al., 2017b), including use of the restriction enzyme PstI to digest DNA. Libraries were sequenced at the Broad Stem Cell Research Center (BSCRC) at UCLA with an Illumina HiSeq 2000 using single-end 100-bp sequencing.

2.2.2 | RADseq mapping and variant discovery

After demultiplexing, the average number of 100-bp reads per sample was 1.74 M, minimum 0.37 M (Supporting information Table S3). Quality was checked with FASTQC (Andrews, 2010). We added sequence data from *Q. alba* and *Q. kelloggii* (courtesy of A. Hipp, Morton Arboretum), which were generated at Floragenex (Eugene, OR) using the same methods used for our samples. We aligned our reads, using BWA-MEM v.0.7.12-r1039 (Li, 2013) to the collapsed version of the *Q. lobata* reference genome v0.5 (Sork et al., 2016a), available from https://doi.org/10.5061/dryad.6pg15js or https://valleyoak.ucla.edu/genomicresources, for which haplotype redundancy had been aggressively removed (for details see Sork et al., 2016a). Variant discovery was done with GATK's HaplotypeCaller v3.6-0-g89b7209 (McKenna et al., 2010) using default parameters. To confirm appropriate parameters for hard filtering, we inspected variant calls along with the read alignments using IGV (Robinson et al., 2011). We chose to only filter the variants by QD (confidence normalized by read depth) as well as requiring them to be biallelic. GATK's VariantFiltration was used to apply a QD < 2.0 filter tag. VCF tools v0.1.15 (Danecek et al., 2011) was used to filter non-biallelic variants. This filtering decreased the number of variants from 382,290 (T_s/T_v = 1.43) to 351,865 (T_s/T_v = 1.48).

TABLE 1 List of ten focal taxa and number of 55 samples and localities used in the phylogenetic analysis of California white oak species (*Quercus* sect. *Quercus*)

Growth form	Species	Number of samples	Number of localities
Scrub oaks	*Q. berberidifolia*	17	11
	Q. cornelius-mulleri	4	4
	Q. dumosa	2	2
	Q. durata var. *durata*	7	5
	Q. durata var. *gabrielensis*	4	2
	Q. john-tuckeri	3	2
	Q. pacifica	6	4
Tree oaks	*Q. douglasii*	3	1
	Q. engelmannii	3	3
	Q. lobata	6	6

Notes. Single samples of *Q. alba* (*Quercus* sect. *Quercus*) *and Q. kelloggii* (*Quercus* sect. *Lobatae)* were included as outgroup species. See Supporting information Table S1 for location information of samples and subsamples used in introgression analyses.

2.2.3 | Phylogenetic inference

Sequences for tree building were from concatenation of RADseq loci for the 56 individuals sampled for this study plus sequence data from one *Q. alba* and two *Q. kelloggii* (*Quercus* sect. *Lobatae*) samples (provided by A. Hipp and P. Manos). We chose this method to allow comparison of our phylogenetic trees with more complete phylogenies published elsewhere (Fitz-Gibbon et al., 2017; McVay et al., 2017b). Although concatenated data analysis does not model the processes underlying gene trees (McVay et al., 2017a,b), our emphasis here is on the relationship of the scrub white oaks with the two focal tree species, and not on a complete phylogeny of *Quercus* sect. *Quercus*.

We identified targeted RADseq loci using GATK's CallableLoci. Genome regions considered callable by CallableLoci's default parameters were kept if they occurred in at least six samples and were >80 bp long, resulting in 36,319 regions with an average size of 135 bp. The average size is larger than the RADseq targets due to merging of overlapping targeted regions on either side of many of the restriction enzyme cut sites. Sample-specific sequences for each of these regions were generated, incorporating variants, using ambiguity codes for heterozygotes and Ns for uncalled variants. Heterozygous indels were recorded as Ns. These retained regions were concatenated to create a single 4,987,878 bp sequence for each sample. Overall there was 33.68% missing data. Simple indel coding (Simmons & Ochoterena, 2000) was done using the 2matrix utility (Salinas & Little, 2014). Maximum-likelihood phylogenies were built using RAxML 8.2.0 (Stamatakis, 2014), with rapid bootstrap support estimated from 443 replicate trees added to the best-scoring of 10 maximum-likelihood trees built from 10 distinct maximum parsimony starting trees. There were 984,503 distinct alignment patterns (973,714 SNPs and 10,789 coded indels treated as binary characters). We used the GTRGAMMA model of nucleotide substitution. The resulting tree was plotted with FIGTREE 1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/).

2.3 | Analysis of population structure

To assess the population structure of the samples associated with each species, we constructed PCA plots using all biallelic variants for each sample set as the input into the R package SNPRelate (Zheng et al., 2012). To gain additional insight about the population structure among the scrub oak species, we conducted STRUCTURE (Pritchard, Stephens, & Donnelly, 2000) on that subset of individuals. Methods are described in Supplemental Information legend of Figure S1.

2.4 | Analysis of introgression

2.4.1 | *D*-statistic

The *D*-statistic is a test of a tree rooted by an outgroup O, with topology (((*P*$_1$,*P*$_2$),*P*$_3$),*O*), which looks at alleles in the configuration (((*A*,*B*),*B*),*A*) (*ABBA*) or (((*B*,*A*),*B*),*A*) (*BABA*). *A* denotes the ancestral allele and is determined by the outgroup while *B* denotes the derived allele.

$$D = \frac{n_{ABBA} - n_{BABA}}{n_{ABBA} + n_{BABA}}$$

In the absence of gene flow between P_3 and either P_1 or P_2, *ABBA* and *BABA* sites occur at a similar frequency due to incomplete lineage sorting. However, if there has been gene flow between P_3 and either P_1 or P_2, *ABBA* and *BABA* sites should appear asymmetrically in the data. Under the null hypothesis of no gene flow, the expectation of the numerator and thus the *D*-statistic is 0.

This statistic counts the allele configuration for a single lineage from each of the four taxa. Alternatively, the allele frequencies in a sample can be used to compute an expected *D* (Patterson et al., 2012), which should increase power by incorporating information from multiple lineages as long as SNP ascertainment is not systematically biased (Durand, Patterson, Reich, & Slatkin, 2011). Denoting the sample allele frequencies in groups P_1, P_2, P_3 and *O* as p_1, p_2, p_3 and p_O, respectively, the *D*-statistic in the above equation can be formulated in terms of allele frequencies. It is then (Patterson et al., 2012):

$$D = \sum_i \frac{\left(p_3^{(i)} - p_O^{(i)}\right)\left(p_2^{(i)} - p_1^{(i)}\right)}{\left(p_3^{(i)} + p_O^{(i)} - 2p_3^{(i)}p_O^{(i)}\right)\left(p_1^{(i)} + p_2^{(i)} - 2p_1^{(i)}p_2^{(i)}\right)}$$

The variance in *D* is estimated by resampling SNPs from the genome and used to compute the distance of *D* from 0 in terms of a *Z*-score, assuming that *D* is approximately normally distributed. In other words, we tested the null hypothesis that $D = 0$.

In the original implementation of *D* (Green et al., 2010), the weighted block jackknife (Busing, Meijer, & Van Der Leeden, 1999) is used to account for the non-independence of SNPs due to linkage by resampling segments of the genome. An alternative approach is to treat each SNP as independent through bootstrap resampling. Our RAD sequences represent an incomplete sampling of loci over the genome, with some linkage among neighbouring SNPs. In addition, the RAD sequences were aligned to a draft genome without chromosomal determination (Sork et al., 2016a). Thus, we estimated the variance of *D* using both the weighted block jackknife and the bootstrap. For the jackknife, we sampled from the concatenated genomic contigs. For the bootstrap, we resampled treating each SNP as an independent observation. The bootstrap is less conservative since it will treat some linked SNPs as independent observations, while the jackknife represents a conservative treatment of the data as independent loci may be sampled together.

The *D*-statistic was computed using sample allele frequencies for all pairwise combinations of scrub oak species (Supporting information Tables S4, S5: *Q. berberidifolia*, *Q. durata* var. *durata*, *Q. durata* var. *gabrielensis*, *Q. pacifica*, *Q. john-tuckeri*, *Q. douglasii*, *Q. dumosa* and *Q. cornelius-mulleri*) in the P_1 and P_2 positions. Individuals from *Q. engelmannii* or *Q. lobata* were used in the P_3 position. Finally, *Q. kelloggii* was used as the outgroup in all analyses. Because the *D*-statistic was computed for many pairwise combinations of clades at the P_1 and P_2 positions, we present our results both with and without multiple testing corrections using the Benjamini–Hochberg procedure (Benjamini & Hochberg, 1995). Note, we tested the hypothesis of $D = 0$ once per unique combination of groups at the P_1 and P_2 positions since the two positions are interchangeable. For these analyses, we used variant data sets generated for the phylogenetic trees, selecting loci with no missing data respective to the test being performed. To retain balanced sample sizes for each test, we did not use all the individuals used in the phylogeny. Lastly, we divided the *Q. berberidifolia* samples into samples from north and south of the Transverse Ranges. Because we have no evidence that *Q. engelmannii* ever ranged north of that topographic barrier, separating the samples allowed us to investigate whether introgression patterns differed between the northern and southern regions.

2.4.2 | Model-based demographic analysis

As a complement to the *D*-statistic, we conducted a model-based analysis of introgression by utilizing the site frequency spectra (SFS) of scrub oaks and tree oaks. We did this by treating groups containing multiple scrub oak species as single populations so that we could make inferences about gene flow based on sites that are polymorphic in white oaks. Although we approximate the allele frequency changes in multiple species with a single population, frequencies of neutral alleles should, on average, not change between clades except due to migration. In other words, any covariance of neutral allele frequencies between two clades, respective to the ancestral allele frequencies, is driven by gene flow events (Patterson et al., 2012; Pickrell & Pritchard, 2012). For these reasons, the migration rates we infer should be at least qualitatively representative of overall gene flow between scrub oaks and tree oaks, despite the fact that any demographic model we fit is only an approximation of the joint demography, as long as the scrub oak and tree oak groups represent two divergent clades.

Based on the phylogenetic analysis as reported in the Results section and denoted in Supporting information Table S2, the scrub oaks and *Q. douglasii* were grouped into two major clades—

cornelius-mulleri clade (Clade C-M), which includes *Q. cornelius-mulleri*, other scrub oak species as well as *Q. douglasii*, and the berberidifolia clade (Clade B), which includes *Q. berberidifolia*, and other scrub white oak species. We constructed four separate 2-dimensional SFSs for each combination of either Clade C-M or Clade B and one of the tree oaks, *Q. engelmannii* or *Q. lobata*. Specifically, each 2D SFS was built by pooling the individuals used in the *D*-statistic analyses into each of their respective major groups, initially into groups of 19, 25, 3 and 3 diploids for Clade C-M, Clade B, *Q. engelmannii* and *Q. lobata*, respectively. Because many sites contained missing data for some individuals, each SNP was projected down to 12, 15, 2 and 2 diploids (respectively) so that the majority of sites could be used. *Q. kelloggii* was used as an outgroup to polarize the SFS.

Next, we used the Python package ∂a∂i (Gutenkunst et al., 2009) to fit a series of demographic models. We allowed for a period of exponential growth in the scrub oaks in our model because single and double size change models under-predicted the rare variation of the scrub oaks. In general, the models we fit were based on a 2-population divergence model (see Figure 2) and were set up as follows. Going forward in time, an equilibrium size population diverges into two populations, which we denote as populations 1 and 2. At this time, each has size N_1 and N_2. After some time T_1, in this state, population 1 begins a period of exponential growth for time T_2 until the current day, when it is at size N_{1B}. The size of population 2 does not change and is N_2 during time intervals T_1 and T_2. Importantly, populations 1 and 2 model the demography of the scrub oaks (Clades B and C-M, separately) and tree oaks (*Q. lobata* and *Q. engelmannii* separately), respectively. In this series of models, migration rates between populations 1 and 2 were varied from no gene flow ($M_{12} = M_{21} = 0$), to a model with symmetric gene flow ($M_{12} = M_{21}$) and lastly to a model with both migration rates M_{12} and M_{21} as free parameters. Therefore, the simplest model (Model 0) had 5 free parameters (N_1, N_{1B}, N_2, T_1 and T_2), and each subsequent model allowed additional flexibility in migration with Model 2 having 7 free parameters (N_1, N_{1B}, N_2, T_1, T_2, M_{12} *and* M_{21}). Lastly, each model was fit by maximizing the multinomial likelihood; that is, the inference was conditioned on the number of SNPs in the data and did not depend on an a priori knowledge of the mutation rate.

Because these demographic models represent a series of nested models, each with one additional free parameter, we tested the improvement in fit between models 0, 1 and 2 in a likelihood-ratio test framework, where the likelihood-ratio test statistic was defined as -2 times the log ratio of the likelihood of the null model (Θ_0) to the alternative model (Θ_A):

$$\Lambda(x) = -2\ln\left(\frac{\mathscr{L}(\Theta_0|X)}{\mathscr{L}(\Theta_A|X)}\right)$$

This likelihood-ratio test statistic should be approximately χ^2 distributed with degrees of freedom equal to the additional number of free parameters in Θ_A compared to Θ_0. Because the improvement in model fit was tested for a series of 3 models (2 tests), a Bonferroni correction was used to compensate for multiple tests.

3 | RESULTS

3.1 | A phylogeny of California white oaks

Our phylogenetic tree utilizes concatenated RADseq data to create a single 4,987,878 bp sequence for each sample of the ten white oak species sampled across California, the northeastern North American oak, *Q. alba* (sect. *Quercus*), and the California red oak, *Q. kelloggii* (sect. *Lobatae*), to root the tree.

The phylogenetic tree revealed that our samples of focal California white oaks fall into five major clades (Figure 3). The first clade

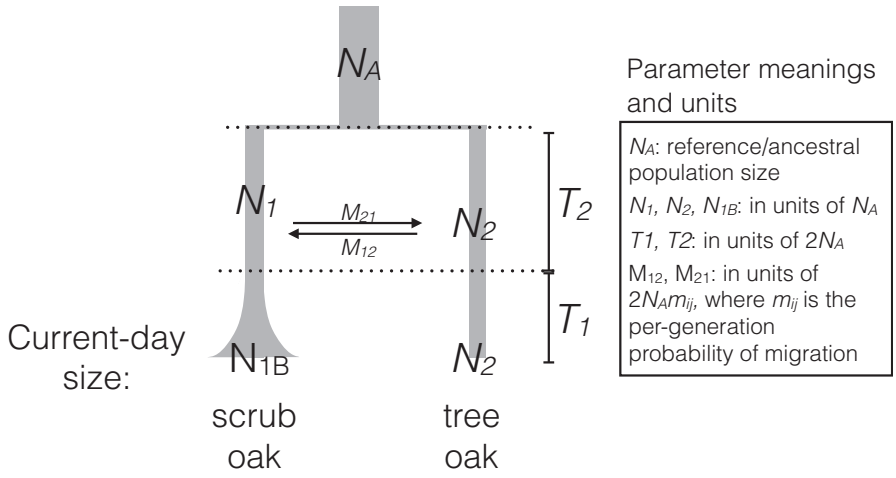

FIGURE 2 The two-population demographic model fit to scrub oak and tree oak groups. This model was fit to the two-dimensional site frequency spectrum of each of the four combinations of pooled scrub oak groups (Clades C-M and B) and tree oaks (*Q. lobata* and *Q. engelmannii*). The parameters of the best-fitting models are presented in Table 3 are: N_1, N_2, N_{1B} representing population sizes relative to the ancestral population; T_1, T_2 representing time in units of $2N_A$ generations; M_{12}, M_{21} representing migration rates in units of $2N_A m$, where m is the proportion of individuals that are migrants each generation; and M_{ij} indicates the migration rate from population j into population i

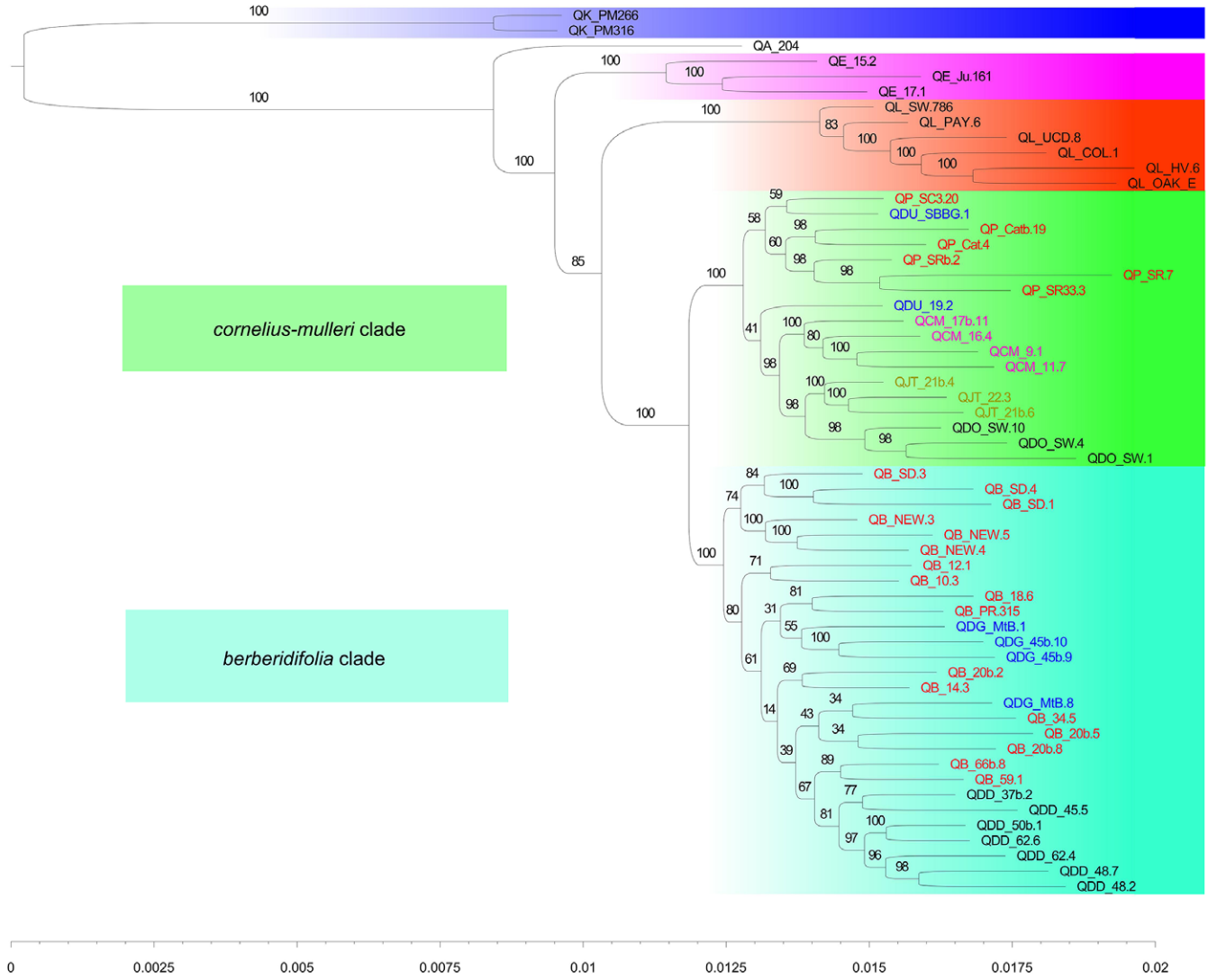

FIGURE 3 Phylogenetic tree of individuals sampled in California belonging to *Quercus* section *Quercus*. The scrub oaks fall into two main clades: the berberidifolia clade (QB, *Q. berberidifolia*; *QDD*, *Q. durata* var. *durata*; QDG; *Q. durata* var. *gabrielensis*) and the cornelius-mulleri clade (QCM, *Q. cornelius-mulleri*; *QJT*, *Q. john-tuckeri*; *QDO*, *Q. douglasii*; *QDU*, *Q. dumosa*; QP, *Q. pacifica*). This phylogenetic tree includes three California tree oaks: *Q. douglasii* (found within Clade C-M), *Q. lobata* (*QL*) and *Q. engelmannii* (*QE*) and *Q. alba*. Numbers following species abbreviations indicate individual identifications (see Supporting information Table S1) [Colour figure can be viewed at wileyonlinelibrary.com]

contains the *Q. engelmannii* individuals. The second clade is comprised of the individuals belonging to *Q. lobata*. The third clade includes all the scrub oak species plus the tree oak, *Q. douglasii*. This clade includes two major subclades: the cornelius-mulleri clade (Clade C-M) and the berberidifolia clade (Clade B). Clade C-M includes *Q. cornelius-mulleri*, *Q. john-tuckeri*, *Q. douglasii*, *Q. dumosa* and *Q. pacifica*. These five species in clade C-M are separated into two groups, with *Q. dumosa* and *Q. pacifica* clustering into one group and the other three species in another group, with *Q. john-tuckeri* and the tree oak, *Q. douglasii*, as sister taxa. *Q. dumosa* is a coastal scrub oak and *Q. pacifica* is found on the California Channel Islands (Figure 1b). Clade B includes *Q. berberidifolia*, *Q. durata* var. *durata* and *Q. durata* var. *gabrielensis*. The most widespread California scrub oak, *Q. berberidifolia*, overlaps in distribution with other two taxa within its subclade and with several species from the other subclade (see Figures 1b,c). The serpentine scrub oak, *Q. durata* var. *durata*, has strong bootstrap support as a separate taxon from *Q. berberidifolia* while *Q. durata* var. *gabrielensis* is interspersed with *Q. berberidifolia*, but not with the other subspecies of *Q. durata*. Because this subspecies has unique morphological traits distinct from *Q. berberidifolia* (Nixon, 2002; Roberts, 1995), it is likely that the *gabrielensis* variety may be its own species or a variety of *Q. berberidifolia*, but not a variety of *Q. durata*.

3.2 | Population structure in the California white oaks

The PCA plots illustrate the genetic relationships among the samples used in the phylogenetic analysis, without the constraints of bifurcation. When including *Q. kelloggii*, the distant outgroup, the tree oaks

—*Q. alba*, *Q. engelmannii* and *Q. lobata*—are a distinctly different cluster from the scrub oak species and blue oak (Figure 4a). When *Q. kelloggii* is excluded (Figure 4b), *Q. alba* and *Q. engelmannii* are more similar to each other than to *Q. lobata*, and *Q. lobata* is only slightly more similar to the cluster of scrub oaks and blue oak than is *Q. engelmannii*. Finally, when we remove those three tree species (Figure 4c), the samples cluster together as observed in the phylogeny (Figure 3). These PCAs provide additional support that *Q. lobata* and *Q. engelmannii* are different genetically from each other and that the scrub oak species are difficult to separate except the two clusters that correspond to the clades in the phylogeny. Within each of the B and C-M clades, it would be very difficult to determine whether the individuals share alleles through introgression or recent divergence.

The analysis of the structure of species within Clades B and C-M clades based on STRUCTURE (Supporting information Figure S1) also illustrates the same pattern of separation between the two clades (statistics described in legend of Supporting information Figure S1) and, in agreement with the PCA findings, shows that the species within each clade do not show clear genetic differentiation. These patterns are the same even when we run STRUCTURE on all the study samples (data not shown). This population structure analysis does not distinguish the species as clearly as the phylogenetic tree, but they also indicate that *Q. durata* var *durata* and *Q. douglasii* are genetically distinct.

3.3 | Ancient introgression between focal tree species with scrub white oak species

To test the hypothesis of pervasive introgression within California oaks, we focused on testing for asymmetric patterns of allele sharing between the tree white oaks *Q. engelmannii* and *Q. lobata* with the scrub oaks of Clades B and C-M using the *D*-statistic, conditioning upon the phylogeny inferred previously (Figure 3).

Overall, our tests for patterns of increased allele sharing between scrub oak lineages and *Q. engelmannii* showed little variation in the amount of *Q. engelmannii* sharing within either of the two major clades (Table 2), but consistently increased allele sharing between *Q. engelmannii* and Clade B relative to Clade C-M (Figure 5; Supporting information Table S4). Nearly all the tests using populations from two different major clades in the P_1 and P_2 positions (i.e., Clade B to Clade C-M, Figure 5b) were significant, and the tests always showed more allele sharing with Clade B. In contrast, the tests where populations from the same major clade were compared (i.e., Clade C-M to Clade C-M, Figure 5a; or Clade B to Clade B, Figure 5c) showed little differences in allele sharing with *Q. engelmannii*, although we point out two significant Clade C-M to Clade C-M contrasts involving the tree oak, *Q. douglasii*. In this case, our tests always favour less allele sharing of *Q. douglasii* with *Q. engelmannii*. These results were largely insensitive to whether both P_1 and P_2 groups were currently or historically sympatric with *Q. engelmannii*. Finally, these findings were

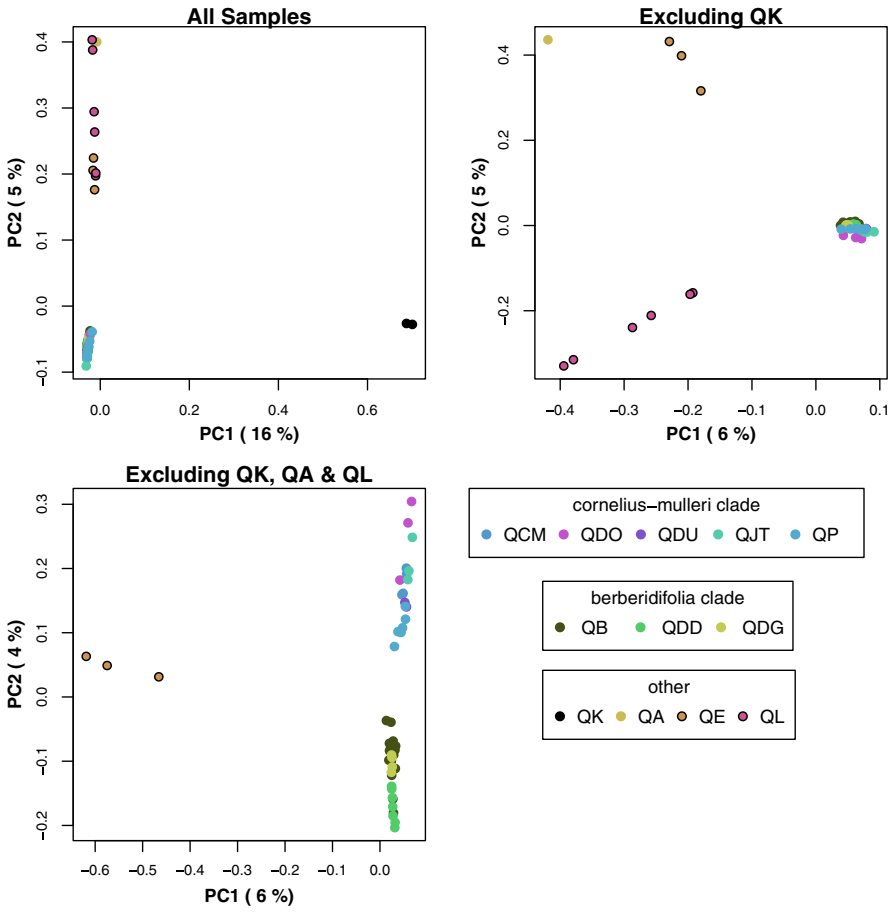

FIGURE 4 Principal component analysis performed on biallelic variants using the R package SNPRelate (Tucker, 1968). The first two components are shown along with the per cent of variation explained by each component in brackets. (a) Plot is based on 351,865 biallelic variants found across all 58 samples. (b) *Quercus kelloggii* variant data were removed along with resulting non-variant sites, leaving 306,739 variants used in this plot. (c) *Quercus kelloggii*, *Quercus alba* and *Quercus lobata* variant data were removed along with resulting non-variant sites, leaving 285,631 variants used in this plot

TABLE 2 Summary of *D*-statistic tests for California scrub white oaks with *Q. engelmannii* and *Q. lobata*, separately, for three sets of pairwise species contrasts: within the berberidifolia clade (Clade B), between the berberidifolia clade (Clade B) and the cornelius-mulleri clade (Clade C-M), and within the cornelius-mulleri clade (Clade C-M)

Position one	Position two	Number of significant tests	Total number of tests
Testing for introgression of *Q. engelmannii*			
Clade B	Clade B	0	6
Sympatric	Parapatric	0	3
Clade B	Clade C-M	19	20
Sympatric	Parapatric	9	10
Clade C-M	Clade C-M	2	10
Sympatric	Parapatric	1[a]	6
Testing for introgression of *Q. lobata*			
Clade B	Clade B	2	6
Sympatric	Parapatric	1	4
Clade B	Clade C-M	6	20
Sympatric	Parapatric	3	10
Clade C-M	Clade C-M	4	10
Sympatric	Parapatric	3[b]	6

Notes. The test determines which clade in positions one versus two shares most loci with *Quercus engelmannii* and *Q. lobata* (position 3), using *Q. kelloggii* as outgroup (position 4) for all tests (see Figure 3). Clade B includes *Q. berberidifolia*, *Q. durata* var. *durata*, *Q. durata* var. *gabrielensis*. Clade C-M includes *Q. cornelius-mulleri*, *Q. douglasii*, *Q. dumosa*, *Q. john-tuckeri* and *Q. pacifica*. Also summarized are the subset of tests conducted when species contrasts in position one are sympatric and position two are parapatric with position 3 species.

[a]*Q. douglasii* shows decreased allele sharing.

[b]*Q. douglasii* always shows increased allele sharing.

robust to the way the data were resampled to assess significance as similar results were obtained when using the bootstrap and jackknife approaches (Figure 5; Supporting information Table S4).

When we performed a similar series of tests for excess allele sharing between scrub oaks and *Q. lobata*, a complex pattern of allele sharing emerged (Figure 6; Supporting information Table S5). Unlike the tests with *Q. engelmannii*, there was no clear pattern of allele sharing dominated by a single major clade. In Clade B to Clade B comparisons, *Q. durata* var. *durata* and northern individuals of *Q. berberidifolia* appeared to share more alleles with *Q. lobata* than did southern individuals of *Q. berberidifolia*, but a simple pattern of increased allele sharing in particular groups was not present (Figure 6a). When species from Clade C-M were compared to other species from Clade C-M, *Q. douglasii* consistently shared more alleles with *Q. lobata* than all other members of Clade C-M (Figure 6c). When species were compared between the major clades, northern populations of *Q. berberidifolia* shared more with *Q. lobata* (Figure 6a) except when compared to *Q. douglasii* (Figure 6b); *Q. durata* var. *durata* always shared more alleles with *Q. lobata* (Figure 6a) except when compared to *Q. douglasii*, but not always significantly; and *Q. douglasii* always shared more alleles with *Q. lobata* except when compared to northern *Q. berberidifolia*, but never significantly more. Furthermore, patterns of allele sharing are generally inconsistent with historic or current patterns of sympatry, even though this inconsistency does not appear to be specific or ancestral to any major clade. Because of this, and because the *D*-statistic is used to test for relative differences in allele sharing, it is possible that patterns of allele sharing within each major clade obfuscated signals of gene flow when comparing species between the major clades. Lastly, the significance of our tests was dependent on the way the data were resampled (Figures 5, 6 and Supporting information Tables S4, S5).

We further investigated the cause of increased allele sharing between *Q. engelmannii* and the scrub oaks by plotting the doubly conditioned frequency spectrum (DCFS) of Clade B, conditioning on sites that are only in the ancestral state in Clade C-M and in the derived state in Clade B and *Q. kelloggii*. Persistent population structure in the common ancestor of Clades C-M and Clade B may produce similar patterns of allele sharing as recent gene flow, but a simple model of ancestral structure should generate a linear DCFS (Yang, Malaspinas, Durand, & Slatkin, 2012). The DCFS of Clade B is non-linear (Supporting information Figure S2) and therefore inconsistent with a model of simple ancestral population structure. Nevertheless, it should be noted that complex models of divergence and structure may produce a non-linear DCFS (Eriksson & Manica, 2012).

Our demographic inference showed that the history of the California white oaks is characterized by older population contractions in both tree and scrub white oak clades, but also that scrub white oaks have recently grown in population size while the effective population size of tree oaks has remained small relative to the ancestral population size (Figure 2; Supporting information Table S6). The marginalized (1-dimensional) and 2-dimensional frequency spectra are provided in Supporting information Figures S3–S7. We note that this relative difference does not preclude the possibility of recent growth in the tree oaks, but it does imply that environmental conditions may have favoured the more drought tolerant scrub oaks over the tree oak form as the region has warmed since the last glacial maximum. Signals of recent growth are found in patterns of rare variation, which we do not capture with our limited sample size of tree oaks. In addition, our best-fit models are most consistent with asymmetric gene flow from scrub to tree white oaks (Table 3) for all 4 possible pairs of scrub and tree oaks ($p < 1 \times 10^{-15}$ in all cases; Supporting information Tables S5, S6). The observed SFS and the expected SFS of these models are compared in Supporting information Figure S2. When our demographic model was fit with *Q. engelmannii* as the tree oak, migration rates from scrub oaks to *Q. engelmannii* were ~15 times larger than migration rates from *Q. engelmannii* to scrub oaks. When our demographic model was fit with *Q. lobata* as the tree oak, the migration rate from scrub oaks to *Q. lobata* was 20–30 times larger than migration rates from *Q. lobata* to scrub oaks. This difference was mostly driven by a decrease in the migration rate from *Q. lobata* into scrub oak. Additionally, the absolute migration rate from Clade B into the tree oaks was always larger than the rate from Clade C-M into the tree oaks, and this relative difference was greater when *Q. engelmannii* was used as the

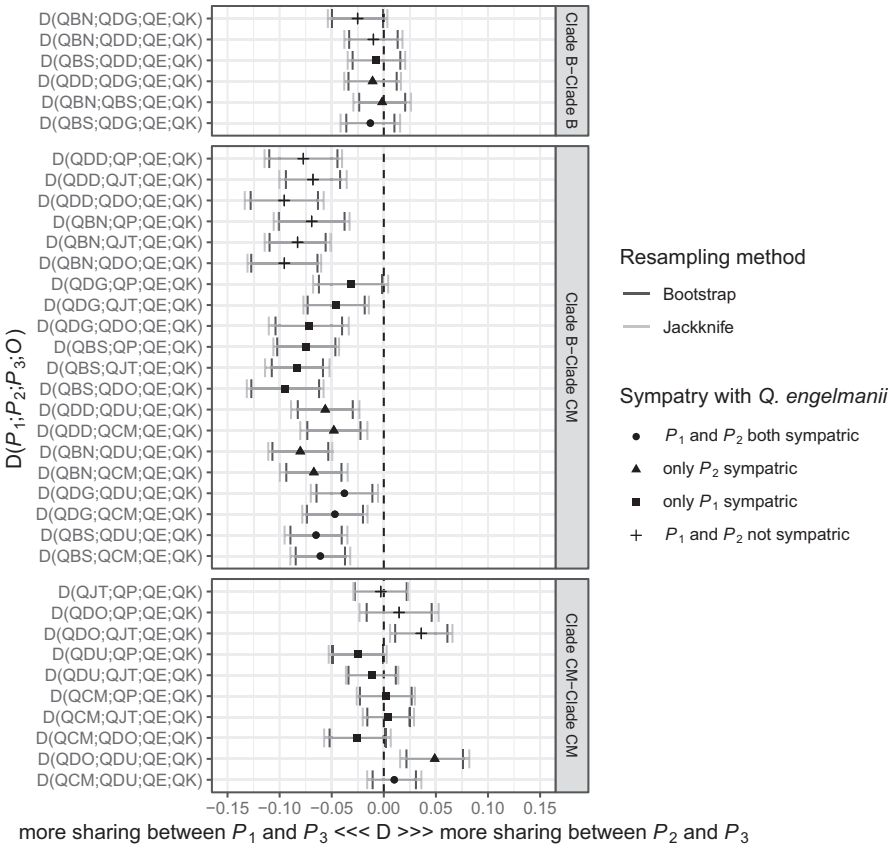

FIGURE 5 *D*-statistics computed with *Q. engelmannii* in the P_3 position and *Q. kelloggii* in the *O* position for species pairs as $D(P_1, P_2, P_3, O)$ within and between both major Californian scrub white oak clades. The tests are sorted by the major clade that the species in positions P_1 and P_2 were drawn from: Clade B and Clade B; Clade B and Clade C-M; or Clade C-M and Clade C-M. The black point represents the *D*-statistic computed for the full data set. Error bars represent 95% confidence intervals estimated by bootstrap (dark grey) or jackknife (light grey) resampling. The black dashed line represents the null expectation and the null hypothesis is rejected when it does not fall within the confidence interval for a particular test. A *D*-statistic skewed to the left (i.e., negative) indicates relatively more allele sharing between P_1 and *Q. engelmannii*; a *D*-statistic skewed to the right (i.e., positive) indicates relatively more allele sharing between P_2 and *Q. engelmannii*

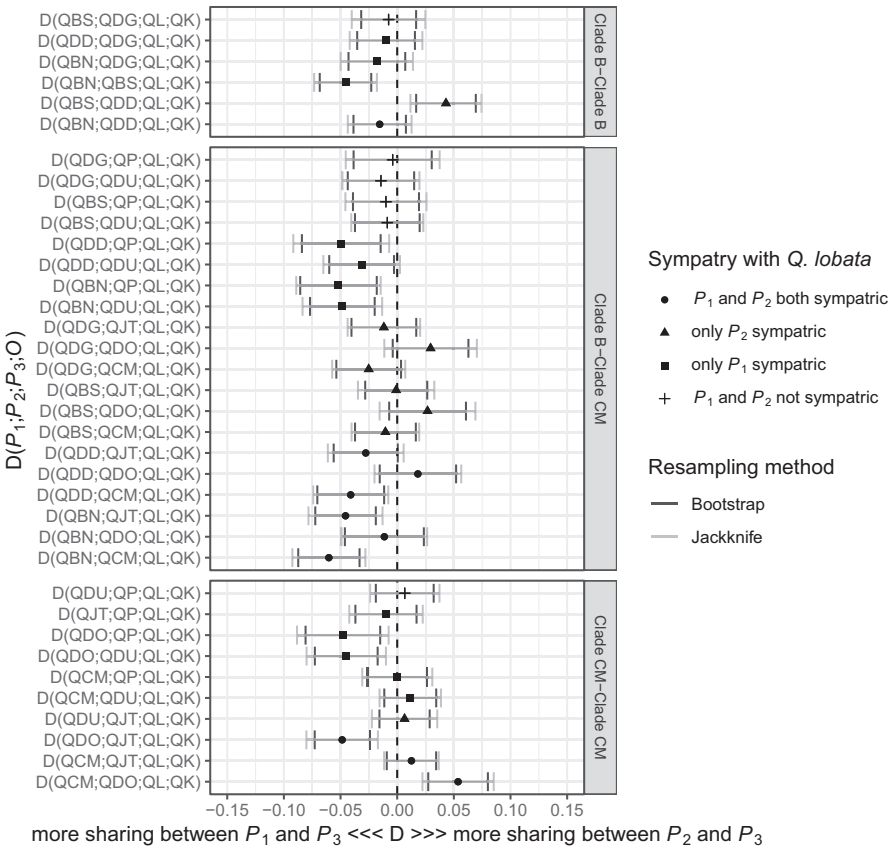

FIGURE 6 *D*-statistics computed with *Q. lobata* in the P_3 position and *Q. kelloggii* in the *O* position for species pairs as $D(P_1, P_2, P_3, O)$ within and between both major Californian scrub white oak clades. The tests are sorted by the major clade that the species in positions P_1 and P_2 were drawn from: Clade B and Clade B; Clade B and Clade C-M; or Clade C-M and Clade C-M. The black point represents the *D*-statistic computed for the full data set. Error bars represent 95% confidence intervals estimated by bootstrap (dark grey) or jackknife (light grey) resampling. The black dashed line represents the null expectation and the null hypothesis is rejected when it does not fall within the confidence interval for a particular test. A *D*-statistic skewed to the left (i.e., negative) indicates relatively more allele sharing between P_1 and *Q. lobata*; a *D*-statistic skewed to the right (i.e., positive) indicates relatively more allele sharing between P_2 and *Q. lobata*

TABLE 3 Parameter estimates[a] and 95% confidence intervals[b] of the best-fitting demographic model (Model 2) to the 2-dimensional site frequency spectrum of pooled scrub oak and tree oak species

Scrub oak[c]	Tree oak[c]	θ	N_1	N_{1B}	N_2	T_1	T_2	M_{12}	M_{21}
B	QE	14779 ± 357	0.199 ± 0.022	2.57 ± 0.094	0.296 ± 0.010	0.0159 ± 0.0347	0.886 ± 0.0235	0.123 ± 0.010	1.948 ± 0.053
C-M	QE	12373 ± 90	0.328 ± 0.057	3.47 ± 0.141	0.417 ± 0.024	0.0238 ± 0.0757	0.928 ± 0.076	0.0818 ± 0.0067	1.136 ± 0.078
B	QL	12166 ± 290	0.367 ± 0.049	3.65 ± 0.135	0.175 ± 0.010	0.00632 ± 0.0420	0.905 ± 0.041	0.0409 ± 0.0045	1.258 ± 0.082
C-M	QL	12375 ± 376	0.376 ± 0.059	3.93 ± 0.176	0.181 ± 0.014	0.00689 ± 0.0478	0.915 ± 0.061	0.0505 ± 0.0051	1.128 ± 0.108

Scrub oak species includes species in Table 1 plus *Q. douglasii* and were pooled into the two major clades (C-M and B) based on the phylogenetic relationships shown in Figure 2. Tree oak species are *Q. engelmannii* (QE) and *Q. lobata* (QL).

[a]Parameter meanings and units. θ: Population-scaled mutation rate, scaled to the ancestral population size ($4N_A\mu l$). N_1, N_2, N_{1B}: population sizes relative to the ancestral population size (N_A). T_1, T_2: time in units of $2N_A$ generations. M_{12}, M_{21}: migration rates in units of $2N_Am$, where m is the proportion of individuals that are migrants each generation. M_{ij} indicates the migration rate from population j into population i.

[b]Confidence intervals are approximate confidence intervals computed with the observed Fisher Information matrix, that is, assuming that the MLE is approximately normally distributed.

[c]Scrub oaks are denoted as population 1 and tree oaks are denoted as population 2.

tree oak (1.71 times larger) than when *Q. lobata* was used as the tree oak (1.12 times larger, Table 3). This pattern supports the striking difference in *D*-statistic analyses between Clade C-M and Clade B with *Q. engelmannii* in position P_3 as compared to the analyses with *Q. lobata* in P_3.

4 | DISCUSSION

4.1 | Phylogenetic relationships of California scrub white oak species

The resolution of the relationships among California white oaks provides an essential context for understanding how introgression may have contributed to the evolutionary relationships among species. The phylogenetic tree indicates that the recently evolved endemic scrub white oak species (including blue oak *Q. douglasii*) belong to one group evolving from a tree common ancestor. In fact, these species along with *Q. garryana* have been grouped together phylogenetically as the Dumosae (McVay et al., 2017a,b). The focal species selected for our introgression analysis are comprised of two subclades—dubbed the berberidifolia (B) and cornelius-mulleri (C-M) clades (Figure 3). The C-M clade includes the widespread tree oak, *Q. douglasii*, which seems to be a sister taxon to the scrub oak, *Q. john-tuckeri*, and may have evolved from a scrub oak that reverted to a tree form. The evolution of the scrub oaks raises the question of whether this emergence or expansion coincides with the mid-Miocene emergence of Mediterranean-type climate in California with dry summers, wet winters and increased fire frequency (Rundel et al., 2016), which would have favoured shrubs over trees and the expansion of a tree, such as *Q. douglasii* with shrub-like traits, throughout the foothill slopes of the Sierra Nevada ranges.

The bootstrap support for species within subclades is not always strong, suggesting recent divergence. The "berberidifolia" clade includes the widespread *Q. berberidifolia*, the mostly serpentine restricted *Q. durata* var. *durata*, and the narrowly distributed *Q. durata* var. *gabrielensis*. The divergence between *Q. durata* var. *durata* and *Q. berberidifolia* is consistent with earlier suggestions that leather oak is a serpentine specialist that evolved through ecological isolation (Forde & Faris, 1962). Individuals of *Q. durata* var. *durata* form a well-supported clade, distinct from *Q. berberidifolia* and *Q. durata* var. *gabrielensis*, a pattern also seen in the STRUCTURE results. *Q. durata* var. *gabrielensis* appears more closely related to *Q. berberidifolia*, which may be due to lack of complete divergence or hybridization. Regardless, these findings indicate that this taxon should not be considered a variety of *Q. durata*.

The second subclade of California scrub oaks comprises two sets of sister taxa, but here also the bootstrap support separating species within the clade is weak. The first set contains two parapatric scrub oak species, *Q. cornelius-mulleri* and *Q. john-tuckeri*, and blue oak, *Q. douglasii*, which are all inland. Two other species in this subclade—*Q. dumosa* and *Q. pacifica*—are coastal and island species, respectively, with the latter being endemic to the California Channel Islands. For both this clade and Clade B, because of the

close relationship among species, it will be difficult to determine whether shared alleles among them represent ancient introgression or a recent common ancestor.

One potential concern is whether the phylogenetic uncertainty described above could have affected our analyses of introgression. Importantly, the presence of the CM and the B clades suggested by the phylogeny are clearly recapitulated in the PCA (Figure 4) and STRUCTURE (Supporting information Figure S1). The existence of the CM and B clades is the basis of the groupings that are used for the subsequent *D* tests. Because these groupings are so well supported by multiple analyses relying on different aspects of the data, our inferences of introgression are robust to the details of the phylogeny.

The phylogenetic tree illustrates that two tree oaks are sufficiently genetically distinct from the Dumosae scrub white oaks and from each other that this study system is ideal for studying ancient introgression. The California endemic tree oak, *Q. lobata*, is a divergent lineage outside of the Dumosae group (also see McVay et al., 2017a). The divergence of *Q. lobata* and *Q. engelmannii* facilitates a robust test for comparing pervasive ancient introgression of these two species with the Dumosae species.

4.2 | Ancient introgression between California tree and scrub white oaks

The patterns of allele sharing we observe arise from the increased relatedness of certain species of scrub oak to tree oak. The fact that the scrub oaks are ~30 million years diverged from the tree oaks, are highly outcrossing, have large N_e, and low F_{ST} and finally that complex patterns of allele sharing exist between contemporary scrub oaks and tree oaks, suggests a complex history of gene flow between California tree and scrub white oaks rather than complex population structure in the common ancestor of scrub oaks. Further, gene flow appears to have been neither pervasive nor representative of complete reproductive isolation. In the case of *Q. engelmannii*, the patterns of allele sharing are explained predominantly by gene flow from the common ancestor of Clade B into the ancestor of *Q. engelmannii*. We believe this explanation is most likely because all the species in Clade B share similar proportions of alleles with *Q. engelmannii* (Figure 5), yet consistently more than the species in Clade C-M, with the directionality of gene flow established by our model-based analyses. While it is possible that *Q. engelmannii* admixed with each species in Clade B individually, this explanation would require that all the taxa in Clade B have independently acquired similar proportions of shared alleles with *Q. engelmannii* regardless of the fact that the ranges of some species in Clade B do not overlap with the range of *Q. engelmannii*.

Within Clade C-M, some species share significantly more alleles with *Q. engelmannii* than other species in Clade C-M, indicative of gene flow between recently sympatric populations of these two clades. The species of Clade C-M that exhibit these patterns are: *Q. dumosa*, *Q. john-tuckeri*, and while not statistically significant *Q. cornelius-mulleri* follows this trend. Both *Q. dumosa* and *Q. cornelius-mulleri* are currently sympatric with *Q. engelmannii*, and F1 hybrids can be found in natural populations. However, *Q. john-tuckeri* is not sympatric with *Q. engelmannii* and thus the reason for the elevated allele sharing is unclear, but may be due to shared alleles derived from a common ancestor to this recently diverged clade. Overall, these patterns suggest recent and independent gene flow events are driving the different amounts of allele sharing within the C-M clade. Additionally, for most the scrub oak species that are likely to have been sympatric with *Q. engelmannii*, we find evidence of asymmetric gene flow from the scrub oaks.

Despite the longevity of opportunity for gene flow between *Q. lobata* with the other white oak species, we do not observe a clear clade-by-clade pattern like what we observed for *Q. engelmannii*. Our model-based demographic analysis does indicate that the overall direction of gene flow appears to go from scrub oaks into *Q. lobata*. Instead of a clear clade-by-clade pattern, it appears that our observations are largely driven by widespread allele sharing between *Q. lobata* and individual scrub oak species, in particular *Q. douglasii*, northern populations of *Q. berberidifolia* and *Q. durata* var. *durata*. These patterns further suggest at least two distinct and relatively contemporary introgression events between scrub white oaks: between *Q. lobata* and *Q. douglasii*; and between *Q. lobata* and either the common ancestor of northern *Q. berberidifolia* and *Q. durata* var. *durata* or into both species after they diverged. However, we do not observe such clear signals of increased allele sharing between *Q. lobata* and scrub oaks such as *Q. john-tuckeri* or *Q. cornelius-mulleri*, which is curious given the history of sympatry of these tree and scrub white oaks. Because our data show some evidence for gene flow between *Q. lobata* and some but not all sympatric oak species, neither a model of reproductive isolation between species nor a model of pervasive introgression appears to be applicable to the California scrub white oaks. Instead, the patterns of allele sharing we observe in white oaks seem to be modulated by complex demographic or selective processes.

Combining our findings of ancient introgression and comparisons of demographic models, we can infer that *Q. engelmannii* has hybridized with the scrub oaks, especially the berberidifolia clade, much more than has *Q. lobata*, which might suggest that *Q. lobata* has evolved stronger reproductive isolating mechanisms with its congeners than has *Q. engelmannii*. The demographic models indicate that for both species the direction of gene flow is largely one-sided from scrub oaks into tree oaks, and minimal from tree oaks into scrub oaks. This pattern might be shaped by the expanding populations of scrub oaks (Figure 2) where their increased densities might be swamping out the pollen pools of the tree oaks, which has been observed in other oaks (Ortego et al., 2014; Petit et al., 2004). Alternatively, the expanding populations of the scrub oaks provide support for the hypothesis that scrub oaks evolved with the emergence of Mediterranean-type climates in California. These climate conditions would have favoured the more drought tolerant scrub oaks creating higher population densities that would create an asymmetry where favourable alleles are being incorporated into the genomes of the tree oaks through a process of adaptive introgression (Anderson

& Stebbins, 1954; Guichoux et al., 2013; Lexer, Heinze, Alia, & Rieseberg, 2004; Ortego et al., 2014). Of course, both might be true but resolving the explanation will require comparative analysis of whole-genome sequences.

This study demonstrates that a complex interaction of demography and reproductive isolation shapes the evolution of the California scrub oak complex, but our work does not reveal what types of genes introgress between species. In the future, whole-genome comparative data sets of polymorphism across species will make it possible to fit models that jointly examine the demographic history and nature of selection in oaks. These models can be used to investigate the number and timing of admixture events. Although we cannot fully reject ancient population structure as an explanation for the allele sharing we observe, future studies with whole-genome sequences can extract information in patterns of linkage disequilibrium to disentangle ancient population structure from more recent gene flow (Sankararaman, Patterson, Li, Pääbo, & Reich, 2012). These data can also be used to determine how natural selection shapes the landscape of introgression in oaks (Schumer et al., 2018).

Hybridization has been widely observed, and introgressive hybridization is widely accepted across many taxa (Anderson, 1953; Arnold, 1997; Harrison & Larson, 2014; Mallet, 2005). However, to understand the extent of pervasive hybridization will require genomewide sequence data and powerful statistical models such as those used here. Certainly, in oaks, these tools have shown the impact of ancient introgression on the phylogenetic tree (McVay et al., 2017a, b). Similar approaches have been used to understand differences in the proportion of Neanderthal ancestry between European and Asian human populations (Harris & Nielsen, 2016; Juric, Aeschbacher, & Coop, 2016; Kim & Lohmueller, 2015; Sankararaman, Mallick, Patterson, & Reich, 2016; Vernot & Akey, 2014, 2015; Wall et al., 2009). In the future, whole-genome sequencing will provide the fine-scale resolution needed to quantify recombination rates across the genome of oaks and to identify the genes that have introgressed or are highly divergent between different oak species. Further, powerful machine-learning approaches are being developed to address these population genetic questions at this scale of data (Schrider & Kern, 2018). This approach in turn should help reveal the genetic mechanisms and ecological factors driving reproductive isolation in the oaks. Lastly, understanding the development of reproductive isolation may be key to understanding the speciation process, but these ideas have yet to be rigorously tested across the tree of life (Cruickshank & Hahn, 2014). Introgression studies across many different species will be necessary to understand and resolve the respective roles of whole-genome reproductive isolation and localized differentiation in the speciation process.

5 | CONCLUSIONS

This study depicts limited signatures of gene flow between two tree oak species and several closely related scrub oaks. Oaks have long been a model for the evolutionary dynamics of gene flow and hybridization, providing a frequently cited challenge to the biological species concept. Yet, even though the tree oaks and scrub oaks have had ample time to hybridize due to their geographic proximity, our current work shows that gene flow is a complex and intricate process. Thus, even in the reputedly promiscuous oaks, an interaction of geographic and intrinsic barriers to gene flow is likely to explain most of the speciation process. Studies such as this one indicate that ancient introgression has influenced the genomes of species which have come into secondary contact, but those events are not pervasive across all species. Future work will reveal the details of the evolutionary processes shaped by interspecific gene exchange.

ACKNOWLEDGEMENTS

We thank Andrew Hipp and Paul Manos for valuable comments and for providing sequences for *Q. alba* and *Q. kelloggii*; Dylan Burge for valuable manuscript review and input on species' distributions; Fred Roberts for taxonomic identification and input on oak species' distributions; and Krista Beckley for DNA extractions and preparation of species maps. Three anonymous reviewers and C. Lexer provided excellent suggestions for this revision. JO was supported by Ramón y Cajal (RYC-2013-12501) and Severo Ochoa (SEV-2012-0262) fellowships; PFG, SFG and laboratory analyses were supported by UCLA seed funding to VLS. The valley oak genome assembly v.0.5 used for this study was supported in part by NSF PGRP to VLS and colleagues (IOS-1444611).

DATA ACCESSIBILITY

Variant data are available from the Dryad Digital Repository: https://doi.org/10.5061/dryad.6pg15js. RADseq data for each of the 55 *Quercus* taxa are available as fastq files from within NCBI SRA accession SRP149281 and for *Quercus alba* from NCBI SRA run SRR5284373. Analysis scripts are available at https://github.com/SorkLab.

AUTHOR CONTRIBUTIONS

B.Y.K designed the study, designed and conducted the analyses of *D*-statistics and demographic inference, and wrote the manuscript. K.E.L. designed and advised on ancient introgression analyses, and contributed to manuscript. X.W. conducted fieldwork, analysed phylogenetic data and contributed to manuscript. S.F.-G. conducted the bioinformatics, analysed phylogenetic data and wrote the manuscript. J.O and P.F.G. conducted fieldwork and commented on the manuscript. V.L.S. designed the study, conducted fieldwork, designed data analysis, wrote the manuscript and procured funding.

ORCID

Kirk E. Lohmueller http://orcid.org/0000-0002-3874-369X
Paul F. Gugger http://orcid.org/0000-0002-4464-8453
Victoria L. Sork http://orcid.org/0000-0003-2677-1420

REFERENCES

Anderson, E. (1953). Introgressive hybridization. *Biological Reviews of the Cambridge Philosophical Society*, *28*, 280–307. https://doi.org/10.1111/j.1469-185X.1953.tb01379.x

Anderson, E., & Stebbins, G. L. (1954). Hybridization as an evolutionary stimulus. *Evolution*, *8*, 378–388. https://doi.org/10.1111/j.1558-5646.1954.tb01504.x

Andrews, S. (2010) FASTQC: A quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Arnold, M. L. (1997). *Natural hybridization and evolution*. New York, NY: Oxford University Press.

Axelrod, D. I. (1983). Biogeography of oaks in the Arcto-Tertiary Province. *Annals of Missouri Botanical Garden*, *70*, 629–657. https://doi.org/10.2307/2398982

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B (Methodological)*, *57*, 289–300.

Burger, W. C. (1975). The species concept in *Quercus*. *Taxon*, *24*, 45–50. https://doi.org/10.2307/1218998

Busing, F. M., Meijer, E., & Van Der Leeden, R. (1999). Delete-m jackknife for unequal m. *Statistics and Computing*, *9*, 3–8. https://doi.org/10.1023/A:1008800423698

Cavender-Bares, J., González-Rodríguez, A., Eaton, D. A. R., Hipp, A. A. L., Beulke, A., & Manos, P. S. (2015). Phylogeny and biogeography of the American live oaks (*Quercus* subsection *Virentes*): A genomic and population genetics approach. *Molecular Ecology*, *24*, 3668–3687. https://doi.org/10.1111/mec.13269

Clausen, J. (1951). *Stages in the evolution of plant species*. Ithaca, NY: Cornell University Press.

Craft, K. J., Ashley, M. V., & Koenig, W. D. (2002). Limited hybridization between *Quercus lobata* and *Quercus douglasii* (Fagaceae) in a mixed stand in central coastal California. *American Journal of Botany*, *89*, 1792–1798. https://doi.org/10.3732/ajb.89.11.1792

Cruickshank, T. E., & Hahn, M. W. (2014). Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology*, *23*, 3133–3157. https://doi.org/10.1111/mec.12796

Cui, R., Schumer, M., Kruesi, K., Walter, R., Andolfatto, P., & Rosenthal, G. G. (2013). Phylogenomics reveals extensive reticulate evolution in Xiphophorus fishes. *Evolution*, *67*, 2166–2179. https://doi.org/10.1111/evo.12099

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., … Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*, 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Darwin, C. R. (1859). *The origin of species by means of natural selection: Or, the preservation of favored races in the struggle for life*. London, UK: John Murray.

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

Dodd, R. S., & Afzal-Rafii, Z. (2004). Selection and dispersal in a multispecies oak hybrid zone. *Evolution*, *58*, 261–269. https://doi.org/10.1111/j.0014-3820.2004.tb01643.x

Durand, E. Y., Patterson, N., Reich, D., & Slatkin, M. (2011). Testing for ancient admixture between closely related populations. *Molecular Biology and Evolution*, *28*, 2239–2252. https://doi.org/10.1093/molbev/msr048

Eaton, D. A. R., Hipp, A. L., 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. https://doi.org/10.1111/evo.12758

Ellstrand, N. C. (2014). Is gene flow the most important evolutionary force in plants? *American Journal of Botany*, *101*, 737–753. https://doi.org/10.3732/ajb.1400024

Eriksson, A., & Manica, A. (2012). Effect of ancient population structure on the degree of polymorphism shared between modern human populations and ancient hominins. *Proceedings of the National Academy of Sciences*, *109*, 13956–13960. https://doi.org/10.1073/pnas.1200567109

Fitz-Gibbon, S., Hipp, A., Pham, K., Manos, P., & Sork, V. L. (2017). Phylogenomic inferences from reference-mapped and de novo assembled short-read sequence data using RADseq sequencing of California white oaks (*Quercus* sect. *Quercus*). *Genome*, *60*, 743–755. https://doi.org/10.1139/gen-2016-0202

Forde, M. B., & Faris, D. G. (1962). Effect of introgression on serpentine endemism of *Quercus durata*. *Evolution*, *16*, 338–347. https://doi.org/10.1111/j.1558-5646.1962.tb03226.x

Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., … Pääbo, S. (2010). A draft sequence of the Neandertal genome. *Science*, *328*, 710–722. https://doi.org/10.1126/science.1188021

Griffin, J. R., andCritchfield, W. B. (1972) *The Distribution of Forest Trees in California*. USDA Forest Service Research Paper, Pacific Southwest Forest and Range Experiment Station PSW-82.

Gugger, P. F., Ikegami, M., & Sork, V. L. (2013). Influence of late Quaternary climate change on present patterns of genetic variation in valley oak, *Quercus lobata* Née. *Molecular Ecology*, *22*, 3598–3612. https://doi.org/10.1111/mec.12317

Guichoux, E., Garnier-Géré, P., Lagache, L., Lang, T., Boury, C., & Petit, R. J. (2013). Outlier loci highlight the direction of introgression in oaks. *Molecular Ecology*, *22*, 450–462. https://doi.org/10.1111/mec.12125

Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H., & Bustamante, C. D. (2009). Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLOS Genetics*, *5*, e1000695.

Harris, K., & Nielsen, R. (2016). The genetic cost of Neanderthal introgression. *Genetics*, *203*, 881–891. https://doi.org/10.1534/genetics.116.186890

Harrison, R. G., & Larson, E. L. (2014). Hybridization, introgression, and the nature of species boundaries. *Journal of Heredity*, *105*, 795–809. https://doi.org/10.1093/jhered/esu033

Hickman, J. C. (1993). *The Jepson manual: Higher plants of California*. Berkeley, CA: University of California Press.

Hipp, A. L., Eaton, D. A., Cavender-Bares, J., Fitzek, E., Nipper, R., & Manos, P. S. (2014). A framework phylogeny of the American oak clade based on sequenced RAD data. *PLoS ONE*, *9*, e93975. https://doi.org/10.1371/journal.pone.0093975

Hipp, A. L., Manos, P. S., González-Rodríguez, A., Hahn, M., Kaproth, M., McVay, J. D., … 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. https://doi.org/10.1111/nph.14773

Juric, I., Aeschbacher, S., & Coop, G. (2016). The strength of selection against Neanderthal introgression. *PLOS Genetics*, *12*, e1006340. https://doi.org/10.1371/journal.pgen.1006340

Kim, B. Y., & Lohmueller, K. E. (2015). Selection and reduced population size cannot explain higher amounts of Neandertal ancestry in East Asian than in European human populations. *The American Journal of Human Genetics*, *96*, 454–461. https://doi.org/10.1016/j.ajhg.2014.12.029

Kruckeberg, A. R. (1984). *California Serpentines: Flora, vegetation, geology, soils, and management problems*. Los Angeles, CA: University of California Publications in Botany.

Lexer, C., Heinze, B., Alia, R., & Rieseberg, L. H. (2004). Hybrid zones as a tool for identifying adaptive genetic variation in outbreeding forest trees: Lessons from wild annual sunflowers (*Helianthus* spp.). *Forest Ecology and Management*, *197*, 49–64. https://doi.org/10.1016/j.foreco.2004.05.004

Lexer, C., Kremer, A., & Petit, R. J. (2006). Shared alleles in sympatric oaks: Recurrent gene flow is a more parsimonious explanation than

ancestral polymorphism. *Molecular Ecology*, *15*, 2007–2012. https://doi.org/10.1111/j.1365-294X.2006.02896.x

Li, H. (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv preprint arXiv:1303.3997.

Mallet, J. (2005). Hybridization as an invasion of the genome. *Trends in Ecology & Evolution*, *20*, 229–237. https://doi.org/10.1016/j.tree.2005.02.010

Mallet, J., Besansky, N., & Hahn, M. W. (2016). How reticulated are species? *BioEssays*, *38*, 140–149. https://doi.org/10.1002/bies.201500149

Matute, D. R. (2010). Reinforcement can overcome gene flow during speciation in Drosophila. *Current Biology*, *20*, 2229–2233. https://doi.org/10.1016/j.cub.2010.11.036

Mayr, E. (1942). *Systematics and the origin of species, from the viewpoint of a zoologist*. Cambridge, MA: Harvard University Press.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., … DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*, 1297–1303. https://doi.org/10.1101/gr.107524.110

McVay, J. D., Hauser, D., Hipp, A. L., & Manos, P. S. (2017a). Phylogenomics reveals a complex evolutionary history of lobed-leaf white oaks in western North America. *Genome*, *60*, 733–742. https://doi.org/10.1139/gen-2016-0206

McVay, J. D., Hipp, A. L., & Manos, P. S. (2017b). A genetic legacy of introgression confounds phylogeny and biogeography in oaks. *Proceedings of the Royal Society B*, *284*, https://doi.org/10.1098/rspb.2017.0300

Nixon, K. C. (1997). *Quercus*. In Flora of North America Editorial Committee (Ed.), *Flora of North America North of Mexico* (pp. 445–447). New York, NY: Oxford University Press.

Nixon, K. (2002). The Oak (*Quercus*) biodiversity of California and adjacent regions. In R. B. Standiford, & K. L. Purcell (Eds.), *Proceedings of the Fifth Symposium on Oak Woodlands: Oaks in California's Changing Landscape, October 22–25, 2001, San Diego, California* (pp. 3–2). San Diego, CA: USDA Forest Service Gen. Tech. Rep.

Novikova, P. Y., Hohmann, N., Nizhynska, V., Tsuchimatsu, T., Ali, J., Muir, G., … Nordborg, M. (2016). Sequencing of the genus *Arabidopsis* identifies a complex history of nonbifurcating speciation and abundant trans-specific polymorphism. *Nature Genetics*, *48*, 1077–1082. https://doi.org/10.1038/ng.3617

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. https://doi.org/10.1111/(ISSN)1438-8677

Ortego, J., Gugger, P. F., Riordan, E. C., & Sork, V. L. (2014). Influence of climatic niche suitability and geographical overlap on hybridization patterns among southern Californian oaks. *Journal of Biogeography*, *41*, 1895–1908. https://doi.org/10.1111/jbi.12334

Ortego, J., Gugger, P. F., & Sork, V. L. (2015a). Climatically stable landscapes predict patterns of genetic structure and admixture in the Californian canyon live oak. *Journal of Biogeography*, *42*, 328–338. https://doi.org/10.1111/jbi.12419

Ortego, J., Noguerales, V., Gugger, P., & Sork, V. (2015b). Evolutionary and demographic history of the Californian scrub white oak species complex: An integrative approach. *Molecular Ecology*, *24*, 6188–6208. https://doi.org/10.1111/mec.13457

Palmer, E. J. (1948). Hybrid oaks of North America. *Journal of the Arnold Arboretum*, *29*, 1–48.

Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., … Reich, D. (2012). Ancient admixture in human history. *Genetics*, *192*, 1065–1093. https://doi.org/10.1534/genetics.112.145037

Pavlik, B. M., Muick, P. C., Johnson, S. G., & Popp, M. (1995). *Oaks of California*. Oakland, California: Cachuma Press.

Payseur, B. A., & Rieseberg, L. H. (2016). A genomic perspective on hybridization and speciation. *Molecular Ecology*, *25*, 2337–2360. https://doi.org/10.1111/mec.13557

Pearse, I. S., & Hipp, A. L. (2009). Phylogenetic and trait similarity to a native species predict herbivory on non-native oaks. *Proceedings of the National Academy of Sciences*, *106*, 18097–18102. https://doi.org/10.1073/pnas.0904867106

Penaloza-Ramirez, J. M., González-Rodríguez, A., Mendoza-Cuenca, L., Caron, H., Kremer, A., & Oyama, K. (2010). Interspecific gene flow in a multispecies oak hybrid zone in the Sierra Tarahumara of Mexico. *Annals of Botany*, *105*, 389–399. https://doi.org/10.1093/aob/mcp301

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

Petit, R. J., Kremer, A., & Wagner, D. B. (1993). Geographic structure of chloroplast DNA polymorphisms in European oaks. *Theoretical and Applied Genetics*, *87*, 122–128.

Petit, R. J., Pineau, E., Demesure, B., Bacilieri, R., Ducousso, A., & Kremer, A. (1997). Chloroplast DNA footprints of postglacial recolonization by oaks. *Proceedings of the National Academy of Sciences of the United States of America*, *94*, 9996–10001. https://doi.org/10.1073/pnas.94.18.9996

Pickrell, J. K., & Pritchard, J. K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *PLOS Genetics*, *8*, e1002967. https://doi.org/10.1371/journal.pgen.1002967

Plagnol, V., & Wall, J. D. (2006). Possible ancestral structure in human populations. *PLoS Genetics*, *2*, e105. https://doi.org/10.1371/journal.pgen.0020105

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

Riordan, E. C., Gugger, P. F., Ortego, J., Smith, C., Gaddis, K., Thompson, P., & Sork, V. L. (2016). Association of genetic and phenotypic variability with geography and climate in three southern California oaks. *American Journal of Botany*, *103*, 73–85. https://doi.org/10.3732/ajb.1500135

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

Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011). Integrative genomics viewer. *Nature Biotechnology*, *29*, 24–26. https://doi.org/10.1038/nbt.1754

Rundel, P. W., Arroyo, M. T. K., Cowling, R. M., Keeley, J. E., Lamont, B. B., & Vargas, P. (2016). Mediterranean biomes: Evolution of their vegetation, floras, and climate. *Annual Review of Ecology, Evolutionary, and Systematics*, *47*, 383–407. https://doi.org/10.1146/annurev-ecolsys-121415-032330

Rushton, B. (1993). Natural hybridization within the genus *Quercus* L. *Annales des Sciences Forestières*, *50*, 73s–90s. https://doi.org/10.1051/forest:19930707

Salinas, N. R., & Little, D. P. (2014). 2matrix: A utility for indel coding and phylogenetic matrix concatenation. *Applications in plant sciences*, *2*, 1300083.

Sankararaman, S., Mallick, S., Dannemann, M., Prüfer, K., Kelso, J., Pääbo, S., … Reich, D. (2014). The genomic landscape of Neanderthal ancestry in present-day humans. *Nature*, *507*, 354–357. https://doi.org/10.1038/nature12961

Sankararaman, S., Mallick, S., Patterson, N., & Reich, D. (2016). The combined landscape of Denisovan and Neanderthal ancestry in present-day humans. *Current Biology*, *26*, 1241–1247. https://doi.org/10.1016/j.cub.2016.03.037

Sankararaman, S., Patterson, N., Li, H., Pääbo, S., & Reich, D. (2012). The date of interbreeding between Neandertals and modern humans. *PLOS Genetics*, *8*, e1002947. https://doi.org/10.1371/journal.pgen.1002947

Schrider, D. R., & Kern, A. D. (2018). Supervised machine learning for population genetics: A new paradigm. *Trends in Genetics*, *34*, 301–312. https://doi.org/10.1016/j.tig.2017.12.005

Schumer, M., Xu, C., Powell, D. L., Durvasula, A., Skov, L., Holland, C., … Przeworski, M. (2018). Natural selection interacts with recombination to shape the evolution of hybrid genomes. *Science*, *360*(6389), 59–660.

Simmons, M. P., & Ochoterena, H. (2000). Gaps as characters in sequence-based phylogenetic analysis. *Systematic Biology*, *49*, 369–338. https://doi.org/10.1093/sysbio/49.2.369

Sork, V. L., Fitz-Gibbon, S. T., Puiu, D., Crepeau, M., Gugger, P. F., Sherman, R., … Salzberg, S. L. (2016a). First draft assembly and annotation of the genome of a California endemic oak. *Quercus lobata* Née (Fagaceae). *G3: Genes, Genomes, Genetics*, *11*, 3485–3495.

Sork, V. L., Riordan, E., Gugger, P. F., Fitz-Gibbon, S., Wei, X., & Ortego, J. (2016b). Phylogeny and introgression of California scrub white oaks (*Quercus* sect. *Quercus*). *International Oak Journal*, *27*, 61–74.

Stamatakis, A. (2014). RAXML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*, 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Stebbins, G. L. (1950). *Variation and evolution in plants*. New York, NY: Columbia University Press.

Tovar-Sanchez, E., & Oyama, K. (2004). Natural hybridization and hybrid zones between *Quercus crassifolia* and *Quercus crassipes* (Fagaceae) in Mexico: Morphological and molecular evidence. *American Journal of Botany*, *91*, 1352–1363. https://doi.org/10.3732/ajb.91.9.1352

Tucker, J. M. (1968). Identity of the Oak-D tree at live oak tank Joshua Tree National Monument California USA. *Madrono*, *19*, 256–266.

Tucker, J. M. (1974). Patterns of parallel evolution of leaf form in New World oaks. *Taxon*, *23*, 129–154. https://doi.org/10.2307/1218095

Tucker, J. (1990). Hybridization in California oaks. *Fremontia*, *18*, 13–19.

Tyler, C. M., Kuhn, B., & Davis, F. W. (2006). Demography and recruitment limitations of three oak species in California. *Quarterly Review of Biology*, *81*, 127–152. https://doi.org/10.1086/506025

Van Valen, L. (1976). Ecological species multi species and oaks. *Taxon*, *25*, 233–239. https://doi.org/10.2307/1219444

Vernot, B., & Akey, J. M. (2014). Resurrecting surviving Neandertal lineages from modern human genomes. *Science*, *343*, 1017–1021. https://doi.org/10.1126/science.1245938

Vernot, B., & Akey, J. M. (2015). Complex history of admixture between modern humans and Neandertals. *The American Journal of Human Genetics*, *96*, 448–453. https://doi.org/10.1016/j.ajhg.2015.01.006

Wall, J. D., Lohmueller, K. E., & Plagnol, V. (2009). Detecting ancient admixture and estimating demographic parameters in multiple human populations. *Molecular Biology and Evolution*, *26*, 1823–1827. https://doi.org/10.1093/molbev/msp096

Whittemore, A. T., & Schaal, B. A. (1991). Interspecific gene flow in sympatric oaks. *Proceedings of the National Academy of Sciences of the United States of America*, *88*, 2540–2544. https://doi.org/10.1073/pnas.88.6.2540

Yang, M. A., Malaspinas, A.-S., Durand, E. Y., & Slatkin, M. (2012). Ancient structure in Africa unlikely to explain Neanderthal and non-African genetic similarity. *Molecular Biology and Evolution*, *29*, 2987–2995. https://doi.org/10.1093/molbev/mss117

Zeng, Y. F., Liao, W. J., Petit, R. J., & Zhang, D. Y. (2011). Geographic variation in the structure of oak hybrid zones provides insights into the dynamics of speciation. *Molecular Ecology*, *20*, 4995–5011. https://doi.org/10.1111/j.1365-294X.2011.05354.x

Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., & Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*, *28*, 3326–3328. https://doi.org/10.1093/bioinformatics/bts606

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Kim BY, Wei X, Fitz-Gibbon S, et al. RADseq data reveal ancient, but not pervasive, introgression between Californian tree and scrub oak species (*Quercus* sect. *Quercus*: Fagaceae). *Mol Ecol*. 2018;27:4556–4571. https://doi.org/10.1111/mec.14869