RESEARCH ARTICLE

American Journal of Botany

Botanical Society of America
leading scientists and educators since 1893

Influence of a climatic gradient on genetic exchange between two oak species

Dylan O. Burge[1,2], V. Thomas Parker[3], Margaret Mulligan[4], and Victoria L. Sork[1,5]

Manuscript received 5 June 2018; revision accepted 27 March 2019.

[1] Department of Ecology and Evolutionary Biology, University of California, Box 957239, Los Angeles, California 90095-7239, USA

[2] 554 Vallombrosa Avenue, P.O. Box 418, Chico, California 95927, USA

[3] San Francisco State University, 1600 Holloway Avenue, San Francisco, California 94132, USA

[4] San Diego Natural History Museum, Balboa Park, 1788 El Prado, San Diego, California 92101, USA

[5] Author for correspondence (e-mail: vlsork@ucla.edu)

Citation: Burge, D. O., V. T. Parker, M. Mulligan, and V. L. Sork. 2019. Influence of a climatic gradient on genetic exchange between two oak species. *American Journal of Botany* 106(6): 864–878.

doi:10.1002/ajb2.1315

PREMISE: In plant groups with limited intrinsic barriers to gene flow, it is thought that environmental conditions can modulate interspecific genetic exchange. Oaks are known for limited barriers to gene flow among closely related species. Here, we use *Quercus* as a living laboratory in which to pursue a fundamental question in plant evolution: Do environmental gradients restrict or promote genetic exchange between species?

METHODS: We focused on two North American oaks, the rare *Quercus dumosa* and the widespread *Q. berberidifolia*. We sampled intensively along a contact zone in California, USA. We sequenced restriction site-associated DNA markers and measured vegetative phenotype. We tested for genetic exchange, the association with climate, and the effect on phenotype.

RESULTS: There is evidence for genetic exchange between the species. Admixed plants are found in areas of intermediate climate, while less admixed plants are found at the extremes of the climatic gradient. Genetic and phenotypic patterns are out of phase in the contact zone; some plants display the phenotype of one species but are genetically associated with another.

CONCLUSIONS: Our results support the hypothesis that a strong climatic gradient can promote genetic exchange between species. The overall weak correlation between genotype and phenotype in the contact zone between the species suggests that genetic exchange can lead to the breakdown of trait combinations used to define species. This incongruency predicts ongoing problems for conservation of *Q. dumosa*, with implications for conservation of other oaks.

KEY WORDS climate; drought; genomic; hybridization; phenotype; population; RADseq; rare.

Selecting speciation criteria and quantifying species boundaries has proven consistently more difficult in plants than in most animals (Carson, 1985; Coyne and Orr, 2004; Rieseberg and Willis, 2007; Abbott et al., 2008, 2016; Soltis and Soltis, 2009). Research suggests that in many plant groups, the speciation process has continued despite extensive and ongoing genetic exchange among closely related species (Arnold, 2006; Rieseberg and Willis, 2007). Although such plant groups are found worldwide and represent a diversity of lineages and life-history strategies, several groups are frequently cited as examples, including sunflowers (*Helianthus* [Asteraceae]; Renaut et al., 2013), monkeyflowers (*Diplacus* [Phrymaceae]; Brandvain et al., 2014), eucalypts (*Eucalyptus* [Myrtaceae]; Rutherford et al., 2018) and oaks (*Quercus* [Fagaceae]; Muir et al., 2000). Here, we focus on oaks. Among woody plants of the Northern Hemisphere, oaks have long been thought of as a poster child for speciation despite ongoing genetic exchange (Burger, 1975; Van Valen, 1976; Muir et al., 2000; Coyne and Orr, 2004), an idea that is supported by extensive empirical data in wild plants (Forde and Faris, 1962; Benson et al., 1967; Hardin, 1975; Whittemore and Schaal, 1991; Nason et al., 1992; Muir et al., 2000; Nixon, 2002; Dodd and Afzal-Rafii, 2004; González-Rodríguez et al., 2004; Petit et al., 2004; Burgarella et al., 2009; Lagache et al., 2013; Eaton et al., 2015; Valencia-Cuevas et al., 2015; Sullivan et al., 2016; Khodewekar and Gailing, 2017; Leroy et al., 2017; McVay et al., 2017a, 2017b; Ortego et al., 2017a, 2017b).

Recent research based on genome-wide DNA sequencing has revealed that while genetic exchange, especially the formation of hybrids, may be a common theme in the evolution of *Quercus*, it has had a limited effect on the genomes of modern species, except

in a few recently characterized cases of ancient introgression (Hipp et al., 2014; McVay et al., 2017a, 2017b; Kim et al., 2018). The apparent lack of a strong signal for widespread genetic exchange during diversification of *Quercus* begs the question: What processes have limited the extent of genetic exchange, especially in geographic regions where many interfertile species come into close contact, such as western North America? Some studies suggest that where species come into contact across environmental gradients, genetic exchange is limited by abiotic selection (Dodd and Afzal-Rafii, 2004; Muir and Schlötterer, 2005; Gailing et al., 2012; Alberto et al., 2013). Other research provides limited evidence for the possibility that hybrids are favored in parts of the environmental gradient under certain circumstances, either in a selective environment that favors the hybrids (Anderson, 1948; Muller, 1952; Mallet, 2005) or where there are differences in population density such that hybrids become favored at the edge of geographic ranges (Muller, 1952; Valbuena-Carabaña et al., 2005; Petit and Excoffier, 2009). These ideas are not mutually exclusive, but they have different consequences for oak biology and it would be useful to distinguish between them. Specifically, research should examine introgression at a local spatial scale along an environmental gradient, employing a large number of genetic markers with sufficient resolution to detect very fine-scale genetic exchange (Boecklen and Howard, 1997). Nevertheless, very few studies have used restriction site-associated DNA (RAD) sequencing or other detailed genomic methods to quantify the distribution of genetic variation across environments in oaks (Eaton et al., 2015; Sork et al., 2016b; Leroy et al., 2017; Ortego et al., 2017a), so it has been difficult to distinguish between the ideas outlined above.

We studied the western North American scrub white oak complex (Tucker, 1952; Nixon and Mueller, 1994, 1997; Nixon, 2002; Rosatti and Tucker, 2012), a group of six species widespread in California, USA, and adjacent parts of Mexico. Within this group, we concentrate on *Quercus dumosa* Nutt. and *Q. berberidifolia* Liebm. Like many of their congeners, these two species are difficult to identify based on morphology, possibly as a result of genetic exchange with each other and with other white oaks (Nixon and Steele, 1981; Nixon, 2002; Ortego et al., 2014). Identification problems are compounded by the phenotypic plasticity that oaks are known for, leading to problems for circumscription of the species (Nixon and Muller, 1994, 1997) and conservation of the rare *Q. dumosa* (AECOM, 2015; CNPS Rare Plant Program, 2016).

Opportunities for genetic exchange between *Q. berberidifolia* and *Q. dumosa* are available in southern California; the region of closest proximity for the species is in coastal San Diego County (Fig. 1). In this region, *Q. dumosa* is common close to the ocean, while *Q. berberidifolia* is found from the coast to the crest of the Peninsular Ranges. Climatic conditions are mild close to the coast but become increasingly hot and dry to the east (Flint et al., 2013). In this region, purported hybrids are most frequently reported in western San Diego County, where the geographic ranges of the species interlock along the climatic gradient (Nixon and Steele, 1981; D. Burge, personal observation). Genetic exchange between *Q. berberidifolia* and *Q. dumosa* on the coast-to-inland climatic gradient is a tantalizing idea. If it occurs, it could influence local adaptation of scrub oak populations along the gradient and potentially lead to long-term genetic consequences for the species (McVay et al., 2017a, 2017b; Ortego et al., 2017b; Kim et al., 2018). However, the extent to which genetic exchange occurs, and the correlation of this exchange with local climatic conditions, has never been tested.

The overall goal of this study is to determine whether genetic exchange between *Q. dumosa* and *Q. berberidifolia* has been shaped by a climate gradient. We sampled the focal species across their geographic range, with intensive sampling on the coast-to-inland climatic gradient of western San Diego County (Fig. 1). We employed RAD sequencing to obtain genome-scale single-nucleotide polymorphism (SNP) data. We inferred climatic conditions for each locale from GIS datasets and also collected data on vegetative phenotype of a subsample of plants within the hybrid zone. We analyzed genetic, climatic, and phenotypic data to find out (1) if genetic exchange is taking place between the focal species along the coast-to-inland climatic gradient in San Diego County, (2) how such genetic exchange is associated with climate, and (3) whether vegetative phenotype corresponds to genotype. We also discuss the implications of these findings for conservation of *Q. dumosa*, a rare species of conservation concern.

MATERIALS AND METHODS

Study system

The western North American scrub white oak complex (Tucker, 1952; Nixon and Mueller, 1994, 1997; Nixon, 2002; Rosatti and Tucker, 2012) is a group of six shrub to tree-like species found in California, USA, and adjacent Baja California, Mexico (Nixon, 2002). The scrub white oaks form part of the Dumosae white oak clade (McVay et al., 2017a, 2017b). Phylogenetic work in the Dumosae white oak clade (Fitz-Gibbon et al., 2017; Kim et al., 2018) revealed two patterns relevant to the present study: (1) the widespread tree *Q. douglasii* Hook. & Arn. is derived from within the group; and (2) the group contains two major clades, with most plants collected to the north of 34°N latitude (most *Q. berberidifolia* Liebm. and *Q. durata* Jeps.) in one clade, and most plants collected to the south of 34°N latitude in a second clade. The two-clade pattern makes the Dumosae scrub white oaks an ideal system for the study of genetic exchange, as the clades are sufficiently isolated that introgression can be distinguished from incomplete lineage sorting (Buckley et al., 2006; Joly et al., 2009; Folk et al., 2018).

Within the Dumosae scrub white oaks, the present study is focused on *Q. berberidifolia* and *Q. dumosa*. *Quercus berberidifolia* is widespread in California, where it is the most frequently encountered scrub white oak; it is also found in northern Baja California (Rosatti and Tucker, 2012). In the southern part of its range, *Q. berberidifolia* comes into contact with *Q. dumosa* (Rosatti and Tucker, 2012). *Quercus dumosa* is found mainly in San Diego County, California, and northern Baja California, with small, outlying occurrences in Orange, Riverside, and Santa Barbara counties in California (Nixon, 2002). *Quercus dumosa* is rare (CNPS Rare Plant Program, 2016) and is therefore the subject of conservation management (AECOM, 2015).

Quercus dumosa and *Q. berberidifolia* are difficult to identify based on morphology, possibly as a result of phenotypic plasticity, which is known to be strong in oaks (Tucker, 1952; Van Valen, 1976; Nixon, 2002). But identification problems may be exacerbated by genetic exchange with each other and with other oaks (Nixon and Steele, 1981; Nixon, 2002; Ortego et al., 2014). This situation has led to problems in circumscribing species in this group, culminating in a drastic taxonomic upheaval in which new species were described and the circumscription of *Q. dumosa* was dramatically narrowed

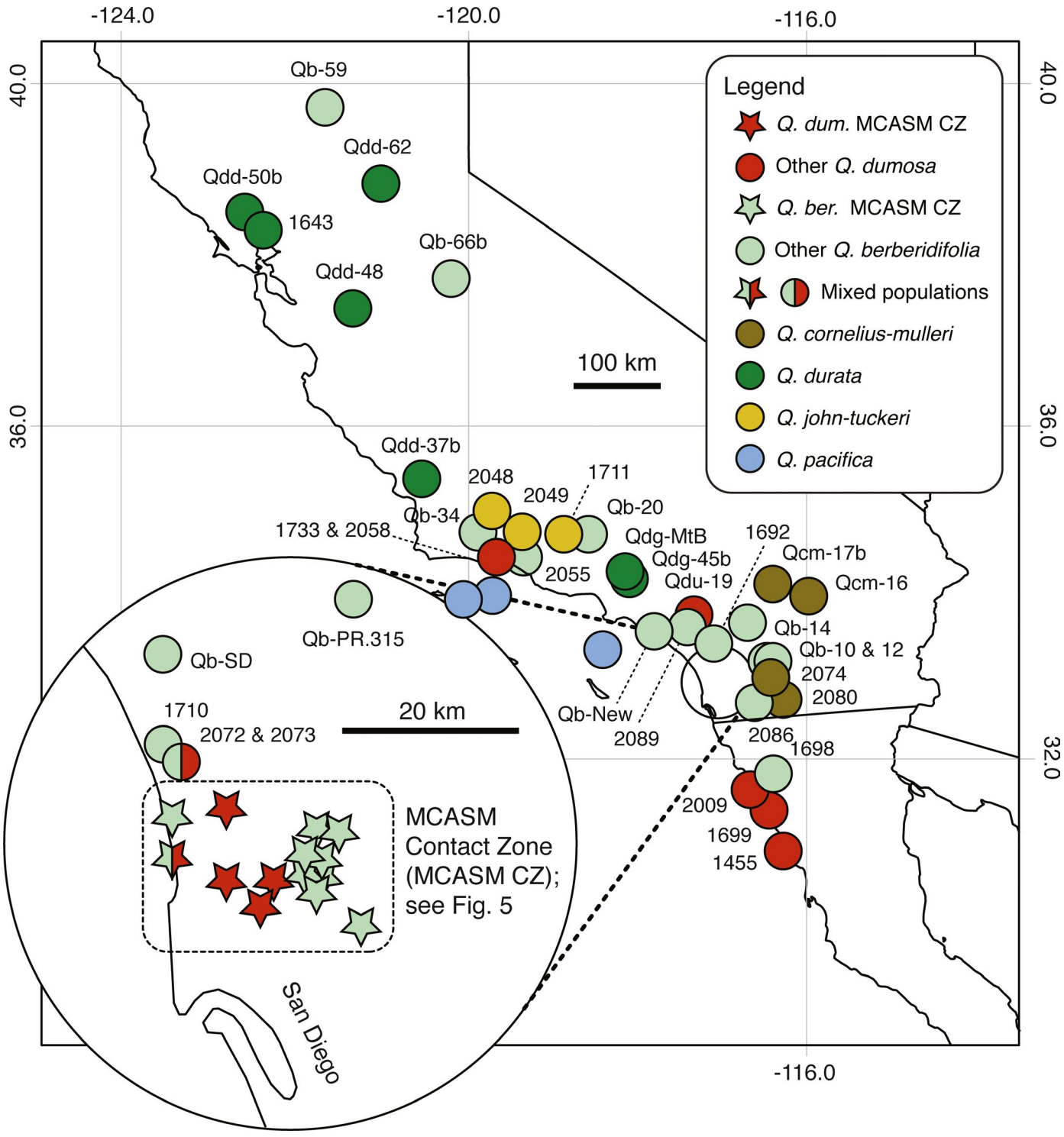

FIGURE 1. Map of sampling locations of *Quercus berberidifolia* and *Q. dumosa* for this study. Grid dimensions are in degrees of latitude and longitude (WGS84 datum). See Table 1 and Dryad Appendix D1 for more information on locales.

in favor of *Q. berberidifolia* (Nixon and Muller, 1994, 1997; Rosatti and Tucker, 2012). Despite reforms, the taxonomic concepts of *Q. berberidifolia* and *Q. dumosa* remain difficult to apply; many populations display intermediate phenotypes and signs of genetic exchange with other oak species. Overall, there is a need for objective means to justify circumscriptions of these species, assess levels of genetic admixture among species, and identify plants, particularly for the rare *Q. dumosa* (CNPS Rare Plant Program, 2016).

Opportunities for genetic exchange between *Q. berberidifolia* and *Q. dumosa* are most prevalent in far southern California and

TABLE 1. Locales sampled for genetic and phenotypic analysis of *Quercus berberidifolia* and *Q. dumosa*. "Taxon" is according to the Jepson Manual (Rosatti and Tucker, 2012). "Code" is the field collection code for the locale where sets of samples were obtained. "Source" indicates samples obtained for the present study by D. Burge ("Burge") and those obtained by the V. Sork lab ("Sork"). For Burge collections, "Code" matches herbarium voucher specimens deposited at DAV (Dryad Appendix D1). Under "Locale name," MCASM CZ = Marine Corps Air Station Miramar Contact Zone (Fig. 1). "Genetic datasets" indicates the number of individuals that were included in each of the three genetic analyses (see text and Dryad Appendices D3–D5).

				Genetic datasets		
Taxon	**Code**	**Locale name**	**Source**	**Broad**	**SDCo A**	**SDCo B**
Q. berberidifolia	1692	Rainbow	Burge	2	0	0
	1698	Arroyo de La Cruz	Burge	2	0	0
	1707	Mission Trails	Burge	2	4	4
	1710	Encinitas Community Center	Burge	2	2	2
	1718	MCASM CZ 9	Burge	2	2	1
	1721	MCASM CZ 10	Burge	2	2	2
	1724	MCASM CZ 11	Burge	2	2	2
	1726	MCASM CZ 12	Burge	2	2	2
	1728	MCASM CZ 14	Burge	2	2	2
	1730	Crest Canyon	Burge	2	2	2
	2055	Casitas Lake	Burge	1	0	0
	2061	MCASM CZ 15	Burge	2	2	2
	2062	MCASM CZ 16	Burge	2	2	2
	2069	Torrey Pines	Burge	2	3	3
	2073	Encinitas	Burge	2	2	2
	2086	Viejas Mountain	Burge	2	2	2
	2089	Santa Ana Mountains	Burge	2	0	0
	Qb-10	Warner Springs	Sork	1	1	0
	Qb-12	Ranchita	Sork	1	1	0
	Qb-14	Hemet	Sork	1	1	0
	Qb-20	Three Points	Sork	2	0	0
	Qb-34	Figueroa Mountain	Sork	1	0	0
	Qb-59	Chico	Sork	1	0	0
	Qb-66b	Coulterville	Sork	1	0	0
	Qb-NEW	Pelican Hill	Sork	2	0	0
	Qb-PR.315	Pala Reserve	Sork	1	1	0
	Qb-SD	Calavera Lake	Sork	2	3	0
Q. cornelius-mulleri	2074	Banner Grade East	Burge	1	0	0
	2080	McCain Valley	Burge	1	0	0
	Qcm-16	Jumbo Rocks	Sork	1	0	0
	Qcm-17b	Pioneertown	Sork	1	0	0
Q. dumosa	1455	Colonet Mesa	Burge	2	0	0
	1699	Rancho Embarcadero	Burge	2	0	0
	1706	Del Mar Mesa	Burge	2	3	3
	1713	MCASM CZ 6	Burge	2	5	5
	1714	MCASM CZ 7	Burge	2	7	4
	1715	MCASM CZ 8	Burge	1	3	3
	1716	Torrey Pines	Burge	2	4	3
	1733	Sta. Barbara Bot. Gard.	Burge	2	0	0
	2009	Punta Banda Peninsula	Burge	1	0	0
	2058	Rattlesnake Canyon	Burge	2	0	0
	2072	Encinitas	Burge	2	2	2
	QDu-19	Perris	Sork	1	0	0
Q. durata var. *durata*	1643	Cavedale Road	Burge	1	0	0
	Qdd-37b	Cuesta Ridge	Sork	1	0	0
	Qdd-48	Del Puerto Canyon	Sork	2	0	0
	Qdd-50b	Santa Rosa	Sork	1	0	0
	Qdd-62	Folsom Lake	Sork	2	0	0
Q. durata var. *gabrielensis*	Qdg-45b	Glendora	Sork	2	0	0
	Qdg-MtB	Mount Baldy	Sork	1	0	0
Q. john-tuckeri	1711	Hungry Valley	Burge	1	0	0
	2048	Caliente Range	Burge	1	0	0
	2049	Cuyama Valley	Burge	1	0	0
Q. pacifica	Qp-Cat	Santa Catalina Island	Sork	2	0	0
	Qp-SC3	Santa Cruz Island	Sork	1	0	0
	Qp-SR33	Santa Rosa Island	Sork	1	0	0
	Qp-SRb	Santa Rosa Island	Sork	1	0	0

northern Baja California. However, only *Q. berberidifolia* is known to come into contact with *Q. dumosa*. The region of closest proximity for *Q. berberidifolia* and *Q. dumosa* is in coastal San Diego County. Climatic conditions are mild close to the coast (little interannual variation in temperature and low drought stress) but become increasingly stringent (high interannual variation in temperature and high drought stress) to the east (Flint et al., 2013). In addition, the region is characterized by complex topography, producing microclimates that lead to high rates of plant community turnover across the landscape (Sawyer et al., 2009). In this region, *Q. dumosa* is usually encountered within 4 km of the sea (Jepson Flora Project, 2018: http://ucjeps.berkeley.edu/jepsonflora), while *Q. berberidifolia* is found from the coast to the crest of the Peninsular Ranges. In a few locations in western San Diego County, *Q. dumosa* grows together in mixed stands with *Q. berberidifolia* (locales 2072 and 2073 [Fig. 1], 1716 and 2069), one of the few cases of overlap observed in Dumosae scrub white oaks (D. Burge, personal observation). Although purported hybrids are frequently reported in San Diego County, they are not widespread, being limited to the contact zone close to the coast (Nixon and Steele, 1981; J. Rebman, San Diego Natural History Museum, personal communication).

Sampling

Our sampling for *Q. dumosa* and *Q. berberidifolia* covered the entire geographic range of these species in California and Baja California, with an emphasis on southern California (Table 1; Fig. 1). We carried out intensive population sampling in San Diego County, where *Q. dumosa* comes into close contact with *Q. berberidifolia*. The majority of this detailed sampling was carried out in a contact zone between the two focal species in and around Marine Corps Air Station Miramar (hereafter "MCASM Contact Zone"; Fig. 1). To place the two focal species in their genetic context within the Dumosae scrub white oaks and to ensure that hybrids were derived from the two focal species, we also sampled populations of the other four species from this group: *Q. cornelius-mulleri* Nixon and K.P. Steele, *Q. durata* Jeps., *Q. john-tuckeri* Nixon & C.H. Mull., and *Q. pacifica* Nixon & C.H. Mull. In San Diego County, we sampled populations that covered large, essentially contiguous areas of the landscape. We refer to locations where we sampled as "locales," which we define as groupings of plants discontinuous from one another at their nearest edges by ≥200 m, with their centers ≥1 km apart. We sampled at a total of 57 locales (12 *Q. dumosa*, 27 *Q. berberidifolia*, and 18 representing other scrub white oak species), 34 of which were sampled by D. Burge and 23 by J. Ortego or V. Sork (Table 1; Fig. 1; Dryad Appendix D1 [data available from the Dryad Digital Repository: https://doi.org/10.5061/dryad.6v03s1s (Burge et al., 2019)]; Sork et al., 2016b). At each locale, ≤10 mature shrubs were sampled. Fresh bud material was kept refrigerated until it could be frozen at −80°C. Plants were identified according to the keys of Rosatti and Tucker (2012), supplemented by revisions (Nixon and Steele, 1981; Nixon and Muller, 1994, 1997).

DNA sequencing

DNA extraction, barcoding, and sequencing efforts were done in two batches, the first carried out in the V. Sork lab (Fitz-Gibbon et al., 2017) and the second in the lab of V. Parker (Table 1, source: Burge). Library preparation and sequencing for batch sets was done at Floragenex (Portland, Oregon, USA) under identical conditions, as described below. For plants sampled by D. Burge (Table 1), DNA was extracted using the DNeasy Plant Mini Kit (Qiagen, Germantown, Maryland, USA) according to the manufacturer's instructions. Plants sampled by other members of the V. Sork lab were extracted as described by Fitz-Gibbon et al. (2017). For both extraction efforts, genomic DNA was checked for size and degradation on a 1% agarose electrophoresis gel. Samples with high-molecular-weight DNA (about 20–50 kilobases) were standardized to ~30 ng/μL, diluted to a volume of 150 μL in pure water, and sent for quality control, library construction, and sequencing at Floragenex. Methods for library preparation follow Lozier (2014), except that we used the restriction enzyme PstI (Hipp et al., 2014) to digest DNA. In brief, the sequence identifier barcodes (a unique one for each sampled plant) and sequence adapters were added to genomic DNA after digestion. Barcoded samples were then combined, and the fragments sequenced outwards from restriction sites using single-end DNA sequencing reads that were 100 base pairs long. Sequencing was done on HiSeq 2500 DNA sequencers (Illumina, San Diego, California). DNA sequences from individual plants were then separated on the basis of their unique barcodes. For information on the sequencing coverage and number of RAD loci successfully sequenced for each sample, refer to Dryad Appendix D1. Raw DNA sequence data for all of the samples newly sequenced for the present work can be found in the NCBI BioProject database under accession PRJNA398401. Raw DNA sequence data for samples sequenced previously (Kim et al., 2018; Table 1) can be found under accession PRJNA473578.

Variant detection and filtering

Detection of variation among individual plants was done using a reference genome approach. A preliminary version of the *Q. lobata* reference genome prepared in the V. Sork lab (Sork et al., 2016a) was used to align reads and call loci. First, the cut-site sequence (TGCAG) was removed from the DNA sequences using the FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit). We then used BWA (Li and Durbin, 2009) to align sequences to the reference genome. Resulting BAM files were indexed and manipulated using SAMTOOLS (Li et al., 2009). Each sample was then genotyped using the HaplotypeCaller function of GATK (McKenna et al., 2010). Genotyping was then done with the GenotypeGVCFs function of GATK. After this, biallelic SNPs were selected using the SelectVariants function of GATK. Variants were filtered again using the VariantFiltration function of GATK, removing all variants with confidence by depth <2.0 and mapping quality <20.0. The resulting dataset contained 913,400 loci for 106 individuals (Dryad Appendix D2).

The SNPs that resulted from the above filtering processes were used to create three datasets (Broad, SDCo A, and SDCo B; Table 1). The Broad dataset was created to reveal the overall relationships among sampled scrub white oaks, so the results of this study can be compared to other studies on scrub white oaks that have made use of genome-scale SNP data (Fitz-Gibbon et al., 2017; Kim et al., 2018). This broad genetic context also ensures that the focal system (*Q. dumosa* and *Q. berberidifolia*) is understood in the context of other scrub white oak species. The SDCo A dataset contains all the samples of the two focal species collected in San Diego County and is thus intended to show relationships between the two focal species in the region where they come into contact. Finally, the SDCo B dataset, a subset of the SDCo A dataset, contains only those samples

of the focal species from coastal San Diego County, the region where genetic exchange is suspected to occur most frequently. Not all samples from coastal San Diego County are included in SDCo B; some populations were subsampled to reduce the number of samples per population to four or less; these samples were selected at random (Table 1).

We used VCFtools (Danecek et al., 2011) to create all three datasets. We removed all loci with a low depth of coverage (min-meanDP = 7) as well as loci with unusually high depth of coverage (max-meanDP = 22). We also removed all sites with >20% missing data (max-missing 0.2) and all singleton loci (singletons) and selected only the biallelic SNPs (min-alleles 2, max-alleles 2). Finally, we thinned the datasets to loci ≥1000 bp apart on the reference genome (thin 1000). The Broad dataset contained 15,384 loci for 89 individuals (Dryad Appendix D3); the final SDCo A dataset contained 14,403 loci for 60 individuals (Dryad Appendix D4), and the SDCo B dataset contained 13,710 loci for 48 individuals (Dryad Appendix D5).

Population genetics

We employed population genetic analysis to identify groups of plants and test for genetic admixture among groups. The methods described below were applied to both the Broad and the SDCo A datasets. We used parametric and nonparametric approaches. For the nonparametric approach, we conducted principal coordinate analysis (PCoA) using an identity-by-state genetic distance matrix. The genetic distance matrix was created in TASSEL version 5.2 (Bradbury et al., 2007). The PCoA was carried out in R version 3.5.0 (R Development Core Team, 2018) using the "ecodist" package version 2.0.1 (Goslee and Urban, 2007). For the parametric approach, we used STRUCTURE version 2.3.4 (Pritchard et al., 2000; Falush et al., 2003, 2007; Hubisz et al., 2009). We used the Evanno et al. (2005) method to infer the most likely number of genetic groups. In each run, we used 50,000 Markov chain Monte Carlo (MCMC) replicates, with an admixture ancestry model. We discarded the first 10,000 replicates of each run as burn-in. For the Broad dataset (Table 1) the analysis was repeated 14 times, varying *K* from 1 to 14, with 10 replicate MCMC chains at each *K*. For the SDCo A dataset (Table 1) the analysis was repeated 10 times, varying *K* from 1 to 10, with 10 replicate MCMC chains at each *K*.

Vegetative phenotype

Phenotypic characters were selected that have been used in the past for systematic distinction of scrub white oak taxa, with an emphasis on the leaves and stems because these characters are visible throughout the year. We used phenotypic data previously collected by D. Burge (Table 1). The phenotypic characters deal mainly with leaf trichomes, which, if present, are stellate but vary in terms of how densely the trichomes are spaced and how erect the rays of the trichome are. We collected data on the following traits: (A) trichome density, the density of the trichomes on the underside of the leaf (1 = sparse, 2 = intermediate, 3 = dense); (B) trichome length, the relative length of the trichomes (1 = short, 2 = medium, 3 = long); (C) ray shape, the amount of waviness in the rays of the trichome (1 = straight, 2 = somewhat wavy, 3 = strongly wavy); (D) ray angle, the angle of the trichome rays in relation to the leaf surface (1 = appressed [parallel to the leaf surface], 2 = weakly erect [somewhat angled in relation to the leaf surface], 3 = strongly erect [stranding perpendicular to the leaf surface]); (E) leaf glands, the presence or absence of glands on the underside of the leaf (0 = absent, 1 = present); (F) petiole trichomes, the presence of trichomes on the petiole (0 = absent, 1 = present); (G) tooth number, the average total number of teeth on five randomly selected leaves; and (H) quantitative leaf shape (length divided by width of the same leaves as in G). Phenotypic data were obtained for 81 plants, including 36 *Q. berberidifolia* and 45 *Q. dumosa* (Dryad Appendix D6).

Phenotypic data were treated in a multivariate framework. We visualized data using principal component analysis (PCA) and tree reconstruction, carried out in R (R Development Core Team, 2018). For PCA, we used the "prcomp" function (R Development Core Team, 2018) to model all eight phenotypic variables. The two quantitative variables (tooth number and quantitative leaf shape) were transformed into *Z*-scores prior to analysis. The first two principal components were visualized in a bivariate plot to examine relationships. The contribution of each phenotypic character to the principal components was determined on the basis of vector loadings. On the basis of preliminary results for the PCA analysis, the variable of tooth number was discarded from the analysis because it was strongly associated with a small group of three plants from several widely separated populations that do not appear, on the basis of genetic data, to be closely related. The distant relationship among these plants suggests that the trait of tooth number does not contain a great deal of information on plant relationships, at least as we measured the trait. Discarding this trait allowed for patterns among the other traits to be more easily examined. Of the remaining traits, it was determined that only A, B, C, and D contributed significantly to the model (absolute value of maximum vector loading on first and second principal components <0.01). All other traits were excluded from the final PCA.

Climate

Average climatic conditions for selected locales were inferred using the GIS software QGIS version 3.0 (QGIS Development Team, 2018). We used latitude and longitude data from all locales included in the SDCo A dataset (Table 1). We obtained eight climatic variables: (1) climatic water deficit, (2) actual evapotranspiration, (3) average minimum temperature in winter, (4) average maximum temperature in summer, (5) temperature seasonality, (6) precipitation seasonality, (7) precipitation of the warmest quarter, and (8) precipitation of the coldest quarter. GIS data used to infer these came from the Basin Characterization Model for California (Flint et al., 2013), which simulates hydrologic response to climate at 270 m resolution.

Using the R package "raster" (Hijmans, 2017), we imported raster GIS data layers from Flint et al. (2013) and extracted data for each locale. Climatic water deficit and actual evapotranspiration were taken directly from GIS layers of Flint et al. (2013); average minimum temperature in winter and average maximum temperature in summer were calculated by taking the average of monthly minimum values for the three winter months (December, January, and February) and for the three summer months (June, July, and August), respectively. Temperature seasonality, precipitation seasonality, precipitation of the warmest quarter, and precipitation of the coldest quarter were calculated using the R package "climates" version 0.1-1.6 (VanDerWal et al., 2018), which uses the methods of Fick and Hijmans (2017). Climatic data were obtained for 24 locales (Dryad Appendix D7).

Association analysis

Associations among climatic, genetic, and phenotypic data were visualized and tested using redundancy analysis (RDA; Forester et al., 2018) and Mantel tests (Mantel, 1967). RDA is a multivariate ordination method that can be used to analyze genome-wide genetic loci and predictor variables (e.g., environment and phenotype) simultaneously. RDA determines how groups of loci covary in response to multivariate predictors, and can also detect processes that result in subtle multilocus molecular signatures (Rellstab et al., 2015; Forester et al., 2018). Both RDA and Mantel tests were implemented in the R package "vegan" version 2.5-1 (Bourret et al., 2014; Oksanen et al., 2018). We did two analyses, the first comparing genetic to climatic patterns using the SDCo A dataset, and the second comparing genetic to phenotypic data using the SDCo B genetic dataset; the SDCo B dataset was used for the analysis of phenotype due to a lack of phenotypic data for some of the oaks included in the SDCo A dataset.

RDA was carried out in "vegan" (Oksanen et al., 2018) according to Forester et al. (2018). For genetics versus climate, we used a custom R script written by Forester et al. (2018) to impute missing data. Based on the resulting dataset, we ran unconstrained RDA to test for collinearity of the climatic variables; we removed all variables with a variance inflation factor >10 (Oksanen et al., 2018), which left cwd, aet, tmin, and tmax. We then carried out a constrained RDA, constraining on latitude and longitude in order to control for the geographic position of the locales. We then ran analyses of variance to test for the significance of the full model, the significance of each RDA axis, and the significance of each of the climatic variables. In all three cases, we used 100 replicates of the permutation test (Oksanen et al., 2018; Forester et al., 2018). We then used the method of Forester et al. (2018) to identify outlier loci associated with climatic variables, selecting only loci with a score on the first three RDA axes at least three times higher (absolute value) than the average for all loci on that axis. A custom R script developed by Forester et al. (2018) was then used to associate these outliers with climatic variables.

As an independent means of assessing the significance of the association revealed by RDA, we also ran partial Mantel tests on the SDCo A genetic dataset. The Mantel tests were run using the Euclidean genetic distance matrices used for the PCoA. Pearson and Spearman correlation coefficients were used, and 100 replicates of the permutation test, controlling for geographic distance (Oksanen et al., 2018).

For genetics versus phenotype, the analysis was analogous to that for genetics versus climate, except that the SDCo B genetic dataset was used, and the analysis was constrained on climate. For the phenotypic data, we created a Euclidean distance matrix using all 11 variables. We used the method described above to identify outlier loci associated with each trait. Mantel tests were carried out as above (Oksanen et al., 2018), including one to test for the influence of environment on plant phenotype.

RESULTS

SNP data

A total of 75 samples were newly sequenced for this project. Quality was high for all samples; none were excluded for low DNA sequencing coverage or excessive amounts of missing data. Sequencing coverage averaged 4,919,971 reads per individual for the 75 samples (Dryad Appendix D1; SD = 2,870,485 reads). These samples were combined with 31 samples obtained and sequenced previously (Table 1: Sork), for a total of 106 samples (Table 1; Dryad Appendix D1). Following genotyping, the complete variant dataset contained 913,400 loci for the 106 included individuals (Dryad Appendix D2). After sub-selection of individuals and filtering of loci, the Broad dataset (Table 1) contained 15,384 loci for 89 individuals (Dryad Appendix D3), the SDCo A dataset contained 14,403 loci for 60 individuals (Dryad Appendix D4), and the SDCo B dataset contained 13,710 loci for 48 individuals (Dryad Appendix D5).

Population genetic patterns

The PCoA of the Broad dataset (Appendix S1), revealed three major groups of scrub white oaks, one containing only samples of *Q. cornelius-mulleri* and *Q. john-tuckeri* (Appendix S1: Group A), one containing all *Q. dumosa* and *Q. pacifica* along with several populations of *Q. berberidifolia* from southern California and northern Baja California (Appendix S1: Group B), and one containing all other *Q. berberidifolia*, as well as *Q. durata* (Appendix S1: Group C). STRUCTURE analysis of the Broad dataset (Appendix S2) largely matches the results of the PCoA, with two groups being identified as optimal; the two groups match those revealed by the PCoA, although in the STRUCTURE results for $K = 2$, plants found in Groups A and B of the PCoA are combined into a single group.

The PCoA of the SDCo A genetic dataset (Fig. 2) split the samples into three major groups. Groups E and F (Fig. 2) match groups B and C, respectively, of the analysis for the Broad dataset (Appendix S1). The PCoA also revealed the presence of three genetic outliers (1714-2, 1714-6, and 1714-7). STRUCTURE analysis of the SDCo A dataset revealed that two ancestral groups are optimal (Fig. 3; Appendix S3). These two major genetic groups correspond, more or less, to *Q. dumosa* and *Q. berberidifolia* (Fig. 3), although there are several plants in which proportions of genetic ancestry are out of phase with the species names applied on the basis of morphology (Fig. 3). For example, the samples obtained at locale 1718 were classified as *Q. berberidifolia* in the field but are genetically allied to the samples classified as *Q. dumosa*. The admixed and misclassified individuals are most common in the MCASM Contact Zone (Fig. 3).

Phenotypic patterns

Patterns of phenotypic variation among species, locales, and individual plants were complex (Fig. 4). Preliminary PCA suggested that only four variables made a strong contribution to the model (quantitative leaf shape, trichome density, trichome length, and ray shape), and so only these variables were retained for the final analysis. In the final model (Fig. 4), the first two principal components account for 87% of the variance. Trichome length and ray shape are both positively correlated with the first principal component (vector loading >0.50). Trichome density and quantitative leaf shape both contribute strongly to the second principal component (vector loadings 0.85 and 0.50, respectively). PCA also shows that most of the plants collected from the MCASM Contact Zone have very few trichomes on the abaxial leaf surface (Fig. 4); this includes plants identified as either focal species.

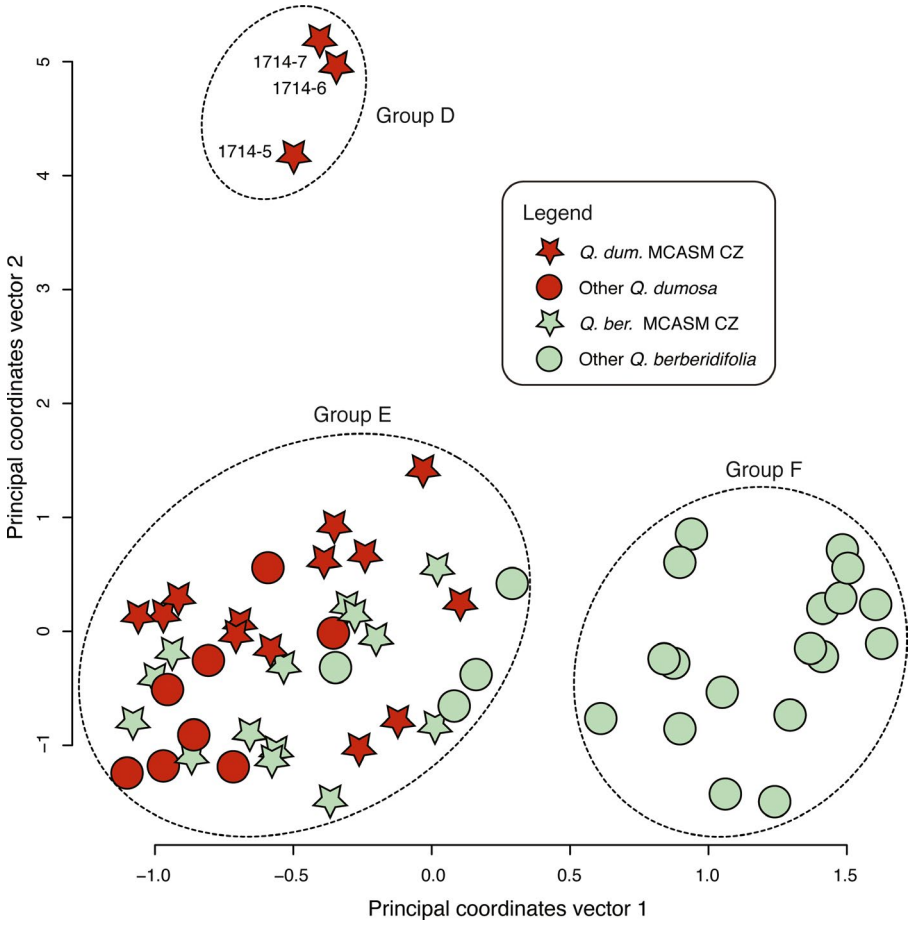

FIGURE 2. Biplot of first two axes from the principal coordinate analysis on the SDCo A dataset for *Quercus berberidifolia* and *Q. dumosa*. Assignment of species names to samples is based on field identification of plants prior to analysis. See text for explanations of Groups D, E, and F. MCASM CZ = Marine Corps Air Station Miramar Contact Zone.

Association between climate and genetics

Quercus dumosa is associated with milder, coastal conditions with a higher average winter minimum temperature (Fig. 5; Appendix S4) and lower average summer maximum temperatures (Fig. 6A); *Q. berberidifolia* is generally found farther from the coast, in areas where winter minima are lower and summer maxima are higher. Most admixed individuals are found near the contact zone between the focal species, 5–15 km from the coast, with decreasing amounts of admixture to the west and east of this zone (Fig. 5C). This area has intermediate climatic conditions in terms of average winter minimum temperature (Fig. 5) and average summer maximum temperature (results not shown).

RDA on climatic versus genetic data, controlling for latitude and longitude, yielded a significant model (F = 1.33, df = 4, P = 0.01). The first three RDA axes were significant (RDA1: F = 1.53, df = 1, P = 0.01; RDA2: F = 1.41, df = 1, P = 0.01; RDA3: F = 1.24, df = 1, P = 0.05), as were all four of the climatic variables (cwd: F = 1.21, df = 1, P = 0.04; aet: F = 1.24, df = 1, P = 0.02; tmin: F = 1.45, df = 1, P = 0.01; tmax: F = 1.29, df = 1, P = 0.01). Results of the Mantel test on climatic versus genetic data partially support RDA, with a significant association between genetics and climate in the Spearman (r = 0.1344, P = 0.026), but not the Pearson test (r = 0.085, P = 0.134).

A plot of the first two axes of RDA (Fig. 6A) mirrors the results of STRUCTURE (Fig. 3), with some separation of *Q. dumosa* from *Q. berberidifolia*, but also evidence of genetic mixture between the species. The plot also shows an association of climatic factors with genetic groups, the primary differentiation occurring between samples of *Q. berberidifolia* from outside the MCASM Contact Zone and those inside it (Fig. 6A); *Q. dumosa* is in areas of high average winter minimum temperature, high average climatic water deficit, low average summer maximum temperature, and low average actual evapotranspiration, while *Q. berberidifolia* is in areas with high average winter minimum temperature, low average climatic water deficit, high average summer maximum temperature, and high average actual evapotranspiration (Fig. 6A). This pattern is most apparent in the MCASM Contact Zone, where all of these variables, particularly average winter minimum temperate, change rapidly from west to east (Figs. 5 and 6A).

Analysis of outlier loci from RDA revealed a total of 82 candidate loci associated with the three significant RDA axes (RDA1: 17 loci; RDA2: 33 loci; RDA3: 32 loci; Appendix S5).

Association between genetics and phenotype

RDA on genetic versus phenotypic data, controlling for geographic location, yielded a significant model (F = 1.24, df = 11, P = 0.01), and the first three RDA axes were significant (RDA1: F = 2.24, df = 1, P = 0.01; RDA2: F = 1.65, df = 1, P = 0.01; RDA3: F = 1.53, df = 1, P = 0.01). Among the 11 phenotypic characters, only trichome density (F = 1.31, df = 1, P = 0.01), trichome length (F = 1.23, df = 1, P = 0.01), and leaf glands (F = 1.25, df = 1, P = 0.03) were significant. A biplot of the first two RDA axes (Fig. 6B) agrees with the groupings of the plants based on phenotype alone (Fig. 4) but reveals that genetic divergence between the focal species is correlated with phenotype, though many samples from MCASM Contact Zone are exceptions to this pattern (Fig. 6B). An RDA controlling for climate did not give substantially different results and identified the same three characters as significant contributors to the model (trichome density: F = 1.16, P = 0.05; trichome length: F = 1.29, P = 0.01; leaf glands: F = 1.24, P = 0.03). Consequently, analysis of outlier loci was carried out using results from the first RDA, constrained on geography.

Analysis of outlier loci from RDA revealed a total of 153 candidate loci associated with the three significant RDA axes (RDA1: 46 loci; RDA2: 37 loci; RDA3: 70 loci; Appendix S6). Results of the partial Mantel test for genetic versus phenotypic data (controlling for geography) revealed a significant association between genetics and phenotype in the Pearson (r = 0.1967, P = 0.009) but not the Spearman test (r = 0.123, P = 0.057); this association was stronger in the partial Mantel test that controlled for climate instead of geography (Pearson: r = 0.190, P = 0.013; Spearman: r = 0.135, P = 0.041).

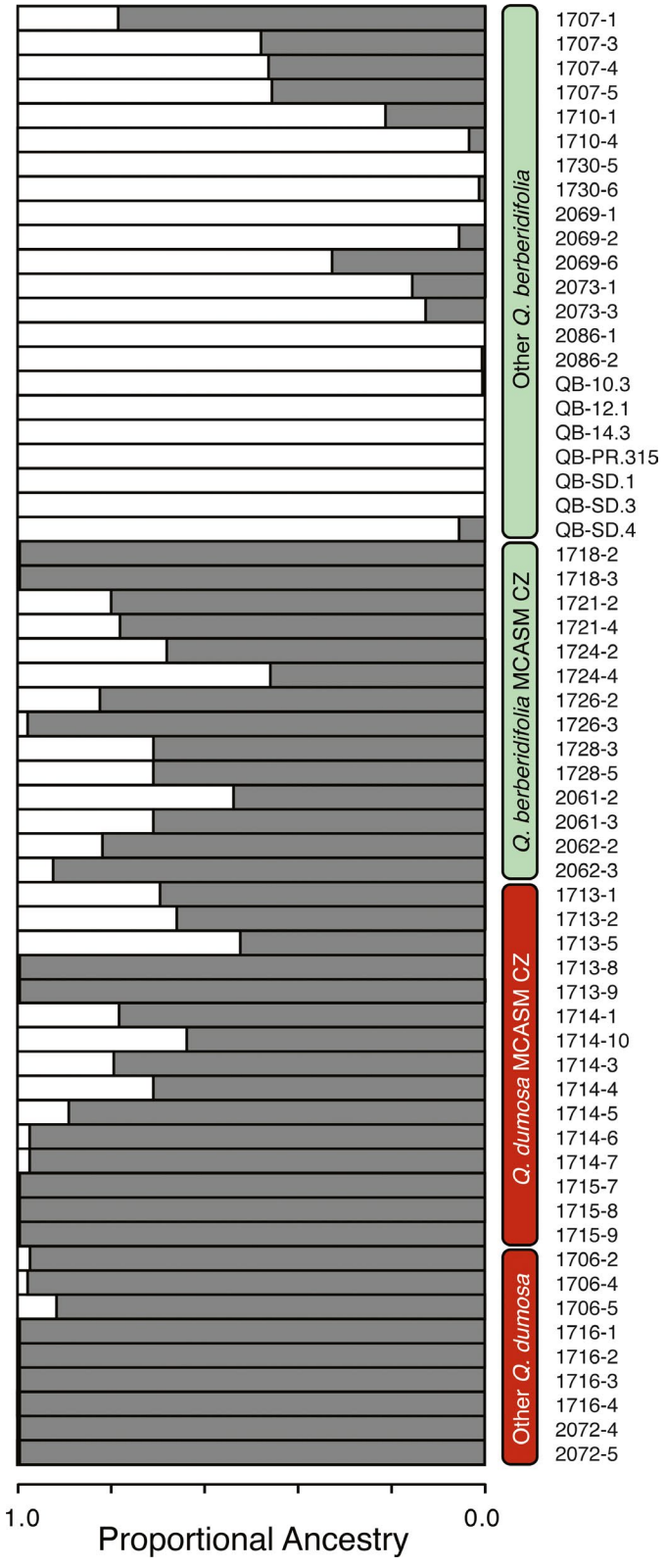

FIGURE 3. STRUCTURE results for the SDCo A dataset (at $K = 2$) for *Quercus berberidifolia* and *Q. dumosa*. Assignment of species names to samples is based on field identification of plants prior to analysis. Numbers are codes for sampling locales (Table 1). MCASM CZ = Marine Corps Air Station Miramar Contact Zone.

DISCUSSION

Genetic admixture and climate

Our results show that (1) hybridization between *Q. dumosa* and *Q. berberidifolia* is common in coastal San Diego County, (2) genetic exchange can be detected across the entire genome of both species in this region, and (3) the balance of introgression can be predicted by a population's position along a strong coast-to-inland climatic gradient (Fig. 5). Specifically, we found that less admixed *Q. dumosa* and *Q. berberidifolia* tend to occur at opposite ends of the temperature and drought stress gradient, the former in milder and the latter in more stringent areas (Fig. 5). Genetically admixed individuals, on the other hand, are generally limited to the area of intermediate climate. Our results are consistent with the idea that natural selection acts against hybrids under either extreme of parental environment (Anderson, 1948; Mallet, 2005; Burge et al., 2013; Abbott and Brennan, 2014; Zhao et al., 2014). We do not find evidence for one of the alternative hypotheses, that a gradient of genetic admixture results from differences in population density, with the hybrids favored at the edge of the geographic range of one or both species (Muller, 1952; Valbuena-Carabaña et al., 2005; Petit and Excoffier, 2009; Bontrager and Angert, 2018; Chhatre et al., 2018). However, future studies could test this more robustly by directly measuring the population density of each species. Furthermore, we were not able to rule out the possibility that the pattern we found is due to differential survival of the hybrids in intermediate environments (Anderson, 1948; Muller, 1952; Mallet, 2005; De La Torre et al., 2015; Mallet et al., 2016). This test would require the use of experiments to determine survival and performance of admixed versus parental genotypes along a climate gradient (reviewed by Lowry et al., 2008; Burge et al., 2013; De La Torre et al., 2014b; Melo et al., 2014; Zhao et al., 2014).

Our findings agree with the recent trend in the study of both plants and animals to view the tree of life as a product of both bifurcation (speciation) and reticulation (hybridization; Yu et al., 2014; Mallet et al., 2016; Goulet et al., 2017), with the balance of these two processes driven by history, ecological opportunity, intrinsic isolating mechanisms, and genomic processes at the level of DNA recombination. The current approach in genome-level analyses of hybridization in plant and animal systems is to reveal which introgressed genetic variants play a causal role in adaptation of populations and species affected by introgression (De La Torre et al., 2014a, 2015; Suarez-Gonzalez et al., 2016; Prunier et al., 2017; Chhatre et al., 2018). The analysis of loci associated with climate along the gradient would be a productive focus for future work. Our work identifying climate-associated loci provides a roadmap for this kind of research in scrub white oaks, and potentially for other closely related oak lineages.

Genotype versus phenotype

Our work also revealed that many individuals assigned to either *Q. dumosa* or *Q. berberidifolia* based on phenotype are nearly 100% genetically associated with the opposite species; this finding was especially true for plants from the Marine Corps Air Station Miramar area identified as *Q. berberidifolia* (e.g., Fig. 3: 1718-2 and 1718-3). On the other hand, both the RDA and some Mantel tests show that genotype and phenotype are significantly correlated, even when controlling for climate ($P < 0.01$ in both tests), which indicates that some combination of the phenotypic traits we analyzed could eventually be used to assign species names or identify hybrids, as was

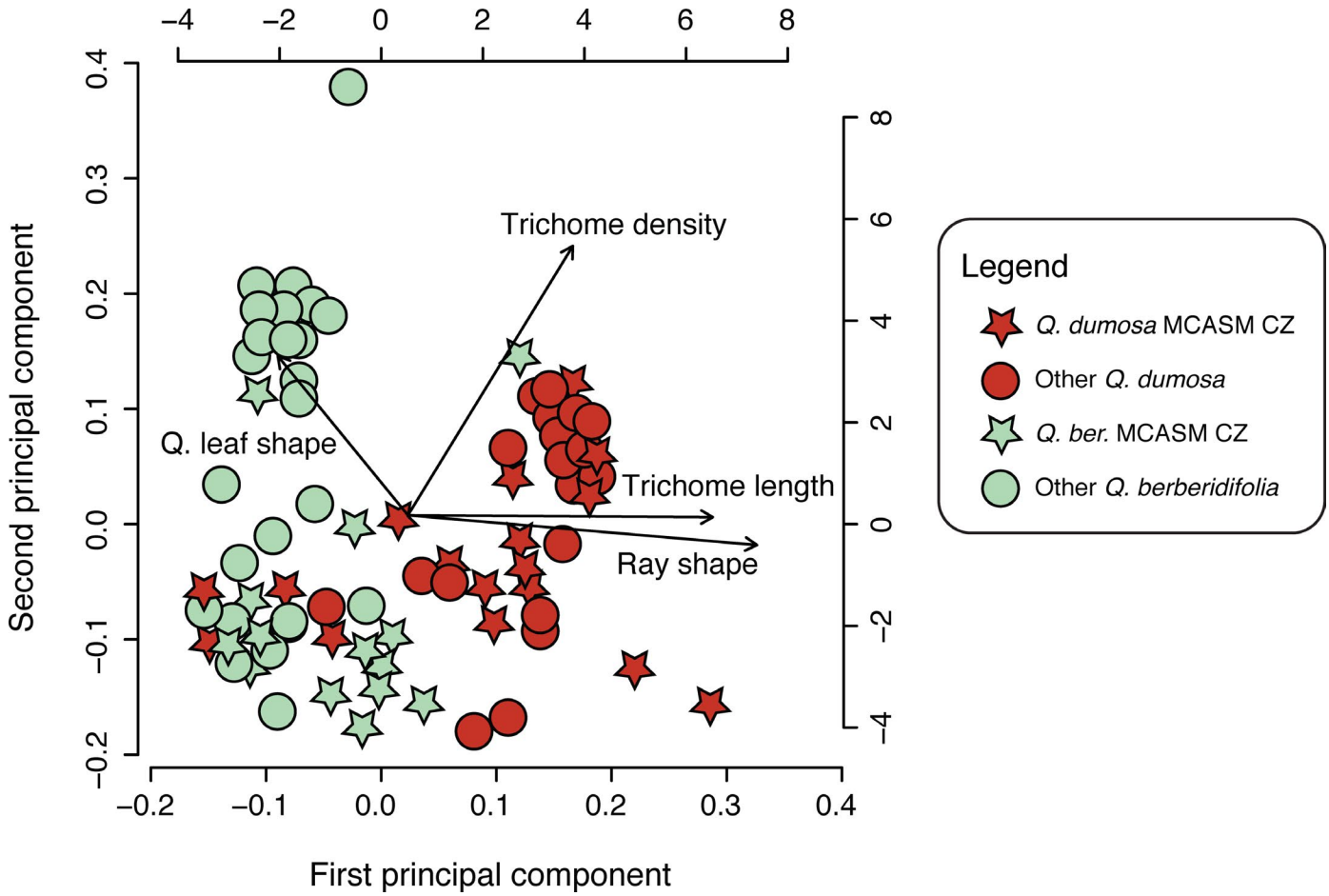

FIGURE 4. Results of principal component analysis on morphometric data for *Quercus berberidifolia* and *Q. dumosa*. Q. leaf shape = quantitative leaf shape; MCASM CZ = Marine Corps Air Station Miramar Contact Zone.

done by Cavender-Bares and Pahlich (2009). Future research on this system should test the utility of a variety of traits to differentiate the species, including those of the acorns and stems, and also explore the geographic region over which genetics is out of phase with phenotype; this region will probably coincide with the zone contact between the species. To test the robustness of those traits to identify species, wild-collected acorns could be grown in a common garden and the seedlings phenotyped. A common-garden approach would ensure that phenotypic plasticity does not produce misleading results. When matched with genome-wide markers, such results could then be used to define species and identify individual plants (Rellstab et al., 2016).

The results of our study mirror those of Riordan et al. (2016), who studied genetic exchange among three oak species in southern California—a tree white oak (*Q. engelmannii*) and two scrub white oaks (*Q. cornelius-mulleri* and *Q. berberidifolia*). Riordan et al. (2016) found that some individual plants at a subset of locales were close to 100% genetically associated with one species while being phenotypically closer to another. Such incongruencies are unlikely to be attributable to phenotypic plasticity alone because they are observed in a single subpopulation, or in a few individuals at a locale, rather than in the majority of the individuals. These incongruent individuals may illustrate the introgression of a few functional loci responsible for phenotype, while the remainder of the genome retains the original species identity. If this is the case, then the results of our study and that of Riordan et al. (2016) suggest that phenotypes currently used to identify scrub white oak species may involve very few genetic loci. As discussed above, a common-garden approach would help test this hypothesis.

Conservation implications

Genetic confirmation of species identity is the gold standard in plant conservation. However, our results illustrate that the genetic criterion for identification is not always straightforward to apply. Our analysis of genome-wide loci indicates that, in the area of contact, a proportion of plants classified as *Q. berberidifolia* contain a significant amount of *Q. dumosa* genetic ancestry (Fig. 3: *Q. berberidifolia* MCASM Contact Zone). Conservation measures for *Q. dumosa*, motivated by the CNPS Rare Plant Program (2016), are difficult to apply to the admixed individuals. A decision should be made as to whether plants that are morphologically assignable to *Q. berberidifolia* but nonetheless contain nearly 100% *Q. dumosa* ancestry (e.g., Fig. 3: 1718-2 and 1718-3) will be treated as rare. A decision should also be made as to how much genetic admixture of *Q. dumosa* by *Q. berberidifolia* is tolerated before a population is no longer managed as rare (Levin et al., 1996; Mallet, 2005; Wayne and Shaffer, 2016). Given that admixture is the result of a natural process, it might be that admixed individuals should receive the same legal protections as the "pure" individuals.

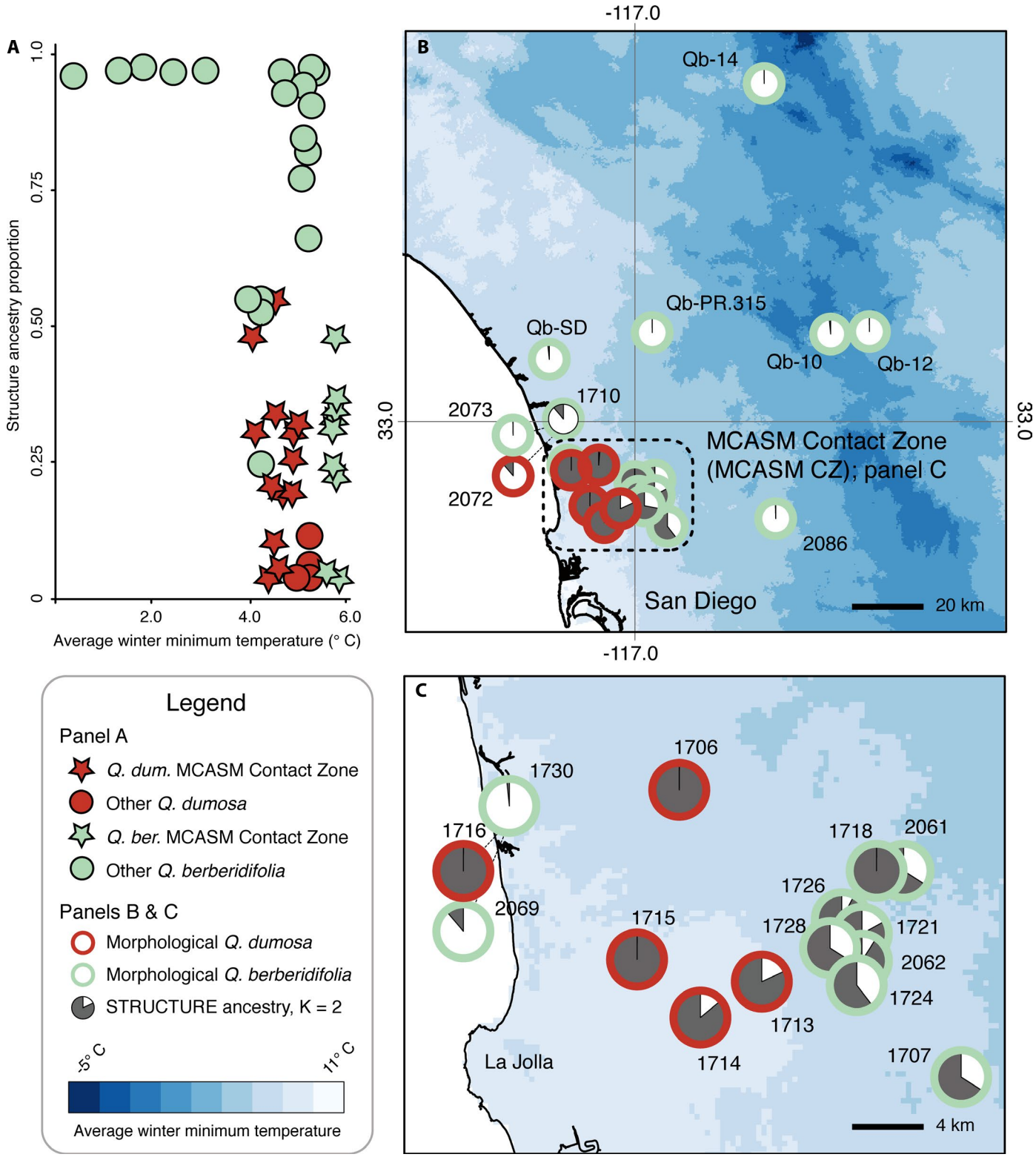

FIGURE 5. Ancestry proportions on a map of average winter minimum temperature for *Quercus berberidifolia* and *Q. dumosa*. Ancestry proportions are based on STRUCTURE analysis for $K = 2$ (average proportion across all sampled plants for a given locale). (A) Biplot of proportional genetic ancestry against average winter minimum temperature. (B) Ancestry proportions mapped on far southern California. (C) Magnification of Marine Corps Air Station Miramar (MCASM) Contact Zone, showing ancestry proportion.

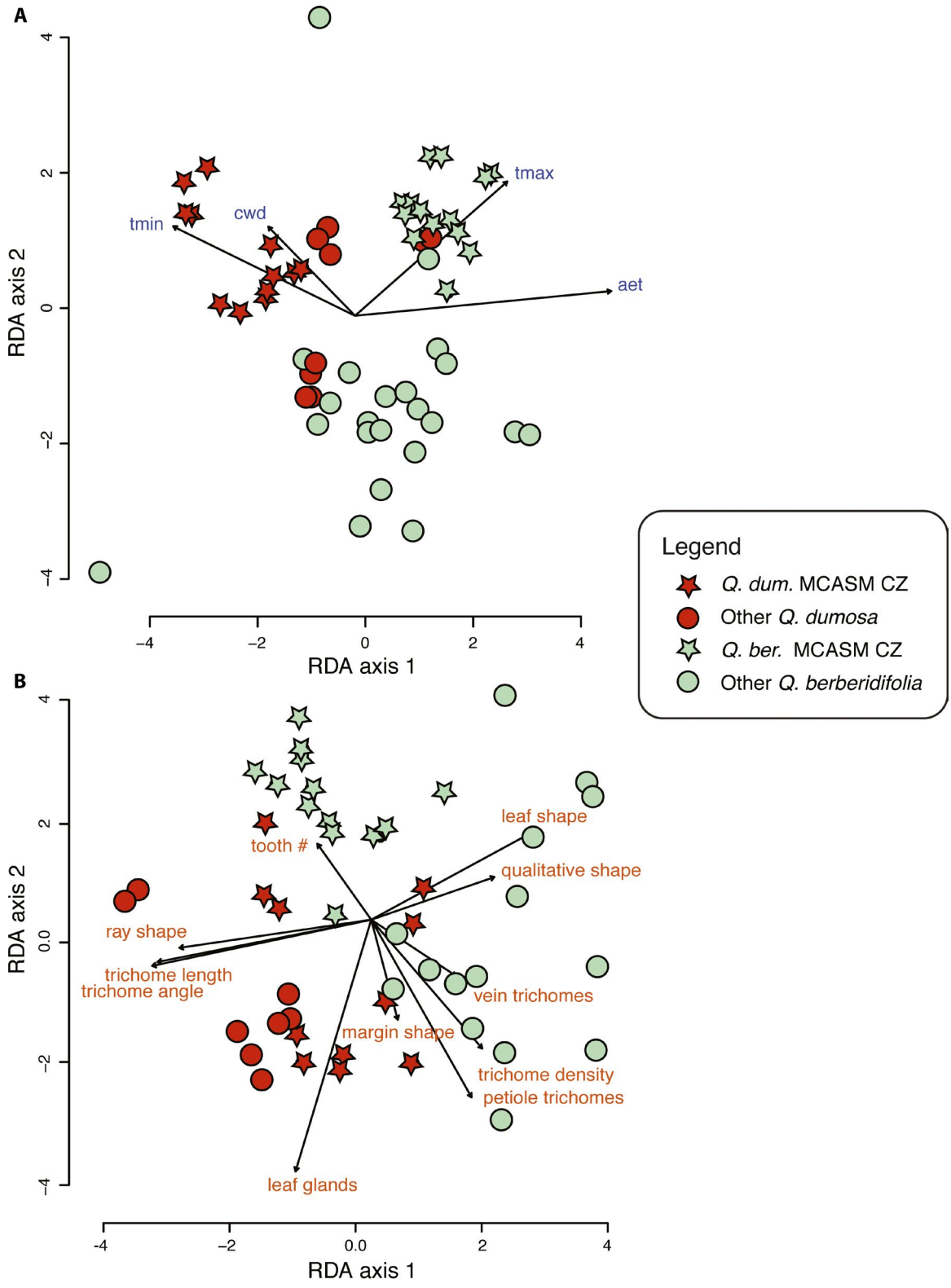

FIGURE 6. Results of redundancy analysis (RDA) of climate and morphology for *Quercus berberidifolia* and *Q. dumosa*. (A) Biplot showing the result of RDA for genetics and climate. (B) Biplot showing the result of RDA for morphology and climate. Both analyses control for geography (latitude and longitude). Small dots at the center represent loci; orange dots are those significantly associated with specific climatic or morphological factors. The position of the sampled plants indicates their genetic relationships with one another; direction and magnitude of the vectors shows how the four analyzed factors are associated with the groupings of plants. MCASM CZ = Marine Corps Air Station Miramar Contact Zone.

Overall, more work will be needed before genetic data can be used to motivate conservation decisions in scrub white oaks. Such work should target more populations of both focal species in southern California, to ascertain the extent of genetic admixture over the region of contact. This work will lead to outcomes that are consistent with the genetics of the plants being conserved rather than the names applied to them based on their visually observable phenotype. Our results imply that conservation of *Q. dumosa*, a rare species from a dwindling habitat type, should take into account the effect of genetic exchange with *Q. berberidifolia*.

CONCLUSIONS

We found that admixed oaks are limited to a zone of rapid climatic transition between mild coastal areas where *Q. dumosa* is most often found and more stringent interior areas where *Q. berberidifolia* is more common. Furthermore, the presence of misclassified individuals in some parts of the contact zone suggests that genetic exchange may lead to the breakdown of phenotypes that are otherwise strongly correlated with genetic patterns. Our work is consistent with the hypothesis that introgression between these two species is shaped by climate. In general, such findings suggest that natural selection associated with climate, not sympatry alone, may explain genetic exchange across species boundaries. Future work could refine these findings by obtaining whole genome sequences along a gradient of admixture, and testing admixed individuals for their performance against "pure" individuals in contrasting climatic conditions. Such work will reveal which genes most often move from one species to another, and whether these genes promote survival under specific environmental conditions. Overall, rather than posing a threat to the biological or taxonomic species concept, genetic exchange among oak species might be better viewed as a source of useful adaptive variation for oak survival.

ACKNOWLEDGEMENTS

Funding was provided by the U.S. Department of the Navy through a cooperative agreement with San Francisco State University (Cooperative Agreement N62473-14-2-0003), by the National Science Foundation via Plant Genome Research Project award to V.L.S. (IOS-1444611), and by University of California, Los Angeles, research funds to V.L.S. The funders played no role in study design, data collection, analysis, or preparation of the manuscript.

DATA ACCESSIBILITY

Raw reads for new DNA sequencing (Table 1: Burge) are deposited in the NCBI Short Read Archive (BioProject PRJNA398401); raw data from sequences collected previously (Table 1: Sork; Kim et al., 2018) were separately deposited to NCBI (BioProject PRJNA473578). The following datasets are available from the Dryad Digital Repository: https://doi.org/10.5061/dryad.6v03s1s (Burge et al., 2019): (1) collection information on individual plants and summary statistics on the raw sequence data for new DNA sequencing (Dryad Appendix D1), (2) raw genetic variant data (Dryad Appendix D2), (3) filtered genetic variant data (Dryad Appendices D3–D5), (4) raw phenotypic data (Dryad Appendix D6), and (5) raw climatic data (Dryad Appendix D7).

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

APPENDIX S1. Biplot of the first two axes of PCoA on the Broad genetic dataset.

APPENDIX S2. Raw STRUCTURE output for the Broad genetic dataset.

APPENDIX S3. Raw STRUCTURE output for the SDCo A genetic dataset.

APPENDIX S4. RDA-based ancestry on a map of average winter minimum temperature for the focal region.

APPENDIX S5. Biplot showing the result of redundancy analysis (including loci) for genetics versus climate.

APPENDIX S6. Biplot showing the result of redundancy analysis (incluing loci) for genetics versus phenotype.

LITERATURE CITED

Abbott, R. J., and A. C. Brennan. 2014. Altitudinal gradients, plant hybrid zones and evolutionary novelty. *Philosophical Transactions of the Royal Society B: Biological Sciences* 369: 20130346.

Abbott, R. J., M. G. Ritchie, and P. M. Hollingsworth. 2008. Introduction. Speciation in plants and animals: pattern and process. *Philosophical Transactions of the Royal Society B* 363: 2965–2969.

Abbott, R. J., N. H. Barton, and J. M. Good. 2016. Genomics of hybridization and its evolutionary consequences. *Molecular Ecology* 25: 2325–2332.

AECOM, Incorporated. 2015. Upland endangered plant species census and monitoring for *Q. dumosa*, MCASM Miramar, Final Report. Prepared for Marine Corps Air Station Miramar, Natural Resources Division.

Alberto, F. J., J. Derory, C. Boury, J.-M. Frigerio, N. E. Zimmermann, and A. Kremer. 2013. Imprints of natural selection along environmental gradients in phenology-related genes of *Quercus petraea*. *Genetics* 195: 495–512.

Anderson, E. 1948. Hybridization of the habitat. *Evolution* 2: 1–9.

Arnold, M. L. 2006. Evolution through genetic exchange. Oxford University Press, Oxford, UK.

Benson, L., E. A. Phillips, and P. A. Wilder. 1967. Evolutionary sorting of characters in a hybrid swarm. I: direction of slope. *American Journal of Botany* 54: 1017–1026.

Boecklen, W. J., and D. J. Howard. 1997. Genetic analysis of hybrid zones: numbers of markers and power of resolution. *Ecology* 78: 2611–2616.

Bontrager, M., and A. L. Angert. 2018. Gene flow improves fitness at a range edge under climate change. *Evolution Letters* 3: 55–68.

Bourret, V., M. Dionne, and L. Bernatchez. 2014. Detecting genotypic changes associated with selective mortality at sea in Atlantic salmon: polygenic multilocus analysis surpasses genome scan. *Molecular Ecology* 23: 4444–4457.

Bradbury, P. J., Z. Zhang, D. E. Kroon, T. M. Casstevens, Y. Ramdoss, and E. S. Buckler. 2007. TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics* 23: 2633–2635.

Brandvain, Y., A. M. Kenney, L. Flagel, G. Coop, and A. L. Sweigart. 2014. Speciation and Introgression between *Mimulus nasutus* and *Mimulus guttatus*. *PLoS Genetics* 10: e1004410.

Buckley, T. R., M. Cordeiro, D. Marshall, and C. Simon. 2006. Differentiating between hypotheses of lineage sorting and introgression in New Zealand alpine cicadas (*Maoricicada dugdale*). *Systematic Biology* 55: 411–425.

Burgarella, C., Z. Lorenzo, R. Jabbour-Zahab, R. Lumaret, E. Guichoux, R. J. Petit, Á. Soto, and L. Gil. 2009. Detection of hybrids in nature: application to oaks (*Quercus suber* and *Q. ilex*). *Heredity* 102: 442–452.

Burge, D. O., R. Hopkins, Y. H. Tsai, and P. S. Manos. 2013. Limited hybridization across an edaphic disjunction between the gabbro-endemic shrub *Ceanothus roderickii* (Rhamnaceae) and the soil-generalist *Ceanothus cuneatus*. *American Journal of Botany* 100: 1883–1895.

Burge, D. O., V. T. Parker, M. Mulligan, and V. L. Sork. 2019. Data from: Influence of a climatic gradient on genetic exchange between two oak species. Dryad Digital Repository. https://doi.org/10.5061/dryad.6v03s1s.

Burger, W. C. 1975. The species concept in *Quercus*. *Taxon* 24: 45–50.

Carson, H. L. 1985. Unification of speciation theory in plants and animals. *Systematic Botany* 10: 380–390.

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

Chhatre, V. E., L. M. Evans, S. P. DiFazio, and S. R. Keller. 2018. Adaptive introgression and maintenance of a trispecies hybrid complex in range-edge populations of *Populus*. *Molecular Ecology* 27: 4820–4838.

CNPS Rare Plant Program. 2016. Inventory of Rare and Endangered Plants. California Native Plant Society, Sacramento, California http://www.rareplants.cnps.org (accessed 28 November 2016).

Coyne, J. A., and H. A. Orr. 2004. Speciation. Sinauer Associates, Sunderland, MA.

Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, et al. 2011. The variant call format and VCFtools. *Bioinformatics* 27: 2156–2158.

De La Torre, A. R., D. R. Roberts, and S. N. Aitken. 2014a. Genome-wide admixture and ecological niche modelling reveal the maintenance of species boundaries despite long history of interspecific gene flow. *Molecular Ecology* 23: 2046–2059.

De La Torre, A. R., T. Wang, B. Jaquish, and S. N. Aitken. 2014b. Adaptation and exogenous selection in a *Picea glauca* × *Picea engelmannii* hybrid zone: implications for forest management under climate change. *New Phytologist* 201: 687–699.

De La Torre, A. R., P. K. Ingvarsson, and S. N. Aitken. 2015. Genetic architecture and genomic patterns of gene flow between hybridizing species of Picea. *Heredity* 115: 153–164.

Dodd, R. S., and Z. Afzal-Rafii. 2004. Selection and dispersal in a multispecies oak hybrid zone. *Evolution* 58: 261–269.

Eaton, D. A. R., A. L. Hipp, A. González-Rodríguez, and J. Cavender-Bares. 2015. Historical introgression among the American live oaks and the comparative nature of tests for introgression. *Evolution* 69: 2587–2601.

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

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

Falush, D., M. Stephens, and J. K. Pritchard. 2007. Inference of sampling locale structure using multilocus genotype data: Dominant markers and null alleles. *Molecular Ecology Notes* 7: 574–578.

Fick, S., and R. Hijmans. 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology* 37: 4302–4315.

Fitz-Gibbon, S. A. L. Hipp, K. K. Pham, P. S. Manos, and V. L. Sork. 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.

Flint, L. E., A. L. Flint, J. H. Thorne, and R. Boynton. 2013. Fine-scale hydrologic modeling for regional landscape applications: the California Basin Characterization Model development and performance. *Ecological Processes* 2: 25.

Folk, R. A., P. S. Soltis, D. E. Soltis, and R. Guralnick. 2018. New prospects in the detection and comparative analysis of hybridization in the tree of life. *American Journal of Botany* 105: 364–375.

Forde, M., and D. Faris. 1962. Effect of introgression on the serpentine endemism of *Quercus durata*. *Evolution* 16: 338–347.

Forester, B. R., J. R. Lasky, H. H. Wagner, and D. L. Urban. 2018. Comparing methods for detecting multilocus adaptation with multivariate genotype-environment associations. *Molecular Ecology* 27: 2215–2233.

Gailing, O., J. Lind, and E. Lilleskov. 2012. Leaf morphological and genetic differentiation between *Quercus rubra* L. and *Q. ellipsoidalis* E.J. Hill populations in contrasting environments. *Plant Systematics and Evolution* 298: 1533–1545.

González-Rodríguez, A., D. M. Arias, S. Valencia, and K. Oyama. 2004. Morphological and RAPD analysis of hybridization between *Quercus affinis* and *Q. laurina* (Fagaceae), two Mexican red oaks. *American Journal of Botany* 91: 401–409.

Goslee, S. C., and D. L. Urban. 2007. The ecodist package for dissimilarity-based analysis of ecological data. *Journal of Statistical Software* 22: 1–19.

Goulet, B. E., F. Roda, and R. Hopkins. 2017. Hybridization in plants: old ideas, new techniques. *Plant Physiology* 173: 65–78.

Hardin, J. W. 1975. Hybridization and introgression in *Quercus alba*. *Journal of the Arnold Arboretum* 56: 336–363.

Hijmans, R. J. 2017. raster: geographic data analysis and modeling. R package version 2.6-7. http://CRAN.R-project.org/package=raster.

Hipp, Al. L., D. A. R. Eaton, J. Cavender-Bares, E. Fitzek, R. Nipper, and P. S. Manos. 2014. A framework phylogeny of the American oak clade based on sequenced RAD data. *PLoS ONE* 9: e102272.

Hubisz, M. J., D. Falush, M. Stephens, and J. K. Pritchard. 2009. Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources* 9: 1322–1332.

Joly, S., P. A. McLenachan, and P. J. Lockhart. 2009. A statistical approach for distinguishing hybridization and incomplete lineage sorting. *The American Naturalist* 174: E54–E70.

Khodewekar, S., and O. Gailing. 2017. Evidence for environment-dependent introgression of adaptive genes between two red oak species with different drought adaptations. *American Journal of Botany* 104: 1088–1098.

Kim, B. Y., X. Wei, S. Fitz-Gibbon, K. E. Lohmueller, J. Ortego, P. F. Gugger, and V. L. Sork. 2018. RADseq data reveal ancient, but not pervasive, introgression between Californian tree and scrub oak species (*Quercus* sect. *Quercus*: Fagaceae). *Molecular Ecology* 27: 4556–4571.

Lagache, L., E. K. Klein, E. Guichoux, and R. J. Petit. 2013. Fine-scale environmental control of hybridization in oaks. *Molecular Ecology* 22: 423–436.

Leroy, T., C. Roux, L. Villate, C. Bodénès, J. Romiguier, J. A. P. Paiva, C. Dossat, et al. 2017. Extensive recent secondary contacts between four European white oak species. *New Phytologist* 214: 865–878.

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

Li, H., and R. Durbin. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25: 1754–1760.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, et al. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* 25: 2078–2079.

Lowry, D. B., J. L. Modliszewski, K. M. Wright, C. A. Wu, and J. H. Willis. 2008. The strength and genetic basis of reproductive isolating barriers in flowering plants. *Philosophical transactions of the Royal Society of London. Series B, Biological Sciences* 363: 3009–3021.

Lozier, J. D. 2014. Revisiting comparisons of genetic diversity in stable and declining species: assessing genome-wide polymorphism in north American bumble bees using RAD sequencing. *Molecular Ecology* 23: 788–801.

Mallet, J. 2005. Hybridization as an invasion of the genome. *Trends in Ecology & Evolution* 20: 229–237.

Mallet, J., N. Besansky, and M. W. Hahn. 2016. How reticulated are species? *BioEssays* 38: 140–149.

Mantel, N. 1967. The detection of disease clustering and a generalized regression approach. *Cancer Research* 27: 209–220.

McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, et al. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* 20: 1297–1303.

McVay, J. D., D. Hauser, A. L. Hipp, and P. S. Manos. 2017a. Phylogenomics reveals a complex evolutionary history of lobed-leaf white oaks in western North America. *Genome* 60: 733–742.

McVay, J. D., A. L. Hipp, and P. S. Manos. 2017b. A genetic legacy of introgression confounds phylogeny and biogeography in oaks. *Proceedings of the Royal Society, B* 284: 20170300.

Melo, M. C., A. Grealy, B. Brittain, G. M. Walter, and D. Ortiz-Barrientos. 2014. Strong extrinsic reproductive isolation between parapatric populations of an Australian groundsel. *New Phytologist* 203: 323–334.

Muir, G., and C. Schlötterer. 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.

Muir, G., C. Fleming, and C. Schlötterer. 2000. Taxonomy: species status of hybridizing oaks. *Nature* 405: 1016.

Muller, C. 1952. Ecological control of hybridization in *Quercus*: a factor in the mechanism of evolution. *Evolution* 6: 147–161.

Nason, J. D., N. C. Ellstrand, and M. L. Arnold. 1992. Patterns of hybridization and introgression in populations of oaks, manzanitas, and irises. *American Journal of Botany* 79: 101–111.

Nixon, K. C. 2002. The oak (*Quercus*) biodiversity of California and adjacent regions. Pp. 3–20 in: Standiford, R. B., D. McCreary, K. L. Purcell, tech. coordinators. Proceedings of the Fifth Symposium on Oak Woodlands: Oaks in California's Challenging Landscape. Gen. Tech. Rep. PSW-GTR-184, Albany, CA: Pacific Southwest Research Station, Forest Service, U.S. Department of Agriculture.

Nixon, K. C., and C. H. Muller. 1994. New names in California oaks. *Novon* 4: 391–393.

Nixon, K. C., and C. H. Muller. 1997. *Quercus* section *Quercus*. *In* Flora of North America north of Mexico, Vol. 3, Flora of North America Editorial Committee, eds. 471–506. New York: Oxford University Press.

Nixon, K. C., and K. P. Steele. 1981. A new species of *Quercus* from southern California. *Madroño* 28: 210–219.

Oksanen, J., F. G. Blanchet, R. Kindt, P. Legendre, P. R. Minchin, R. B. O'Hara, G. L. Simpson, et al. 2018. vegan: community ecology package. R package version 2.5-1. http://CRAN.R-project.org/package=vegan.

Ortego, J. V., P. F. Gugger, E. C. Riordan, and V. L. Sork. 2014. Influence of climatic niche suitability and geographical overlap on hybridization patterns among southern Californian oaks. *Journal of Biogeography* 41: 1895–1908.

Ortego, J. V., P. F. Gugger, and V. L. Sork. 2017a. Impacts of human-induced environmental disturbances on hybridization between two ecologically differentiated Californian oak species. *New Phytologist* 213: 942–955.

Ortego, J. V., P. F. Gugger, and V. L. Sork. 2017b. Genomic data reveal cryptic lineage diversification and introgression in Californian golden cup oaks (section Protobalanus). *New Phytologist* 218: 804–818.

Petit, R. J., and L. Excoffier. 2009. Gene flow and species delimitation. *Trends in Ecology and Evolution* 24: 386–393.

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

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

Prunier, J., S. Caron, and J. MacKay. 2017. CNVs into the wild: screening the genomes of conifer trees (*Picea* spp.) reveals fewer gene copy number variations in hybrids and links to adaptation. *BMC Genomics* 18: 1.

QGIS Development Team. 2018. QGIS Geographic Information System. Open Source Geospatial Foundation Project. http://qgis.osgeo.org.

R Development Core Team. 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Rellstab, C., F. Gugerli, A. J. Eckert, A. M. Hancock, and R. Holderegger. 2015. A practical guide to environmental association analysis in landscape genomics. *Molecular Ecology* 24: 4348–4370.

Rellstab, C., A. Buhler, R. Graf, C. Folly, and F. Gugerli. 2016. Using joint multivariate analyses of leaf morphology and molecular-genetic markers for taxon identification in three hybridizing European white oak species (*Quercus* spp.). *Annals of Forest Science* 73: 669–679.

Renaut, S., C. J. Grassa, S. Yeaman, B. T. Moyers, Z. Lai, N. C. Kane, J. E. Bowers, et al. 2013. Genomic islands of divergence are not affected by geography of speciation in sunflowers. *Nature Communications* 4: 1827.

Rieseberg, L. H., and J. H. Willis. 2007. Plant Speciation. *Science* 317: 910–914.

Riordan, E. C., P. F. Gugger, J. Ortego, C. Smith, K. Gaddis, P. Thompson, and V. L. Sork. 2016. Association of genetic and phenotypic variability with geography and climate in three southern California oaks. *American journal of botany* 103: 73–85.

Rosatti, T. J., J. M. Tucker. 2012. *Quercus*. *In* B. G. Baldwin et al., eds. The Jepson Manual: Vascular Plants of California, 803–808. University of California Press, Berkeley, California.

Rutherford, S., M. Rossetto, J. G. Bragg, H. McPherson, D. Benson, S. P. Bonser, and P. G. Wilson. 2018. Speciation in the presence of gene flow: population genomics of closely related and diverging *Eucalyptus* species. *Heredity* 121: 126–141.

Sawyer, J. O., T. Keeler-Wolf, and J. M. Evens. 2009. A manual of California vegetation, 2nd edition. California Native Plant Society, Sacramento, California.

Soltis, P. S., and D. E. Soltis. 2009. The role of hybridization in plant speciation. *Annual Review of Plant Biology* 60: 561–588.

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

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

Suarez-Gonzalez, A., C. A. Hefer, C. Christie, O. Corea, C. Lexer, Q. C. B. Cronk, and C. J. Douglas. 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.

Sullivan, A. R., S. A. Owusu, J. A. Weber, A. L. Hipp, and O. Gailing. 2016. Hybridization and divergence in multi-species oak (Quercus) communities. *Botanical Journal of the Linnean Society* 181: 99–114.

Tucker, J. M. 1952. Taxonomic interrelationships in the *Quercus dumosa* complex. *Madroño* 11: 234–251.

Valbuena-Carabaña, M., S. C. González-Martínez, V. L. Sork, C. Collada, A. Soto, P. G. Goicoechea, and L. Gil. 2005. Gene flow and hybridisation in a mixed oak forest (*Quercus pyrenaica* Willd. and *Quercus petraea* (Matts.) Liebl.) in central Spain. *Heredity* 95: 457–465.

Valencia-Cuevas, L., P. Mussali-Galante, D. Piñero, E. Castillo-Mendoza, G. Rangel-Altamirano, and E. Tovar-Sánchez. 2015. Hybridization of *Quercus castanea* (Fagaceae) across a red oak species gradient in Mexico. *Plant Systematics and Evolution* 301: 1085–1097.

VanDerWal, J., L. Beaumont, N. Zimmerman, P. Lorch, and D. Blodgett. 2018. climates: methods for working with weather & climate. R package version 0.1-1.6. https://rforge.net/climates/index.html.

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

Wayne, R. K., and H. B. Shaffer. 2016. Hybridization and endangered species protection in the molecular era. *Molecular Ecology* 25: 2680–2689.

Whittemore, A. T., and B. A. Schaal. 1991. Interspecific gene flow in sympatric oaks. *Proceedings of the National Academy of Sciences* 88: 2540–2544.

Yu, Y., J. Dong, K. J. Liu, and L. Nakhleh. 2014. Maximum likelihood inference of reticulate evolutionary histories. *Proceedings of the National Academy of Sciences of the United States of America* 111: 16448–16453.

Zhao, W., J. Meng, B. Wang, L. Zhang, Y. Xu, Q. Y. Zeng, Y. Li, et al. 2014. Weak crossability barrier but strong juvenile selection supports ecological speciation of the hybrid pine *Pinus densata* on the Tibetan plateau. *Evolution* 68: 3120–3133.