Three Scales of Gene Flow in California White Oaks

By

Prahlada D. Papper

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Integrative Biology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor David D. Ackerly, chair
Professor Richard S. Dodd
Professor Paul V. A. Fine

Summer 2021

Abstract

Three Scales of Gene Flow in California White Oaks

by

Prahlada D. Papper

Doctor of Philosophy in Integrative Biology

University of California, Berkeley

Professor David D. Ackerly, Chair

Gene flow between populations is one of the primary mechanisms of evolution. In plants, it can occur either by the dispersal and establishment of seeds into a local population from outside or by the dispersal of pollen and successful reproduction. Either of these sources of genetic variation must then be followed by successful survival and reproduction of the immigrant or offspring so that the introduced genetic variation continues to contribute to the local population over time. This time-integration of the genetic contribution constitutes the realized gene flow between populations or lineages. The rate of gene flow and constraints on it are crucially important to population genetic structure and phylogenetic lineage divergence as well as patterns of local adaptation and genetic variation.

In this dissertation I study gene flow at three evolutionary scales in the western North American clade of white oaks (genus *Quercus*, section *Quercus s.s.*, series *Dumosae*).

First, in chapter one, I use two established common garden plantings of blue oak (*Q. douglasii*), together with surveys of the provenance field populations that provided acorns for the gardens, to investigate gene flow and local adaptation among populations of a single species. I show that there are both environmental and genetic components to variation in spring phenological timing among these blue oaks (as well as a very small genotype×environment effect). There are significant differences in phenological timing associated with the different provenance sites even for trees growing in a common garden environment, reflecting a genetic component of their phenological variation. This genetic variation is correlated with the climate experienced by trees at the provenance sites. In particular, I identify notable influences of spring maximum temperature and fall and spring precipitation on genetic variation in spring phenology. Additional genetic variation, at the individual level, may be reflected by the phenological variation observed among trees from the same provenance. This individual variation is high relative to the variation that can be associated with provenance sites, suggesting that even though there may be local adaptation for this trait, there is also a great deal of genetic variation within populations. This may be due to high rates of gene flow among populations of blue oak, balanced

by a moderate effect of selection on local variation, but may also be due to temporally fluctuating selection within populations together with a more moderate rate of gene flow.

Next, in the second chapter, I focus on gene flow between pairs of oak species occurring within hybrid zones where their ranges overlap. Blue oak is again the central species, hybridizing in the northern part of its range with Oregon white oak (*Q. garryana* var. *garryana*) and in the southern part of its range with Tucker's scrub oak (*Q. john-tuckeri*). My research reveals the evolutionary and biogeographic contexts of these two hybrid zones by measuring and partitioning landscape-scale barriers to gene flow within them. I explain pairwise genetic dissimilarity between individuals as a function of their geographic separation (isolation-by-distance), environmental difference (isolation-by-environment), and phenological asynchrony (isolation-by-time). Even though it is commonly considered a basic control on genetic structure, I do not find evidence for isolation-by-distance in either of the hybrid zones, nor in a third geographic data set consisting of only blue oaks. Instead, I find that genetic dissimilarity can be partially explained as isolation-by-environment, both across the hybrid zones and within blue oaks alone. This signal is especially strongly associated with winter temperatures, and to a lesser extent with summer high temperature and aridity. In addition, in the southern hybrid zone only, between blue oak and Tucker's scrub oak, there is a strong signal of isolation-by-time. Using variation partitioning models to separate the effects of these three isolating factors into their independent and overlapping contributions, I suggest the differences in flowering phenology that contribute to genetic structure in the southern hybrid zone result from both environmental differences and genetic differences. This is much like the overall influences on phenology identified in chapter one, but in this case they represent extrinsic and intrinsic reproductive isolating mechanisms, respectively. This pattern is found only in the southern hybrid zone and not in the northern hybrid zone, which may be a reflection of the closer phylogenetic relationships between the hybridizing species in the southern zone and their biogeographic histories.

In chapter three, I place gene flow in its full phylogenetic context within this clade of oaks. I identify two evolutionary scales of gene flow: contemporary hybridization, as already discussed in chapter two, and ancient hybridization. I use methods drawn from both population genetics and phylogenetics to identify individual samples that show signs of contemporary hybridization. Removing these samples from the tips of a maximum likelihood phylogenetic tree dramatically improves resolution of groups within the clade. Using this well-resolved tree, I then turn to phylogenetic invariants methods (D-statistics, ABBA-BABA) to investigate patterns of gene flow deeper within the phylogeny. This reveals a particularly strong signal of hybridization involving the common ancestor of California scrub oak (*Q. berberidifolia*) and leather oak (*Q. durata*) and either the common ancestor of the entire southern California scrub oak clade (including *Q. douglasii*) or the *Quercus garryana* clade. There are also potentially signals of additional episodes of hybridization at medium depth within the phylogeny. The implications of hybridization at multiple depths in the phylogeny and lasting impacts of these periods are discussed in relation to evolutionary models of oaks as a whole.

Dedicated to Robin Carney
who was the first one to tell me about
the balance between Anarchy and Animism.

Later, the oaks told me about the same thing.

Anything within these pages that doesn't reflect that tension
is surely wrong.

TABLE OF CONTENTS

ACKNOWLEDGMENTS

First, I would like to thank the oaks of California for allowing me to share some of their secrets. I don't think they always understood exactly what I was after, which is probably my fault more than anything, but we had a lot of good times together.

Among humans, none is greater than David Ackerly. Thank you for all the time and attention you gave me, pulling it from who knows where. Thank for the countless ideas you planted and waited for and watched grow. I'll miss having frequent conversations with you; they were always startling and meaningful. Thank you for supporting my wandering and natural history and showing me how it connected.

Thanks also to all my other advisors and committee members over the years. To Richard Dodd, who consistently pushed me forward in the most practical ways when I could have stalled and who stimulated so much of my thinking about gene flow in oaks. To Paul Fine, who showed me how it can be done in the ecosystems of California and probably other places, too. To Brent Mishler, who let me argue with him from time to time about species and other things that don't exist. To Cindy Looy, who helped me dip my toe into deep time even if I retreated to the most recent few million years. To Joe McBride, who planted those dang trees almost 30 years ago now and gave me great advice and insights all these years later.

I also thank many of my advisors from before Berkeley. First and foremost, thanks to Steve Barnhart for lighting a fire in me that burned me through the hills of Northern California where these strange and beautiful lobey blue-green oaks grow. Thanks to Shawn Brumbaugh for advice and mentorship on how to approach research and teaching. Thanks to Glenn McGourty for many good times in the field and really being the first to show me how great doing science outdoors can be.

Thanks to all my fellow travelers in Integrative Biology. To Rachael Olliff-Yang, who I met such a long time ago now, to Meagan Oldfather and Matt Kling and Andrew Weitz and Kyle Rosenblad and Louise Barton. Thanks also to Chris DiVittorio for talking late into the night one time back in 2014 and making me think I could do this. Thanks to Will Freyman and Dori Contreras and Gabriel Damasco for conversations and advice over and over along the way. And to Seema Sheth and Rob Skelton and Lee Anderegg and Melina Kozanitas.

I also owe debt to many people who helped immeasurably with lab work, especially Lydia Smith, Karen Lundy (FGL), and Shanna McDevitt (GSL). Special thanks to Shirley Kim and Aastha Shah who were really the ones that made this stuff work. I don't know what kind of nonsense I would have come up with without you. Double extra special thanks to Chris McCarron, who contributed data from his own project o *Quercus durata* to chapter three.

Love back to Ron and Bobbi Tyler who showed me love and put their confidence in me when I felt low and wasn't sure about so much.

I want to extend special gratitude to the many people and groups that helped fund my research. Especially to Carol Baird and the Carol Baird Graduate Student Award for Field Research for funding me through crucial stages of fieldwork and writing. Also to the many groups who awarded me research grants that funded the genomic sequencing portion of my work: NorCal Botanists, the Heckard Fund of the Jepson Herbarium, Sigma Xi, SANDS for Sustainable Future, the Marin County chapter of CNPS, the Santa Clara Valley chapter of CNPS, the East Bay chapter of CNPS, the Milo Baker chapter of CNPS, and the Sanhedrin chapter of CNPS. I have been especially proud over the years that I was able to put together funding for the work I did with support and encouragement from so many local groups of botanists and naturalists.

Finally, I have to recognize all of the facilities and assistance I received from organizations and land managers around the state, where I carried out my fieldwork. Especially to Hopland and Sierra Foothills RECs and all the people there who've helped including Jeremy James, John Bailey, Kim Rodrigues, Bob Kiefer, Hannah Bird, Troy McWilliams, Dustin Flavell. Also, the entire UC Natural Reserve System, especially staff at Angelo Coast Range Reserve, Blue Oak Ranch Reserve, Hastings Natural History Reservation, Sedgwick Reserve, Santa Cruz Island Reserve, and Landells Hill-Big Creek Reserve. To Pepperwood Preserve and the staff there who gave me early opportunities to do fieldwork and encouraged the growth of my project. To the staff of the Tejon Ranch Conservancy for helping with access and recommendations. Thanks to Ryan O'Dell for help accessing remote BLM sites and great conversations in the field, they were always a highlight of my trips. And to the many people who helped with sites within National Forest land (Six Rivers, Shasta-Lassen, Stanislaus, Sequoia, Los Padres, and Angeles) and the San Joaquin Experiment Range, California State Parks, California Department of Fish & Wildlife, the parks departments of the cities of Chico and Ukiah, Marin County Parks & Open Space, and Sonoma County Landpaths. What an incredible list and what an unbelievable privilege it was to be able to do work across all of these places!

INTRODUCTION

Three scales of gene flow in California white oaks

In this dissertation, I present data primarily drawn from the western North American white oak clade and analyses that investigate gene flow at several evolutionary scales and contexts within the group.

The western North American white oaks are only recently confirmed as a monophyletic group within section *Quercus sensu stricto* of genus *Quercus* (Hipp et al. 2017). The clade is provisionally treated at the taxonomic series level, as series *Dumosae* (Hipp et al. 2019), and includes all of the white oaks native to California and the Pacific Northwest with the exception of *Q. engelmannii* and *Q. turbinella*, which have been shown to belong to series *Leucomexicanae*, a clade otherwise found in the American Southwest, Sonoran region, and Sierra Madre. Also excluded from the western North American white oak clade is *Q. sadleriana*, which has sometimes been considered a white oak, but is more recently treated in section *Ponticae* along with its sister species *Q. pontica* (Denk et al. 2017).

The oaks have a long history as a model in evolutionary and ecological studies (Cavender-Bares 2019). In particular, the frequently observed ability of species in the same taxonomic section to hybridize has been widely studied (Muller 1952, Van Valen 1976, Rushton 1993, Petit et al. 2004, Curtu et al. 2007, McVay et al. 2017a).

Gene flow and hybridization in oaks is influenced by mechanisms operating at several stages within the oak life cycle (figure 1). The geographic distance between individuals or populations must first of all be overcome by dispersal, which occurs at two stages in the oak life cycle: gametophyte dispersal of pollen and embryo dispersal of acorns. The majority of acorns may only disperse as far as gravity takes them, but longer distances can result from animal dispersal. Scatter-hording corvids are likely responsible for the most long-distance dispersal of acorns, with maximum dispersal distance between 1 and 4 km (Pesendorfer et al. 2016). Pollen dispersal in oaks may involve even longer distances, but the details remain controversial. Some parentage studies have shown that pollen dispersal distances are generally short (Pluess et al. 2009) while studies in geographically isolated stands have demonstrated that at least in some cases it may be dozens or hundreds of kilometers (Kremer et al. 2012). Pollen monitoring stations that have been established to assist with allergy forecasting may represent a novel source of data on pollen dispersal that has not been utilized in ecological and evolutionary studies. Records from these stations show distinct pulses of oak pollen associated with phenological activity occurring as much as 16 to 24 km from the collector (Fairley and Batchelder 1986, Kasprzyk 2009).

Pollen dispersal can only successfully contribute to gene flow when flowering is synchronous between individuals. My observation has been that female flowers become receptive shortly after male catkins on the same tree begin shedding pollen, which agrees with previous published results (Knapp et al. 2001, Williams et al. 2001). Thus, overlap of the pollen release interval

between trees can be used as an effective estimate of their overall flowering synchrony.

If flowering is synchronous and female flowers are within the geographic dispersal range of released pollen, then genomic incompatibilities form a series of potential reproductive barriers that may still prevent gene flow. These incompatibilities may include pollen incompatibility (inhibition of pollen tube development), gamete incompatibility (failure of fertilization), or developmental incompatibility (unsuccessful growth of offspring). Experimental pollination studies capable of quantifying pre-zygotic and early developmental genome incompatibilities in oaks are rare. What data we do have from hybridizing species suggest that interspecific pollination generally results in viable acorn production at a rate comparable to intraspecific pollination. However, in some cases there may be asymmetric incompatibilities and self-pollination rate is often reduced (Steinhoff 1993, Williams et al. 2001).

Finally, comprehensive estimates of realized gene flow must take into account the results of natural selection acting on the fitness of immigrants, whether they arrive as pollen or acorns. Those effects may appear as differential growth or survival through demographic transitions, especially juvenile stages, but can also manifest as reduced flowering or seed set even in mature trees.

Gene flow and Local Adaptation

In chapter one, I use data from established common gardens (McBride et al. 1997) to investigate genetic structure and local adaptation among 26 populations within a single oak species, blue oak (*Quercus douglasii*). Population structure can only exist when gene flow between populations is limited. Limited gene flow between populations may most simply be attributed to dispersal limitation. Even the longest estimates of pollen dispersal distance described above, which also likely only apply to tiny proportion of dispersed pollen grains, suggest that it could be a cause of limited gene flow, at least between distant populations.

However, as I show in my second chapter, geographic distance does not in fact appear to be an important explanation for genetic dissimilarity within blue oaks. Instead, the crucial factor may be selection acting on immigrants, whether they arrive as acorns or pollen. Thus, the local adaptation in spring phenology I found among blue oak populations growing in the common gardens points to the importance of this trait for survival, probably through effects of herbivory, freeze damage, and carbon acquisition (Augspurger 2013, Pearse et al. 2015). So it seems to be realized gene flow, as a long-term contribution to the local gene pool, that is limited among these populations of oaks rather than simply arrival of alleles.

Selection for greater flowering synchrony within populations is an additional possible explanation for the population signal of spring phenology observed in the common gardens. Though the blue oaks growing the common gardens were not yet reproductive at the time of my surveys and the data I collected was thus only on vegetative shoot phenology, I show in chapter two that shoot emergence correlates almost perfectly with flowering phenology in these oaks.

Flowering synchrony has been shown to increase fecundity in oaks (Knapp et al. 2001, Koenig et al. 2014) and so it represents a potential target of sexual selection (Wilson and Burley 1983, Moore and Pannell 2011). Synchrony of spring phenology would thus be both the outcome of local population adaptation as well as a mechanism contributing to reduced gene flow between populations. Other mechanisms of selection, on survival and growth of individuals, are likely to still contribute as well and help explain the relationship that I found between historical climate and locally adapted phenology among the provenances.

Gene Flow and Reproductive Isolation

In chapter two, I turn from intraspecific to interspecific gene flow to study the patterns and processes of hybridization in two hybrid zones involving blue oak, with Oregon white oak (*Q. garryana*) in the north and Tucker's scrub oak (*Q. john-tuckeri*) in the south.

In this chapter, I focus more directly on the reproductive isolation that acts as a barrier to gene flow. Reproductive isolation is generally thought of as a series of separate barriers, which can be divided into mechanisms that are pre-zygotic or post-zygotic with reference to the production of offspring (Epling 1947, Coyne 1992, Moyle et al. 2004) or alternatively into ones with intrinsic or extrinsic causes (figure 2). Intrinsic reproductive barriers depend on genetic differences between populations, including several forms of genomic incompatibility discussed above, which can be either pre-zygotic or post-zygotic (developmental). On the other hand, extrinsic reproductive barriers are mediated by the environment and include the geographic separation between populations or individuals as well as natural selection against hybrid offspring. Extrinsic isolation is capable of enforcing boundaries between populations even if they have full intrinsic reproductive compatibility (Rice and Hostert 1993). Behavioral barriers such as reproductive timing or mate choice, may be either intrinsic (genetic) or extrinsic (environmental) or, perhaps most commonly, an interaction between the two (Wilson and Burley 1983).

Among these many potential barriers to gene flow, the ones that play important roles in any particular case of reproductive isolation will depend in part on the biogeographic history of the nascent lineages involved. The classical allopatric speciation model of divergence involves extrinsic geographic separation as the primary or initial barrier between populations. This extrinsic barrier is followed by a period of isolated genetic drift, which may secondarily lead to the evolution of degrees of intrinsic isolation (Mayr 1947, Endler 1977). If lineages that have diverged under the allopatric model come into geographic contact again before the evolution of intrinsic barriers has become complete, a hybrid zone can form at their geographic interface, referred to as a zone of secondary contact. In these cases, the process of reinforcement may strengthen the intrinsic isolation between nascent lineages as an adaptive response to reduced fitness of hybrid offspring (Dobzhansky 1937, Servedio and Noor 2003). Alternatively, the lineages coming into secondary contact may undergo fusion into a single indistinguishable population if gene flow does not reduce fitness. A third possible outcome in a secondary contact hybrid zone involves balancing selection between divergence dependent on ecological

differences and the potentially adaptive maintenance of gene flow (Van Valen 1976, Moore and Buchanan 1985, DiVittorio et al. 2021).

Alternatives to the allopatric speciation model describe the potential for lineages to diverge even without geographic separation. Populations with partial geographic overlap or even complete sympatry might still develop isolating mechanisms driven by extrinsic post-zygotic selection against intermediate genotypes (Smith 1966, Davies and Snaydon 1976, Devaux and Lande 2008). Gene flow in hybrid zones may also occur in cases of sympatric divergence, representing areas of primary and continuous contact between the lineages, maintained by spatially or temporally varying selection.

Any long term barrier to gene flow is expected to lead to identifiable genetic structure among populations (Wright 1931). For populations across a range of geographic separation, this structuring can follow a pattern of genetic isolation-by-distance (IBD) in which geographic distance predicts genetic dissimilarity (Wright 1943, Slatkin 1993, Ishida 2009). Isolation-by-distance is often treated as the null expectation of genetic divergence in landscape genetics. When populations occur along environmental gradients, a pattern of isolation-by-environment (IBE) may arise, primarily via natural selection on offspring fitness (Fisher 1950, Wang and Bradburd 2014). Isolation-by-distance and isolation-by-environment can coexist, but environmental difference, rather than geographic distance, may better characterize genetic dissimilarity in some cases, particularly when the patterns of environmental variation are complex and topographically replicated, in what has been called a 'mosaic hybrid zone' (Harrison 1986). The occurrence of hybrids in a mosaic zone will appear widespread when the area is viewed coarsely, leading to an expectation of isolation-by-distance, but finer scale characterization of the environmental and genetic structure will reveal that genotypes actually occupy distinct habitats and parental types remain largely distinct outside of relatively narrow hybrid bands.

Like geographic distance and environmental dissimilarity, behavioral isolating mechanisms can be characterized across a landscape and associated with genetic dissimilarity. In particular, isolating mechanisms related to the timing of reproductive activity have been studied. Isolation-by-time (IBT) has been developed as a metric analogous to isolation-by-distance and isolation-by-environment to characterize the effects of asynchrony in timing of reproduction on genetic structure (Hendry and Day 2005). Among angiosperm plants, flowering time is the most widely studied behavioral isolating mechanism and is widely found to depend on both genetic (intrinsic) and environmental (extrinsic) differences (Wilczek et al. 2010, Gaudinier and Blackman 2020).

In oaks, isolation-by-distance (IBD) results from the geographic separation between adult trees, thus requiring dispersal to achieve gene flow. As previously discussed, dispersal occurs at two distinct stages for oaks: pollen and acorns. Previous research suggests that dispersal distances in oaks, particularly in the case of pollen, may be large enough for isolation-by-distance to be identifiable at scales larger than several kilometers (Dow and Ashley 1996, Valbuena-Carabaña

et al. 2005, but see also Pluess et al. 2009 suggesting shorter dispersal distances). Isolation-by-time (IBT) in oaks is associated with flowering synchrony among individuals as well as between separate male and female flowers on the same or different trees. As I showed in the first chapter, onset of spring phenology in oaks, including flowering, responds to both extrinsic environmental influences (e.g. winter or spring temperatures) and intrinsic genetic factors. This result is also supported by previous research in other groups of oaks (Derory et al. 2010, Ramírez-Valiente et al. 2010). Finally, isolation-by-environment (IBE) will be generated by post-zygotic natural selection occurring across a series of demographic transitions beginning with acorn germination, through to the survival and reproduction of mature trees in their environment.

Gene Flow and Phylogenetic Inosculation

In chapter three, I focus on gene flow in the western North American white oaks as an entire clade. In this phylogenetic context, gene flow produces lineage reticulation or what may be called inosculation between the branches of a phylogenetic tree (figure 3). Gene flow at this scale is increasingly recognized as a widespread phenomenon across the entire tree of life. It has certainly been identified previously in oaks (Eaton et al. 2015, McVay et al 2017a, 2017b) as well as an ever-growing number of other taxonomic groups, including common examples in pines (Millar 1983, Menon et al. 2018), *Helianthus* sunflowers (Rieseberg 1991), bears (Miller et al. 2012, Cahill et al. 2018), canids (Hailer and Leonard 2008, VonHoldt et al. 2016), *Heliconius* butterflies (Mallet 2008), Darwin's finches (Grant and Grant 1996, Lamichhaney et al. 2017), and even at several points within our own genus *Homo* (Hammer et al. 2011, Racimo et al. 2015, Scerri et al. 2018).

Yet, despite its near ubiquity, gene flow between the branches of a phylogenetic tree fundamentally violates the most commonly used evolutionary models used to reconstruct trees, which are absolutely constrained to show only divergence between lineages. Network phylogenetic models have been developed to address this short-coming, such as the multispecies network coalescent (Wen and Nakhleh 2018) and species networks applying quartets (SNaQ) (Solís-Lemus and Ané 2016). However, because of their computational requirements, these methods are not tractable for genome-scale data sets, especially when a large number of tips are included. As a result, analytic methods that do not specifically infer a network phylogeny must be used. One option is to analyze the set of trees generated by traditional (divergence-only) phylogenetic models. Included here would be the CV of pairwise phylogenetic distances method I develop in chapter three as well as clustering and concordance methods (e.g. Hillis et al. 2005, Jombart et al. 2017). The phylogenetic invariants framework and derived D-statistics, which I also use in chapter three, are another option. These methods are particularly elegant in the way they distinguish gene flow between lineages from incomplete lineage sorting, which can otherwise produce similar patterns (Patterson et al. 2012, Kubatko and Chifman 2019).

Whatever methods are used to infer gene flow across a phylogeny, the evolutionary significance of that gene flow remains a separate question. Inosculation of a phylogenetic tree may sometimes

only be relevant to taxonomy and phylogenetic reconstruction. However, this scale of gene flow takes also on a central role as adaptive exchange in evolutionary models like the ecological multispecies concept developed by Van Valen (1976). The evidence I present in chapter three supports Van Valen's model conceptually by showing that hybridization occurs at many depths in the phylogeny. However, it does not demonstrate the adaptive outcome of these hybridization events, though the similar life history traits (e.g. tree vs. shrub growth form or evergreen vs. deciduous leaf habit) of the lineages that are involved is tantalizing. In any case, a clearer picture of the relationships among taxa, taking into account their history of gene flow even after divergence, is crucial as a first step toward understanding any adaptive significance of that gene flow.

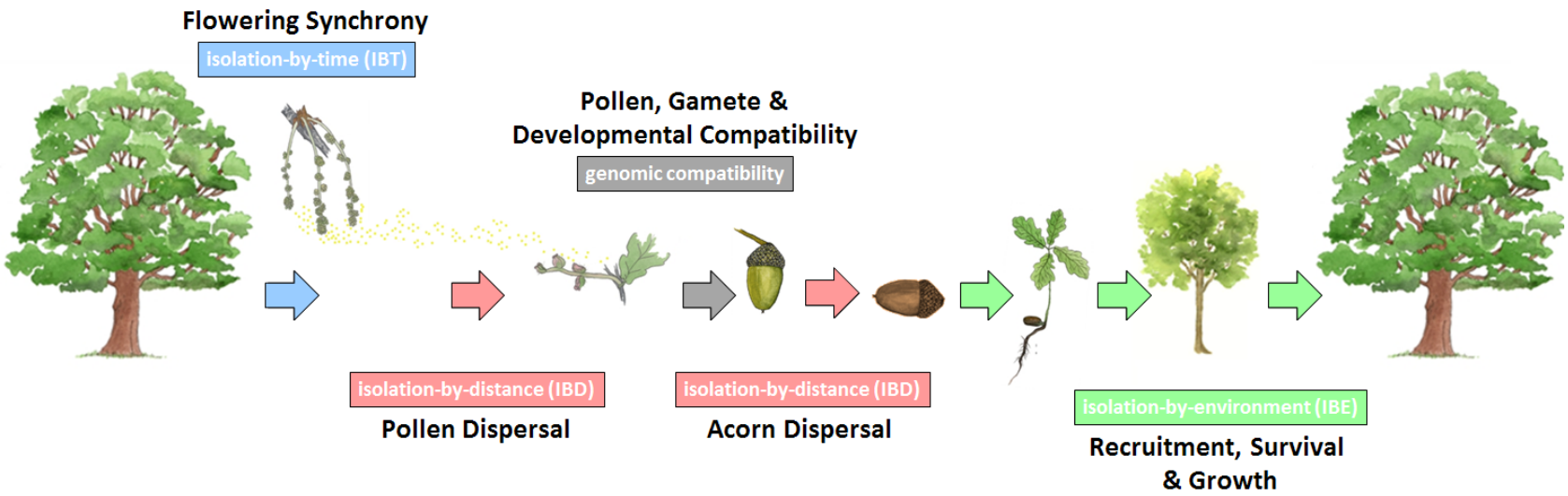


Figure 1: Stages in the life cycle of an oak, illustrating the timing and context of concepts discussed in this dissertation as sources of gene flow or reproductive isolation. Illustration courtesy of NorArte Studio (CC-BY).

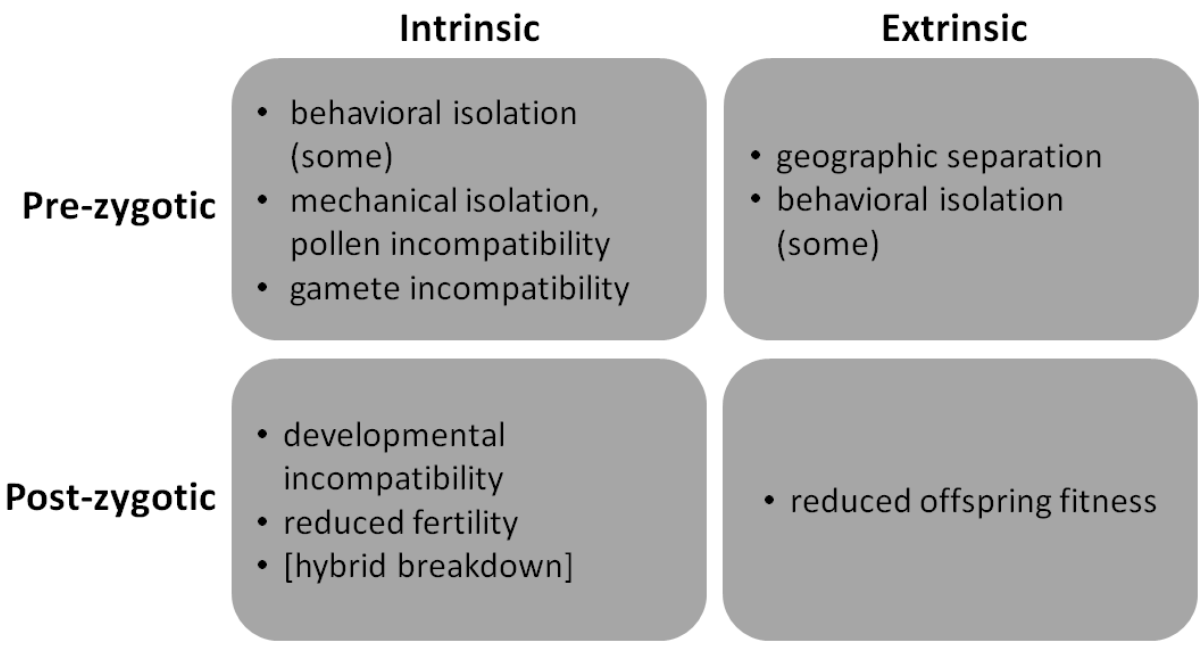


Figure 2: Division of reproductive isolation into pre-zygotic or post-zygotic and intrinsic or extrinsic mechanisms, with some examples of each.

Figure 3: Inosculation of tree branches, caused by damage followed by wound-healing and secondary growth, serves as an analogy for reticulate branches of a phylogenetic tree.

CHAPTER ONE

Partitioning genetic and environmental components of phenological variation in *Quercus douglasii* (Fagaceae)

ABSTRACT

Oaks (genus *Quercus*) often display a large range of phenotypic variation across many of their traits. The contribution of genetic and environmental sources, and their interaction, to this variation can be partitioned experimentally using common garden plantings in which several genotypes are grown in a single location and the phenotype of interest is measured. Due to their slow growth and complex genetic structure, oaks have rarely been grown to maturity in experimental conditions that would allow partitioning phenotypic variation in this traditional way. Here I present results from trees growing in two established experimental gardens of *Quercus douglasii* planted in 1991. Data from these common gardens are combined with additional data collected at the original provenance populations that served as acorn source locations. I surveyed phenological progression through the spring at both garden and field sites and found significant associations between fall and spring minimum temperatures and spring phenology, represented here as the date of bud break. A genetic component of phenological variation associated with the provenance sites was identified at both common gardens and accounts for 16.4% of the total variation observed among trees, while 68.2% of the variation can be attributed to environmental plasticity, plus a genetic × environmental interaction that accounts for about 1% of bud break variation. I discuss the implications of these components of phenological variation in blue oak, especially with respect to climate change, local adaptation, restoration, and assisted gene flow.

INTRODUCTION

California oaks show large variation in the timing of bud break and flowering across their range in any given year. This variation is commonly associated with both geographic differences among sites and local topographic variation within sites. Attempts to explain phenological variation in temperate trees tend to focus on analyses of weather variables such as temperature or precipitation, cumulative metrics such as growing-degrees, or geographic proxies such as latitude or elevation (Polgar and Primack 2011, Roberts et al. 2015). Among California oaks, previous research has shown that winter and spring temperature and precipitation are important drivers of spring phenology (Koenig et al. 2014, Gerst et al. 2017). Conflicting results among these studies concerning the effects of particular predictors suggest that phenological responses in these oaks may result from complex interactions, potentially integrated over more than one growing year. Additionally, the combination of multiple highly correlated weather variables, or especially weather and geographic variables, in a single phenological model makes identifying their independent effects even more difficult.

Temperature and the availability of water or sunlight prior to or early in the growing season are direct environmental drivers of phenology. Genetic phenological adaptation of populations to longer-term local conditions has often been found in previous research as well (Savolainen et al. 2007, Wilkinson et al. 2017, Dixit et al. 2020). Genetic differentiation among populations is expected in a variable environment when environmental, population, or community interactions that depend on phenology have fitness effects. The timing of spring phenology (bud break and flowering) has several potential fitness effects for trees, via pollen limitation (Knapp et al. 2001), freezing damage (Augspurger 2013), and herbivory (Pearse et al. 2015). Direct influences of resource availability in the environment on phenological timing are a plastic response that allows individuals to respond to variation from year to year, including responses to progressive directional changes associated with climate change. Genetic adaptation to historical climate, on the other hand, limits the potential range of responses to the immediate environment. These limits can be adaptive if they dampen responses to short-term anomalies, but they also constrain the potential to respond to directional change and can result in non-adaptive climate change responses and evolutionary lags (Browne et al. 2019).

Climate change simultaneously challenges California oaks with increased mortality from extreme events at the trailing edge of their distribution (Brown et al. 2018) and a potential dispersal lag at the leading edge that would prevent them from expanding into newly suitable locations (Serra-Diaz et al. 2014). The ability for oaks to respond to both of these challenges crucially depends on acorn production, which has been shown to be influenced by spring phenological timing, both as a result of flowering synchrony (Koenig et al. 2015) and herbivore pressure early in the season (Pearse et al. 2015). High fecundity in local populations at the trailing edge could preserve the adaptive potential of those populations against dieback due to extreme weather events by increasing genotypic diversity. Simultaneously, at the leading edge, greater acorn production can increase the rate of longer distance dispersal.

In the contexts of conservation and ecological restoration, climate change has given new impulse to debates over how to preserve the evolutionary potential of local populations while also encouraging migration into suitable habitats. Climate-informed assisted gene flow recommends sourcing seeds for restoration from sites that have historical climates more similar to the restoration site's predicted future climate, on the assumption that local adaptation to the historical location will make them better adapted in the new location compared to locally sourced seed (Aitken and Whitlock 2013). More recently, genome-informed assisted gene flow has also been suggested, which involves selecting seed sources based on empirically confirmed relationships between source genotypes and fitness measures such as growth rate or reproductive success at a target site (Browne et al. 2019). The restoration community has responded with skepticism to the suggestions for assisted gene flow due to a long-established preference for local seed sources in order to avoid disrupting existing patterns of local adaptation (McKay et al. 2005, Vander Mijnsbrugge et al. 2010). A lack of information concerning how much local adaptation there is among oaks, what the natural rates of gene flow are, and whether assisted gene flow is likely to disrupt locally adapted genotypes has left restoration specialists uncertain how to best proceed with restoration of degraded oak ecosystems in light of climate change (Spotswood et al. 2017, Baumgarten et al. 2020).

Reciprocal transplant and common garden experiments, in which individuals from several different local populations are grown and observed in one or more common environment, provide the clearest ways to address questions about the relative importance of local environment and genetic adaptation on phenotype (Clausen, Keck, and Hiesey 1941, Howe et al. 2003). Here I present results from established common garden plantings of the California endemic blue oak (*Quercus douglasii*), along with observations of adult trees at the original acorn provenance field sites sampled for the common gardens. Using spring bud break phenology as a readily observed measure of phenotypic differentiation, I examine the relative roles of local environment and genetics on the timing of bud break. I also investigate the roles of local weather at the garden sites and climate history at the provenance locations to identify how patterns of temperature and precipitation variation affect bud break phenology in blue oaks over both long and short timescales.

METHODS

Acorns for the common garden plantings were originally collected in 1990 at 26 provenance sites spanning the range of *Quercus douglasii* Hook. and Arn. (blue oak), including the species' entire latitudinal range as well as five transects across its elevational range in the Sierra Nevada (figure 1, table S1). Acorns were collected from 10 maternal trees at each provenance site and grown together for the first year at a greenhouse location in Magalia, CA. In 1992, the one-year-old seedlings were then planted out at two common gardens, one in Mendocino County at the Hopland UC Research & Extension Center (N39.0137, W123.0959) and the other in Yuba County at the Sierra Foothills UC Research & Extension Center (N39.2467, W121.3173). The plantings followed a randomized complete block design consisting of ten blocks, each with nine

sub-replicate seedlings from each of the 26 provenance sites. Plots were weeded, but not irrigated during the first two years of seedling growth and were protected by deer fences. In 2001-2002, six of the ten original blocks at each garden were thinned from nine to three sub-replicates per provenance. For additional details on the establishment of the garden plantings, refer to McBride et al. (1997). Trees were more than 25 years old at the time of this study and only trees growing in the thinned blocks were surveyed as growth in the blocks that had not been thinned was clearly reduced due to crowding. Mean tree height across all populations in the thinned blocks at Hopland was 3.79 m (sd 1.15 m) and at Sierra Foothills was 3.67 m (sd 0.93 m). Of the 468 trees originally planted in the six survey blocks at each garden, 13 had died by 2017 at Hopland and 105 had died Sierra Foothills. Despite their age and size, flower production was only sporadically seen on some trees over several years of observation.

In addition to the common gardens, 21 of the original 26 acorn provenance sites were re-identified and surveyed for this study in 2017. In 2018, three additional provenance sites were identified for a total of 24. The exact original maternal trees could not be located using the collection records, but the sites selected to represent the provenance populations are within 4 km of the original collection trees and +/-50 m elevation. At each provenance field site, 7 to 12 trees at least the same size as the trees currently growing in the common gardens were selected for phenological observation.

Spring phenology was surveyed on each tree at the common garden and provenance sites in 2017 and 2018 using a bud break index that ranged from 0 to 8 (see table S3 for descriptions of each index stage). Common garden sites were surveyed every 7 to 14 days between mid-February and late-April. Provenance field sites were surveyed a minimum of two times during the same period.

The phenological index was designed to show a linear progression during leaf out across index values 1-7. This was confirmed using common garden trees surveyed at high frequency. As a result, a linear model could be fitted to the periodic survey data and used to interpolate a common phenological stage for all trees even though they were not all surveyed at the same stage. Phenological index value 3 (bud break) is used as the common stage for all analyses reported here.

Using the phenological variation observed in the common garden plantings, I ran two-way ANOVA (R 3.5.2, R Core Team) to partition the phenotypic variation in the date of bud break into separate environmental and genetic components. Common garden location × observation year (four levels in total) represents the environmental component while acorn provenance location (26 levels) represents the genetic component. The interaction between them is a genotype-by-environment (G × E) effect and was included in the model.

Next, combining phenological data from both common gardens and the 24 provenance field sites, I sought to explain the environmental and genetic components of variation in terms of weather and climate variations using linear regression models. For the environmental

component, weather variables at the individual growing location (common gardens or field sites) were used from the current water year, which runs from October of the previous year to September of the growth year in order to capture precipitation patterns in California's Mediterranean-type climate. To identify the genetic component explained by adaptation to local climate, I used 30-year climate means from 1951-1980 (the 20th century climate baseline) at the provenance collection sites.

Gridded weather and climate data were extracted from the Basin Characterization Model (BCM) for California, which models weather and hydrology using the PRISM weather models topographically downscaled to 270 m resolution (Flint et al. 2013). Monthly estimates of minimum and maximum temperature and precipitation (36 total raster layers) were used for both the study period 2017-2018 as well as the climate baseline 1951-1980 (table S2). To maximize explanatory power (adjusted R^2) while reducing multicollinearity among the 72 weather and climate layer, I used automated AIC-based stepwise model selection (function stepAIC in the R package MASS). This was followed up by manually removing individual remaining variables to reduce multicollinearity (as measured by the variable inflation factor, VIF) and highlight important variables.

RESULTS

More phenological variation was found across the provenance field sites than was found within either of the gardens. Mean date of bud break for populations at the provenance field sites in 2017 spanned 32 days from March 6 to April 7 in 2017 and 78 days in 2018 from February 18 to May 7. At the Sierra Foothills common garden, the mean date of bud break among the provenance populations spanned 19 days from February 26 to March 17 in 2017 and 35 days from February 10 to March 17 in 2018. At the Hopland common garden, mean date of bud break spanned 23 days from March 12 to April 4 in 2017 and 25 days from March 15 to April 9 in 2018. The longer period covered by bud break in 2018 compared to 2017, especially in the Sierra Foothills common garden and provenance sites located in the Sierra Nevada foothills, may be associated with a cold snap with freezing temperatures in early-March of 2018. When data from the two common gardens are combined, the variation in bud break timing in the gardens covered approximately the same range of dates as the field provenance sites (figure 2).

In addition to the difference in bud break timing between the Hopland and Sierra Foothills common gardens and some differences between the 2017 and 2018 seasons, consistent phenological variation associated with individual provenance populations is readily identifiable in the common garden trees. Reaction norms plotted as mean bud break of trees from each provenance between the two garden locations show little difference in slope (lines are parallel) for the 26 provenance populations, indicating a consistent genetic effect of provenance site and similar degrees of plastic response across all of them (figure 3). This is confirmed by ANOVA models, which identify major components of phenological variation associated with both environment (two garden locations across two years for four categories) ($p << 0.0001$) and

genetics (acorn provenance site) ($p << 0.0001$) for the trees growing in the common gardens, as well as a significant but very small interaction effect between environment and genetics ($p = 0.001$). Omega-squared (ω^2) calculated from the sums of squares in the ANOVA model shows 68.2% of the observed phenological variation at the common gardens can be associated with environment (garden location × observation year) and 16.4% is associated with genetics (provenance location) while only about 1% of the observed variation is a G × E interaction (table 1).

Using stepwise model selection on linear regressions with 36 weather variables for the 2017-2018 water years (monthly minimum temperature, monthly maximum temperature, and monthly precipitation), I found that only a handful of these were needed to explain a large portion of the observed phenological variation. First, for bud break phenology observed at the provenance field sites, the full set of water year 2017-2018 weather variables explains 59.5% of the variation ($p << 0.0001$). Fall and spring monthly minimum temperatures have the most important effect in the models. November minimum temperature and March minimum temperature in the current water year alone explain 33.3% of the observed phenological variation at the provenance sites in 2017 and 2018 ($p << 0.0001$).

The single variable November minimum temperature of the current water year (i.e. prior calendar year) was particularly powerful in explaining observed bud break variation, across both provenance field sites and the two common gardens, with an R^2 of 42.0% across all sites in 2017 and 2018 and an effect size of -10.5 days for every degree-Celsius ($p << 0.0001$) (figures 4A and 5A). Multicollinearity between weather variables is common, though, and despite variance inflation factor (VIF) between November and March minimum temperature of < 2.0, indicating very low collinearity, parameter estimates of the effect size may still be unreliable. What is clear, though, is that higher minimum temperatures, especially in fall and spring months, advanced the timing of bud break at these sites.

The remaining phenological variation observed among trees at the common gardens, which cannot be explained by the current year weather they experienced, is partly explained by the climate of the acorn provenance sites sampled to plant them. The same 36 monthly temperature and precipitation variables, but as 30-year climate mean from 1951-1980 rather than the current water year, in a combined model explain (adjusted R^2) 12.9% of the variation observed among the common garden trees ($p << 0.0001$). Multicollinearity was again a problem for these climate-based analyses, but maximum temperatures appear stronger in the climate effect. March maximum temperature for 1951-1980 was the single best predictor from the provenance populations of bud break in the common garden locations, although it explains only 2.4% of the observed variation with a fairly consistent effect size of one day of bud break advancement for every degree-Celsius ($p = 0.0004$) (figures 4B and 5B). In addition, 1951-1980 average March and September precipitation at the provenance sites were also identified as significant predictors, associated with later bud break in the common gardens. In combined models with March

maximum temperature, all three climate variables were able to explain a total of 4.8% of phenological variation in the common gardens.

Finally, these findings of genetic effects associated with provenance climate in the common gardens were applied back to both the garden and provenance field sites by including both the current water year weather variables (November and March minimum temperature) and the 1951-1980 30-year mean of March maximum temperature and March and September precipitation in a combined model (table 2). Warmer November and March minimum temperatures experienced in the current water year advance bud break date by about 8 and 3 days per degree-Celsius, respectively, while warmer March maximum temperatures in the historical climate advance bud break by an additional day per degree-Celsius. These variables together explain 51.5% of the total observed phenological variation across all sites ($p << 0.0001$). However, previous warnings about the effects multicollinearity among variables should still be applied here, in spite of VIF < 2.0.

DISCUSSION

Using an established common garden experiment planted from acorn collections across the range of *Quercus douglasii* (blue oak), with trees growing at two locations and more than 25 years old at the time of the study, I found significant effects of both environment (common garden location and year) and population level genetics (acorn provenance site) on the timing of spring bud break (table 1). The genetic × environmental interaction effect in the model is significant, but very weak, meaning that individuals responded uniformly based on their genetic background and the local environment. The remaining residual variation, similar in size to the provenance genetic effect, may be largely due to individual genetic variation among trees from the same provenance, though the effect of microsite soil and moisture variability within the common gardens is likely to be important as well. A large role for environmentally-driven phenotypic plasticity in phenology is expected (García-Mozo et al. 2002, Parmesan and Yohe 2003, Vitasse et al. 2009) while the significant genetic effect I found demonstrates a role of local evolutionary forces in shaping the phenological timing of blue oak as well.

Previous studies of oaks in both North America and Europe have reported high within population phenotypic variation across a range of traits and while a consistent local genetic effect can sometimes also be identified, its importance varies widely (Derory et al. 2010, Ramírez-Valiente et al. 2010, Cavender-Bares and Ramírez-Valiente 2017). Previous results even from this same common garden experiment illustrate how the magnitude of genetic effects may depend on what aspect of phenotype is studied. Shoot growth over the first four years in the common gardens was found to be strongly influenced by provenance site (McBride et al. 1997), but more recent ecophysiological work on the same trees found no variation among provenances in their susceptibility to vascular embolism (Skelton et al. 2019).

In my study, the environmental component of phenological variation across all sites could largely be explained by minimum temperatures in the months preceding bud break, especially November and March (figures 4A and 5A), while the genetic component, as revealed in common garden conditions, was in part associated with climate history at the provenance locations. A composite climate effect in regression models with only common garden trees explained 12.9% of the bud break variation, which compares well with the 16.4% explained by provenance site as a categorical variable in ANOVA. No single monthly climate variable was found to explain a major portion of this genetic effect, but March maximum temperature was consistently significant, though alone it only explained a small amount of the observed phenological variation (figures 4B and 5B).

The magnitude of the effect of temperature forcing on spring phenology I report in these oaks (table 2) is similar to figures that have previously been reported for oaks in the Mediterranean region (Sanz-Pérez et al. 2009) and northern European temperate forest trees, including *Quercus robur* (Roberts et al. 2015). Between 5 and 10 days of phenological advance per degree-Celsius of temperature increase is commonly found. However, I must note again that, despite VIF in all my models < 2.0 (indicating very low multicollinearity), the parameter estimates remained sensitive to exactly which monthly temperature and precipitation variables were included together in a model. As a result, I do not believe that the parameter estimates for the effects of November or March minimum temperature in my models and reported here should be considered definitive. Instead, the responses I highlight involving these particular months point to a more general importance of low temperature, in both fall and spring, as an environmental influence on phenological timing in blue oaks. A summary statistic like monthly minimum temperature is ultimately not likely to be the direct environmental factor influencing the timing of phenology. Similar to geographic proxies that are often used (e.g., elevation or latitude), monthly minimum temperatures are correlated with mechanistically proximate factors. Variable selection and parameter estimates in models will depend on the strength and interactions involved in those correlations.

I did not find that winter cold temperatures advanced the timing of bud break (i.e. chilling hours). In fact, the relationship between winter temperature and spring phenology was negative, so that colder temperatures in winter delayed bud break just as they did in spring. This finding agrees with previously reported results in European oaks from both Mediterranean and continental climates (García-Mozo et al. 2002, Roberts et al. 2015, Wilkinson et al. 2017). Oaks do not seem to require chilling-hour accumulation during dormancy to set buds as seen for some other temperate forest trees (e.g., birch or hawthorn). Instead, temperature has a uniform forcing effect, with warmer temperature advancing bud break throughout the entire dormant season.

My results suggest that the debate over assisted gene flow versus using local seed sources should not in fact be a major concern in the case of blue oaks, at least when considering spring phenology with its important implications for pollination success, freeze-damage, productivity, and herbivory. While a significant signal of local adaptation in phenological timing was

observed, there was a large amount of variation among trees collected from the same provenance site, often encompassing almost the total variation observed across all provenance sites (figures 4B and 5B, open shapes). Thus, while climate-informed assisted gene flow seems unlikely to disrupt patterns of local adaptation, it also may not be particularly beneficial in the case of blue oaks. As I discuss in chapter 2 of this dissertation, local populations of blue oak harbor large amounts of genetic variation, a finding that has previously been made as well (Rice et al. 1997). This within-population genetic variation, together with the microsite variation within both gardens and provenance sites, likely explains much of the residual phenological variation that was not explained in this study. If the local phenological variation within sites is comparable to variation across the species' entire range, as it seems to be, that existing local variation already provides the potential to allow populations to track changing climate and assisted gene flow would add little to their adaptive potential. This agrees with results found in other oaks, for which within population variation has been suggested as important for seedling establishment (González-Rodríguez et al. 2012). The critical question is whether cycles of seedling recruitment, maturation and mortality will be able to keep pace with the rate of changing climate. This process can be aided by restoration efforts, with or without involving assisted gene flow. For restoration purposes, the most important recommendation may simply be to plant more acorns, in order to better capture the adaptive potential of whatever population is chosen as a seed source. The advantage of assisted gene flow should be seen as only a somewhat increased probability of planting well-adapted genotypes, in which case fewer acorns may be necessary. This could warrant consideration in terms of cost- and labor-saving in restoration projects.

By building on the legacy of the two established common garden plantings, combined with new field work tracking phenological activity and high-resolution monthly temperature and precipitation model data from PRISM/BCM, I have been able to tease out a significant component of genetic phenological adaptation in blue oaks and associate it with the temperature and precipitation history of the provenance sites that provided acorns for the common garden plantings. This study illustrates a powerful combination of data sources with the potential to be replicated in many other cases. A rich legacy of mature common garden plantings, involving a variety of woody species, can be found languishing, largely unknown, across California and outside the state. The availability of modeled climate data has the potential to make these gardens very valuable if the seed source location are or can be established, as they were in this case. Broad efforts should be made to document and catalog historical common garden plantings such as these and collate whatever data is available about them, so that they can be utilized in a new generation of research informing climate adaptation strategies.

TABLES AND FIGURES

Table 1: ANOVA partition of phenological variation among oak trees growing at two common garden locations into components explained by provenance site (genotype, G), garden location×observation year (environment, E) and the interaction term (G×E).

	Df	SS	F-statistic	p-value	ω^2
G	25	17203	29.314	2.136×10^{-82}	0.1642
E	3	69091	981.116	1.483×10^{-211}	0.6822
G×E	75	2859	1.624	0.00143	0.0109
Residual	511	11995			0.1427

Table 2: Multiple regression results of selected climate and weather variables on date of bud break across both gardens and all provenance field sites for 2017 and 2018. Historical climate (1951-1980) at the acorn provenance site represents a genetic component of variation in this model while the water year variables represent an environmental component. Adjusted $R^2 = 51.45\%$, $p << 0.0001$.

	Parameter Estimate	Standard Error	t-value	p-value
March max temp (1951-1980)	-1.0646	0.2291	-5.646	3.845×10^{-6}
March precip (1951-1980)	0.0586	0.0147	3.868	1.169×10^{-4}
September precip (1951-1980)	0.4700	0.0754	6.234	6.744×10^{-10}
November min temp (water year)	-8.0369	0.4142	-18.963	1.461×10^{-68}
March min temp (water year)	-2.7827	0.2653	-10.525	1.283×10^{-24}

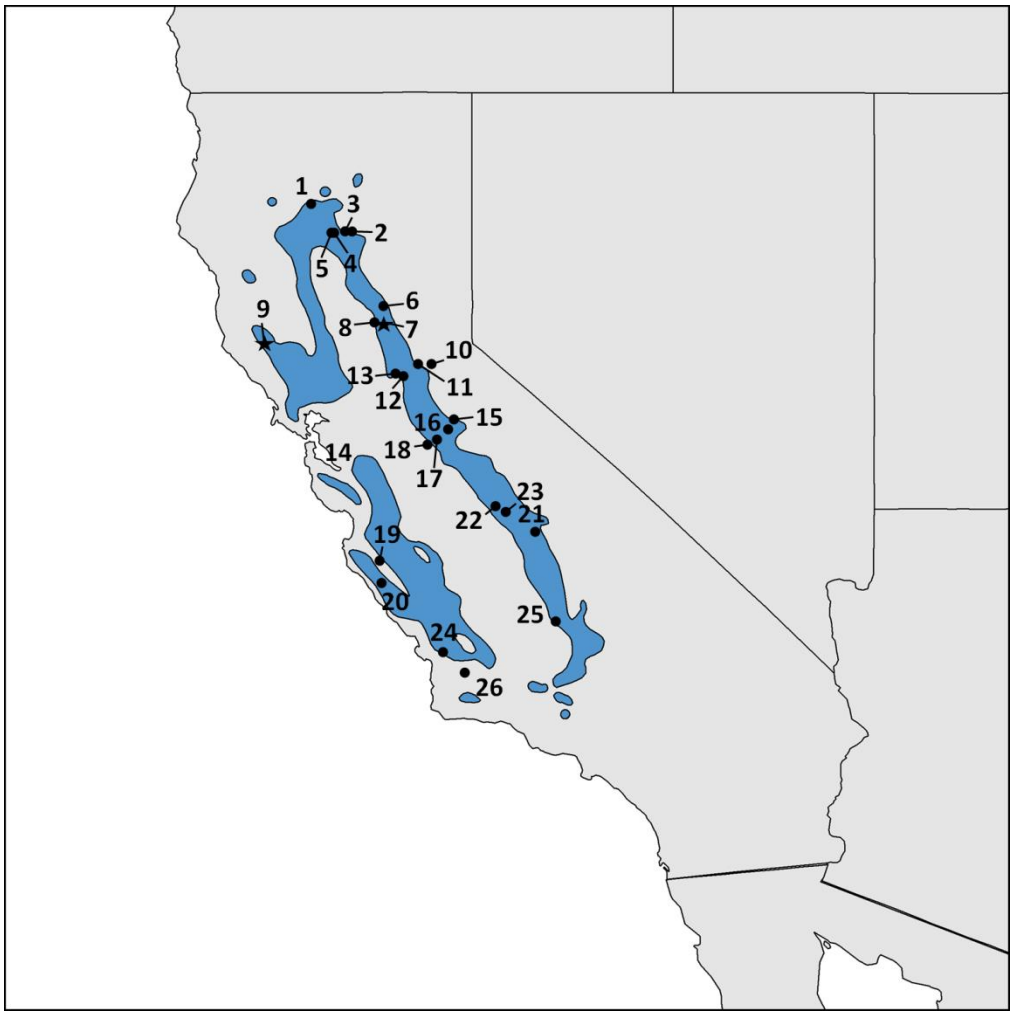


Fig. 1: The distribution of *Quercus douglasii* in California. The 26 acorn provenance locations and field survey sites used in this study are shown with black markers. Two common gardens are planted close to provenance 9 (Hopland) and provenance 7 (Sierra Foothills), indicated with stars on the map.

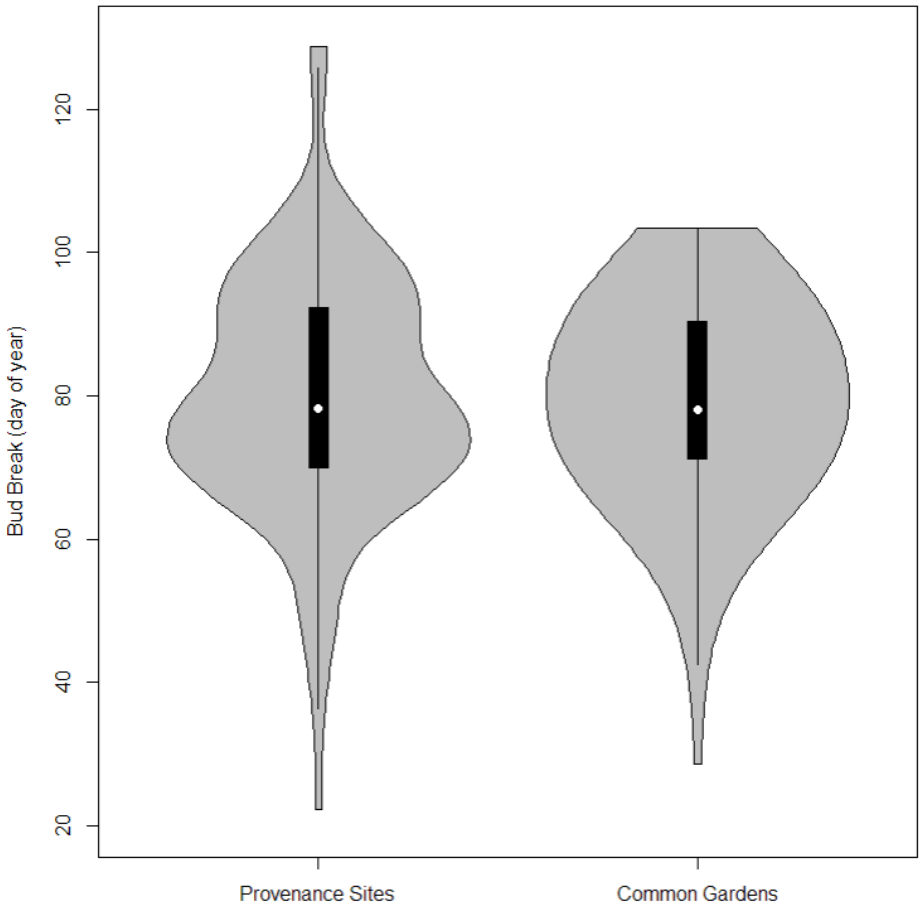


Fig. 2: Variation in the date of bud break in *Quercus douglasii* observed in 2017 and 2018. Wild trees growing at 24 field sites throughout the species' range (left) compared to study trees growing at two common gardens (right). Field sites served in 1990 as the original acorn provenances for the common garden plantings.

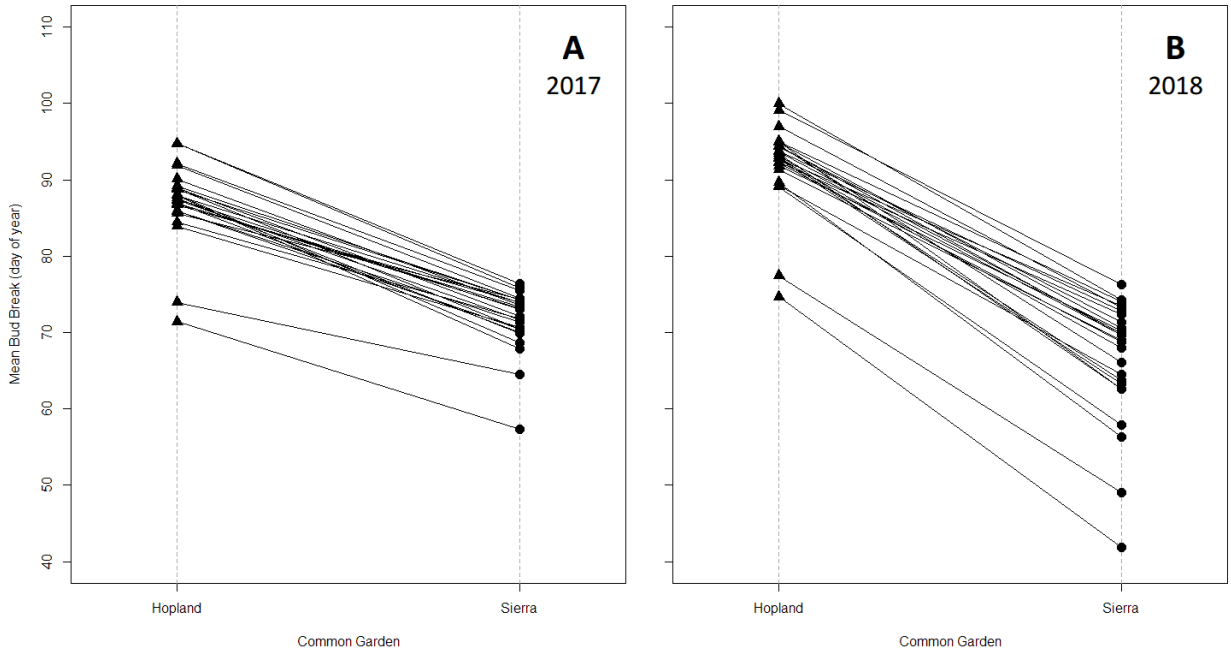


Fig. 3: Mean date of bud break in 2017 (A) and 2018 (B) for 26 provenance populations of *Quercus douglasii* growing at two common garden locations.

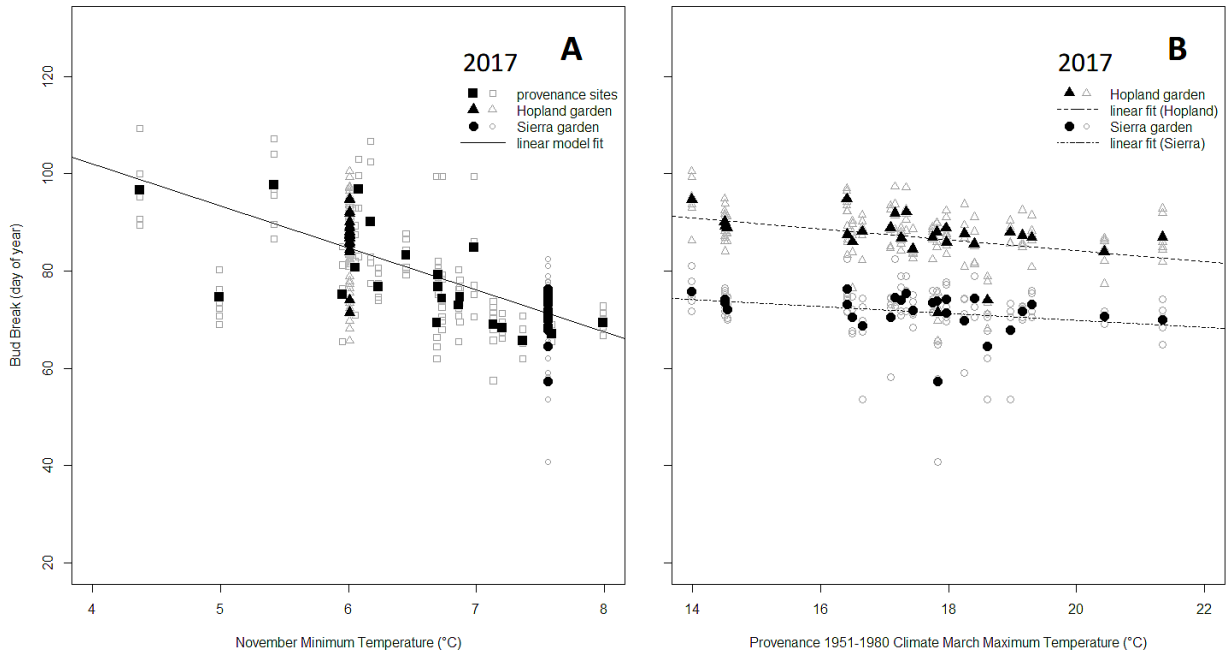


Fig. 4: 2017 bud break phenology for *Quercus douglasii*. (A) Bud break for all surveyed trees across two common gardens and 21 provenance field sites as a response to minimum temperature the previous November. (B) Bud break in the two common gardens (Hopland and Sierra Foothills) as a response to 1951-1980 climate March maximum temperature at the original acorn collection provenance sites. Open shapes in grey represent individual trees, while closed shapes are population means.

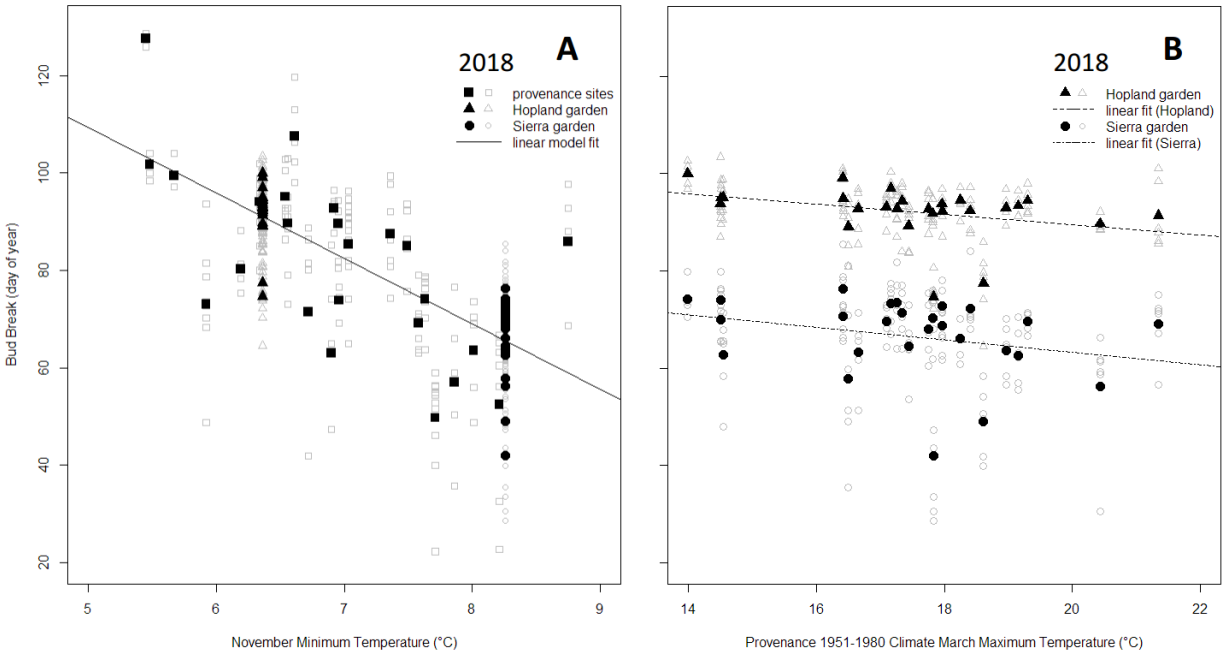


Fig. 5: 2018 bud break phenology for *Quercus douglasii*. (A) Bud break for all surveyed trees across two common gardens and 24 provenance field sites as a response to minimum temperature the previous November. (B) Bud break in the two common gardens (Hopland and Sierra Foothills) as a response to 1951-1980 climate March maximum temperature at the original acorn collection provenance sites. Open shapes in grey represent individual trees, while closed shapes are population means.

SUPPLEMENTAL MATERIAL

Table S1: Latitude, longitude, and elevation of the 26 acorn provenance sites used for the common garden planting and surveyed for this study. Acorn provenances marked with an asterisk could not be re-identified from the original records, but the coordinates and elevation noted at the time of collection are given in parentheses.

	Latitude	Longitude	Elevation (m)
Hopland garden	39.014	-123.096	277
Sierra Foothills garden	39.247	-121.317	198
Provenance 1	40.666	-122.372	230
Provenance 2	40.346	-121.825	870
Provenance 3	40.347	-121.879	658
Provenance 4	40.313	-122.024	263
Provenance 5	40.301	-122.098	203
Provenance 6	39.452	-121.283	631
Provenance 7	39.265	-121.302	405
Provenance 8	39.267	-121.413	108
Provenance 9	39.017	-123.091	387
Provenance 10	(38.77)	(-120.63)	(1010)
Provenance 11	(38.77)	(-120.78)	(435)
Provenance 12	38.648	-121.028	341
Provenance 13	38.663	-121.099	186
Provenance 14	37.915	-121.892	238
Provenance 15	38.117	-120.242	879
Provenance 16	37.982	-120.368	638
Provenance 17	37.847	-120.56	303
Provenance 18	37.816	-120.655	104
Provenance 19	36.280	-121.322	110
Provenance 20	36.073	-121.369	409
Provenance 21	36.752	-119.132	610
Provenance 22	37.102	-119.743	348
Provenance 23	37.006	-119.638	150
Provenance 24	35.287	-120.441	491
Provenance 25	35.707	-118.774	923
Provenance 26	35.087	-120.066	610

Table S2: Selected weather and climate variables from the California Basin Characterization Model for common garden and provenance survey locations. All temperatures are reported as degrees-Celsius. "n.d." indicates provenance sites that were not surveyed in that year.

	Water Year 2017		Water Year 2018		1951-1980
	Nov minimum temp	March minimum temp	Nov minimum temp	March minimum temp	March maximum temp
Hopland garden	7.56	7.59	8.26	5.51	(18.24)
Sierra Foothills garden	6.01	5.23	6.36	3.57	(18.20)
Provenance 1	6.44	6.34	6.55	3.98	17.95
Provenance 2	5.42	4.92	5.48	2.24	13.98
Provenance 3	6.05	5.76	6.19	3.20	16.40
Provenance 4	6.69	6.63	6.96	4.42	17.75
Provenance 5	6.69	6.59	6.94	4.46	18.39
Provenance 6	6.07	6.09	6.61	3.47	16.50
Provenance 7	7.13	6.94	7.71	4.78	17.82
Provenance 8	7.36	7.61	8.21	5.50	18.61
Provenance 9	6.17	5.32	6.34	3.61	17.15
Provenance 10	(n.d.)	(n.d.)	(n.d.)	(n.d.)	14.51
Provenance 11	(n.d.)	(n.d.)	(n.d.)	(n.d.)	16.64
Provenance 12	6.86	6.46	7.57	4.59	17.09
Provenance 13	7.19	7.09	8.01	5.15	17.81
Provenance 14	(n.d.)	(n.d.)	7.86	5.48	17.44
Provenance 15	(n.d.)	(n.d.)	5.44	2.11	14.53
Provenance 16	4.98	4.65	5.92	2.77	16.42
Provenance 17	6.86	6.42	7.48	4.67	17.96
Provenance 18	7.59	6.88	7.63	5.30	18.25
Provenance 19	(n.d.)	(n.d.)	6.71	4.67	21.34
Provenance 20	6.23	5.30	6.53	3.97	20.43
Provenance 21	6.98	6.15	7.36	4.53	17.34
Provenance 22	6.73	6.23	7.03	5.09	19.29
Provenance 23	6.69	6.11	6.90	4.92	19.14
Provenance 24	5.94	5.46	6.92	3.97	18.95
Provenance 25	4.36	4.57	5.67	2.90	14.50
Provenance 26	7.98	6.51	8.75	5.03	17.26

Table S3: Description of the phenological index used to track spring leaf phenology of blue oaks in this study. Index values 0 and 8 were not used for analysis because they last for the entire off or growing season, respectively.

Index	Description
0	Overwintering buds, no swelling is visible yet. These are often dark because of fungal growth on the bud scales.
1	Swelling buds. As buds begin to swell, the lighter brown parts of the bud scale that were protected from mold become visible.
2	Some green is showing at the edges of the bud scales, but leaf tips haven't emerged yet.
3	Breaking buds. Leaf tips begin to emerge from the buds, but the petiole is not visible and internodes are not elongated.
4	Stem internodes begin to extend, but the leaves are not fully unfolded and the petioles are not easily visible.
5	Young leaves 1. Youngest fully unfolded leaves. In blue oaks these are often about 1cm long.
6	Young leaves 2. In blue oaks these leaves are often about 2cm long.
7	Young leaves 3. They have reached their full size, but are still lighter green and softer in texture than mature leaves.
8	Mature leaves. They are full size, mature blue-green color, and leathery texture. This stage persists for the remainder of the growing season.

CHAPTER TWO

Evidence for reproductive isolation due to environment or flowering time, but not geographic distance, in two oak hybrid zones

ABSTRACT

Interspecific hybridization most commonly occurs in specific geographic areas, called hybrid zones, where the ranges of hybridizing species overlap. The evolutionary dynamics within hybrid zones will serve to either limit or enable the exchange of genes between lineages and thus define the influence hybridization has on them. These dynamics include several potential isolating mechanisms that limit the flow of genes, including geographic distance, environmental selection, and reproductive behavior. In this study, I investigate hybridization and reproductive isolation within two contrasting and geographically distinct hybrid zones, involving three species of oak (genus *Quercus*). I first characterize pairwise genetic dissimilarity between individuals across the both hybrid zones using a set of 4,790 genome-wide, unlinked SNPs derived from double-digest RADseq. I then use multiple matrix regression to explain this dissimilarity using a series of important landscape-scale isolating mechanisms and partition their effects. I find that geographic distance, as isolation-by-distance (IBD), is not a significant factor in either hybrid zone. On the other hand, climate variation, as isolation-by-environment (IBE), is significant in both hybrid zones including effects associated with both summer temperatures/aridity and winter temperature. In addition, I find a strong effect of phenological asynchrony, as isolation by time (IBT), on genetic dissimilarity, but only in one of the two hybrid zones. The ways in which these patterns may reflect the differing phylogenetic relationships and historical biogeography of the species in each of the hybrid zones are discussed, as are some potential evolutionary outcomes of gene flow within the zones.

INTRODUCTION

Reproductive isolation between populations is a measure of realized gene flow that varies from panmixis to complete isolation, including a full range of partial isolation between them. Intermediate levels of reproductive isolation represent a reversible "grey zone" of lineage divergence and highlight that speciation is a process rather than an event (Dobzhansky 1940). If lineages have become diverged enough that they are treated as separate taxa, gene flow between them may be termed hybridization, but degrees of reproductive isolation occurring at all evolutionary scales show similar patterns and can be analyzed as a continuous process (Hewitt 1988). Identifying and disentangling the mechanisms of reproductive isolation that maintain at least a semblance of lineage identity in spite of this kind of hybridization have been important goals of evolutionary biology and ecology.

Reproductive isolating mechanisms are often divided into pre-zygotic and post-zygotic (Epling 1947, Coyne 1992, Moyle et al. 2004). This may also simply be viewed as an ordered life history sequence centered on reproduction, from geographic separation between individuals to behavioral mating isolation to gametic and developmental incompatibilities to reduced hybrid fitness (see additional discussion in the introduction to this dissertation). This same series of isolating mechanisms may also be divided into intrinsic versus extrinsic mechanisms, depending on whether they have a genetic or environmental basis.

Many of these traditional conceptions of reproductive isolation have been incorporated into landscape genetic studies of genetic dissimilarity and structure via measures based on the model of the earlier isolation-by-distance (IBD) (Wright 1943, Slatkin 1993, Ishida 2009). In particular, isolation-by-environment (IBE) may explain genetic dissimilarity in space, primarily as a result of natural selection on offspring fitness in contrasting environments (Fisher 1950, Wang and Bradburd 2014). Isolation-by-time (IBT) has also been developed as a metric analogous to IBD and IBE to characterize asynchrony in the timing of reproductive behavior as an isolating mechanism (Hendry and Day 2005). Among plants, flowering time is the most widely studied such behavioral isolating mechanism and is widely found to depend on both genetic and environmental differences (Wilczek et al. 2010, Gaudinier and Blackman 2020).

Building on the idea that several mechanisms of reproductive isolation may often act in combination to determine the rate of gene flow between particular populations (Coyne and Orr 1989), several conceptual frameworks have been developed to distinguish their interacting effects in the context of landscape genetics. Geographic distance tends to be correlated with environmental difference at multiple scales, a widely recognized issue in ecological data (Legendre et al. 2005). In the context of landscape genetics, this has been addressed through partitions of IBD and IBE (Wang 2013, Wang and Bradburd 2014). Other analyses have taken this partitioning approach further, by also separating out additional effects such as dispersal limitation or colonization history (Orsini et al. 2013, Nadeau et al. 2016). Because behavioral reproductive isolation, including isolation-by-time, may be associated with both intrinsic genetic

factors and extrinsic environmental factors, IBT can be partitioned from IBE in a similar way, to disentangle these influences (Deacon and Cavender-Bares 2015, Suni and Whiteley 2015).

Here, I investigate mechanisms of reproductive isolation in three well-recognized species of oak, across two geographically separated hybrid zones, in the Western North American white oak clade (*Quercus* series *Dumosae*). Several potential mechanisms of reproductive isolation are included in the analysis in order to determine the major evolutionary and biogeographic processes occurring within each of the hybrid zones and to compare between them. I test whether geographic isolation-by-distance (IBD) is present as a primary or contributing factor maintaining species boundaries across the hybrid zones. In particular, by partitioning geographic IBD and isolation-by-environment (IBE) along climatic gradients, I ask whether local selection acting in contrasting environments may be an important constraint on gene flow in the hybrid zones. Finally, by partitioning isolation-by-time (IBT), dependent on asynchrony in flowering phenology, into two separate components, one associated with IBE and the other independent of it, I explore the extent to which the effects of reproductive timing on genetic structure in these species is due to intrinsic genetic factors versus extrinsic environmental factors.

METHODS

Study System

The Western North American white oaks are a recently confirmed monophyletic group within the white oaks (genus *Quercus*, section *Quercus*, series *Dumosae*), most commonly delineated as nine species-level taxa, all of which have the potential to hybridize (Hipp et al. 2017). There are both evergreen and deciduous taxa within the clade, as well as both trees and shrubs. The recognized taxa span a broad environmental gradient from the arid Mojave Desert to the wet Pacific Northwest. Blue oak (*Quercus douglasii* Hook. & Arn.) is endemic to Mediterranean-climate California, with a range near the geographic and environmental center of the clade. At opposite ends of its range, blue oak is sympatric with other taxa in the clade, producing hybrid zones with named hybrids (figure 1). In the northern hybrid zone, blue oak overlaps with Oregon white oak (*Quercus garryana* Hook.), which has a range that extends from coastal Northern California, through Oregon and Washington into southern British Columbia. The hybrid of blue oak and Oregon white oak is known as Epling's oak (*Quercus* × *eplingii* Mull.), a deciduous oak with small, lobed leaves found forming mixed woodland/perennial grassland on mesic hillsides of Northern California. In the southern hybrid zone, blue oak overlaps with Tucker's scrub oak (*Quercus john-tuckeri* Nixon & C.H. Mull.), an evergreen shrub oak found on dry slopes above the western Mojave and San Joaquin deserts. The hybrid produced in the southern hybrid zone, Alvord oak (*Quercus* × *alvordiana* Eastw.), is a semi-deciduous tree or large shrub that occurs in California's inner South Coast Ranges.

Sampling Design and Surveys

I sampled individuals from 28 sites across the entire study area (figure 1). Between 3 and 6 individuals were sampled from each species present at each site (based on field identifications), yielding a total of 157 trees in the full data set. For analysis, the sampled individuals are divided into three sample partitions. Northern and southern hybrid zone partitions are mutually exclusive subsets with a geographic dividing line in central California, from the Monterey Bay in the Coast Ranges to the Yosemite region in the Sierra Nevada (figure 1). The northern hybrid zone sample partition contains 92 individuals from 18 sites. The southern hybrid zone sample partition contains 62 trees from 10 sites. For comparison, a blue oak core sample partition was also created, consisting of all individuals identified as having > 99% membership in a *Quercus douglasii* cluster by fastSTRUCTURE analysis (described below). The blue oak core sample partition contains 70 trees from 19 sites, all of which are also part of either the northern or southern hybrid zone partitions. GPS coordinates were collected for all individuals and geographic distance was calculated as geodesic distance in the "geosphere" R package (R Core Team, v4.0.5), yielding a pairwise matrix used in isolation-by-distance (IBD) analyses.

In spring 2018, I surveyed spring shoot emergence phenology at all sites. Individuals were scored according to a phenological index ranging from 0 to 8 that covered progressive stages from bud dormancy to mature leaves (table S1). Each individual at each site was surveyed between two and four times during spring leaf flushing and maturation, allowing linear regression to interpolate the date at which they reached a particular phenological stage. Index scores of 0 (bud dormancy) and 8 (mature leaves) were discarded as uninformative and the progression of the remaining stages from index score 1 to 7 was confirmed as linear through time using weekly surveys of mature oak trees planted at two common garden locations (see chapter one for additional details). The phenological regressions constructed for each individual at the sampling sites allowed me to estimate specific dates for the onset of phenological stages even when they were not directly observed. The flowering interval, as the dates of first pollen release to dried catkin, was estimated from the regression models based on the observed correlation between catkin dehiscence (pollen release) and leafout indices 5, 6, 7 (figure S1). My observations and previous research suggests that female flowers of oaks become receptive during the same period that male catkins are releasing pollen on the same tree (Knapp et al. 2001, Williams et al. 2001).

The per individual flowering interval was used to calculate a pairwise index of flowering asynchrony using the formula:

$$\frac{phenological}{asynchrony} = 1 - \left(\frac{2 \times \left(\min\left(LAST_i, LAST_j\right) - \max\left(FIRST_i, FIRST_j\right) \right)}{INTERVAL_i + INTERVAL_j} \right)$$

where *i* and *j* are the two individuals whose flowering intervals are being compared, *FIRST* is the first flower date, *LAST* is the last flower date, and *INTERVAL* is the total flowering interval in days. The value of this index:

- equal 0 when the flowering intervals of the two individuals overlap perfectly
- fall between 0 and 1 for partially overlapping flowering intervals
- equal exactly 1 when the flowering intervals abut and do not overlap (i.e. flowering of individual 2 begins just as flowering of individual 1 ends)
- are greater than 1 for non-overlapping flowering intervals with increasing gaps between them

The raw index was log+1 transformed to reduce left-skewness in the observed flowering asynchrony values. This log transformed pairwise matrix of flowering asynchrony was used to characterize isolation-by-time (IBT) among pairs of trees in relation to their genotypes.

Climate variables were extracted for the location of each individual from downscaled 1951-1980 climate summary layers in the Basin Characterization Model (BCM) at 270 m resolution covering the entire study area (Flint et al. 2013). Eight BCM climate layers were selected for use, including annual average and December-January-February minimum temperature (TMN and DJF), annual average and June-July-August maximum temperature (TMX and JJA), annual precipitation (PPT), actual and potential evapotranspiration (AET and PET), and climatic water deficit (CWD). These eight environmental measures were combined in a principal components analysis and the first two PC axes (accounting for 75% of the variation) were retained. Varimax rotation of these two axes yielded a rotated-PC1 that largely associated with variation in summer maximum temperatures and water deficit and a rotated-PC2 mostly associated with winter minimum temperatures (figure S2). Pairwise distance matrices were calculated from each of the rotated PC axes as measures of isolation-by-environment (IBE) along these two components of climatic variation.

Genomic Methods

DNA of sampled individuals was extracted from silica-dried leaf material using a CTAB-chloroform extraction protocol (Doyle and Doyle 1987). Illumina sequencing libraries were prepared following a modified double-digest restriction-enzyme associated DNA (ddRADseq) protocol based on the original ddRAD (Peterson et al. 2012) and gdGBS (Peterson et al. 2014) protocols and also incorporating a qPCR step before size selection to standardize sequenceable fragment concentrations across samples. Prepared libraries were sequenced on an Illumina HiSeq 4000 platform with 150 base-pair paired-end reads. Sequence data were aligned to the valley oak, *Quercus lobata* Née, reference genome (v3.0) (Sork et al. 2016a) using bowtie2 and SAMtools (Li et al. 2009, Langmead and Salzberg 2012). Variants were called from the aligned reads in FreeBayes (Garrison and Marth 2012) and the resulting variants file was filtered to remove low quality reads or paralogs as well as linked sites within 50,000bp. After filtering, a final data set of 4,790 SNPs was analyzed in fastSTRUCTURE (Raj et al. 2014) to confirm distinct genetic clusters as well as identify putative hybrid individuals. A genetic dissimilarity matrix was also calculated from the same SNP data, as one minus the proportion of SNP alleles

shared by each pair of two individuals, calculated using the propShared function in the "adegenet" R package (Jombart and Ahmed 2011).

Matrix Regression Models

Multiple matrix regression was performed in R (v.4.0.5) with the pairwise genetic dissimilarity matrix as the response to the separate matrices of geographic distance, phenological asynchrony, and the principal components of climate dissimilarity. Random permutations of the genetic dissimilarity matrix were used to generate a null distribution for significance testing using the multiple matrix regression with randomization (MMRR) method (Wang et al. 2013). Additionally, 95% confidence intervals for the effect size of each predictor matrix were estimated based on the resampling method implemented in the "gdm" R package (Fitzpatrick and Keller 2015, Shryock et al. 2015).

To disentangle the effects of the different isolating mechanisms in the presence of multicollinearity among the predictor matrices (figure S3), the variation explained in the full model was partitioned using all possible single, multiple, and partial regressions and comparison of the adjusted R^2 values in these reduced models (Peres-Neto et al. 2006), implemented as the varpart function in the "vegan" R package (Oksanen et al. 2020).

RESULTS

The analysis of SNP data using fastSTRUCTURE supports three distinct genetic clusters that largely align to the field identification of the study species: blue oak, Oregon white oak, and Tucker's scrub oak (figure 2A). Many hybrid individuals are also identified in this analysis, including ones that appear to be evenly admixed (potential F_1 hybrids) through a range of backcrosses to individuals with only a trace of identifiable admixture. The fastSTRUCTURE results are generally concordant with my field identification of hybrids, though hybrids appear to have been somewhat over-identified in the field compared to fastSTRUCTURE cluster assignment (figure S4), a common concern among naturalists and ecologists (Craft et al. 2002).

Overall genetic dissimilarity (1 – the proportion of shared alleles) ranged from 8 to 27% in the complete data set (figure 2B). This full range of pairwise dissimilarities occurred in the northern hybrid zone, which was also found to have a notably bimodal distribution of dissimilarities (figure 3A), consistent with distinct comparisons between intra- and interspecific genotypes. In the southern hybrid zone, genetic dissimilarity ranged from 9% to 17% and the distribution was unimodal, though with a slight left skew (figure 4A), suggesting that blue oak and Tucker's oak are not as clearly differentiated. The blue oak core data partition had a narrow and symmetrical unimodal distribution of genetic dissimilarities ranging from 8 to 14% (figure 5A).

For the northern hybrid zone, neither geographic distance (IBD) nor phenological asynchrony (IBT) were significant predictors of genetic dissimilarity, while both of the environmental PC axes (IBE) were significant (figure 3). The full model with all four matrices explained 11.07% of

the variation in genetic dissimilarity (table 1). On the other hand, phenological asynchrony was the most important predictor of genetic dissimilarity in the southern hybrid zone (figure 4). Environmental PC2 was also a significant predictor in this hybrid zone, while environmental PC1 was only marginally significant ($p = 0.042$) and isolation by distance was not significant. The full model accounted for 30.70% of genetic dissimilarity in the southern hybrid zone (table 1). Finally, in the blue oak core data partition, only 5.42% of the smaller amount of observed genetic dissimilarity was explained by the IBD, IBT, and IBE matrices, with environmental PC2 the most significant predictor (figure 5), though PC1 was also marginally significant ($p = 0.04$, table 1).

The results of variation partitioning confirm that the effect sizes found in the multiple matrix regression models are generally accurate, but that many of the predictors appear to have overlapping effects. In the northern hybrid zone, most of the model variation is uniquely attributed to isolation-by-environment (IBE), but there are smaller proportions overlapping with both isolation-by-distance (IBD) and isolation-by-time (IBT) as well as the combination of all three (figure 6A). Isolation-by-time (IBT) is confirmed as independently significant in the southern hybrid zone, though about a third of the variation in the model is overlapping between IBT and IBE (figure 6B). In the blue oak core partition, IBE and IBD had independent effects, with a small overlapping portion (figure 6C).

DISCUSSION

By focusing analyses on isolation-by-distance (IBD), isolation-by-time (IBT), and isolation-by-environment (IBE), I was able to disentangle mechanisms of reproductive isolation and associate them with genetic dissimilarity across two different interspecific oak hybrid zones involving blue oak (*Quercus douglasii*) and its congeners Oregon white oak (*Quercus garryana*) and Tucker's scrub oak (*Quercus john-tuckeri*). Isolation-by-distance was not shown to be a significant factor in either hybrid zone, but I did find significant isolation-by-environment within both zones, especially associated with mean annual and winter minimum temperatures (PC axis 2). In addition, isolation-by-time associated with flowering synchrony was significant in the southern hybrid zone, both as an overlapping effect with IBE and independently. These significant isolating mechanisms together explained 11 and 31% of the observed genetic dissimilarity in the northern and southern hybrid zones, respectively (table 1). Even though this represents only a fraction of the total genetic differentiation between the species, contrasts in the importance of the isolating mechanisms between the two hybrid zones point to evolutionary differences between them.

The northern hybrid zone, involving blue oak, Oregon white oak and their named hybrid *Quercus* × *eplingii*, had the greatest maximum genetic distance of any of the three sample partitions. This is likely a reflection of the deeper phylogenetic divergence between the hybridizing taxa involved compared to the southern hybrid zone (Fitz-Gibbon et al. 2017, Hipp et al. 2017). Pairwise comparisons of trees growing in the northern hybrid zone also exhibited a

bimodal distribution of genetic dissimilarity values (figure 3A). A bimodal dissimilarity distribution results from a strong distinction between intraspecific and interspecific comparisons in this hybrid zone. There is a low dissimilarity mode of intraspecific comparisons and a high dissimilarity mode of interspecific ones. This suggests that there are significant reproductive barriers that have maintained the genetic identity of parental types (Harrison and Bogdanowicz 1997). However, I found no evidence for intrinsic asynchrony in flowering time (IBT) in this hybrid zone (figure 3). Isolation-by-time in blue oak's northern hybrid zone is only associated with extrinsic environmental influences on flowering time, as revealed by the overlap between the genetic variation that can be explained by IBT and IBE (figure 6A). There may be genomic incompatibilities between parental genotypes within this hybrid zone, making hybrids less common, though evolution of these incompatibilities does not seem to be common in oaks (Steinhoff 1993, Williams et al. 2001) and this has been suggested as an unlikely explanation for bimodal hybrid zones in the absence of other isolating mechanisms (Jiggins and Mallet 2000).

An alternative explanation for the maintenance of genetic structure in the northern hybrid zone is suggested by the significant effect of isolation-by-environment (IBE) and the non-significance of isolation-by-distance (IBD). This pattern suggests that northern hybrid zone is a mosaic zone with steep environmental gradients that separate habitat suitable for the parental genotypes. IBD would not be identified in this case because contrasting habitat patches are spatially close to one another (e.g., on opposing north and south slope aspects, or hilltops vs. cold air pools). However, significant IBE suggests that natural selection against intermediate phenotype hybrids acts as a filter on gene flow between adjacent patches (Harrison 1986). Thus, the hybrid individuals (*Quercus* × *eplingii*) that are frequently identified within this hybrid zone will be limited to narrow bands of ecotone habitats. Informal descriptions often cite hybrids as common and widespread across this zone as a whole, but a combination of complex topographic environmental variation and the possible over-identification of hybrids in the field may be a better description of what is occurring in this hybrid zone. Over-identification of hybrids is supported by comparison of the author's field identification to the genetic assignments made by fastSTRUCTURE (figure S4A). Previous studies have also reported similar issues with over-identification and mis-identification of hybrids (Craft et al. 2002, Ortego et al. 2014). Misidentification of the parental types as hybrids in this hybrid zone could be due to phenotypic plasticity that results in convergent phenotypes or possibly to a pattern of introgression specifically of alleles affecting the traits commonly used for identification despite selection against more pervasive gene flow.

Compared to the northern hybrid zone, the maximum observed genetic dissimilarity was considerably lower in the southern hybrid zone (figure 4A), reflecting closer phylogenetic relationship of blue oak and Tucker's oak. Perhaps because of the importance of isolation-by-time in the southern hybrid zone, more of the total variation could be explained in my models. The unimodal distribution of genetic distances observed in this hybrid zone suggests an underlying flat distribution of sampled genotypes, with parental types and hybrids occurring at

similar frequencies (Jiggins and Mallet 2000). While this pattern could occur due solely to isolation-by-distance without other isolating mechanisms, my analysis does not support IBD as a significant factor in blue oak's southern hybrid zone (figure 4). However, while IBE was again found to be a significant as in the northern hybrid zone, most of the explained variation in the southern hybrid zone was associated with IBT (phenological asynchrony). This may include both intrinsic and extrinsic factors, as suggested by the independent significance of IBT in addition to the overlapping significance of IBE and IBT (figure 6B). If flowering time is an intrinsic isolating mechanism in this hybrid zone, it suggests that species identity could be maintained even in direct sympatry and despite common hybridization.

An important role for environmental selection with little or no isolation-by-distance has been previously reported in oaks and attributed to large scale mixing by pollen dispersal, especially integrated over the long demographic history of oak populations (Dodd and Afzal-Rafii 2004, Craft and Ashley 2007, Sork et al. 2016b). The significant role for isolation-by-time I identified in the southern hybrid zone of this study is more unique. One previous study that attempted to partition the role of flowering time on genetic structure in an oak, using a similar framework, found no IBT effect. However, that study involved a single species, rather than a hybrid zone, and the species' distribution was partitioned between seasonal lowlands and aseasonal uplands, potentially making any partition of geography, climate, and flowering time impossible (Deacon and Cavender-Bares 2015). Several other studies comparing flowering time between hybridizing species in temperate climates have not found significant differences either (Jensen et al. 2009, Wilkinson et al. 2017). The role of flowering time as an isolating mechanism between blue oak and Tucker's scrub oak may be related to their relationship as recently diverged sister taxa (Hipp et al. 2017), suggesting that their hybrid zone may be a primary one, reinforced by the evolution of an intrinsic isolating mechanism in the form of flowering time (Devaux and Lande 2008). A role for flowering asynchrony has similarly been suggested for the closely related *Quercus virginiana* and *Q. geminata* in the U.S. southeast, though this has not been explicitly tested against the effects of IBD or IBE (Cavender-Bares and Pahlich 2009).

Among the climate variables that were included in the IBE portion of my models, it is especially interesting that low temperature variables associated with the second principal component of the climate PCA consistently explained more of the genetic structure than the high temperature and water availability variables associated with first principal component (figure S2). Water availability and evolutionary mechanisms to maintain transpiration throughout the growing season are generally considered strong selective forces on oaks, especially in Mediterranean-type climates such as California (Cavender-Bares et al. 2004, Skelton et al. 2021). The three species included in this study widely are often thought of as partitioned along an aridity gradient, from the mesic-adapted Oregon white oak found in wet coastal Northern California and the Pacific Northwest to the more drought-tolerant blue oak of inland California foothills to the still more arid-adapted Tucker's scrub oak of desert margins. While a high temperature/aridity gradient associated with PC axis 1 was indeed significant in both hybrid zones, it seems that minimum

temperatures may have a greater effect on the realized gene flow both between hybridizing species and between populations of blue oak alone. This environmental isolating effect remains significant independent of its effect on asynchrony in flowering time which is also influenced by winter and spring minimum temperatures (figure 6), suggesting that there may be an important selective role for frost hardiness or associated traits (Gugger et al. 2016). This possibility is also discussed in the first chapter of this dissertation for populations of blue oak specifically.

The remaining genetic variation within each of the sample partitions which could not be explained in my models is likely explained by a variety of factors. The largest component may simply be unstructured variation resulting from genetic drift and persistence of variation associated with stochastic dispersal, both of which are expected to be important in the population genetics of long-lived, sedentary organisms like oaks (Petit and Hampe 2006). This source of genetic variation can be highlighted in the blue oak core sample partition, which as expected had comparatively low genetic dissimilarity, but also an even smaller amount that was explained by IBD, IBT, and IBE in my models (table 1, figure 6C).

In addition to the influence of random processes on the distribution of genetic variation, some of the unexplained genetic dissimilarity in the hybrid zones could be associated with intrinsic genomic isolating mechanisms, such as pollen or developmental incompatibilities, which I was not able to quantify here. While genomic incompatibilities are often treated as unimportant in hybridizing oaks, there is some experimental evidence that they can occur. In particular, there may be asymmetric effects in which hybridization may have reduced success in one direction (e.g. species A pollen to species B female flower), while there is full interfertility in the other direction (Steinhoff 1993, Williams et al. 2001).

Other axes of environmental variation, not captured by the climate variables included here, might also help to explain more of the genetic dissimilarity. Paleoclimatic variation could leave a genetic signal that persists in modern populations distinct from the effects of contemporary climate (Papper 2019). Additionally, variation in soil or bedrock characteristics and fine-scale topography can modify the experienced environment in ways that would not be captured by the 270 m topographically downscaled climate layers used here (Ackerly et al. 2020, McLaughlin et al. 2020).

In conclusion, by using multiple matrix regression on pairwise measures of geographic, environmental, and phenological dissimilarity in combination with variation partitioning analyses, I have shown both similarities and differences in the mechanisms of reproductive isolation maintaining genetic dissimilarity across two hybrid zones involving blue oak in California. There is no signal of independent isolation-by-distance in either hybrid zone. Instead, isolation-by-environment is found to play an important role, in particular minimum temperatures. In the northern hybrid zone involving Oregon white oak, isolation-by-environment seems to be more complex, involving both variation in minimum temperatures and separately a gradient of high temperatures and aridity. In blue oak's southern hybrid zone involving Tucker's scrub oak,

aridity was of surprisingly small importance to the general pattern of isolation-by-environment, but there was a significant role for isolation-by-time. The effect of IBT as flowering synchrony was partitioned into a portion overlapping with IBE, interpreted as an extrinsic influence on flowering time, and a portion that was independent of climate, interpreted as an intrinsic genetic mechanism of reproductive isolation between these species.

TABLES AND FIGURES

Table 1: Results of multiple matrix regression with randomization (MMRR) explaining genetic dissimilarity as a function of four explanatory dissimilarity matrices. In this framework, the effect of geographic distance represents isolation-by-distance, phenological asynchrony represents isolation-by-time, and the two environmental PC axes represent isolation-by-environment. Individuals are partitioned into three partitions analyzed separately as described in the text. Significant and marginally significant ($p < 0.1$) model terms are bolded and p-values and r^2 are given for the full models in each sample partition analysis.

Northern Hybrid Zone	coefficient	t-value	p-value	r^2
Intercept	0.1466			
Geographic Distance (km)	0.000003389	0.487	0.6338	
Phenological Asynchrony Index	0.0001084	0.907	0.4232	
Environmental PC Axis 1 (max temp/aridity)	**0.004003**	**8.851**	**< 0.00001**	
Environmental PC Axis 2 (min temp)	**0.008370**	**16.669**	**< 0.00001**	
Full Model			**< 0.00001**	**0.1107**
Southern Hybrid Zone				
Intercept	0.1222			
Geographic Distance (km)	-0.000009164	-2.881	0.1999	
Phenological Asynchrony Index	**0.01477**	**20.470**	**< 0.00001**	
Environmental PC Axis 1 (max temp/aridity)	**0.001054**	**3.024**	**0.04208**	
Environmental PC Axis 2 (min temp)	**0.003224**	**8.828**	**0.00008**	
Full Model			**< 0.00001**	**0.3070**
Blue Oak Core				
Intercept	0.1170			
Geographic Distance (km)	-0.000003843	-0.453	0.8401	
Phenological Asynchrony Index	-0.0004314	-1.244	0.7406	
Environmental PC Axis 1 (max temp/aridity)	**0.0007737**	**5.226**	**0.0400**	
Environmental PC Axis 2 (min temp)	**0.001731**	**9.847**	**0.0038**	
Full Model			**0.00826**	**0.0542**

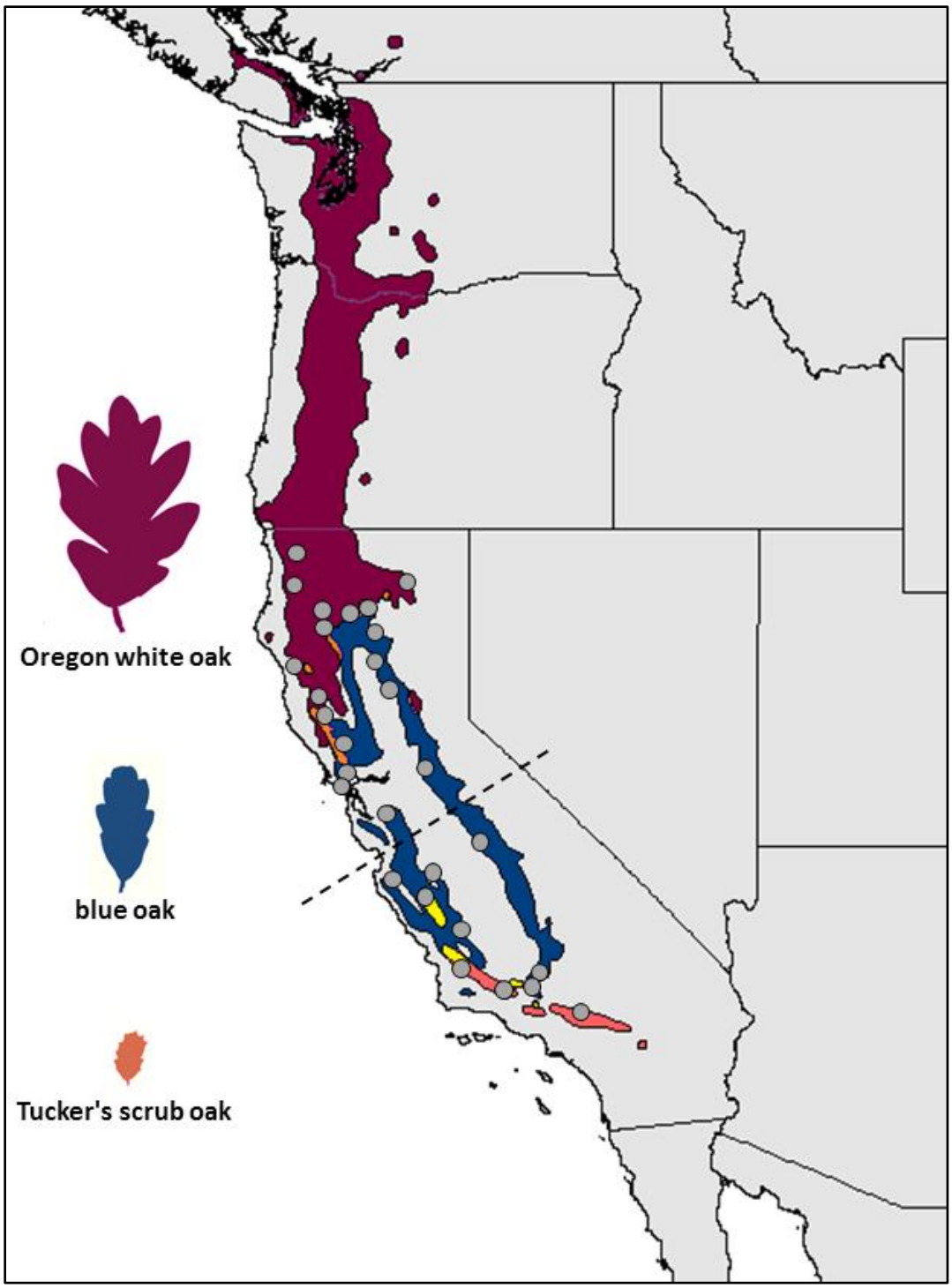


Figure 1: Range maps of the three oak taxa included in this study. Oregon white oak (purple), blue oak (blue), and Tucker's scrub oak (red), with the northern and southern zones of range overlap and hybridization shown in orange and yellow, respectively. Grey dots indicate sampling localities and the dashed line indicates the partition used for separate analyses of the northern and southern hybrid zones described in the text.

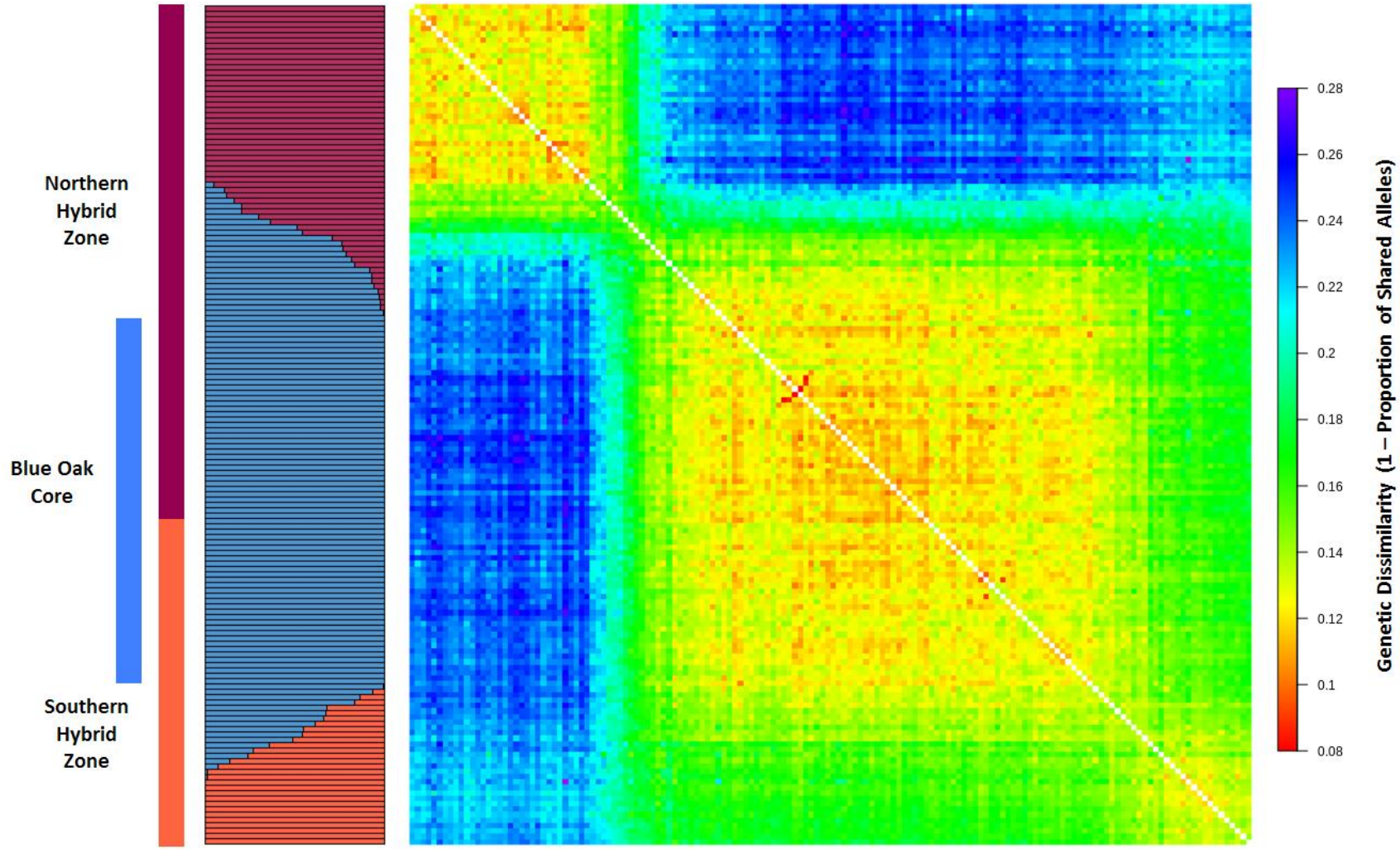


Figure 2: Genetic analyses of the total oak data set (154 individuals). fastSTRUCTURE results (left) showing three genetic clusters that align with the three study species (with same colors as distributions in figure 1) and hybrid individuals as mixed bars. Individual genetic distance matrix (right) showing pairwise patterns of genetic dissimilarity (as 1 – the proportion of shared alleles) with rows matching the fastSTRUCTURE result.

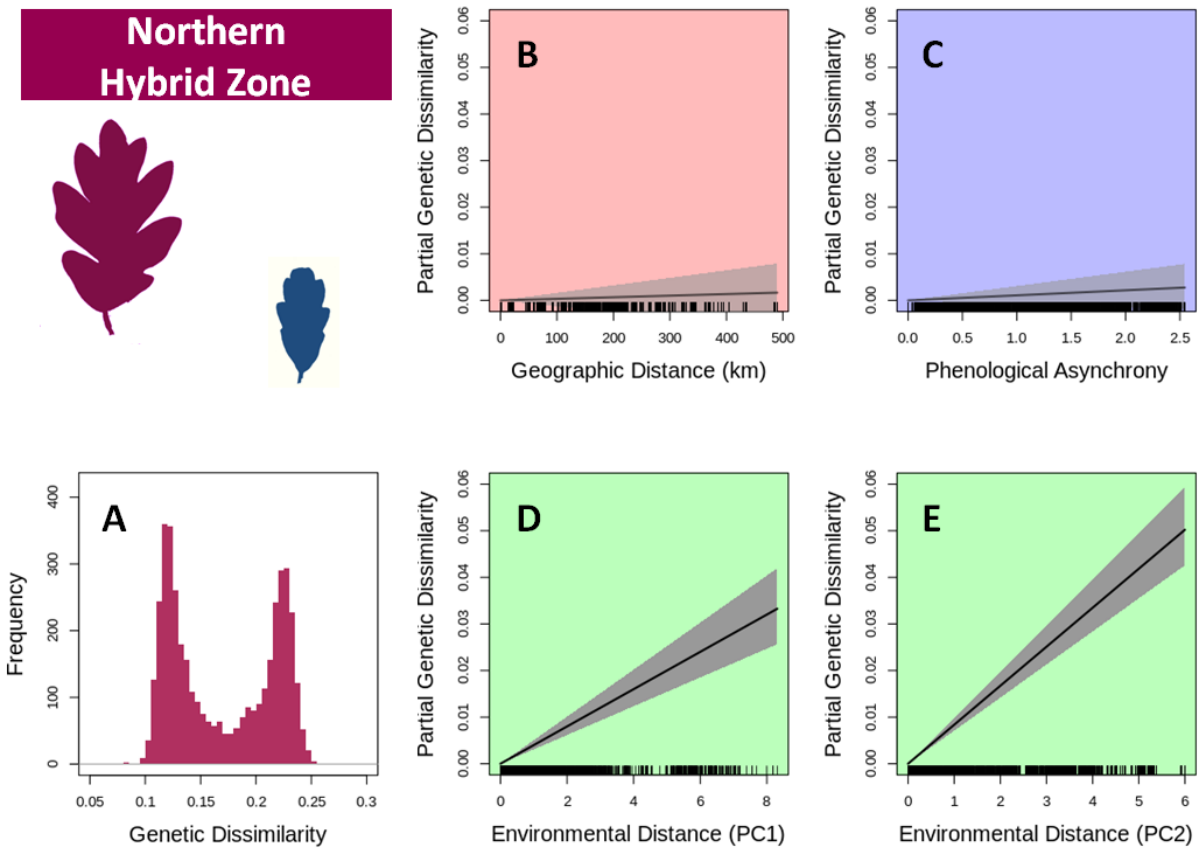


Figure 3: The northern hybrid zone between blue oak and Oregon white oak. Showing the distribution of pairwise genetic dissimilarity among individuals (A) and multiple matrix regression results on that dissimilarity for isolation-by distance (B), isolation-by-time as an index of flowering asynchrony (C), and isolation-by-environment using the first two principle components of climate variation (D and E).

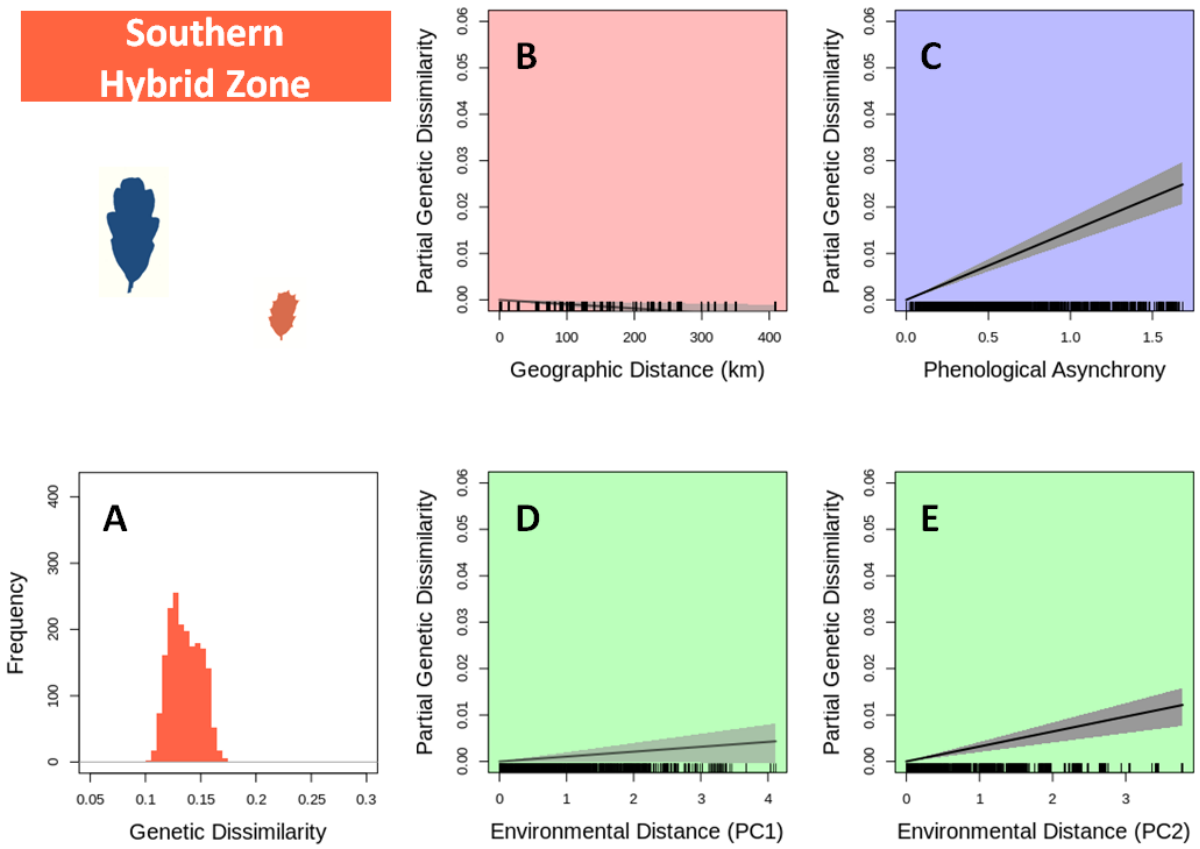


Figure 4: The southern hybrid zone between blue oak and Tucker's scrub oak. Showing the distribution of pairwise genetic dissimilarity among individuals (A) and multiple matrix regression results on that dissimilarity for isolation-by distance (B), isolation-by-time as an index of flowering asynchrony (C), and isolation-by-environment using the first two principle components of climate variation (D and E).

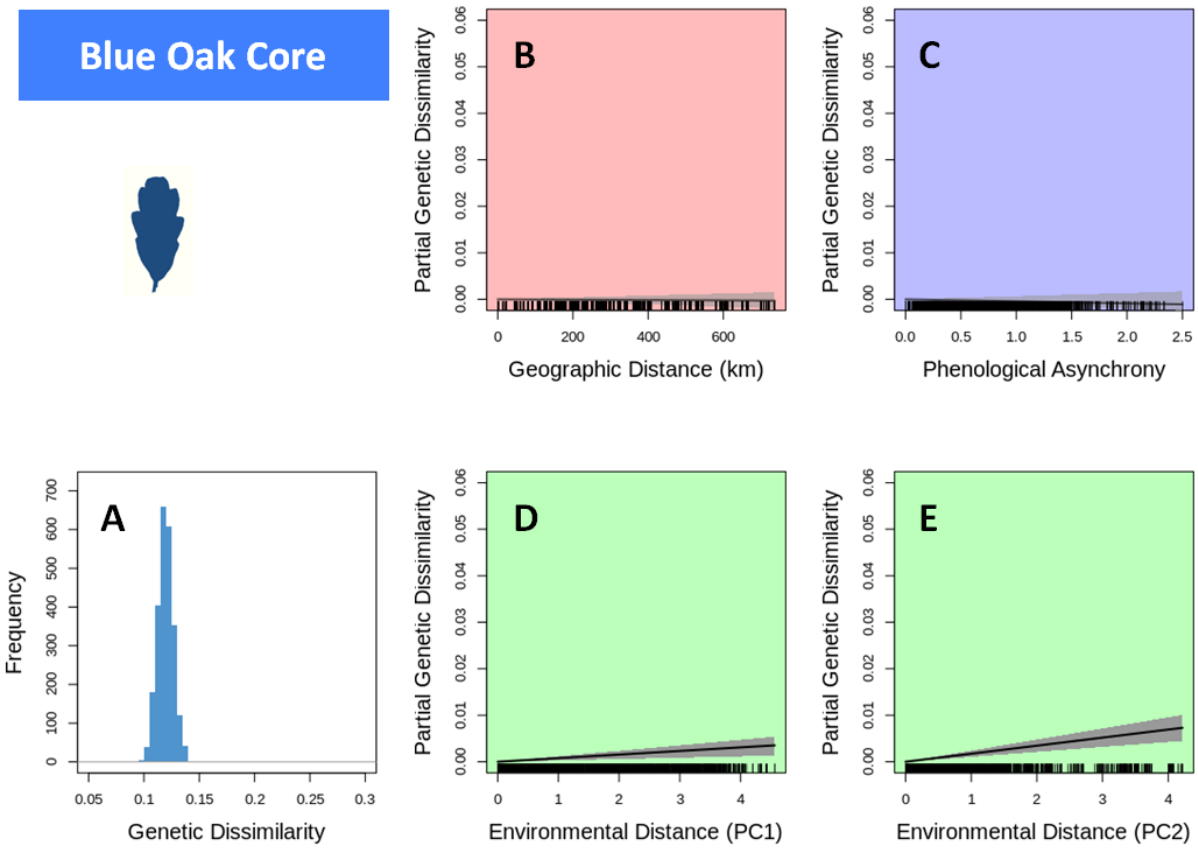


Figure 5: The blue oak core range. Showing the distribution of pairwise genetic dissimilarity among individuals (A) and multiple matrix regression results on that dissimilarity for isolation-by distance (B), isolation-by-time as an index of flowering asynchrony (C), and isolation-by-environment using the first two principle components of climate variation (D and E).

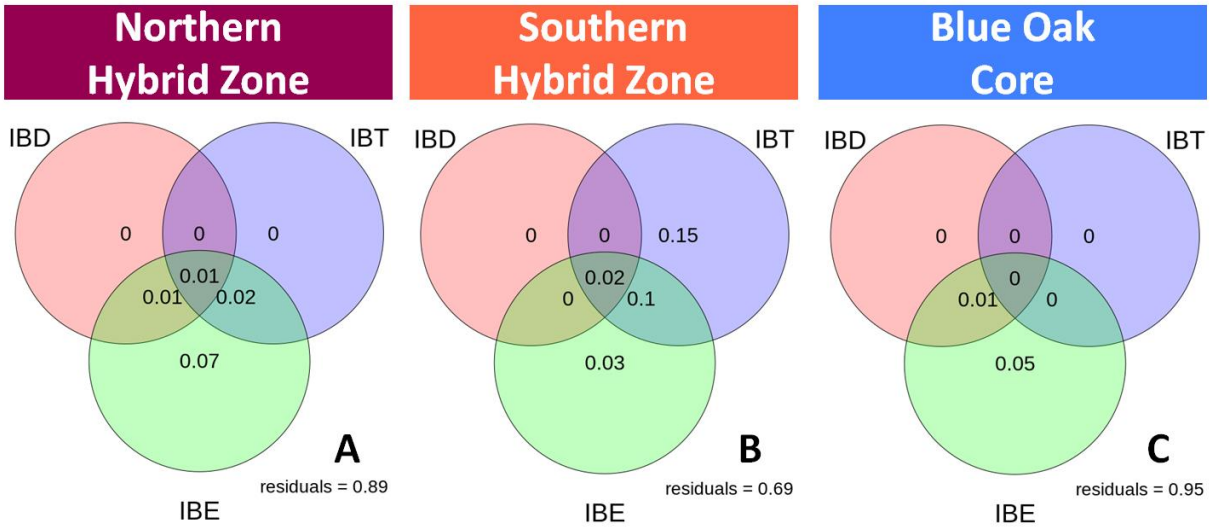


Figure 6: Variation partitioning diagrams for the three analyzed data sets, showing independent and combined effects of isolation-by-distance (IBD), isolation-by-time (IBT) as flowering time, and isolation-by-environment (IBE) as principle components of environmental distance, for the northern hybrid zone (A), southern hybrid zone (B), and blue oak core data partitions (C).

SUPPLEMENTAL MATERIAL

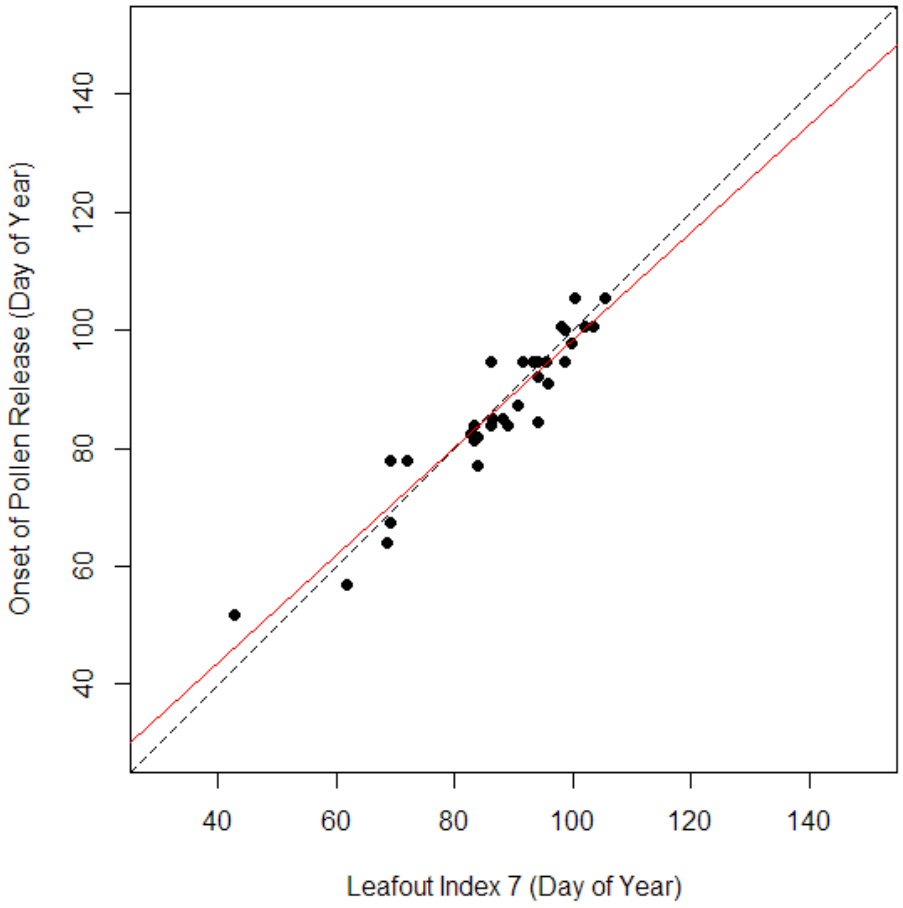


Figure S1: Correlation between leafout index 5 (not 7) and the onset of pollen release for a subset of individuals in the phenological survey for which flowering data was collected. The dashed grey line is a 1-to-1 date line.

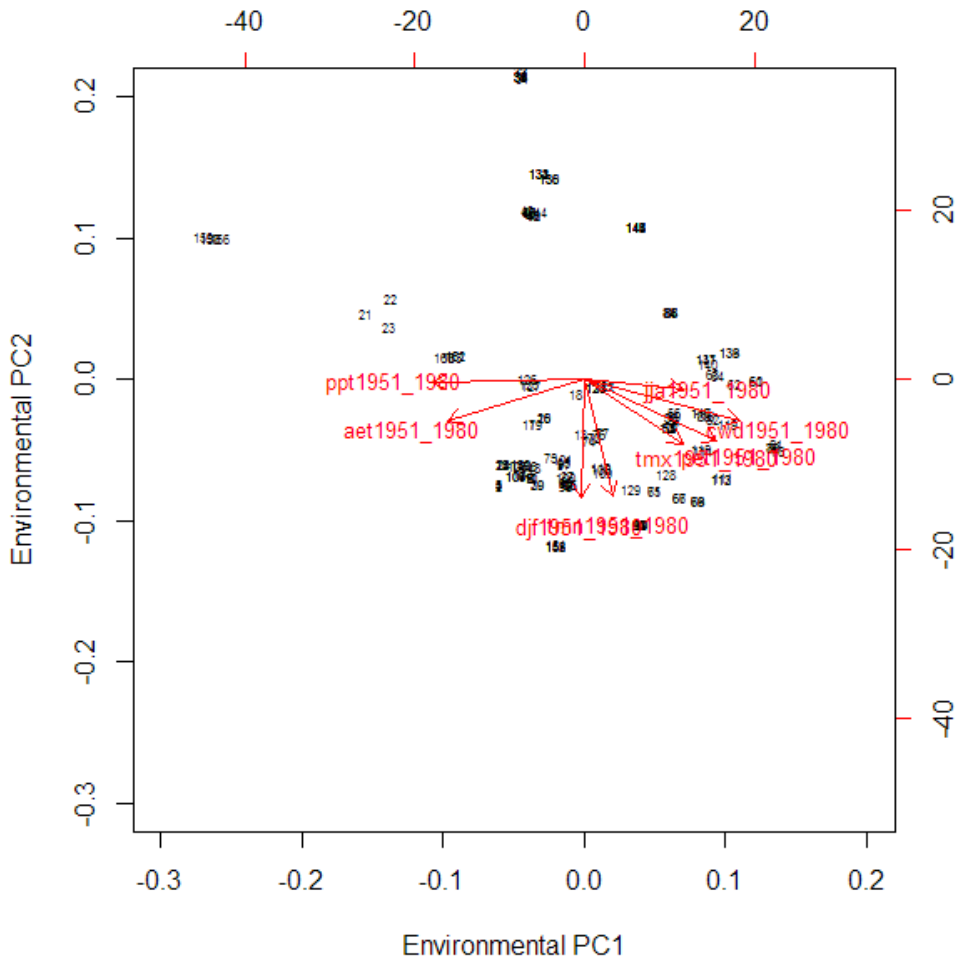


Figure S2: Varimax rotation of the first two principal components of eight 1951-1980 climate variables for sampled individuals, accounting for 74% of the total variation. Rotated PC1 includes variation in summer maximum temperatures (JJA) as well as precipitation (PPT) and water availability (AET, CWD). Rotated PC2 includes winter and mean annual minimum temperatures (DJF, TMN).

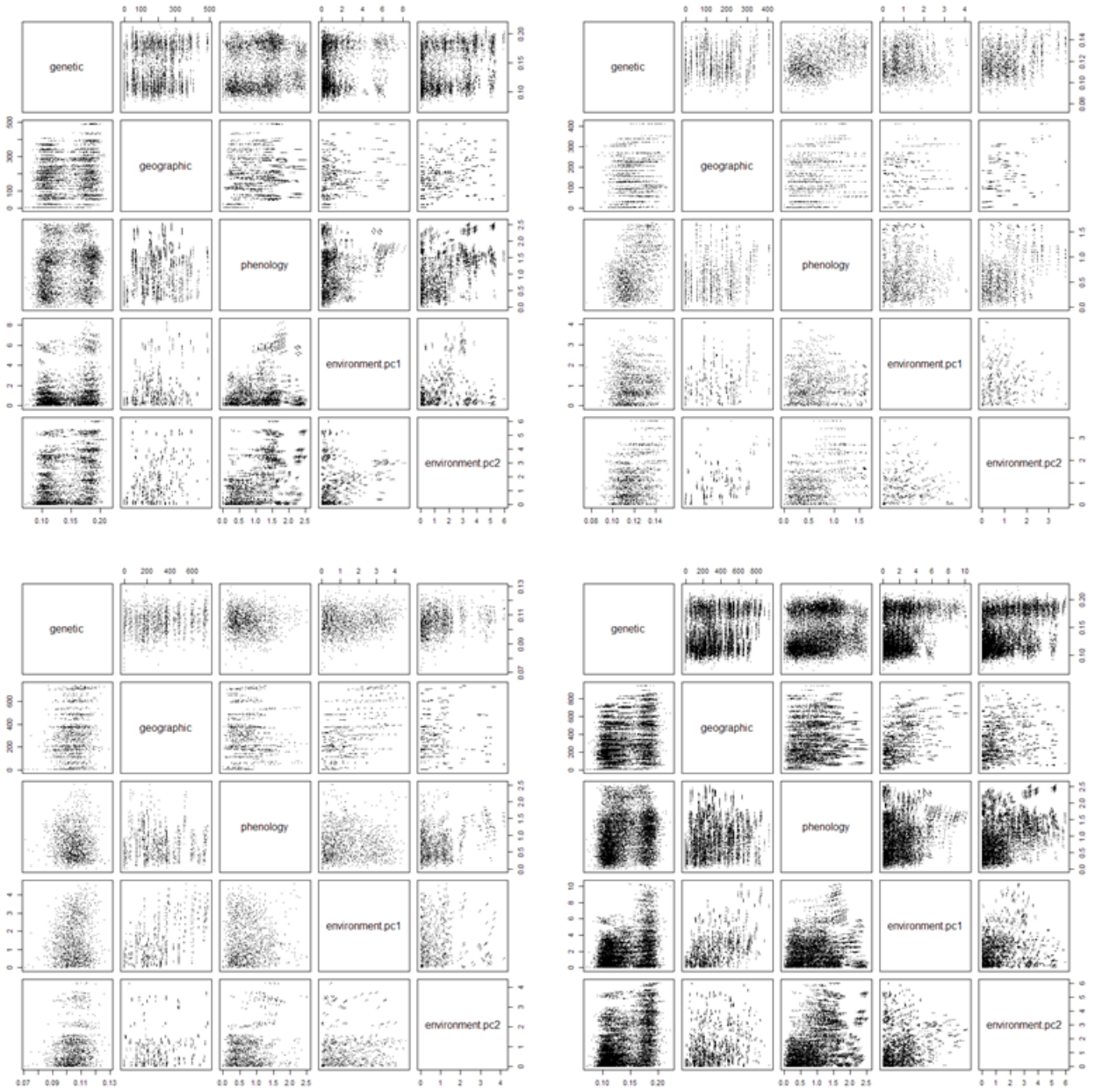


Figure S3: Full correlation matrices between genetic dissimilarity and the four explanatory distance matrices for the northern hybrid zone (A), the southern hybrid zone (B), the blue oak core (C), and the full data set (D). Note the clustering of genetic dissimilarity in the northern hybrid zone partition.

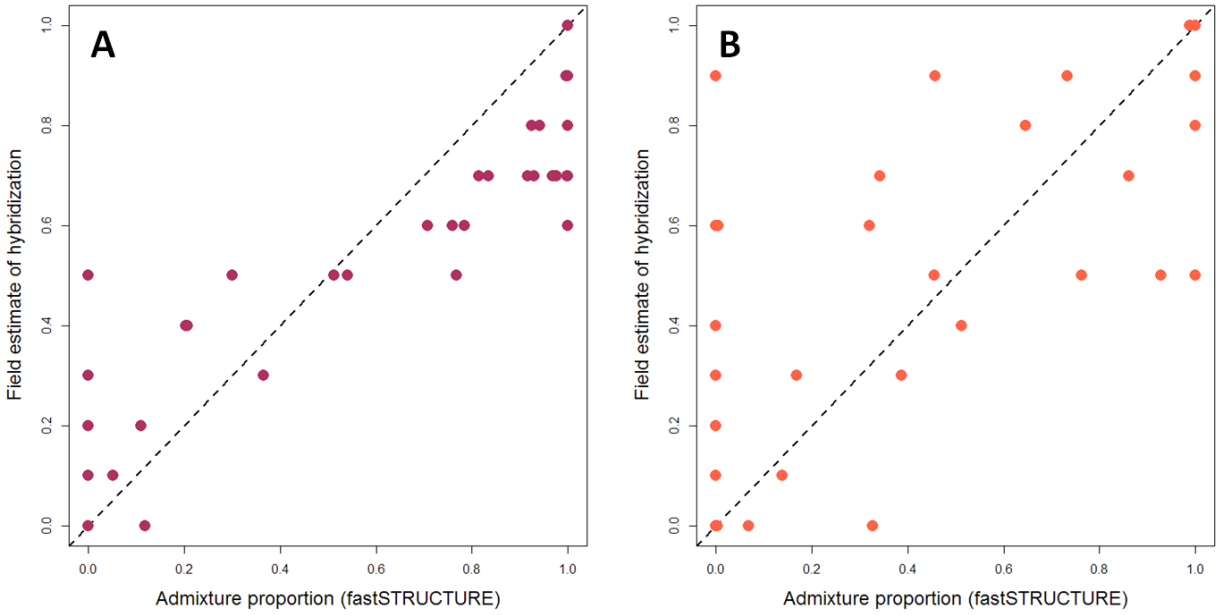


Figure S4: Correlation between the author's identification of species and hybridization in the field and estimation of admixture proportion using fastSTRUCTURE for sampled individuals from the northern hybrid zone (A) and southern hybrid zone (B). Blue oak is shown as 0.0 in both plots and Oregon white oak or Tucker's scrub oak as 1.0 in their respective hybrid zones.

CHAPTER THREE

Identifying complex patterns of ancient and contemporary hybridization in a phylogenetic context

ABSTRACT

Hybridization is a topic of enduring interest in evolutionary studies. Frequently, research has been divided between studies of relatively contemporary hybridization within population genetics, where it manifests as population structure and admixture, and studies of more ancient hybridization within phylogenetics, where it is characterized as reticulation between lineages. In many taxonomic groups, however, both ancient and contemporary hybridization are likely to be found. Here I combine methods from population genetics and phylogenetics to study hybridization within a lineage of oaks expected to show complex patterns of hybridization. I show that the method of phylogenetic rogues applied to a set of maximum likelihood bootstrap trees identifies many of the same individuals that are inferred as admixed by the STRUCTURE model commonly used in population genetic studies. These individuals are interpreted as resulting from contemporary hybridization and pruning them from the phylogeny increases resolution of major clades within the group dramatically. With this well-resolved phylogeny, I am able to use phylogenetic invariants methods, specifically ABBA-BABA D-statistics and partitioned D-statistics, to identify clades that enrichment of certain site patterns throughout the genome interpreted as evidence of ancient hybridization. Specifically, the clade comprised of California scrub oak (*Quercus berberidifolia*) and leather oak (*Q. durata*) shows evidence of extensive hybridization and possibly of hybrid speciation. Using partitioned D-statistics, this hybridization is shown to have occurred deep within the phylogeny, between the common ancestors of several diverged lineages. In addition, *Q. berberidifolia* appears to have also undergone a later episode hybridization with just some descendants of one of those lineages. Meanwhile, other descendants of these lineages also show contemporary hybridization, as identified using the phylogenetic rogues method and STRUCTURE model. This demonstrates the power of the methods used here to identify ancient and contemporary hybridization even in extremely complex situations.

INTRODUCTION

Hybridization can be found in extant taxa in two distinct forms. First, when there is reproductive compatibility between lineages, contemporary and ongoing hybridization will be reflected at a landscape scale in locations where reproductive timing and dispersal ability allow it. Gene flow through contemporary hybridization is a source of novel genetic variation, which may be adaptive (Suarez-Gonzalez et al. 2018, Leroy et al. 2019). Such hybridization may even stimulate lineage divergence, in a process called hybrid speciation (Gross and Rieseberg 2005), but it can also lead to a loss of diversity if incipient lineages eventually re-merge as a result of extensive hybrid gene flow (Petit et al. 2004). In addition to these effects of contemporary hybridization, though, even modern taxa that do not hybridize in the present may still show the effects of ancient hybridization as introgressed genetic variation. Ancient introgression reflects details of the historical biogeography and ecology of the lineages involved that continue to affect them down to the present, long after gene flow has ceased. The genomic effects of ancient hybridization are also important in phylogenetic reconstruction where they can lead to false or low support for topologies that are model constrained to only show divergence between lineages. Reticulate or inosculated phylogenetic trees are necessary in these cases to properly account for ancient hybridization but remain computationally challenging at genomic scales (Solís-Lemus and Ané 2016, Wen and Nakhleh 2018). In taxonomic groups that could be considered long-term stable multispecies complexes (Van Valen 1976) or what has also been called a syngameon (Lotsy 1925, Hipp 2015), both contemporary and ancient hybridization are to be expected. Such complex signals of gene flow at multiple temporal scales present additional challenges to phylogenetic and population genetic analyses.

Oaks (genus *Quercus*) have long been an important model group in the study of hybridization. They were the group used by Van Valen (1976) in the development of the ecospecies and multispecies complex concept and so it is not surprising that previous research has identified introgression at both population genetic (Muller 1952, Dodd and Afzal-Rafii 2004, Curtu et al. 2007) and phylogenetic scales (Tucker 1961, Eaton et al. 2015, McVay et al. 2017a, 2017b, Kim et al. 2018). Oaks' ability to hybridize within taxonomic sections is notorious and seems to have broad importance for their evolution.

Here, I focus on the western North American (WNA) clade of white oaks (section *Quercus*, series *Dumosae*) and present analyses that straddle the line between population genetics and phylogenetics to identify signals of both contemporary and ancient introgression. This group is a particularly interesting one for the study of hybridization due to the range of life histories among its species, including evergreen shrubs, deciduous shrubs, and deciduous trees of small, medium, and large stature. This life history variation also reflects the broad ecological variation covered by the clade's species, which range from the Mojave Desert and arid inland Baja California, through Mediterranean California, including coastal, interior, and mid-elevation montane habitats, and as far north as southern British Columbia. Though at least one species in the clade (valley oak, *Quercus lobata*) may be partially reproductively isolated from others (Craft et al.

2002, Kim et al. 2018), this and other species in the clade do present evidence of extensive contemporary hybridization, with varying frequency, in most locations where their ranges overlap. These zones of overlap and hybridization may be broad relative to the species' ranges (e.g. Tucker 1952) or narrow (Burge et al. 2019). Meanwhile, there is evidence from both fossils and paleoclimate reconstructions that the range limits of species in the group have been quite different in the past, which would have created opportunities for hybridization between lineages whose ranges do not currently overlap (Mensing 2014, Papper 2019).

In this study, I investigate gene flow among nine white oak taxa to understand the importance of hybridization in the evolution and divergence of western white oaks. I explore whether dense sampling of species traditionally associated with population genetic studies, but analyzed in an explicitly phylogenetic context, is able to identify and disentangle ancient introgression even in the presence of widespread contemporary hybridization. In some cases, these two signals of hybridization may involve the same parent lineages. In particular, I will address the evolutionary and phylogenetic history of California scrub oak (*Quercus berberidifolia*) and its sister taxon leather oak (*Q. durata*), which have been variously reconstructed by recent research (Fitz-Gibbon et al. 2017, Hipp et al. 2017). While there are many explanations for this type of disagreement in phylogenetic reconstruction, even when using similar data and inference methods, I will explore whether ancient hybridization can explain this lineage's uncertain placement.

METHODS

Samples and Sequence Data

I sampled 310 trees and shrubs across 9 taxa, representing all of the named taxa in the western North American white oak clade with the exception of the rare and range restricted *Quercus dumosa* and a named variety of *Quercus durata* restricted to the San Gabriel Mountains. Nine individuals of *Quercus kelloggii*, which belongs to the red oak section, are also included in these analyses as an outgroup, for a total sample size of 319 (figure 1).

For all samples, DNA was extracted from silica-dried leaf material following a standard CTAB-chloroform method (Doyle and Doyle 1987). Four separate sequencing libraries were prepared using a double-digest RADseq technique with PstI and MspI restriction enzymes (Peterson et al. 2012, Peterson et al. 2014). Final size selection was made using Pippin Prep (Sage Science, Beverly, MA, USA) with a target insert size of 500-bp +/- 75-bp. These libraries were sequenced on an Illumina platform HiSeq-4000 as 150-bp paired-end reads.

SNPs were called and consensus sequences generated using the ipyrad pipeline (Eaton and Overcast 2020) with reference alignment to the completed *Quercus lobata* genome (ver. 3.0, Sork et al. 2016). ipyrad was run with a cluster threshold of 0.85 and the maximum number of heterozygous sites per sequence set to 10%. A minimum of 10% samples were also required in order to call a variant per site, though a threshold of 50% missing data was used for downstream

analyses of the SNP data set. Default ipyrad settings (for ver. 0.9.68) were kept for all other parameters.

fastSTRUCTURE Analysis

To identify signals of admixture usually interpreted as contemporary hybridization, I first analyzed the ipyrad SNP dataset using fastSTRUCTURE (Raj et al. 2014) for all values of K between 2 and 10. These analyses were run using the simple prior setting with a convergence diagnostic of 10^{-9} delta marginal likelihood and 10-fold cross validation.

Phylogenetic Inference and Rogues

Maximum likelihood phylogenetic inference was conducted in RAxML 8.2.13 (Stamatakis 2014) with a concatenation of ipyrad consensus loci under the GTRCAT model and with 100 rapid bootstraps.

To further identify samples showing signs of contemporary hybridization, I used the concept of phylogenetic rogues (Aberer et al. 2013) to identify tips that tended to change position most dramatically in the set of RAxML bootstrap trees. I used a method based on the coefficient of variation (CV) among pairwise distances among tips in the bootstrap trees, similar to the Taxonomic Instability Index implemented in Mesquite (Maddison & Maddison 2010). I first calculated pairwise distance between all tips for each of the 100 RAxML bootstrap trees (with internode branch lengths set to 1) to produce a num_taxa × num_taxa × num_trees three-dimensional array. The CV among bootstrap trees for each pairwise comparison is an estimate of instability in the phylogenetic relationship between that pair of samples. The mean of these pairwise CV values across a row (or down a column) gives an estimate of the instability per individual sample in comparison to all others in the data set, identifying which taxa are moving around among trees.

D-statistics

Signals of more ancient introgression were identified using the method of phylogenetic invariants (Lake 1987, Kubatko and Chifman 2019), related to the D-statistic/ABBA-BABA method (Durand et al. 2011, Patterson et al. 2012). This method identifies introgression among three ingroup populations (clades or, nominally, taxa) using the expectations of a specific phylogenetic topology that allows separate estimation of incomplete lineage sorting and introgression (figure 2). The site pattern 'BBAA' on this topology is expected to be the most common pattern with two tips sharing the derived allele if the hypothesized phylogeny is correct. The site patterns 'ABBA' or 'BABA' are discordant with the phylogenetic hypothesis and could result from either incomplete lineage sorting or introgression. A significant excess count of either pattern is a signal of introgression between P3 and either P1 or P2. In case there is introgression between P3 and both P1 and P2, the excess of site pattern counts in the ABBA-BABA test shows the populations with a larger amount of introgression. Significance of the excess site counts can

be assessed with a chi-squared test (Lake 1987, Green et al. 2010), which should be corrected for multiple tests, as with the Bonferroni correction method.

Site patterns were counted for all SNPs in the ipyrad output using the HyDe python package (Blischak et al. 2018) and the best ML tree from the RAxML analysis with rogue tips pruned out. This tree was used to group individual samples into clades that maximally approximate the currently recognized species and named varieties for the ingroup populations. *Quercus kelloggii*, a red oak (*Quercus* section *Lobatae*), was used as the outgroup for all tests presented here. The RAxML best tree also served as the proposed topology for the HyDe input, though these initial topologies were compared to the 'BBAA' site pattern counts to confirm that the RAxML topology was the best one for each set of three populations.

First, site pattern counts were made at the whole population (clade) level. Then the same comparisons were re-run, but with the topology rearranged to give the highest BBAA (best topology) and ABBA greater than BABA (i.e. introgression between P2 and P3 rather than P1 and P3). This labeling simplifies the rest of the analyses by minimizing both the number of negative ABBA – BABA differences and estimates of introgression (gamma) that come out greater than 50%. Additionally, in the second round of HyDe run, each sample from the P2 population (putative hybrids) was tested individually against the P1 and P3 populations in order to provide an estimate for the amount of variation in introgression.

From the results of this second run of HyDe, the proportion of introgression between P2 and P3 was calculated using the method described by Kubatko and Chifman (2019). Their calculation takes into account both the difference between the counts of ABBA and BABA to quantify the amount of introgression versus incomplete lineage sorting (as in the most common D-statistic) as well as the difference between BBAA and BABA to quantify incomplete lineage sorting versus the underlying phylogenetic signal. The result is a robust estimate of gamma (γ), which is the percent of admixture between introgressed populations (figure 2).

In order to constrain the timing of introgressions identified by these phylogenetic invariants tests, I next used a five-taxon expansion of the four-taxon D-statistic tests. These tests essentially add a sister population to P3, called P4 in some implementations (e.g. Pease and Hahn 2015) though I will use the more common $P3_1$ and $P3_2$ labels (Eaton and Ree 2013). In these tests, an excess of the ABBBA site pattern suggests that the introgression involved the common ancestor $P3_1$ and $P3_2$; alternatively, if the ABBAA and ABABA patterns are more common than ABBBA, it suggests introgression involving the terminal clades themselves.

RESULTS

The ipyrad pipeline returned 11,374 consensus loci for the 319 input samples. When aligned, this consisted of 7,022,283 total base pairs with 77.69% undetermined characters (Ns), which includes both missing sample data and a variable sized gap between the 150-bp paired-end mate reads. A total of 148,868 genome-wide SNPs were called across these loci in the original ipyrad

output. The complete concatenated alignment was used for phylogenetic analysis in RAxML. The SNP data set was further filtered to include only bi-allelic SNPs and sites with a maximum of 50% missing data. This filtering yielded a final set of 63,870 SNPs used for the fastSTRUCTURE and HyDe analyses.

The chooseK.py script in fastSTRUCTURE's indicated that 6 clusters (K=6) are optimal to explain the structure in the data (figure 3). These 6 genetic clusters broadly match species identifications made in the field, though with the full data set fastSTRUCTURE did not differentiate between *Quercus berberidifolia* and *Q. durata*, nor between *Q. john-tuckeri* and *Q. cornelius-mulleri*, nor between the two named varieties of *Q. garryana*, var. *garryana* and var. *breweri*. However, in subsetted fastSTRUCTURE analyses of each of these clusters separately, each of them was divided into two sub-clusters that matched field identification of these taxa (analyses not shown).

Additionally, fastSTRUCTURE did not identify *Q. pacifica* as a unique cluster, instead depicting all of the samples as admixed between the *Q. berberidifolia/durata* cluster and the *Q. john-tuckeri/cornelius-mulleri* cluster. Overall, almost one third (99 out of 319) of samples were determined by fastSTRUCTURE to have more than 5% admixture.

Phylogenetic Inference and Rogues

The RAxML best tree identified clades and a topology largely congruent with expectations from previous published phylogenies of the WNA white oak group (Fitz-Gibbon et al. 2017, Hipp et al. 2017, 2019). However, bootstrap branch supports were very poor on this tree with only a handful of clades reaching a 95% consensus among the 100 bootstraps (figure S1A) and most not even reaching 50% consensus. This indicates considerable phylogenetic instability, most likely resulting from the large number of inferred hybrids acting as phylogenetic rogues.

Rogue identification by the CV of pairwise distances across the bootstrap trees revealed many samples with high mean CV values (figure 4). The 50 samples with highest mean CV of pairwise phylogenetic distance were selected as rogues and pruned out of the RAxML phylogeny. Forty-one of the 50 samples selected as rogues had also been determined as more than 5% admixed in fastSTRUCTURE while another 5 showed between 1 and 5% admixture. The remaining four samples selected as rogues, but not determined as admixed in fastSTRUCTURE all had large proportions of missing data, which is another common use of rogue identification, particularly in the construction of supermatrices.

After pruning rogues, the RAxML inferred tree with 269 remaining tips had considerably higher bootstrap support, Most named taxa were identified as clades with better than 95% consensus (figure S1B, S2). The RAxML tree search was also re-run with the 50 rogue taxa culled, but this produced a best tree without major differences compared to the pruned tree. Since some authors have suggested that tree inference using all available data, including rogue tips, is preferable

(Aberer et al. 2013), I opted to proceed with the original RAxML best tree with the 50 rogues pruned out (figure 5).

D-statistics

All 9 named WNA white oak taxa that were sampled were inferred as clades after pruning phylogenetic rogues, though some still with low bootstrap support (figure S2). HyDe results of all 84 possible triplets of these 9 populations revealed a significant and consistent signal of introgression in all triplets that had either *Q. berberidifolia* or *Q. durata* in the P2 position (table 1). HyDe results also showed that all evidence or lack of evidence for introgression was identical for *Q. berberidifolia* and *Q. durata* as well as for *Q. garryana* var. *garryana* and *Q. garryana* var. *breweri* (table 2). Because of this pattern and the low bootstrap support for *Q. berberidifolia* and *Q. garryana* var. *breweri* (fig S2), these taxon pairs were collapsed into single clades for the next round of HyDe comparisons. The remaining D-statistic tests involving other taxa were not significant after Bonferroni correction, though some are close (table S1).

Gamma values for the proportion of introgression calculated from the second round of HyDe results, placing individual samples in the P2 hybrid position, further confirmed a signal of introgression between the *Q. berberidifolia*/*durata* clade and species in the blue oak-southern scrub oak clade (figure 6). Estimates of gamma are close to or in some cases greater than 50% for these tests, indicating a high proportion of ancestral gene flow that obscures the "true" sister clade to *Q. berberidifolia*/*durata*. *Quercus garryana* appears to be the most likely sister as also found in my RAxML topology, but gamma values greater than 50% when comparing to *Q. pacifica* or in some cases *Q. john-tuckeri* explain the clade's phylogenetic instability in different analyses.

Further tests for patterns of introgression revealed elevated gamma values for tests involving *Q. berberidifolia*/*durata* in the P3 position and some species of southern California scrub oak in the P2 position. This pattern was especially notable when *Q. douglasii* was in P1 and strongest for *Q. pacifica* (figure 7).

Additional gamma values for tests with *Q. lobata* in the P3 position show very weak signals of introgression, particularly with *Q. berberidifolia*/*durata* as has previously been reported (Kim et al. 2018) (figure S3). I also found a stronger signal of introgression between *Q. lobata* and *Q. garryana* (figure S4). However, none of these comparisons had significant p-values as calculated by the conservative chi-squared test after Bonferroni correction for multiple tests.

Next, I compared the invariants-based method of calculating gamma values to admixture proportions inferred by fastSTRUCTURE, which should be an estimate of the same value. This involved a series of D-statistic tests that placed the 50 identified phylogenetic rogue samples in the P2 position, with *Q. garryana* in P1 and *Q. douglasii* in P3. This configuration was selected because many of the identified rogues are believed to be hybrids of those two parental lineages (a hybrid named *Q.* × *eplingii*). These tests resulted in a wide range of gamma values among the 50

rogue samples, reflecting the variability and frequent backcrossing expected in cases of contemporary hybridization without fertility barriers. Comparison to the proportional cluster membership in a predominantly *Q. douglasii* cluster as inferred by fastSTRUCTURE shows that gamma values calculated using the phylogenetic invariants method agree very closely with fastSTRUCTURE inference when the correct parental P1 and P3 populations are used (figure 8).

The five-taxon D-statistic, which places two distinct, but more closely related populations in the P3 position and tests for introgression involving either of them or their common ancestor with P2, showed that introgression with *Q. berberidifolia*/*durata* did not happen independently with species in the blue oak-southern scrub oak clade, but with their common ancestor (table 3). However, there is again a weak signal of further introgression, specifically with *Q. pacifica*.

DISCUSSION

A sequence of analyses conducted with dense sampling of 9 named taxa in the western North American white oak clade reveals consistent patterns of both contemporary and ancient introgression. The California scrub oak (*Quercus berberidifolia*) and its sister taxon leather oak (*Q. durata*) show descent from both the *Q. garryana* clade (at a point ancestral to the extant named varieties) and the common ancestor of the southern California scrub oaks plus *Q. douglasii* clade (figure 6). The extent of introgression obscures reconstruction of the proper historical sister relationship for this lineage, though the balance of genetic evidence suggests they are sister to *Q. garryana*. Meanwhile two modern descendants of these same ancestral lineages, namely Oregon white oak (*Q. garryana* var. *garryana*) and blue oak (*Q. douglasii*) also show extensive contemporary hybridization, as identified separately from the ancient signal using a combination of fastSTRUCTURE analyses, rogue identification, and gamma values (figures 4, 8).

Contemporary hybridization between *Q. garryana* var. *garryana* and *Q. douglasii* is not an unexpected result. These hybrids were recognized early by naturalists and have been discussed widely and even given the taxonomic name, *Quercus* × *eplingii* (Muller 1938, Dobzhansky 1940) and are the subject of the second chapter of this dissertation. However, ancient hybridization involving the ancestors of those taxa and *Q. berberidifolia* does represent a novel result. While it had not previously been suspected based on morphological or ecological evidence, it is a strong and consistent signal across multiple tests. In fact, given the high values estimated for gamma of almost 50% (figure 6), it seems possible that this is an example of ancient hybrid speciation (Gross and Rieseberg 2005, Meng and Kubatko 2009). However, it is also possible that a proto-*Quercus berberidifolia* lineage had already diverged from its sister lineage before undergoing extensive introgression with another lineage. Under this scenario, we may not be able to accurately determine which ancestral lineage *Q. berberidifolia* truly belonged to prior to introgression. My maximum likelihood analysis places *Q. berberidifolia/durata* sister to *Q. garryana* with high bootstrap support (figure S2), but other analyses have placed them

sister to the clade composed of *Q. douglasii* and the southern California scrub oaks, also with similarly high support (Fitz-Gibbon et al. 2017, Hipp et al. 2019).

Several scenarios have been suggested that could create significant D-statistic results even in the absence of introgression. The most widely discussed possibility is a model of ancestral populations structure in which two separate descendant lineages arise sequentially from the same subpopulation of a highly structured ancestral lineage (Durand et al. 2011, Eriksson and Manica 2012). Methods based on the site frequency spectrum and linkage disequilibrium have been used to differentiate ancestral population structure from more recent hybridization (Yang et al. 2012, Theunert and Slatkin 2017), but would not be useful in this case because it involves very ancient introgression. Therefore, this possibility cannot be ruled out, though the degree of structure in the ancestral population would have to be extremely strong to explain the values of gamma found here. As demonstrated by widespread contemporary hybridization also seen in these same oaks, the line between extreme population structure and phylogenetic divergence may not be clear. Under a model of ancestral population structure, it is not known what the divergence order would be for these lineages. Based on their modern distributions and life history, though, it could be suggested that from a highly structured ancestral population with a north-to-south genetic gradient, the ancestor of *Q. douglasii* and the southern California scrub oaks first diverged and then later the ancestor of *Q. berberidifolia*/*durata* diverged from the lineage that would become *Q. garryana*.

Convergent selection could also explain the significant D-statistic results and gamma values I reported here (Durand et al 2011), especially if it was acting on shared genetic polymorphisms in the ancestral gene pool. Given the incomplete nature of the genome annotation for oaks, this explanation cannot be ruled out either. Similarities in physiology and life history among the various evergreen shrub species do make this scenario seem possible, though again the amount of the genome inferred to be affected seems very large.

A third scenario, not discussed in the existing literature, but potentially relevant in this case, is divergent selection. Under this scenario, *Q. berberidifolia*/*durata* would be the earliest of the three lineages to diverge before *Q. garryana* and *Q. douglasii*/southern scrub oaks diverge from one another. This offers an explanation for the approximately equal relatedness of *Q. berberidifolia*/*durata* to the other two lineages, as it would be sister to them both. This is not the preferred explanation because members of those two lineages are never shown as sister. They are always placed as P1 and P3 in HyDe tests, representing the lowest 'BABA' site pattern and are also not reconstructed as sister lineages in maximum likelihood phylogenies. However, if there was strong divergent selection between them, this might explain the results. *Q. berberidifolia* would thus represent a more complete sample of the ancestral genetic variation in the group, possibly reflecting a more generalist ecological strategy.

Thus, though ancient hybridization is not the only possible explanation for the high values of gamma shown in my results, each of the alternative explanations would be surprising and

evolutionarily novel in its own right. Further work in this clade may help to resolve some of these open questions. Improved annotation of the oak genome and more intensive sampling, particularly in the southern portion of *Q. berberidifolia*'s range and of *Quercus dumosa*, will be especially important.

Beyond the pattern interpreted as ancient introgression and described above, all sampled individuals of the Channel Islands endemic *Quercus pacifica* were also determined as 10 - 20% admixed with *Quercus berberidifolia* in the fastSTRUCTURE results (figure 3). These samples were not identified as phylogenetic rogues using the CV of pairwise distances method, suggesting that it may represent a further case of ancient introgression in this clade. HyDe tests with *Q. pacifica* in the P2 position and *Q. berberidifolia* or *Q. durata* in P3 did have high gamma estimates, especially with *Q. douglasii* in P1 (figure 7). These tests were also significant even after Bonferroni correction (table S1). The five-taxon partition D-statistic also showed a signal of additional introgression between *Q. pacifica* and *Q. berberidifolia*/*durata* even after accounting for the more ancient introgression involving the ancestor of the southern California scrub oak clade (table 3). However, previous simulations have shown that D-statistics may have a high false positive rate in phylogenetic situations like this involving the descendant of one of the parental lineages hybridizing again with a lineage that was involved in an earlier period of introgression (Kong and Kubatko 2021). Nevertheless, agreement from multiple analyses does suggest there may be additional hybridization here between *Q. pacifica* and *Q. berberidifolia*. It is possible that this represents a signal of the unsampled taxon *Quercus dumosa*, acting as a ghost lineage in my analysis. Previous research has found frequent hybridization between *Q. dumosa* and *Q. berberidifolia* (Burge et al. 2019) while *Q. pacifica and Q. dumosa* have been reconstructed as sister taxa in some previous phylogenetic work (Fitz-Gibbon et al. 2017, Hipp et al. 2019).

Interestingly, there is little other evidence of more contemporary hybridization involving either *Quercus berberidifolia* or *Q. durata* with any of the other taxa descended from the lineages involved in the ancient hybridization identified here. In my geographically broad sample, only three individuals appear to be recent hybrids between *Q. berberidifolia* and *Q. douglasii* in fastSTRUCTURE results, two as putative F1s and just one potential backcross (figure 3). No hybrids were identified between *Q. berberidifolia* and *Q. john-tuckeri* despite their extensive overlap in range, which I sampled intensively. In a similarly broad sample of 144 individuals focused on *Quercus durata* (data not shown here), no hybrids of that taxon with *Q. douglasii* were found and three were identified as admixed with *Q. garryana* (with again only one apparent backcross). In a prior study, acorn viability tests showed extremely poor germination of acorns collected from putative hybrids between *Q. garryana* and both *Q. berberidifolia* and *Q. durata* (Tucker 1953).

In general, fastSTRUCTURE was found to be accurate for estimating contemporary hybridization and the results of phylogenetic rogue identification using the CV of pairwise distances method agreed well with it, at least for samples determined as more than 10% admixed

in fastSTRUCTURE. The exception is samples that were identified as admixed between southern scrub oaks (probably *Quercus john-tuckeri*) and *Q. douglasii* in fastSTRUCTURE. Few of these were selected as phylogenetic rogues in my analysis (see unpruned black terminal branches in figure 4). This is also a well-studied and named hybrid, *Quercus* × *alvordiana* (Tucker 1952), but unlike *Q.* × *eplingii*, its parent taxa are themselves immediate sister lineages. The phylogenetic rogue identification method used here is not expected to identify them well because they do not ever move as far in alternate topologies.

Similarly, in comparison to gamma values estimated using HyDe, fastSTRUCTURE also performs well (figure 8). STRUCTURE-type admixture inference is widely used, but often treated as a form of exploratory analysis (Meirmans 2015) due to uncertainties for the applicability of the underlying model (Marie et al. 2011, Toyama et al. 2020). This is why I used an alternative method, based on a maximum likelihood phylogeny and phylogenetic rogue identification, to identify lineages (populations) for the D-statistic tests in this study. However, the overall congruence among the methods suggests that the STRUCTURE model is identifying a similar pattern, at least in the case of contemporary hybridization. STRUCTURE may also be able to identify more ancient admixture, as in the example of *Quercus pacifica* discussed above, but not reliably and does not differentiate between them as was possible in this case using four- and five-taxon D-statistics.

In addition to recommendations to use many genetic markers or, as in the case of this study, genomic scale SNP data, STRUCTURE and other population genetic analyses benefit from dense sampling of individuals within taxa or lineages. I have also shown here that this kind of sampling also benefits analysis in a phylogenetic context when both ancient and contemporary hybridization violate the standard bifurcating phylogenetic model. Rogue identification using the CV of pairwise phylogenetic distances method was able to reveal a well-resolved underlying phylogeny despite the very large number of contemporary hybrids in the sample set. Meanwhile, D-statistics and gamma inference show consistent evidence for one or possibly two ancient reticulation events deeper within the phylogeny that help to clarify previous conflicts in phylogenies inferred for the western North American white oak clade.

Thus, there is evidence for hybridization at multiple points throughout the western North American white oak clade's history. This hybridization is not simply a transient phenomenon that becomes selected out over time, without having a long-term impact on the evolution of the lineages involved. Instead, ancient hybridization has left lasting effects on modern lineages even as they continue to hybridize in the present. This long-term pattern of hybridization strongly supports the multispecies or syngameon model of evolution for this group of oaks.

TABLES AND FIGURES

Table 2: Partial results of four-taxon D-statistic tests on combinations of 9 named taxa in the western North American white oak clade. Proportions of the ABBA, BABA, and BBAA site patterns are shown along with chi-squared test p-values on the D statistic, indicating introgression between the P2 and P3 populations. Significant p-values after Bonferroni correction for multiple tests are highlighted in red. See table 2 and table S1 for additional tests.

P1	P2	P3	OUT	pABBA	pBABA	pBBAA	D p-value
Quegar	Queber	Quedou	Quekel	0.01467	0.00974	0.01636	**1.32E-05**
Quegar	Queber	Quejoh	Quekel	0.01604	0.00909	0.01520	**1.24E-09**
Quegar	Queber	Quecor	Quekel	0.01551	0.00906	0.01599	**3.63E-09**
Quegar	Queber	Quepac	Quekel	0.01629	0.00919	0.01442	**3.27E-09**
Quegar	Quedur	Quedou	Quekel	0.01407	0.00980	0.01673	**0.000126**
Quegar	Quedur	Quejoh	Quekel	0.01545	0.00914	0.01547	**2.05E-08**
Quegar	Quedur	Quecor	Quekel	0.01503	0.00911	0.01626	**3.42E-08**
Quegar	Quedur	Quepac	Quekel	0.01588	0.00921	0.01460	**1.98E-08**
Quebre	Queber	Quedou	Quekel	0.01402	0.00967	0.01708	**4.77E-05**
Quebre	Queber	Quejoh	Quekel	0.01523	0.00918	0.01577	**2.09E-08**
Quebre	Queber	Quecor	Quekel	0.01480	0.00916	0.01655	**3.86E-08**
Quebre	Queber	Quepac	Quekel	0.01544	0.00934	0.01494	**9.05E-08**
Quebre	Quedur	Quedou	Quekel	0.01348	0.00979	0.01738	**0.000448**
Quebre	Quedur	Quejoh	Quekel	0.01472	0.00931	0.01600	**3.5E-07**
Quebre	Quedur	Quecor	Quekel	0.01440	0.00928	0.01677	**3.71E-07**
Quebre	Quedur	Quepac	Quekel	0.01511	0.00944	0.01505	**5.2E-07**

Table 2: Partial results of four-taxon D-statistic tests on combinations of 9 named taxa in the western North American white oak clade. Proportions of the ABBA, BABA, and BBAA site patterns are shown along with chi-squared test p-values on the D statistic, indicating introgression between the P2 and P3 populations. Significant p-values after Bonferroni correction for multiple tests are highlighted in red. See table 1 and table S1 for additional tests.

P1	P2	P3	OUT	pABBA	pBABA	pBBAA	D p-value
Quedur	Queber	Quedou	Quekel	0.01045	0.00966	0.02037	0.42365
Quedur	Queber	Quejoh	Quekel	0.01084	0.01007	0.01814	0.441017
Quedur	Queber	Quecor	Quekel	0.01058	0.00995	0.01911	0.504288
Quedur	Queber	Quepac	Quekel	0.01092	0.01037	0.01701	0.599647
Quedur	Queber	Quegar	Quekel	0.00992	0.01001	0.01786	0.931087
Quedur	Queber	Quebre	Quekel	0.01034	0.01028	0.01705	0.954909
Quegar	Quebre	Quedou	Quekel	0.00968	0.00906	0.02304	0.524127
Quegar	Quebre	Quejoh	Quekel	0.00979	0.00888	0.02233	0.351408
Quegar	Quebre	Quecor	Quekel	0.00968	0.00887	0.02292	0.389846
Quegar	Quebre	Quepac	Quekel	0.00990	0.00894	0.02128	0.348501
Quegar	Quebre	Queber	Quekel	0.01088	0.00968	0.01665	0.233307
Quegar	Quebre	Quedur	Quekel	0.01086	0.00982	0.01671	0.296059

Table 3: Results of five-taxon D-statistic tests (partitioned D-statistic) with Quekel as an outgroup. P-values for the D_1, D_2, and D_{12} statistics are based on chi-squared tests on introgression between P2 and $P3_1$, P2 and $P3_2$, or P2 and the common ancestor of $P3_1$ and $P3_2$, respectively. Significant p-values after Bonferroni correction for multiple tests are highlighted in red.

P1	P2	$P3_1$	$P3_2$	OUT	pABBAA	pABABA	pABBBA	D_1 p-value	D_2 p-value	D_{12} p-value
Quegarb	Queberd	Quejoh	Quecor	Quekel	0.00495	0.00452	0.0102	0.064768	0.161054	**1.33E-07**
Quegarb	Queberd	Quepac	Quecor	Quekel	0.00572	0.00481	0.00977	0.02844	0.182983	**7.03E-07**
Quegarb	Queberd	Quecor	Quedou	Quekel	0.00529	0.00449	0.00941	0.020856	0.772875	**4.09E-06**
Quegarb	Queberd	Quejoh	Quedou	Quekel	0.00526	0.00399	0.00988	0.011041	0.993875	**6.28E-06**
Quegarb	Queberd	Quepac	Quedou	Quekel	0.00631	0.00456	0.00921	**0.005374**	0.928661	**1.62E-05**
Quegarb	Queberd	Quepac	Quejoh	Quekel	0.00548	0.00497	0.01003	0.047396	0.12951	**7.58E-07**

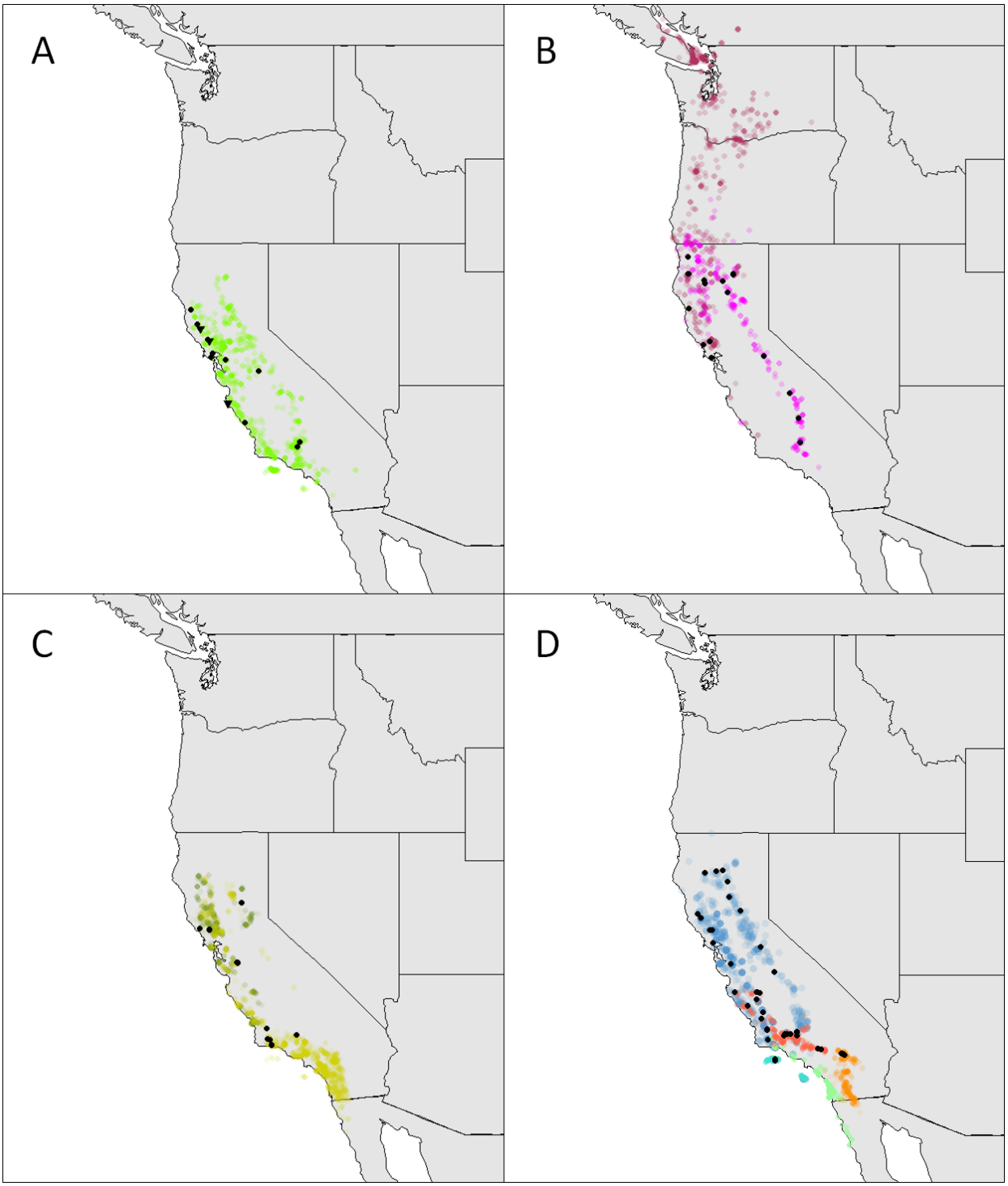

A
B
C
D

Figure 1 (previous page): Maps of sampling localities for individual samples included in the analyses. Colored dots show taxon ranges from herbarium specimen databases (CCH2, the Consortium of California Herbaria; and CPNWH, the Consortium of Pacific Northwest Herbaria). Colors match figure 5. (A) Valley oak (*Quercus lobata*), plus sampling localities for the outgroup black oak (*Quercus kelloggii*) shown with inverted triangles. (B) Oregon white oak (*Quercus garryana* var. *garryana*) and shrub-form Brewer's oak (*Quercus garryana* var. *breweri*), together referenced with the clade name "Quegarb" in the text. (C) California scrub oak (*Quercus berberidifolia*) and serpentine endemic leather oak (*Quercus durata*), together referred to with the clade name "Queberd" in the text. (D) Blue oak (Quercus douglasii) and the southern California scrub oaks (*Quercus cornelius-mulleri*, *Q. john-tuckeri*, *Q. pacifica*, and *Q. dumosa*). Note that *Q. dumosa* was not sampled for this study, but is shown in pale green on map D.

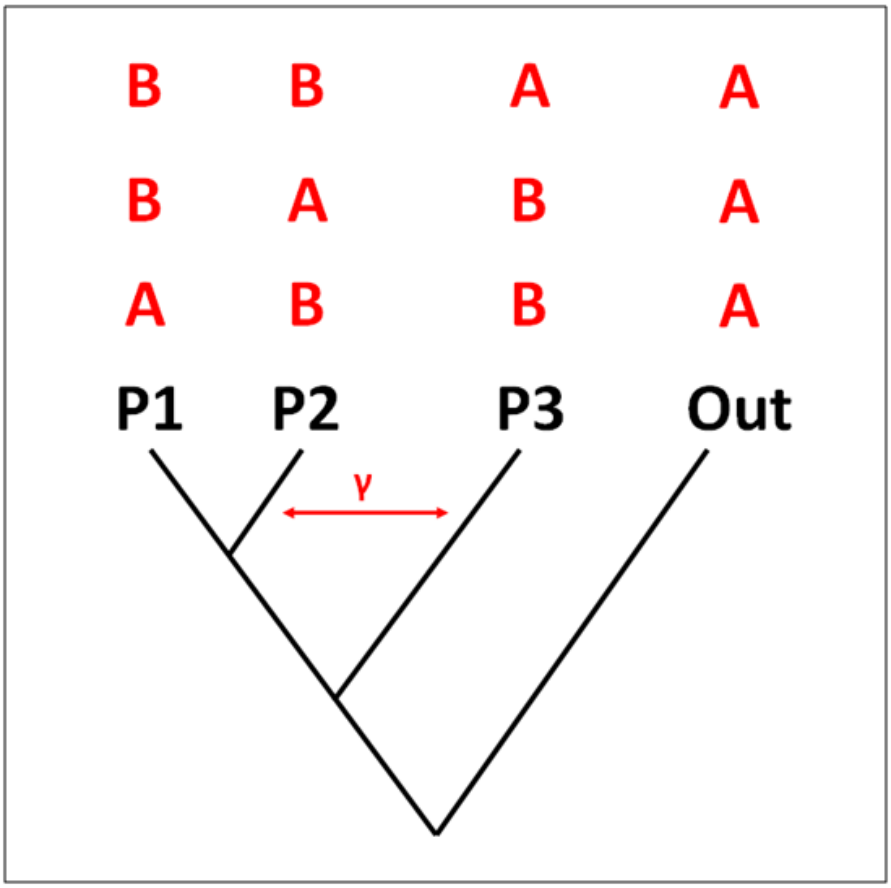


Figure 2: Schematic of the four-taxon D-statistic. Comparisons use loci with two alleles and the ancestral allele ('A') present in the outgroup and one ingroup tip, while the derived allele ('B') is present in the other two ingroup tips. Site pattern 'BBAA' is congruent with the phylogenetic hypothesis and expected to have the largest count. Site patterns 'ABBA' and 'BABA' are incongruent with the phylogeny and expected to be approximately equal (phylogenetically invariant) under incomplete lineage sorting. An excess of 'ABBA' patterns over 'BABA' indicates introgression between P2 and P3, which can be quantified as gamma (γ) the proportion of introgressed genome.

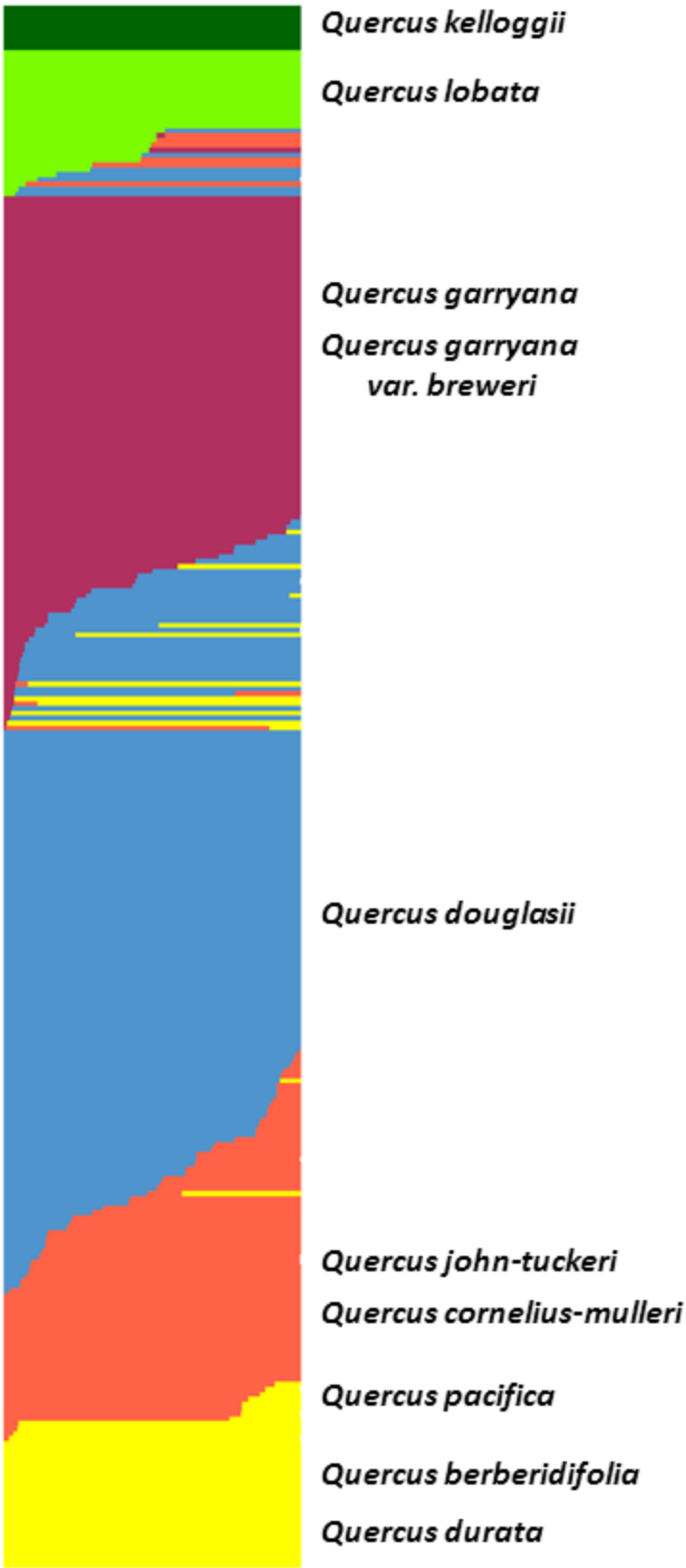

Quercus kelloggii
Quercus lobata
Quercus garryana
Quercus garryana var. breweri
Quercus douglasii
Quercus john-tuckeri
Quercus cornelius-mulleri
Quercus pacifica
Quercus berberidifolia
Quercus durata

Figure 3 (previous page): fastSTRUCTURE results at optimal K of six. The six clusters have been tentatively identified with one or more of the 9 named taxa (plus the outgroup *Quercus kelloggii*) based on field identification of the majority of samples included in the cluster.

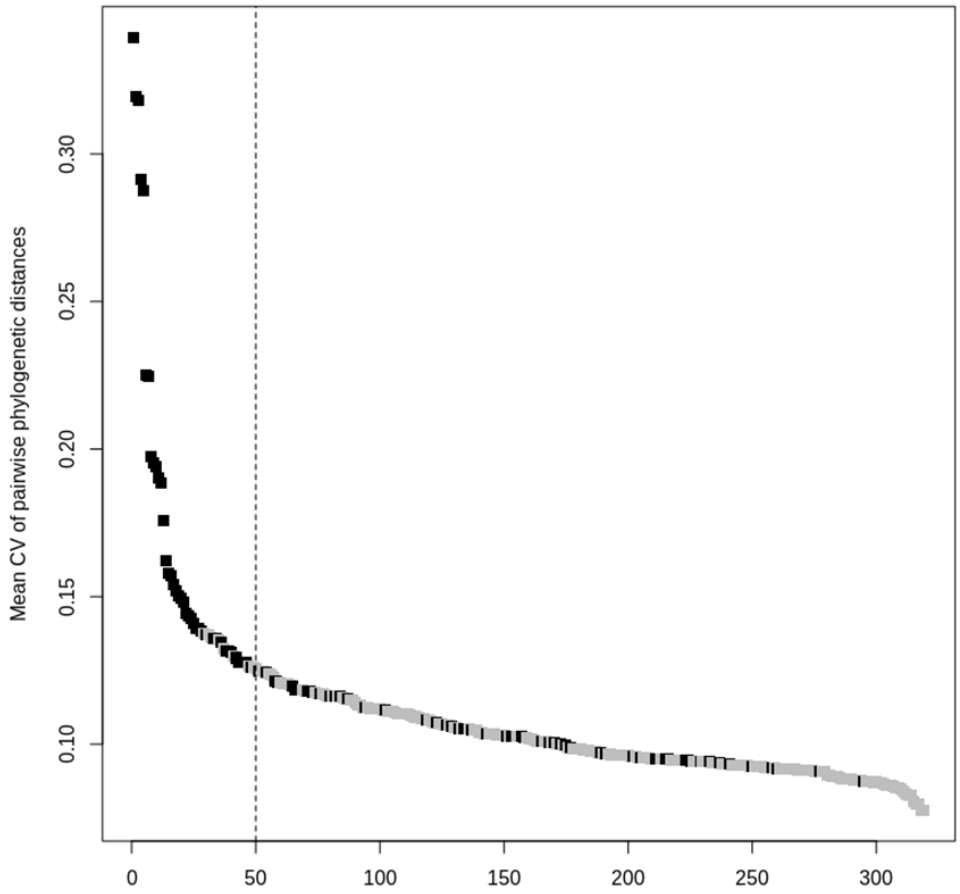


Figure 4: Identification of phylogenetic rogues in the sample set. Mean CV of pairwise phylogenetic distance across 100 RAxML bootstrap trees is sorted for 319 samples. Dotted line at 50 shows the chosen threshold for rogue identification. Black squares were determined as introgressed in the fastSTRUCTURE analysis.

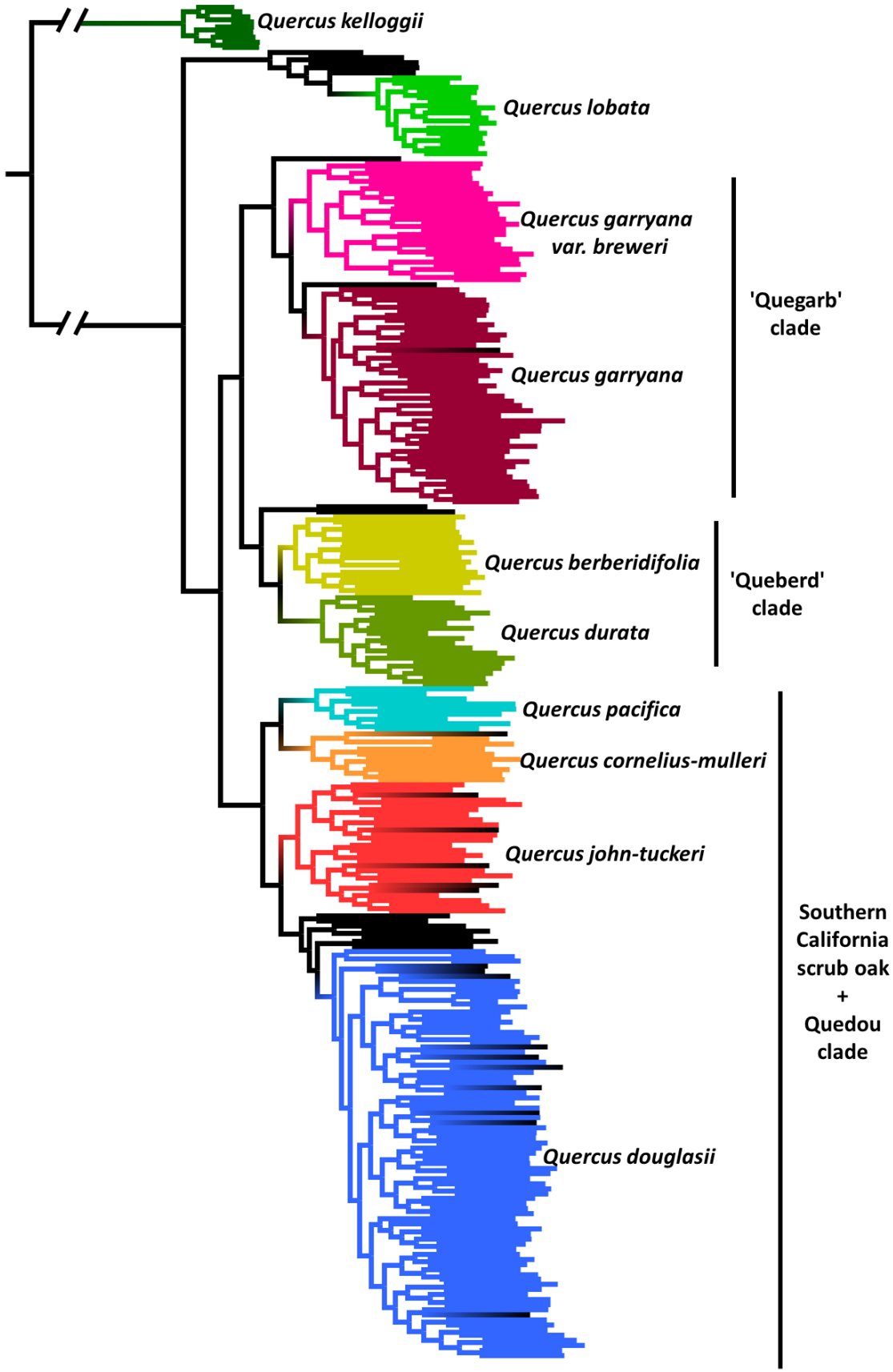

Quercus kelloggii
Quercus lobata
Quercus garryana var. breweri
Quercus garryana
'Quegarb' clade
Quercus berberidifolia
Quercus durata
'Queberd' clade
Quercus pacifica
Quercus cornelius-mulleri
Quercus john-tuckeri
Quercus douglasii
Southern California scrub oak + Quedou clade

Figure 5 (previous page): Maximum-likelihood phylogeny with rogue tips pruned. Samples determined as admixed by fastSTRUCTURE, but not selected as rogues, are shown with black terminal branches. Differently colored clades show the populations (taxa) used in D-statistic comparisons. Clades indicated on the right are the collapsed clades used in some analyses.

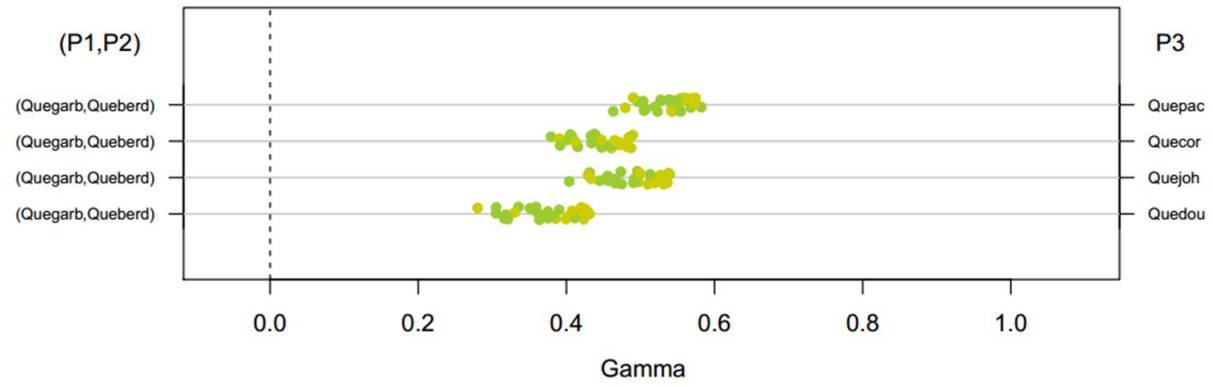


Figure 6: Gamma values from the phylogenetic invariants method for comparisons with *Quercus berberidifolia*/*durata* clade in the P2 position, *Q. garryana* clade in P1, and various species of the southern California scrub oaks and *Q. douglasii* in P3. Markers represent different samples within P2, colored as in figure 4.

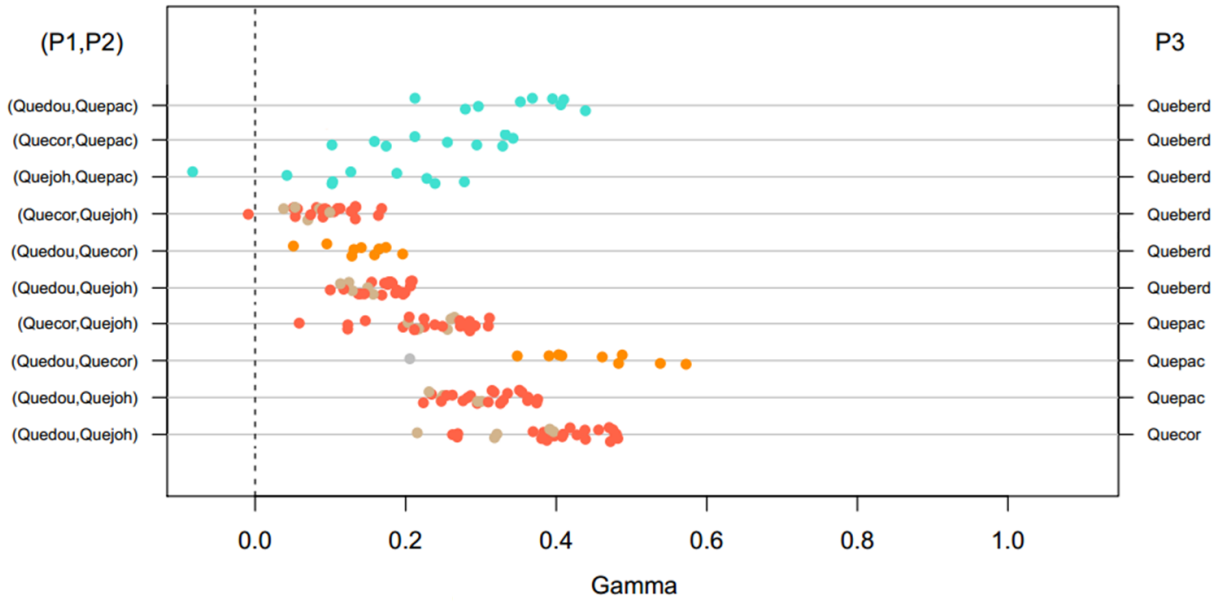


Figure 7: Gamma values from the phylogenetic invariants method for comparisons with various species of the southern California scrub oaks and *Q. douglasii* in P1 and P2 positions and *Q. berberidifolia*/*durata* or another southern scrub oak in P3. Markers represent different samples within P2, colored as in figure 4.

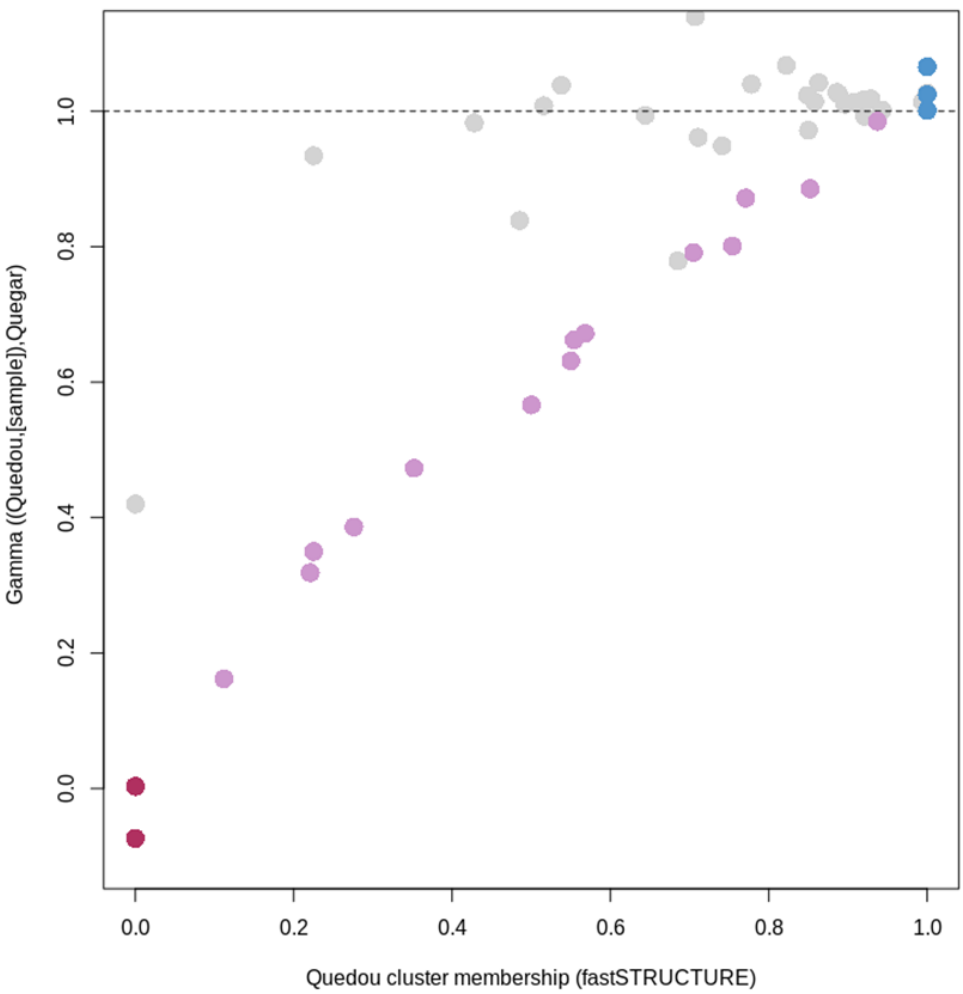


Figure 8: Comparison of gamma values calculated using the phylogenetic invariants method to cluster membership determined in fastSTRUCTURE. Gamma values were calculated from a comparison with *Quercus garryana* as P1 and *Q. douglasii* as P3. Dots are colored according to fastSTRUCTURE cluster membership: maroon for *Q. garryana* clade, blue for *Q. douglasii* clade, pink for their hybrids (*Q.* × *eplingii*), and grey for all other hybrids.

SUPPLEMENTAL MATERIAL

Table S1: Partial results of four-taxon D-statistic tests on combinations of 9 named taxa in the western North American white oak clade. Proportions of the ABBA, BABA, and BBAA site patterns are shown along with chi-squared test p-values on the D statistic, indicating introgression between the P2 and P3 populations. Significant p-values after Bonferroni correction for multiple tests are highlighted in red. See tables 1 and 2 for additional tests.

P1	P2	P3	OUT	pABBA	pBABA	pBBAA	D p-value
Quedou	Quejoh	Quecor	Quekel	0.01286	0.010437	0.014071	0.021737
Quedou	Quejoh	Quepac	Quekel	0.012506	0.00999	0.015798	0.024166
Quedou	Quecor	Quepac	Quekel	0.013221	0.011669	0.013733	0.177386
Quecor	Quejoh	Quepac	Quekel	0.012202	0.01125	0.01422	0.389824
Quedou	Quejoh	Queber	Quekel	0.011573	0.009195	0.020687	0.019989
Quedou	Quecor	Queber	Quekel	0.012032	0.010816	0.018521	0.240619
Quedou	Quepac	Queber	Quekel	0.014056	0.010761	0.016537	**0.00485**
Quecor	Quejoh	Queber	Quekel	0.01129	0.010157	0.019227	0.255223
Quejoh	Quepac	Queber	Quekel	0.012052	0.01111	0.01714	0.401067
Quecor	Quepac	Queber	Quekel	0.01283	0.010709	0.016859	0.054319
Quedou	Quejoh	Quedur	Quekel	0.011514	0.009099	0.021235	0.017015
Quedou	Quecor	Quedur	Quekel	0.012063	0.010621	0.019069	0.158297
Quedou	Quepac	Quedur	Quekel	0.01409	0.010432	0.016938	**0.001604**
Quecor	Quejoh	Quedur	Quekel	0.011121	0.010164	0.01983	0.32962
Quejoh	Quepac	Quedur	Quekel	0.012111	0.010857	0.017636	0.26003
Quecor	Quepac	Quedur	Quekel	0.012834	0.01058	0.017344	0.03903
Quedou	Quejoh	Quegar	Quekel	0.009476	0.009152	0.025823	0.745281
Quedou	Quecor	Quegar	Quekel	0.010241	0.01053	0.023397	0.778136
Quedou	Quepac	Quegar	Quekel	0.011602	0.010556	0.021435	0.356235
Quecor	Quejoh	Quegar	Quekel	0.009834	0.009135	0.024893	0.472865
Quejoh	Quepac	Quegar	Quekel	0.01042	0.009729	0.022883	0.52039
Quecor	Quepac	Quegar	Quekel	0.01099	0.009571	0.022308	0.182337
Quedou	Quejoh	Quebre	Quekel	0.009782	0.009117	0.025253	0.492248
Quedou	Quecor	Quebre	Quekel	0.010491	0.010468	0.022891	0.981996
Quedou	Quepac	Quebre	Quekel	0.011935	0.010495	0.020756	0.192619
Quecor	Quejoh	Quebre	Quekel	0.009992	0.009303	0.02428	0.461444
Quejoh	Quepac	Quebre	Quekel	0.010678	0.009889	0.022076	0.452916
Quecor	Quepac	Quebre	Quekel	0.01128	0.00975	0.021614	0.139597
Quepac	Quecor	Quelob	Quekel	0.009083	0.008757	0.02724	0.741459
Quecor	Quejoh	Quelob	Quekel	0.008418	0.008202	0.029721	0.810718
Quepac	Quejoh	Quelob	Quekel	0.008909	0.008265	0.027908	0.518202

Quejoh	Quedou	Quelob	Quekel	0.008587	0.007774	0.030431	0.383398
Quecor	Quedou	Quelob	Quekel	0.009709	0.008616	0.027824	0.251544
Quepac	Quedou	Quelob	Quekel	0.010316	0.008789	0.026103	0.148578
Quedou	Queber	Quelob	Quekel	0.011022	0.010781	0.020778	0.821291
Quejoh	Queber	Quelob	Quekel	0.010527	0.009524	0.022596	0.32497
Quecor	Queber	Quelob	Quekel	0.011105	0.009895	0.021961	0.22697
Quepac	Queber	Quelob	Quekel	0.010587	0.00897	0.023362	0.125808
Quedou	Quedur	Quelob	Quekel	0.011456	0.010856	0.020126	0.576016
Quejoh	Quedur	Quelob	Quekel	0.010798	0.009575	0.021993	0.2299
Quecor	Quedur	Quelob	Quekel	0.011337	0.009967	0.021554	0.169547
Quepac	Quedur	Quelob	Quekel	0.010738	0.008901	0.022958	0.081643
Queber	Quegar	Quelob	Quekel	0.010731	0.009825	0.02094	0.388702
Quedur	Quegar	Quelob	Quekel	0.010476	0.009814	0.021126	0.523097
Quedou	Quegar	Quelob	Quekel	0.013038	0.012	0.016772	0.379409
Quejoh	Quegar	Quelob	Quekel	0.013046	0.011245	0.017008	0.120442
Quecor	Quegar	Quelob	Quekel	0.013571	0.011439	0.016863	0.060719
Quepac	Quegar	Quelob	Quekel	0.012976	0.010573	0.017794	0.043763
Queber	Quebre	Quelob	Quekel	0.01068	0.009561	0.021989	0.261641
Quedur	Quebre	Quelob	Quekel	0.010483	0.009592	0.022047	0.364519
Quedou	Quebre	Quelob	Quekel	0.01307	0.011805	0.017099	0.262617
Quejoh	Quebre	Quelob	Quekel	0.012967	0.010971	0.017588	0.070211
Quecor	Quebre	Quelob	Quekel	0.013472	0.011259	0.017396	0.039584
Quepac	Quebre	Quelob	Quekel	0.012855	0.010276	0.018362	0.023675

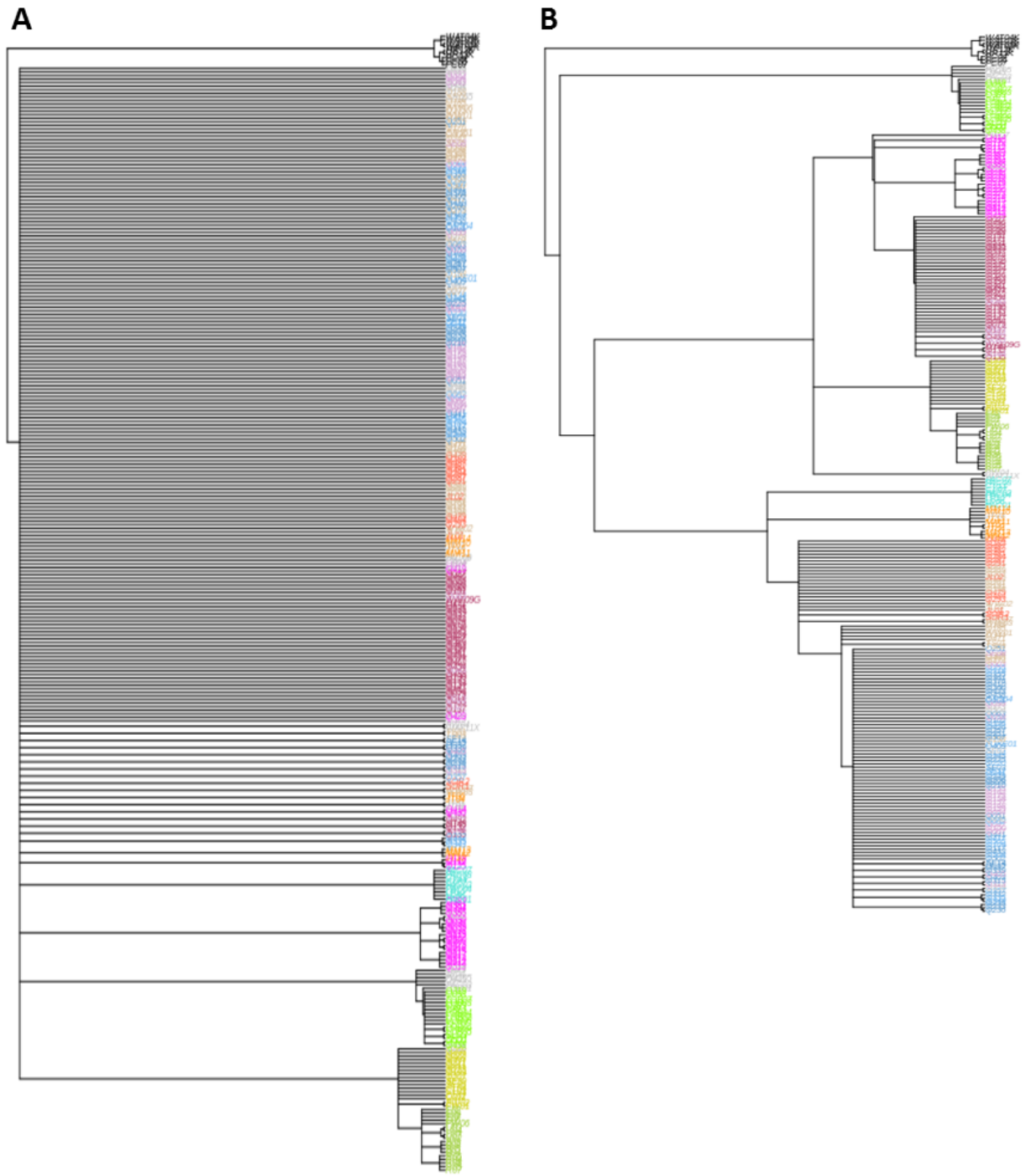
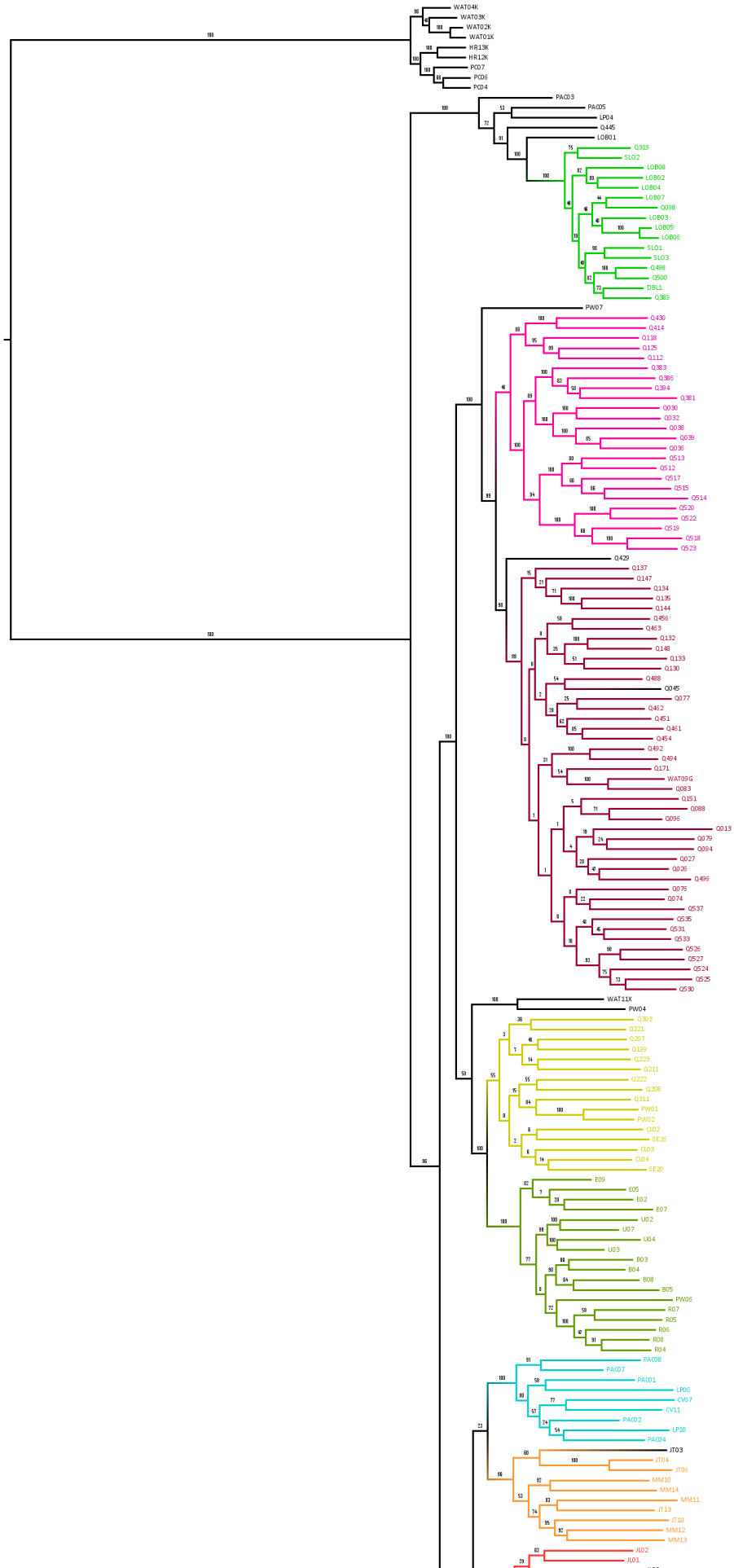


Figure S1: 95% majority rule consensus of 100 RAxML bootstrap trees before (A) and after (B) pruning 50 phylogenetic rogue tips.


WAT04K
WAT03K
WAT02K
WAT01K
HR13K
HR12K
PC07
PC06
PC04
PAC03
PAC05
LP04
Q445
LOB01
Q319
SLO2
LOB08
LOB02
LOB04
LOB07
Q098
LOB03
LOB05
LOB06
SLO1
SLO3
Q498
Q500
DBL1
Q389
PW07
Q430
Q414
Q118
Q125
Q112
Q383
Q386
Q394
Q381
Q030
Q032
Q038
Q039
Q036
Q513
Q512
Q517
Q515
Q514
Q520
Q522
Q519
Q518
Q523
Q429
Q137
Q147
Q134
Q135
Q144
Q456
Q463
Q132
Q148
Q133
Q130
Q488
Q045
Q077
Q462
Q451
Q461
Q454
Q492
Q494
Q171
WAT09G
Q083
Q151
Q088
Q096
Q013
Q079
Q094
Q027
Q026
Q496
Q076
Q074
Q537
Q535
Q531
Q533
Q526
Q527
Q524
Q525
Q530
WAT11X
PW04
Q303
Q221
Q207
Q199
Q223
Q211
Q222
Q208
Q311
PW01
PW02
CL02
SE26
CL03
CL04
SE20
E09
E05
E02
E07
U02
U07
U04
U03
B03
B04
B08
B05
PW06
R07
R05
R06
R08
R04
PAC08
PAC07
PAC01
LP06
CV07
CV11
PAC02
LP10
PAC04
JT03
JT04
JT06
MM10
MM14
MM11
JT13
JT10
MM12
MM13
JL02
JL01


R07
R05
R06
R08
R04
PAC08
PAC07
PAC01
LP06
CV07
CV11
PAC02
LP10
PAC04
JT03
JT04
JT06
MM10
MM14
MM11
JT13
JT10
MM12
MM13
JL02
JL01
JL03
WW03
WW07
WW02
Q293
SOR1
SOR2
Q398
Q484
Q479
Q487
Q478
Q482
Q481
Q261
Q263
Q269
Q267
Q191
Q183
Q291
Q282
Q287
Q284
Q344
Q471
Q188
TS03
TS01
Q185
WW01
Q251
SE15
SE14
Q474
Q465
Q305
Q255
Q339
Q337
Q332
Q343
Q335
Q433
Q502
Q508
Q505
Q417
Q357
Q359
Q047
Q062
Q436
Q051
Q054
Q067
Q100
Q441
Q313
Q315
SE12
Q003
Q225
Q159
Q173
Q162
Q155
Q168
Q157
Q158
Q245
Q109
Q117
Q124
SE03
SE07
Q210
Q235
Q236
Q233
SE08
Q206
Q249
Q228
Q410
Q361
Q241
Q220
Q200
Q403
Q405
Q423
Q257
Q426
Q063
Q102
Q058
LUKE01
Q376
Q507
Q409
Q198
Q504
Q373
Q226
Q099
Q438
Q101
Q318
CR564
Q351
Q355

Figure S2 (previous pages): Maximum-likelihood phylogeny with rogue tips pruned as in figure 4, but expanded, with tip labels, and bootstrap branch supports shown.

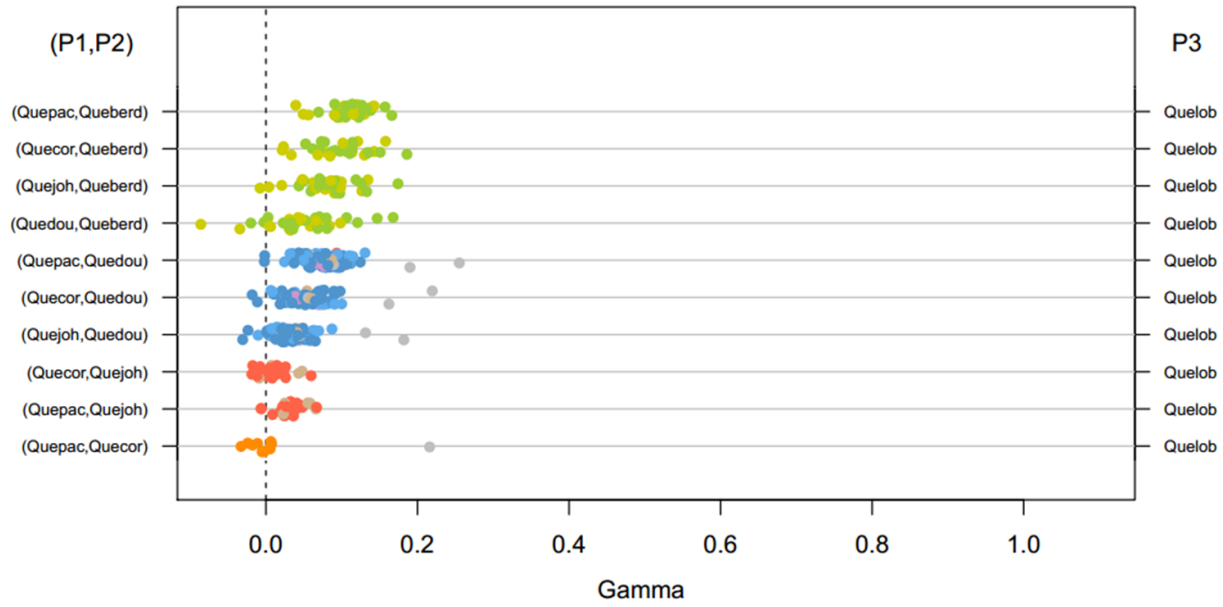


Figure S3: Gamma values from the phylogenetic invariants method for comparisons with *Q. lobata* in P3. Markers represent different samples within P2, colored as in figure 4.

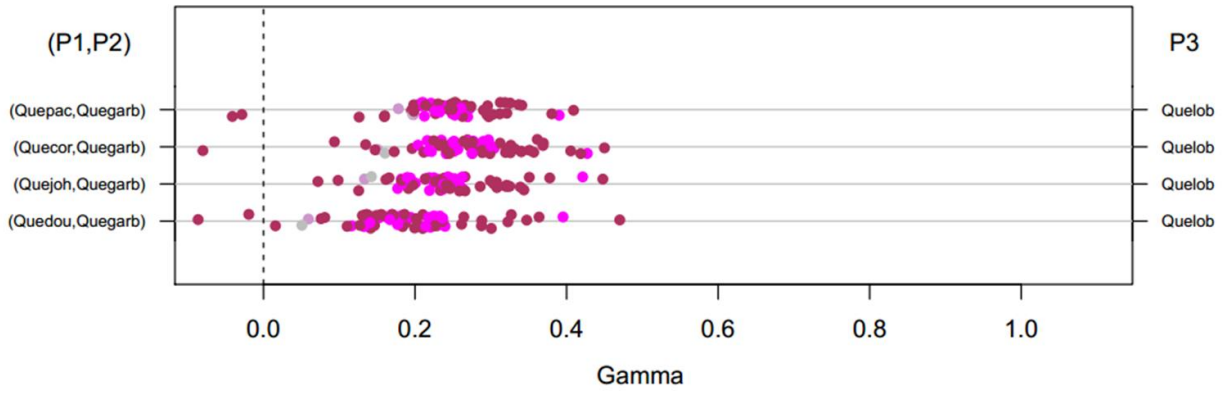


Figure S4: Gamma values from the phylogenetic invariants method for comparisons with *Q. garryana* in P2 and *Q. lobata* in P3. Markers represent different samples within P2, colored as in figure 4.

REFERENCES

Aberer, A. J., D. Krompass, and A. Stamatakis. 2013. Pruning rogue taxa improves phylogenetic accuracy: An efficient algorithm and webservice. *Syst. Biol.* 62:162–166.

Ackerly, D. D., M. M. Kling, M. L. Clark, P. Papper, M. F. Oldfather, A. L. Flint, and L. E. Flint. 2020. Topoclimates, refugia, and biotic responses to climate change. *Front. Ecol. Environ.* 18:288–297.

Aitken, S. N. and M. C. Whitlock. 2013. Assisted gene flow to facilitate local adaptation to climate change. *Annu. Rev. Ecol. Evol. Syst.* 44:367–388.

Augspurger, C. K. 2013. Reconstructing patterns of temperature, phenology, and frost damage over 124 years: Spring damage risk is increasing. *Ecology* 94:41–50.

Blischak, P. D., J. Chifman, A. D. Wolfe, and L. S. Kubatko. 2018. HyDe: A Python package for genome-scale hybridization detection. *Syst. Biol.* 67:821–829.

Browne, L., J. W. Wright, S. Fitz-Gibbon, P. F. Gugger, and V. L. Sork. 2019. Adaptational lag to temperature in valley oak (*Quercus lobata*) can be mitigated by genome-informed assisted gene flow. *Proc. Natl. Acad. Sci. U. S. A.* 116:25179–25185.

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. *Am. J. Bot.* 106:864–878.

Cahill, J. A., P. D. Heintzman, K. Harris, M. D. Teasdale, J. Kapp, A. E. Soares, I. Stirling, D. Bradley, C. J. Edwards, K. Graim, and A. A. Kisleika. 2018. Genomic evidence of widespread admixture from polar bears into brown bears during the last ice age. *Mol. Biol. Evol.* 35:1120-1129.

Cavender-Bares, J. 2019. Diversification, adaptation, and community assembly of the American oaks (*Quercus*), a model clade for integrating ecology and evolution. *New Phytol.* 221:669–692.

Cavender-Bares, J., K. Kitajima, and F. A. Bazzaz. 2004. Multiple trait associations in relation to habitat differentiation among 17 Floridian oak species. *Ecol. Monogr.* 74:635–662.

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

Cavender-Bares, J. and J. A. Ramírez-Valiente. 2017. Physiological evidence from common garden experiments for local adaptation and adaptive plasticity to climate in American live oaks (*Quercus* section Virentes): Implications for conservation under global change. Pages 107–135 in Oaks Physiological Ecology: Exploring the Functional Diversity of Genus Quercus L.

Clausen, J., D. D. Keck, and W. M. Hiesey. 1941. Regional differentiation in plant species. *Am. Nat.* 75:231–250.

Coyne, J. A. 1992. Genetics and speciation. *Nature* 355:511–515.

Coyne, J. A. and H. A. Orr. 1989. Patterns of speciation in *Drosophila*. *Evolution* 43:361–381.

Craft, K. J. and M. V. Ashley. 2007. Landscape genetic structure of bur oak (*Quercus macrocarpa*) savannas in Illinois. *For. Ecol. Manage.* 239:13–20.

Craft, K. J., M. V. Ashley, and W. D. Koenig. 2002. Limited hybridization between *Quercus lobata* and *Quercus douglasii* (Fagaceae) in a mixed stand in Central Coastal California. *Am. J. Bot.* 89:1792–1798.

Curtu, A. L., O. Gailing, and R. Finkeldey. 2007. Evidence for hybridization and introgression within a species-rich oak (*Quercus* spp.) community. *BMC Evol. Biol.* 7:218.

Davies, M. S. and R. W. Snaydon. 1976. Rapid population differentiation in a mosaic environment. III. Measures of selection pressures. *Heredity* 1:59–66.

Deacon, N. J. and J. Cavender-Bares. 2015. Limited pollen dispersal contributes to population genetic structure but not local adaptation in *Quercus oleoides* forests of Costa Rica. *PLoS One* 10:e0138783.

Denk, T., G. W. Grimm, P. S. Manos, M. Deng, A. Hipp, A. S. Ørsted, W. Trelease, O. Karl, A. Schwarz, A. Camus, Y. L. Menitsky, and K. C. Nixon. 2017. An updated infrageneric classification of the oaks: Review of previous taxonomic schemes and synthesis of evolutionary patterns. Pages 13–38 in Oaks Physiological Ecology. Exploring the Functional Diversity of Genus *Quercus* L.

Derory, J., C. Scotti-Saintagne, E. Bertocchi, L. Le Dantec, N. Graignic, A. Jauffres, M. Casasoli, E. Chancerel, C. Bodénès, F. Alberto, and A. Kremer. 2010. Contrasting relations between diversity of candidate genes and variation of bud burst in natural and segregating populations of European oaks. *Heredity (Edinb).* 105:401–411.

Devaux, C. and R. Lande. 2008. Incipient allochronic speciation due to non-selective assortative mating by flowering time, mutation and genetic drift. *Proc. R. Soc. B Biol. Sci.* 275:2723–2732.

DiVittorio, C. T., S. Singhal, A. B. Roddy, F. Zapata, D. D. Ackerly, B. G. Baldwin, C. R. Brodersen, A. Búrquez, P. V. A. Fine, M. P. Flores, E. Solis, J. Morales-Villavicencio, D. Morales-Arce, and D. W. Kyhos. 2021. Natural selection maintains species despite frequent hybridization in the desert shrub *Encelia*. *Proc. Natl. Acad. Sci. U. S. A.* 117:33373–33383.

Dixit, A., T. Kolb, and O. Burney. 2020. Provenance geographical and climatic characteristics influence budburst phenology of southwestern ponderosa pine seedlings. *Forests* 11:1–10.

Dobzhansky, T. 1937. Genetic nature of species differences. *Am. Nat.* 71:404–420.

Dobzhansky, T. 1940. Speciation as a stage in evolutionary divergence. *Am. Nat.* 74:312–321.

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

Dow, B. D. and M. V. Ashley. 1996. Microsatellite analysis of seed dispersal and parentage of saplings in bur oak, *Quercus macrocarpa*. *Mol. Ecol.* 5:615–627.

Doyle, J. J. and J. F. Doyle. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem. Bull.* 19:11–15.

Durand, E. Y., N. Patterson, D. Reich, and M. Slatkin. 2011. Testing for ancient admixture between closely related populations. *Mol. Biol. Evol.* 28:2239–2252.

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 supplementary information. *Evolution* 69:2587–2601.

Eaton, D.A. and I. Overcast. 2020. ipyrad: Interactive assembly and analysis of RADseq datasets. *Bioinformatics*, *36*(8):2592–2594.

Eaton, D. A. and R. H. Ree. 2013. Inferring phylogeny and introgression using RADseq data: An example from flowering plants (*Pedicularis*: Orobanchaceae). *Syst. Biol.* 62:689–706.

Endler, J. A. 1977. Geographic variation, speciation, and clines. Princeton University Press.

Epling, C. 1947. Actual and potential gene flow in natural populations. *Am. Nat.* 81:104–113.

Eriksson, A. and A. Manica. 2012. Effect of ancient population structure on the degree of polymorphism shared between modern human populations and ancient hominins. *Proc. Natl. Acad. Sci.* 109:13956–13960.

Fairley, D. and G. L. Batchelder. 1986. A study of oak-pollen production and phenology in northern California: Prediction of annual variation in pollen counts based on geographic and meterologic factors. *J. Allergy Clin. Immunol.* 78:300–307.

Fisher, R. A. 1950. Gene frequencies in a cline determined by selection and diffusion. *Biometrics* 6:353–361.

Fitz-Gibbon, S., A. Hipp, K. Pham, P. 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* subgenus Quercus). *Genome* 60:743–755.

Fitzpatrick, M. C. and S. R. Keller. 2015. Ecological genomics meets community-level modelling of biodiversity: mapping the genomic landscape of current and future

environmental adaptation. *Ecol. Lett.* 18:1–16.

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. *Ecol. Process.* 2:25.

García-Mozo, H., C. Galán, M. J. Aira, J. Belmonte, C. Díaz De La Guardia, D. Fernández, A. M. Gutierrez, F. J. Rodriguez, M. M. Trigo, and E. Dominguez-Vilches. 2002. Modelling start of oak pollen season in different climatic zones in Spain. *Agric. For. Meteorol.* 110:247–257.

Garrison, E. and G. Marth. 2012. Haplotype-based variant detection from short-read sequencing. BioRxiv [preprint] 1–9.

Gaudinier, A. and B. K. Blackman. 2020. Evolutionary processes from the perspective of flowering time diversity. *New Phytol.* 225:1883–1898.

Gerst, K. L., N. L. Rossington, and S. J. Mazer. 2017. Phenological responsiveness to climate differs among four species of *Quercus* in North America. *J. Ecol.* 105:1610–1622.

González-Rodríguez, V., I. C. Barrio, and R. Villar. 2012. Within-population variability influences early seedling establishment in four Mediterranean oaks. *Acta Oecologica* 41:82–89.

Green, R. E., J. Krause, A. W. Briggs, T. Maricic, U. Stenzel, M. Kircher, N. Patterson, H. Li, W. Zhai, M. H. Y. Fritz, N. F. Hansen, E. Y. Durand, A. S. Malaspinas, J. D. Jensen, T. Marques-Bonet, C. Alkan, K. Prüfer, M. Meyer, H. A. Burbano, J. M. Good, R. Schultz, A. Aximu-Petri, A. Butthof, B. Höber, B. Höffner, M. Siegemund, A. Weihmann, C. Nusbaum, E. S. Lander, C. Russ, N. Novod, J. Affourtit, M. Egholm, C. Verna, P. Rudan, D. Brajkovic, Ž. Kucan, I. Gušic, V. B. Doronichev, L. V. Golovanova, C. Lalueza-Fox, M. De La Rasilla, J. Fortea, A. Rosas, R. W. Schmitz, P. L. F. Johnson, E. E. Eichler, D. Falush, E. Birney, J. C. Mullikin, M. Slatkin, R. Nielsen, J. Kelso, M. Lachmann, D. Reich, and S. Pääbo. 2010. A draft sequence of the Neandertal genome. *Science* 328:710–722.

Gross, B. L. and L. H. Rieseberg. 2005. The ecological genetics of homoploid hybrid speciation. *J. Hered.* 96:241–252.

Gugger, P. F., S. J. Cokus, and V. L. Sork. 2016. Association of transcriptome-wide sequence variation with climate gradients in valley oak (*Quercus lobata*). *Tree Genet. Genomes* 12:1–14.

Hailer, F. and J. A. Leonard. 2008. Hybridization among three native North American *Canis* species in a region of natural sympatry. *PLoS One* 3.

Hammer, M. F., A. E. Woerner, F. L. Mendez, J. C. Watkins, and J. D. Wall. 2011. Genetic

evidence for archaic admixture in Africa. *Proc. Natl. Acad. Sci. U. S. A.* 108:15123–15128.

Harrison, R. G. 1986. Pattern and process in a narrow hybrid zone. *Heredity* 56:337–349.

Harrison, R. G. and S. M. Bogdanowicz. 1997. Patterns of variation and linkage disequilibrium in a field cricket hybrid zone. *Evolution* 51:493.

Hendry, A. P. and T. Day. 2005. Population structure attributable to reproductive time: Isolation by time and adaptation by time. *Mol. Ecol.* 14:901–916.

Hewitt, G. M. 1988. Hybrid zones-natural laboratories for evolutionary studies. *Trends Ecol. Evol.* 3:158–167.

Hillis, D. M., T. A. Heath, and K. St. John. 2005. Analysis and visualization of tree space. *Syst. Biol.* 54:471–482.

Hipp, A. L. 2015. Should hybridization make us skeptical of the oak phylogeny? *Int. Oaks* 26:9–18.

Hipp, A. L., P. S. Manos, A. González-Rodríguez, M. Hahn, M. Kaproth, J. D. McVay, S. V. Avalos, and J. Cavender-Bares. 2017. Sympatric parallel diversification of major oak clades in the Americas and the origins of Mexican species diversity. *New Phytol.* 217:439–452.

Hipp, A. L., P. S. Manos, M. Hahn, M. Avishai, C. Bodénès, J. Cavender-Bares, A. Crowl, M. Deng, T. Denk, S. Fitz-Gibbon, O. Gailing, M. S. González-Elizondo, A. González-Rodríguez, G. W. Grimm, X.-L. Jiang, A. Kremer, I. Lesur, J. D. McVay, C. Plomion, H. Rodríguez-Correa, E.-D. Schulze, M. C. Simeone, V. L. Sork, and S. Valencia-Avalos. 2019. Genomic landscape of the global oak phylogeny. *New Phytol.* 226:1198–1212.

Howe, G. T., S. N. Aitken, D. B. Neale, K. D. Jermstad, N. C. Wheeler, and T. H. H. Chen. 2003. From genotype to phenotype: unraveling the complexities of cold adaptation in forest trees. *Can. J. Bot.* 81:1247–1266.

Ishida, Y. 2009. Sewall Wright and Gustave Malécot on isolation by distance. *Philos. Sci.* 76:784–796.

Jensen, J., A. Larsen, L. R. Nielsen, and J. Cottrell. 2009. Hybridization between *Quercus robur* and *Q. petraea* in a mixed oak stand in Denmark. *Ann. For. Sci.* 66:706–706.

Jiggins, C. D. and J. Mallet. 2000. Bimodal hybrid zones and speciation. *Trends Ecol. Evol.* 15:250–255.

Jombart, T. and I. Ahmed. 2011. Adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics* 27:3070–3071.

Jombart, T., M. Kendall, J. Almagro-Garcia, and C. Colijn. 2017. Treespace: Statistical exploration of landscapes of phylogenetic trees. *Mol. Ecol. Resour.* 17:1385–1392.

Kasprzyk, I. 2009. Forecasting the start of quercus pollen season using several methods - the evaluation of their efficiency. *Int. J. Biometeorol.* 53:345–353.

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). *Mol. Ecol.* 27:4556–4571.

Knapp, E. E., M. A. Goedde, and K. J. Rice. 2001. Pollen-limited reproduction in blue oak: Implications for wind pollination in fragmented populations. *Oecologia* 128:48–55.

Koenig, W. D., J. M. H. Knops, W. J. Carmen, and I. S. Pearse. 2014. What drives masting? The phenological synchrony hypothesis. *Ecology* 96:140702180423006.

Kong, S. and L. S. Kubatko. 2021. Comparative performance of popular methods for hybrid detection using genomic data. *Syst. Biol.* NN:nnn–nnn.

Kremer, A., O. Ronce, J. J. Robledo-Arnuncio, F. Guillaume, G. Bohrer, R. Nathan, J. R. Bridle, R. Gomulkiewicz, E. K. Klein, K. Ritland, A. Kuparinen, S. Gerber, and S. Schueler. 2012. Long-distance gene flow and adaptation of forest trees to rapid climate change. *Ecol. Lett.* 15:378–392.

Kubatko, L. S. and J. Chifman. 2019. An invariants-based method for efficient identification of hybrid species from large-scale genomic data. *BMC Evol. Biol.* 19:1–13.

Lake, J. A. 1987. A rate-independent technique for analysis of nucleic acid sequences: Evolutionary parsimony. *Mol. Biol. Evol.* 4:167–191.

Lamichhaney, S., F. Han, M. T. Webster, L. Andersson, B. R. Grant, and P. R. Grant. 2017. Rapid hybrid speciation in Darwin's finches. *Science* 359:224–228.

Langmead, B. and S. L. Salzberg. 2012. Fast gapped-read alignment with bowtie-2. *Nat. Methods* 9:357–360.

Legendre, P., D. Borcard, and P. R. Peres-Neto. 2005. Analyzing beta diversity: Partitioning the spatial variation of community composition data. *Ecol. Monogr.* 75:435–450.

Leroy, T., J. M. Louvet, C. Lalanne, G. Le Provost, K. Labadie, J. M. Aury, S. Delzon, C. Plomion, and A. Kremer. 2019. Adaptive introgression as a driver of local adaptation to climate in European white oaks. *New Phytol.* 226:1171–1182.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, G. P. Data, and T. Sam. 2009. The sequence alignment / map format and samtools. *Bioinformatics* 25:2078–2079.

Lotsy, J. P. 1925. Species or linneon? *Genetica* 7:487–506.

Maddison, W. P. and D. R. Maddison. 2008. Mesquite: A modular system for evolutionary

analysis. *Evolution* 62:1103-1118.

Mallet, J. 2008. Hybridization, ecological races and the nature of species: Empirical evidence for the ease of speciation. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 363:2971–2986.

Marie, A. D., L. Bernatchez, and D. Garant. 2011. Empirical assessment of software efficiency and accuracy to detect introgression under variable stocking scenarios in brook charr (*Salvelinus fontinalis*). *Conserv. Genet.* 12:1215–1227.

Mayr, E. 1947. Ecological factors in speciation. *Evolution* 1:263–288.

McBride, J. R., E. Norberg, J. Bertenshaw, S. Kloss, and A. Mossadegh. 1997. Genetic variation in shoot growth, phenology, and mineral accumulation of northern and central Sierra Nevada foothill populations of blue oak. Pages 117–126 in Fourth Symposium on Oak Woodlands.

McKay, J. K., C. E. Christian, S. Harrison, and K. J. Rice. 2005. How local is local? A review of practical and conceptual issues in the genetics of restoration. *Restor. Ecol.* 13:432–440.

McLaughlin, B. C., R. Blakey, A. P. Weitz, X. Feng, B. J. Brown, D. D. Ackerly, T. E. Dawson, and S. E. Thompson. 2020. Weather underground: subsurface hydrologic processes mediate tree vulnerability to extreme climatic drought. *Glob. Chang. Biol.* 26:3091–3107.

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* 10:1–10.

McVay, J. D., A. L. Hipp, and P. S. Manos. 2017b. A genetic legacy of introgression confounds phylogeny and biogeography in oaks. *Proc. R. Soc. London B* 284:20170300.

Meirmans, P. G. 2015. Seven common mistakes in population genetics and how to avoid them. *Mol. Ecol.* 24:3223–3231.

Meng, C. and L. S. Kubatko. 2009. Detecting hybrid speciation in the presence of incomplete lineage sorting using gene tree incongruence: A model. *Theor. Popul. Biol.* 75:35–45.

Menon, M., J. C. Bagley, C. J. Friedline, A. V Whipple, A. W. Schoettle, A. Leal-Saenz, C. Wehenkel, L. Molina-Freaner, Francisco Flores-Renteria, M. S. Gonzalez-Elizondo, R. A. Sniezko, S. A. Cushman, K. M. Waring, and A. J. Eckert. 2018. The role of hybridization during ecological divergence of southwestern white pine (*Pinus strobiformis*) and limber pine (*P. flexilis*). *Mol. Ecol.* 27:1245–1260.

Mensing, S. 2014. The paleohistory of California oaks. Pages 35–47 in Proceedings of the Seventh Symposium on Oak Woodlands and Management.

vander Mijnsbrugge, K., A. Bischoff, and B. Smith. 2010. A question of origin: Where and how to collect seed for ecological restoration. *Basic Appl. Ecol.* 11:300–311.

Millar, C. I. 1983. A steep cline in *Pinus muricata*. *Evolution* 37:311–319.

Miller, W., S. C. Schuster, A. J. Welch, A. Ratan, O. C. Bedoya-Reina, F. Zhao, H. L. Kim, R. C. Burhans, D. I. Drautz, N. E. Wittekindt, and L. P. Tomsho. 2012. Polar and brown bear genomes reveal ancient admixture and demographic footprints of past climate change. *Proc. Natl. Acad. Sci.* 109:E2382-E2390.

Moore, J. C., and J. R. Pannell. 2011. Sexual selection in plants. *Curr. Biol.* 21:R176–R182.

Moore, W. S. and D. B. Buchanan. 1985. Stability of the northern flicker hybrid zone in historical times: Implications for adaptive speciation theory. *Evolution* 39:135.

Moyle, L. C., M. S. Olson, and P. Tiffin. 2004. Patterns of reproductive isolation in three Angiosperm genera. *Evolution* 58:1195–1208.

Muller, C. H. 1938. Further studies in southwestern oaks. *Am. Midl. Nat.* 19:582–588.

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

Nadeau, S., P. G. Meirmans, S. N. Aitken, K. Ritland, and N. Isabel. 2016. The challenge of separating signatures of local adaptation from those of isolation by distance and colonization history: The case of two white pines. *Ecol. Evol.* 6:8649–8664.

Oksanen, J., F. G. Blanchet, M. Friendly, R. Kindt, P. Legendre, D. McGlinn, P. R. Minchin, R. B. O'Hara, G. L. Simpson, P. Solymos, M. H. H. Stevens, E. Szoecs, and H. Wagner. 2020. VEGAN: Community Ecology Package. R package version 2.5-7.

Orsini, L., J. Vanoverbeke, I. Swillen, J. Mergeay, and L. De Meester. 2013. Drivers of population genetic differentiation in the wild: isolation by dispersal limitation, isolation by adaptation and isolation by colonization. *Mol. Ecol.* 22:5983–5999.

Ortego, J., 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. *J. Biogeogr.* 41:1895–1908.

Papper, P. D. 2019. Modeling ancient potential for gene flow in California white oaks. *Int. Oaks* 30:185–195.

Parmesan, C. and G. Yohe. 2003. A globally coherent fingerprint of climate change impacts across natural systems. *Nature* 421:37–42.

Patterson, N., P. Moorjani, Y. Luo, S. Mallick, N. Rohland, Y. Zhan, T. Genschoreck, T. Webster, and D. Reich. 2012. Ancient admixture in human history. *Genetics* 192:1065–1093.

Pearse, I. S., J. H. Baty, D. Herrmann, R. Sage, and W. D. Koenig. 2015. Leaf phenology

mediates provenance differences in herbivore populations on valley oaks in a common garden. *Ecol. Entomol.* 40:525–531.

Pease, J. B. and M. W. Hahn. 2015. Detection and polarization of introgression in a five-taxon phylogeny. *Syst. Biol.* 64:651-662.

Peres-Neto, P. R., P. Legendre, S. Dray, and D. Borcard. 2006. Variation partitioning of species data matrices: Estimation and comparison of fractions. *Ecology* 87:2614–25.

Pesendorfer, M. B., T. S. Sillett, W. D. Koenig, and S. A. Morrison. 2016. Scatter-hoarding corvids as seed dispersers for oaks and pines: A review of a widely distributed mutualism and its utility to habitat restoration. *Condor* 118:215–237.

Peterson, B. K., J. N. Weber, E. H. Kay, H. S. Fisher, and H. E. Hoekstra. 2012. Double digest RADseq: An inexpensive method for de novo snp discovery and genotyping in model and non-model species. *PLoS One* 7:e37135.

Peterson, G. W., Y. Dong, C. Horbach, and Y. B. Fu. 2014. Genotyping-by-sequencing for plant genetic diversity analysis: A lab guide for SNP genotyping. *Diversity* 6:665–680.

Petit, R. J. and A. Hampe. 2006. Some evolutionary consequences of being a tree. *Annu. Rev. Ecol. Evol. Syst.* 37:187–214.

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

Pluess, A. R., V. L. Sork, B. Dolan, F. W. Davis, D. Grivet, K. Merg, J. Papp, and P. E. Smouse. 2009. Short distance pollen movement in a wind-pollinated tree, *Quercus lobata* (Fagaceae). *For. Ecol. Manage.* 258:735–744.

Polgar, C. A. and R. B. Primack. 2011. Leaf-out phenology of temperate woody plants: From trees to ecosystems. *New Phytol.* 191:926–941.

Racimo, F., S. Sankararaman, R. Nielsen, and E. Huerta-Sánchez. 2015. Evidence for archaic adaptive introgression in humans. *Nat. Rev. Genet.* 16:359–371.

Raj, A., M. Stephens, and J. K. Pritchard. 2014. fastSTRUCTURE: Variational inference of population structure in large SNP data sets. *Genetics* 197:573–589.

Ramírez-Valiente, J. A., D. Sánchez-Gómez, I. Aranda, and F. Valladares. 2010. Phenotypic plasticity and local adaptation in leaf ecophysiological traits of 13 contrasting cork oak populations under different water availabilities. *Tree Physiol.* 30:618–627.

Rice, K. J., J. H. Richards, and S. L. Matzner. 1997. Patterns and processes of adaptation in blue oak seedlings. Pages 109–116 in Fourth Symposium on Oak Woodlands.

Rice, W. R. and E. E. Hostert. 1993. Laboratory experiments on speciation: What have we

learned in 40 years? *Evolution* 47:1637–1653.

Rieseberg, L. H. 1991. Homoploid reticulate evolution in *Helianthus* (Asteraceae): Evidence from ribosomal genes. *Am. J. Bot.* 78:1218-1237.

Roberts, A. M., C. Tansey, R. J. Smithers, and A. B. Phillimore. 2015. Predicting a change in the order of spring phenology in temperate forests. *Glob. Chang. Biol.* 21:2603–2611.

Rushton, B. S. 1993. Natural hybridization within the genus *Quercus* L. *Ann. For. Sci.* 50:73–90.

Sanz-Pérez, V., P. Castro-Díez, and F. Valladares. 2009. Differential and interactive effects of temperature and photoperiod on budburst and carbon reserves in two co-occurring Mediterranean oaks. *Plant Biol.* 11:142–151.

Savolainen, O., T. Pyhäjärvi, and T. Knürr. 2007. Gene flow and local adaptation in trees. *Annu. Rev. Ecol. Evol. Syst.* 38:595–619.

Scerri, E. M. L., M. G. Thomas, A. Manica, P. Gunz, J. T. Stock, C. Stringer, M. Grove, H. S. Groucutt, A. Timmermann, G. P. Rightmire, F. d'Errico, C. A. Tryon, N. A. Drake, A. S. Brooks, R. W. Dennell, R. Durbin, B. M. Henn, J. Lee-Thorp, P. deMenocal, M. D. Petraglia, J. C. Thompson, A. Scally, and L. Chikhi. 2018. Did our species evolve in subdivided populations across Africa, and why does it matter? *Trends Ecol. Evol.* 33:582–594.

Servedio, M. R. and M. A. F. Noor. 2003. The role of reinforcement in speciation: Theory and data. *Annu. Rev. Ecol. Evol. Syst.* 34:339–364.

Shryock, D. F., C. A. Havrilla, L. A. DeFalco, T. C. Esque, N. A. Custer, and T. E. Wood. 2015. Landscape genomics of *Sphaeralcea ambigua* in the Mojave Desert: A multivariate, spatially-explicit approach to guide ecological restoration. *Conserv. Genet.* 16:1303–1317.

Skelton, R. P., L. D. L. Anderegg, J. Diaz, M. M. Kling, P. Papper, L. J. Lamarque, S. Delzon, T. E. Dawson, and D. D. Ackerly. 2021. Evolutionary relationships between drought-related traits and climate shape large hydraulic safety margins in western North American oaks. *Proc. Natl. Acad. Sci.* 118.

Skelton, R. P., L. D. L. Anderegg, P. Papper, E. Reich, T. E. Dawson, M. Kling, S. E. Thompson, J. Diaz, and D. D. Ackerly. 2019. No local adaptation in leaf or stem xylem vulnerability to embolism, but consistent vulnerability segmentation in a North American oak. *New Phytol.* 223:1296–1306.

Slatkin, M. 1993. Isolation by distance in equilibrium and non-equilibrium populations. *Evolution* 47:264–279.

Smith, J. M. 1966. Sympatric speciation. *Am. Nat.* 100:637–650.

Solís-Lemus, C. and C. Ané. 2016. Inferring phylogenetic networks with maximum pseudolikelihood under incomplete lineage sorting. *PLoS Genet.* 12:1–21.

Soltis, P. S., X. Liu, D. B. Marchant, Cl. J. Visger, and D. E. Soltis. 2014. Polyploidy and novelty: Gottleib's legacy. *Philos. Trans. R. Soc. B Biol. Sci.* 369:20130351.

Sork, V. L., S. T. Fitz-Gibbon, D. Puiu, M. Crepeau, P. F. Gugger, R. Sherman, K. Stevens, C. H. Langley, M. Pellegrini, and S. L. Salzberg. 2016a. First draft assembly and annotation of the genome of a California endemic oak *Quercus lobata* Nee (Fagaceae). *G3 Genes|Genomes|Genetics* 6:3485–3495.

Sork, V. L., K. Squire, P. F. Gugger, S. E. Steele, E. D. Levy, and A. J. Eckert. 2016b. Landscape genomic analysis of candidate genes for climate adaptation in a California endemic oak, *Quercus lobata*. *Am. J. Bot.* 103:1–14.

Stamatakis, A. 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312–1313.

Steinhoff, S. 1993. Results of species hybridization with *Quercus robur* L. and *Quercus petraea* (Matt) Liebl. *Ann. For. Sci.* 50:137s-143s.

Suarez-Gonzalez, A., C. Lexer, and Q. C. B. Cronk. 2018. Adaptive introgression: A plant perspective. *Biol. Lett.* 14.

Suni, S. S. and A. R. Whiteley. 2015. Genetic structure of a montane perennial plant: The influence of landscape and flowering phenology. *Conserv. Genet.* 16:1431–1442.

Theunert, C., and M. Slatkin. 2017. Distinguishing recent admixture from ancestral population structure. *Genome Biol. Evol.* 9:427–437.

Toyama, K. S., P. A. Crochet, and R. Leblois. 2020. Sampling schemes and drift can bias admixture proportions inferred by structure. *Mol. Ecol. Resour.* 20:1769–1785.

Tucker, J. M. 1952. Evolution of the California oak *Quercus alvordiana*. *Evolution* 6:162–180.

Tucker, J. M. 1953. Two new oak hybrids from California. *Madroño* 12:119–127.

Tucker, J. M. 1961. Studies in the *Quercus undulata* complex. I. A preliminary statement. *Am. J. Bot.* 48:202–208.

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.

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

Vitasse, Y., S. Delzon, C. C. Bresson, R. Michalet, and A. Kremer. 2009. Altitudinal differentiation in growth and phenology among populations of temperate-zone tree species growing in a common garden. *Can. J. For. Res.* 39:1259–1269.

VonHoldt, B. M., J. A. Cahill, Z. Fan, I. Gronau, J. Robinson, J. P. Pollinger, B. Shapiro, J. Wall, and R. K. Wayne. 2016. Whole-genome sequence analysis shows that two endemic species of North American wolf are admixtures of the coyote and gray wolf. *Sci. Adv.* 2:e1501714.

Wang, I. J. 2013. Examining the full effects of landscape heterogeneity on spatial genetic variation: a multiple matrix regression approach for quantifying geographic and ecological isolation. *Evolution* 67:3403–3411.

Wang, I. J. and G. S. Bradburd. 2014. Isolation by environment. *Mol. Ecol.* 23:5649–5662.

Wang, I. J., R. E. Glor, and J. B. Losos. 2013. Quantifying the roles of ecology and geography in spatial genetic divergence. *Ecol. Lett.* 16:175–182.

Wen, D. and L. Nakhleh. 2018. Coestimating reticulate phylogenies and gene trees from multilocus sequence data. *Syst. Biol.* 67:439–457.

Wilczek, A. M., L. T. Burghardt, A. R. Cobb, M. D. Cooper, S. M. Welch, and J. Schmitt. 2010. Genetic and physiological bases for phenological responses to current and predicted climates. *Philos. Trans. R. Soc. B Biol. Sci.* 365:3129–3147.

Wilkinson, M., E. L. Eaton, and J. I. L. Morison. 2017. Variation in the date of budburst in *Quercus robur* and *Q. petraea* across a range of provenances grown in southern England. *Eur. J. For. Res.* 136:1–12.

Williams, J. H., W. J. Boecklen, and D. J. Howard. 2001. Reproductive processes in two oak (*Quercus*) contact zones with different levels of hybridization. *Heredity* 87:680–690.

Wright, S. 1931. Evolution in mendelian populations. *Genetics* 16:97–159.

Wright, S. 1943. Isolation by distance. *Genetics* 28:114–138.

Yang, M. A., A. S. Malaspinas, E. Y. Durand, and M. Slatkin. 2012. Ancient structure in Africa unlikely to explain Neanderthal and non-African genetic similarity. *Mol. Biol. Evol.* 29:2987–2995.