Title

Evidence of Hybridization and Introgression in Two Distantly Related, Sympatric Californian White Oaks (Quercus sect. Quercus)

Author

O'Donnell, Scott

Publication Date

2023

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
Los Angeles

Evidence of Hybridization and Introgression in Two Distantly Related, Sympatric Californian White Oaks (*Quercus* sect. *Quercus*)

A dissertation submitted in partial satisfaction

of the requirements for the degree

Doctor of Philosophy in Biology

by

Scott O'Donnell

2023

ABSTRACT OF THE DISSERTATION

Evidence of Hybridization and Introgression in Two Distantly Related, Sympatric Californian White Oaks (*Quercus* sect. *Quercus*)

by

Scott O'Donnell

Doctor of Philosophy in Biology

University of California, Los Angeles, 2023

Professor Victoria Sork, Chair

Climate regimes are changing rapidly at a global scale and the rate of change often outpaces the ability of many species to evolve in response to changing selective pressures. It is possible that other evolutionary processes, such as introgressive hybridization, allow species to exchange alleles across species boundaries at rates that exceed the creation of novel adaptive alleles through mutation alone. By analyzing genomic sequence data for evidence of ancient adaptive introgression, we can characterize the role that this process played in the evolutionary history of these species and its potential under predicted future climate regimes. Here, we test for evidence of adaptive introgression in two hybridizing white oaks (*Quercus* sect. *Quercus*): the drought tolerant California shrub oak (*Quercus berberidifolia*) and the comparatively drought intolerant tree Engelmann oak (*Quercus engelmannii*).

In Chapter 1, we use single nucleotide polymorphisms (SNPs) generated by reduced-representation sequencing to test the likelihood of various ancient demographic models to identify the presence, timing, and direction of ancient introgression in this system. We discovered evidence of asymmetric introgression from *Q. berberidifolia* during the advent of a

Mediterranean-type climate, indicating a potential association between ancient gene flow between these species and large-scale climatic change.

In Chapter 2 we used whole genome sequence data to characterize landscape-scale distribution of genetic diversity in both *Q. engelmannii* and *Q. berberidifolia.* We also identified and compared candidate SNPs significantly associated with climatic gradients to neutral genetic diversity under current and future predicted climate models. Evidence showed that *Q. engelmannii* is likely well-adapted to current local climate conditions throughout its range and is more likely to become maladapted under future climate scenarios.

In Chapter 3 we test for evidence of introgression in both species and identified any functional genes found in physical regions of introgression. We discovered significant genome-wide evidence of introgression of functional genes associated with stress response in *Q. engelmannii but* did not find any significant evidence of introgression in *Q. berberidifolia*. This combined evidence suggests that ancient adaptive introgression with *Q. berberidifolia* may have allowed *Q. engelmannii* to adapt to past large-scale climate change.

The dissertation of Scott O'Donnell is approved.

Kirk Edward Lohmueller

Sriram Sankararaman

Felipe Zapata Hoyos

Victoria Sork, Committee Chair

University of California, Los Angeles

2023

Contents

List of Figures

List of Tables

Acknowledgements

Chapter 1 was published October 2021 in the Journal of Heredity with the authors Scott O'Donnell, Sorel T. Fitz-Gibbon, and Victoria Sork. Samples were collected with permission from the Bureau of Land Management (Permit CA930-F-2020-012) and Cleveland National Forest (Permit 003653). We thank Kirk Lohmueller and Bernard Kim for their advice and input on the analyses and valuable comments on early drafts of this manuscript and members of Sork lab for feedback on the study. Sorel Fitz-Gibbon performed variant calling and provided advice on genome analyses. We thank Ingrid Lara for her help in sample collection and Claudia Henriquez for her help in lab work and library preparation. This project was supported in part by National Science Foundation PGRP grant (IOS-1444661) to V.L.S.

Chapter 2 is in preparation for publication with the authors Scott O'Donnell, Sorel Fitz-Gibbon, and Victoria Sork. Sorel T. Fitz-Gibbon performed variant calling and aided with file formatting and genomic analyses. Victoria Sork provided guidance on study design, methods, and interpretation of analyses. This work was supported by the California Conservation Genomics Project, with funding provided to the University of California by the State of California, State Budget Act of 2019 [UC Award ID RSI-19-690224].

Chapter 3 is also in preparation with the authors Scott O'Donnell, Sorel Fitz-Gibbon, and Victoria Sork. Sorel T. Fitz-Gibbon performed variant calling and aided and advice with file formatting and genomic analyses. Victoria Sork provided guidance on study design, methods, and interpretation of analyses. This work was supported by the California Conservation Genomics Project, with funding provided to the University of California by the State of California, State Budget Act of 2019 [UC Award ID RSI-19-690224]. We acknowledge the native peoples of California for their conservation and stewardship of oak habitats.

This dissertation truly was the work of so many people who contributed in some way, large or small, to finally getting this finished and ready for submission.

First and foremost, I could not have ever hoped to finish this work without the support and mentorship provided by my advisor Victoria Sork. No matter how greatly I struggled throughout this process, you were always there to remind me that lack of confidence was not lack of competence. You believed in me when I could scarcely believe in myself, and I can only hope to emulate your abilities as a scientist and mentor in my future endeavors. Without you taking a chance on admitting me to your lab, I would not be where I am today. I am eternally grateful for your help and guidance.

Thank you to my committee members who always provided critical feedback and served as a sounding board as I designed and conducted each experiment in this dissertation. Thank you to Kirk Lohmueller for providing his near boundless knowledge of population genetics methodology and its limitations. Thank you to Sriram Sankararaman for his help in designing and implementing introgression detection methods and proper sampling design. Finally, thank you to Felipe Zapata for your insight into plant evolution and systematics and its influence on methods designed to detect ancient evolutionary processes.

I cannot fathom how I would have made it through to the end of this process without my colleagues in the Sork Lab. I cannot overstate how grateful I am to have lucked into such a welcoming and helpful environment when starting this process. In no specific order, I thank Rachel Prunier, Sorel Fitz-Gibbon, Alayna Mead, Brandon MacDonald, Marissa Ochoa, Berenice Badillo, Heidi Yang, Diego Zapata, Luke Browne, Dylan Burge, Stephanie Steele,

Claudia Henriquez, Krista Beckley, Zhizhong Li, Yao Li, Ingrid Lara, and Jason Leung, you each managed to help me when I was lost in some analysis or tagged along with me on some extended field trip at some point. While it may have been bonding at a professional level at first, I am happy and honored to call most of you friends. Thank you for being accepting and welcoming of me for this entire journey. I learned so much about science and about being a good friend and colleague from all of you.

To all my Los Angeles friends, I truly never could have hoped to find such lovely supportive people to share my time in such a big city. Chris, Sarah, Winston, and Alex, words cannot give justice to how grateful I am to have met you and gotten to spend as much time together as we did. You all were my crutch when I needed it and my gentle shove in the right direction when it felt like I had no momentum. I am so lucky to have stumbled into such a great support network after moving across the entire country. I am so happy to have made such lifelong friends. Thank you all for everything. To Zack Gold, what might have started out as chance putting us in the same cohort in grad school truly blossomed into a lifelong friendship and I am honored to call you one of my best friends. You helped inspire me to work harder and fight through difficult times when I did not think I could. Hopefully I was able to provide even a fraction of that same support to you through the ups and downs of grad school.

To my grad student support network. Again, I do not understand how lucky I was to meet such kind and supportive people. I have met so many great people in my time at UCLA whose friendship and support got me over any hurdles science, or life in general, threw at me. Leila, Guarav, Camila, Rachel, Marcel, Christian, Grace, and Jonathan, thank you all for the late nights and long afternoons spent working or venting about work with me.

To all my friends back at home in Pennsylvania, thank you for always being a release valve when the pressure of grad school got to be too much and for sitting through me venting even when it made little sense to you. Zac, Nick, Arjay, Noah, Charles, Karl, Kyle, Daylan, and Donald, your friendship and support has never wavered throughout the years. Thank you.

Finally, I could not have hoped to achieve this milestone without the support of my family. As difficult as it was to let me move to the opposite end of the country, you were always prepared to support and encourage me, even at my lowest points. Mom and Dad, I love you and thank you for making sure I had the opportunity to figure out what life I wanted to live and how I would manage to live it. I made it, and now it is done.

VITA

EDUCATION

BS Biology, *suma cum laude,* 2015

East Stroudsburg University of Pennsylvania

AWARDS AND GRANTS

2018 Vavra Research Grant for Plant Studies

2015 Commonwealth of Pennsylvania University Biologists Outstanding Student Award

2015 East Stroudsburg University Chapter of Sigma Xi Outstanding Academic Achievement award in Integrative Organismal Biology

2015 Commonwealth of Pennsylvania University Biologists Annual Meeting Outstanding Poster 2nd place

PUBLICATIONS

O'Donnell, S. T., Fitz-Gibbon, S. T., & Sork, V. L. (2021). Ancient Introgression Between Distantly Related White Oaks (Quercus sect. Quercus) Shows Evidence of Climate-Associated Asymmetric Gene Exchange. *Journal of Heredity*, *112*(7), 663-670.

RESEARCH EXPERIENCE

2015-2023 Graduate research with Dr. Victoria Sork

2013-2015 Undergraduate research at the Northeast Wildlife DNA Lab with Dr. Jane Huffman

CONFERENCE PRESENTATIONS

O'Donnell ST, S Fitz-Gibbon, and VL Sork. November 2022. *Detecting introgression and determining the potential for adaptive gene flow between Californian white oaks (*Quercus *sect*. Quercus*).* Oral presentation, California Oak Symposium, San Luis Obispo, California.

O'Donnell ST, S Fitz-Gibbon, and VL Sork. September 2022. *Patterns of adaptive and neutral genetic population structure in two hybridizing Californian white oaks* (Quercus *sect.* Quercus). Oral presentation, International Oak Society Meeting, Las Cruces, New Mexico.

O'Donnell ST, S Fitz-Gibbon, and VL Sork. June 2019. *The extent and timing of ancient introgression among two distantly related white oak species* (Quercus sect. Quercus) *in southern California.* Poster presentation, Western Forest Genetics Association, Placerville, California.

O'Donnell ST, S Fitz-Gibbon, and VL Sork. September 2018. *Ancient introgression among two distantly related white oak species* (Quercus *sect.* Quercus) *in southern California: when and how much?* Poster presentation, International Oak Society Meeting, Davis, California.

Chapter 1

Ancient introgression between distantly related white oaks (*Quercus* sect *Quercus*) shows evidence of climate-associated asymmetric gene exchange

Abstract

Ancient introgression can be an important source of genetic variation that shapes the evolution and diversification of many taxa. Here, we estimate the timing, direction, and extent of gene flow between two distantly related oak species in the same section (*Quercus* sect. *Quercus*). We estimated these demographic events using genotyping by sequencing data (GBS), which generated 25,702 single nucleotide polymorphisms (SNPs) for 24 individuals of California scrub oak (*Quercus berberidifolia*) and 23 individuals of Engelmann oak (*Q. engelmannii*). We tested several scenarios involving gene flow between these species using the diffusion approximation-based population genetic inference framework and model-testing approach of the Python package *DaDi*. We found that the most likely demographic scenario includes a bottleneck in *Q. engelmannii* that coincides with asymmetric gene flow from *Q. berberidifolia* into *Q. engelmannii*. Given that the timing of this gene flow coincides with the advent of a

Mediterranean-type climate in the California Floristic Province, we propose that changing precipitation patterns and seasonality may have favored the introgression of climate-associated genes from the endemic into the non-endemic California oak.

Introduction

Hybridization has long been recognized as an evolutionarily significant, naturally occurring phenomenon when related species come into contact (Anderson, 1949; Anderson & Stebbins, 1954; Arnold, 1997; Rieseberg & Wendel, 1993). The exchange of genetic variation between closely related taxa can influence patterns of diversification and evolution in hybridizing taxa (Abbott et al., 2016; Anderson & Stebbins, 1954; Harrison & Larson, 2016; Mallet, 2005). In addition, the incorporation of alleles into novel genomic backgrounds through recurrent interspecific gene flow (otherwise known as introgression) may enhance adaptation in hybridizing taxa (Abbott et al., 2016) as proposed by Anderson and Stebbins (1954). Contemporary evidence in support of the hypothesis of adaptive introgression has been observed in mosquitoes (Crawford et al., 2015), birds (Bay & Ruegg, 2017; Ruegg et al., 2014), postman butterflies (Heliconius spp.Dasmahapatra et al., 2012; Edelman et al., 2018) and canids (vonHoldt et al., 2016). Moreover, the formation of contemporary hybrid zones between plant taxa with varying degrees of introgression may result in adaptive introgression with examples from grasses (Becker et al., 2013), monkey flowers (Mimulus spp., Brandvain et al., 2014; Vallejo-Marin et al., 2016), irises (Iris spp., Arnold et al., 1991; Martin et al., 2006), poplar (Populus spp., Christe et al., 2016; Christe et al., 2017; Suarez-Gonzalez et al., 2016; Suarez-Gonzalez et al., 2018), spruce (De La Torre et al., 2014; Hamilton et al., 2013), and sunflowers, (Helianthus spp. Rieseberg et al., 2003; Rieseberg et al., 1999; Whitney et al., 2015). With the recent increase in scrutiny on hybridization as an evolutionary process it is likely that it will be implicated in the local adaptation of a growing number of species of both animals and plants (Abbott et al., 2016; Anderson, 1949). This makes it critical to understand the influence of

hybridization in the evolution of taxa that are well known for their ability to interbreed, such as oaks (*Quercus* spp.).

Oaks are notorious examples of hybridization among species (Van Valen 1976). Studies using nuclear markers (e.g., Dodd & Afzal-Rafii, 2004; Howard et al., 1997; Kremer, 2016; Rémy J Petit et al., 2004; Van Valen, 1976; Whittemore & Schaal, 1991) and chloroplast markers (Whittemore & Schaal, 1991) indicate that interspecific gene exchange is common between closely related oak species where their distributions overlap. Moreover, prevalence of hybridization between species when conspecific densities are low (R. J. Petit et al., 2004(a); Rémy J Petit et al., 2004(b)) suggests that hybridization allows reproduction when conspecific pollen is less available. If rates of gene flow are influenced primarily by the proximity and density of conspecific individuals, hybridization would likely be localized, sharing both neutral and potentially adaptive genes. Yet, other evidence suggests that hybridization between oaks may be more complex than can be explained by local population densities alone. For instance, differential strength of pre-zygotic reproductive isolation between different pairs of oak species indicates contemporary asymmetric gene flow between species is possible when species are sympatric (Lepais et al., 2013). Additionally, contemporary hybridization between oak species has also been shown to have a stronger correlation with fine-scale environmental variables than spatial distance between conspecific individuals (Dodd & Afzal-Rafii, 2004), which implies that natural selection may shape patterns of gene exchange through introgression. Recent studies have shown an association between the direction and extent of gene flow with specific selective pressures such as climatic niche overlap (Ortego et al., 2014; Ortego et al., 2015; Riordan et al., 2016), fire regime (Ortego et al., 2017) and resistance to herbivory (Pearse & Hipp, 2009). These studies indicate that genetic variation introduced through hybridization between species is likely

shaped by selection caused by the environmental conditions at both the pre-zygotic and post-zygotic stage of reproduction.

One difficulty in identifying evidence of adaptive introgression is distinguishing between shared alleles due to a recent common ancestor or due to ancient hybridization events. Therefore, this study of ancient introgression will utilize two distantly related Californian oak species, California scrub oak (*Quercus berberidifolia*) and Engelmann oak (*Q. engelmannii*), which are within the same section (white oaks; *Quercus* sect. *Quercus*) but in different sub-clades estimated to have diverged more than 40 MYA (Hipp et al., 2019). Currently, these two species form hybrid zones in southern California where they come into contact (Ortego et al., 2014; Riordan et al., 2016). However, it is not known when *Q. engelmannii* and its closest relatives from the Mexican clade of white oaks, now found in Arizona and New Mexico, extended into areas west of the Mojave Desert. The contemporary center of diversity of that clade is found in the sub-tropical forests of Mexico and Central America, leaving *Q. engelmannii* as its lone representative in the California Floristic Province (Hipp et al., 2018). Thus, the ancestral population or species of *Q. engelmannii* would have evolved under conditions of different seasonality of rainfall and temperature than what the species now experiences in the Mediterranean-type climate of southern California and northern Baja California, which has been associated with wide-spread diversification in plant taxa throughout the California Floristic Province (CFP) (Rundel et al., 2016). Conversely, *Q. berberidifolia* is a member of the California white oak clade that has been endemic to western North America since its divergence from the rest of the North American oak clade at ~40mya (Hipp et al., 2019). While their climatic niches broadly overlap, *Q. berberidifolia* is associated with hotter, drier conditions typical of chaparral ecosystems found throughout the CFP and *Q. engelmannii* is restricted to

more mesic conditions found at mid-elevation plateaus in the Peninsular Ranges and foothills of the San Gabriel Mountains in southern California and mid-elevation regions of northern Baja California (Figure 1A) (Ortego et al., 2014; Riordan et al., 2015; Riordan et al., 2016). Indeed, previous work demonstrates ancient introgression shows asymmetric patterns of gene flow from the shrub oak (*Q. berberidifolia*) into the tree oak, *Q. engelmannii* (Kim et al., 2018), but limited sample sizes prevented further testing of the hypothesis that asymmetrical gene flow from the currently widespread California endemic species into the Mexican-evolved species was due to adaptation to changing environmental conditions. If the California endemic shrub oak is more adapted to the local climate conditions, then we would expect that retained genetic variants in *Q. engelmannii* from ancient hybridization would be biased towards those from *Q. berberidifolia*, rather than the other way around.

In this study, we test whether ancient introgression is symmetric or asymmetric. If selection has influenced ancient introgression, we would predict that the species that is more adapted to local environmental conditions would contribute genes disproportionately to the other species. Consequently, we expect that *Quercus engelmannii,* the species that arrived more recently in southern California would benefit from genes from the native endemic *Quercus berberidifolia*. We test these predictions by estimating the timing, extent and direction of introgression using a demographic modeling framework using the Python package *DaDi* (Gutenkunst et al., 2010). Specifically, we test the opposing hypotheses that (i) gene flow between species was low with little significant influence on the genetic composition of either species or (ii) levels of gene flow between species were extensive and played a large role in shaping genetic composition of one or both taxa. If the second hypothesis is supported, the direction of gene flow (i.e., symmetric vs. asymmetric) will determine whether each species

affected the evolutionary dynamics of the other equally or if the tree oak *Q. engelmannii* versus shrub oak *Q. berberidifolia* had a greater impact. We generated 25,702 single nucleotide polymorphisms (SNPs) as genetic loci through a method of reduced representation library sequencing (genotyping by sequencing (GBS), Elshire et al., 2011) of samples collected from both species throughout southern California where the ranges of the two species currently overlap (Figure 1B). We compared demographic models to address these questions: (1) What is the direction of ancient introgression between these two species? (2) When did this introgression occur? (3) Does the timing of this introgression coincide with periods of climate fluctuation in the California Floristic Province?

Methods

Study Species and Sample Design:

The two study species are currently sympatric in southern California, and both are endemic to the California Floristic Province of western North America. Engelmann oak (*Quercus engelmannii*) is a tree oak representative of the clade of subtropical Mexican white oaks, while California scrub oak (*Quercus berberidifolia*) is a wide-spread, medium-to-large shrub in the California white oak clade with a range that encompasses the entire CFP from northern California south into the Baja California peninsula (Figure 1.1B). According to ecological niche modelling, *Q. engelmannii* is largely restricted to the cooler, wetter conditions found on mesas between 450-1200m in elevation in Riverside, Orange and San Diego counties in southern California (Riordan et al., 2016). The distribution and ecological niche of *Q. engelmannii* broadly overlap throughout its range with that of the widespread *Q. berberidifolia,* which is associated with drier climatic and microhabitat conditions (Figure 1.1A) (Riordan et al., 2016). Past evidence indicates

that the genetic diversity and genetic structure of *Q. engelmannii* are associated with past climatic suitability, which points to a core distribution in the mountains of the Peninsular Ranges in modern San Diego and Riverside counties (Ortego et al., 2012) and sampling was focused on this region (Figure 1.1A, 1.1B). In regions where their distribution overlaps, they form contemporary hybrid zones along microhabitat gradients as they share broadly similar climatic and habitat niche preferences (Riordan et al., 2016). Evidence of ancient introgression has also been shown between these species using reduced library sequencing data generated by the RADseq method (Kim et al., 2018). This combination of factors makes them a unique study system to examine the history of gene flow between oak taxa.

Leaf tissue was collected from two mature trees per species from twelve sites (24 total sites, 48 total samples), distributed throughout the range of overlap between *Q. engelmannii* and *Q. berberidifolia* in southern California, United States of America (Figure 1.1). Individuals were first identified to species in the field based on diagnostic morphological characteristics including leaf shape, trichome shape and trichome density (Roberts, 1995) following the methods described in previous work on this study system (Ortego et al., 2014; Ortego et al., 2017; Ortego et al., 2012). Any morphologically intermediate individuals were excluded from sampling to minimize the impacts of contemporary hybridization on this study. Additionally any genetically intermediate samples identified using both STRUCTURE (Pritchard et al., 2000) (Figure 1.1C) and PCA analyses (Zheng et al., 2012) were also removed from any other downstream analyses (Figure 1.2).

DNA Extraction and Sequencing

Leaf tissue was frozen in liquid nitrogen and processed using a mortar and pestle until ground into a fine powder. DNA extraction procedures followed a modified version of the Qiagen DNEasy plant extraction protocol. DNA was prepared using GBS sequencing protocols (Elshire et al., 2011) including a bead clean up step to ensure proper sized fragments. Libraries were sequenced at the Technology Center for Genomics and Bioinformatics at the University of California, Los Angeles on an Illumina Hiseq3000 (Illumina Inc. 5200 Illumina Way San Diego, CA) in August of 2017. To equalize coverage between species, samples were re-sequenced in July 2018. Samples were sequenced using two lanes (48 samples per lane) and the combined sequence data for each individual was used for this project.

Variant calling and filtering

Sequences were filtered for quality and sequencing depth and aligned to the *Quercus lobata v.3.0* reference genome. We used GATK to call the SNPs according to the methods described in Fitz-Gibbon et al. (2017). To remove confounding effects of contemporary hybridization, we excluded samples identified as genetically intermediate using the principal component analysis methods implemented in the R package SNPRelate (Zheng et al., 2012)(Figure 1.2). Additionally, we performed an analysis of population substructure using STRUCTURE (Pritchard et al., 2000). STRUCTURE analysis was performed with a burn-in of 10,000 reps and 10,000 MCMC runs for a hypothesized K of 1:5. Each run was replicated 15 times for each K. The most likely number of clusters was determined using the Evanno method (Evanno et al., 2005) (Fig 1.1C). For these species, the most likely number of clusters was 2 with none of the individuals having over 15% proportion of ancestry from the other species after

initial filtering using SNPRelate (Figure 1C, 1D). Of the remaining samples, SNPs were stringently filtered for linkage disequilibrium (LD) using an r^2 value of 0.4, a sliding window size of 500 bp and steps of 1 bp using plink v. 2.0 (Purcell et al., 2007). It is necessary to filter for LD to ensure that the results of model testing result in a true likelihood value to determine the best-fit model and minimize uncertainty associated with the estimated parameters of the hypothesized models (see Demographic Modeling below) (Coffman et al., 2015). After filtering, 47 samples (23 *Q. engelmannii* and 24 *Q. berberidifolia* samples respectively) and 25,702 SNPs were used to determine the best fit demographic models. A 2D site frequency spectrum was generated using easySFS.py (available at https://github.com/isaacovercast/easySFS). For the final analysis and demographic model reconstruction, the folded SFS was used.

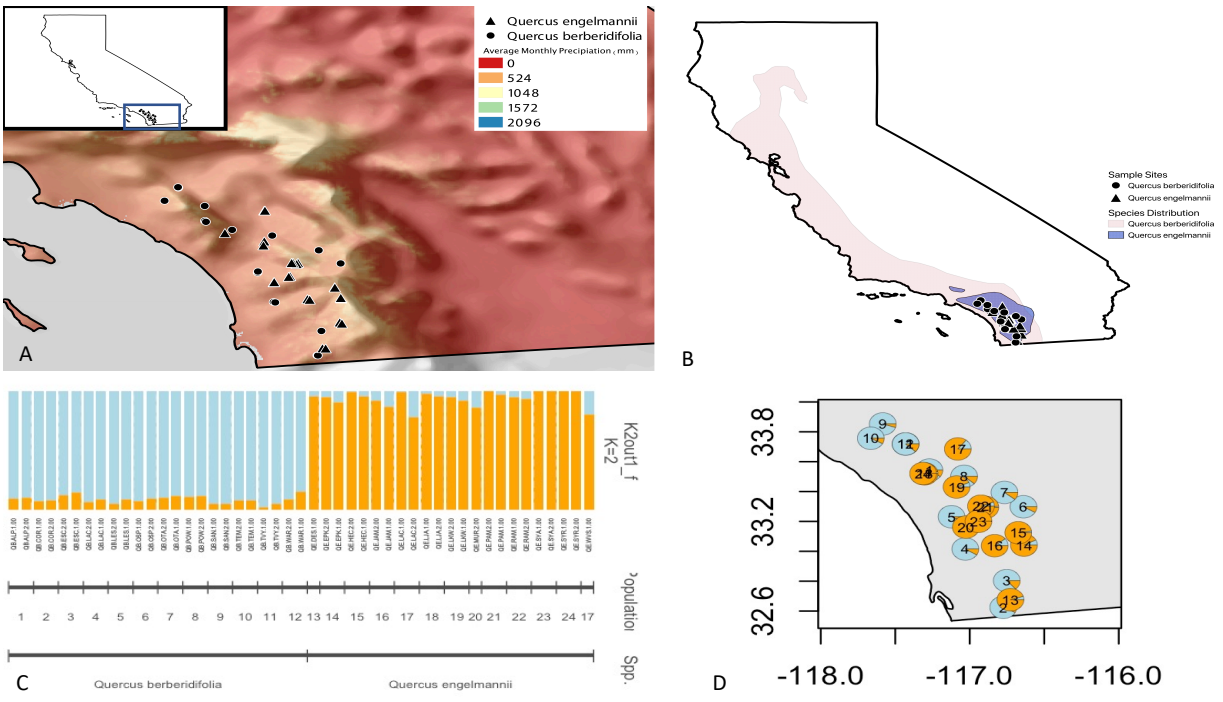

Figure 1.1 Background information on the distribution, sampling regime and population substructure of the system. **(A)** Sample sites for both *Q. berberidifolia* (circles) and *Q. engelmannii* (triangles) (24 individuals per species, 48 total individuals) along an average monthly precipitation gradient produced using average monthly precipitation data from the Worldclim database (Fick SE & Hijmans, 2017) , **(B)** sample sites with respect to the range of *Q. berberidifolia* (pink) and *Q. engelmannii* (blue) in California, **(C)** a STRUCTURE plot of the most likely number of genetic clusters (K=2) for the 47 sequenced individuals and their population assignments and **(D)** average proportion of ancestry (as pie charts) for each of the 24 populations labeled in **(C)**

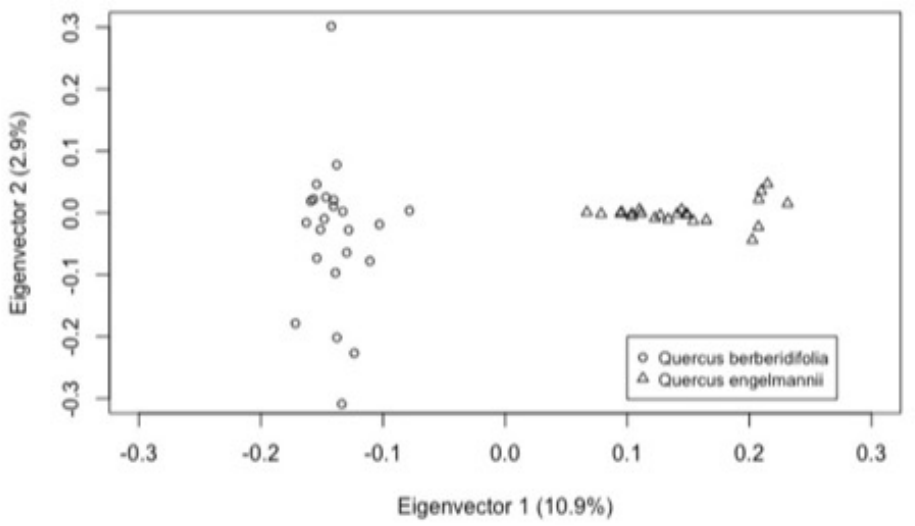

Figure 1.2: SNPRelate results for the final dataset. *Q. berberidifolia* (circles) and *Q. engelmannii* (triangles) showing separate clusters of individuals for each species. Any genetically intermediate individuals were removed for final analysis.

Demographic Modeling

Demographic models were fit with the Python package *DaDi* (Gutenkunst et al., 2009). This method generates simulated site frequency spectra (SFS) according to specific hypothesized demographic scenarios that are then compared to SFS from sampled genetic data to determine the most likely demographic model. Models are specified using a series of free variables during model optimization that are associated with changes in demography, such as migration rates or population size expansion. The values of these parameters from the most likely demographic model can then be used as estimates for those parameters in each hypothesized scenario. For each iteration, the simulated SFS were compared to the observed SFS and ranked according to their log likelihood scores. Model optimization was achieved by running 100 replicate simulations using the Inference.Optimize_log() command in *DaDi* that generates simulated SFS for the free parameter values used to define each model to determine which values for those

parameters that will result in the highest likelihood of the specified model. Models were compared using log likelihood values based on the number of free parameters used to define each model to determine the most likely demographic model for this system. Each model was designed in a nested fashion with the simplest models being treated as a specialized version of the more complex models with the results from simpler models informing the development of the more complex models (see Appendix 1.1 for the complete list of hypotheses and their descriptions). A total of 15 hypotheses were tested and compared based on their number of free parameters. The parameter values of the global best fit model were then used to calculate effective population size of each species with respect to the ancestral population ($\nu 1$ and $\nu 2$), the timing of events such as divergence, secondary contact or bottleneck events (T), the number of effective migrants per generation (m in symmetric migration scenarios, m_{12} and m_{21} in the case of asymmetric migration) and changes in effective population size during both bottlenecks or population expansion (νB and νF respectively). Because contemporary hybridization occurs between the two species, gene flow was considered ongoing from the point of secondary contact and the time point at which gene flow stopped was not included as a free parameter in the tested models. Uncertainties for each parameter were calculated using a Fisher's Information Matrix through the dadi.Godambe.FIM_uncert() command in *DaDi* (Coffman et al., 2015).

Timing of events in the best fit model was estimated using a mutation rate of 1.01×10^{-8} estimated from the analysis of heterozygosity in the valley oak genome (Sork et al., 2021) and total length of sequence (L) of 25,767 after filtering for LD. All timing estimates were calibrated using estimated divergence times of a maximum of ~40mya as reported in the latest time-calibrated global oak phylogeny to account for unknown variables such as variation in generation times between species (Hipp et al., 2019). Timing of specific events within each model was

estimated using an average generation time of 5 years, which is possible for shrub oaks and fast-growing tree oaks such as Engelmann oak (VL Sork, personal observation). Generation times of 10 and 20 years were also used to estimate the timing of these events, but these values resulted in the estimated divergence time between these species to be before the proposed crown age of the genus (see table S1 in Supplementary Materials). As surveys of *Q. engelmannii* have shown, most young trees do not persist in the environment or reach large sizes or advanced age classes, so it is likely that most individuals have relatively short life spans (Lathrop et al., 1991).

Results

All subsets of models that included ancient gene flow performed better than those that did not include gene flow (Table 1.1). Of the simple set of models with an ancient divergence event coinciding with gene flow between species, the models that included ancient gene flow performed the best (Table 1.1). In all models, regardless of increasing complexity, those that included asymmetric gene flow performed the best, indicating that ancient gene flow was likely critical in the past demography of these species.

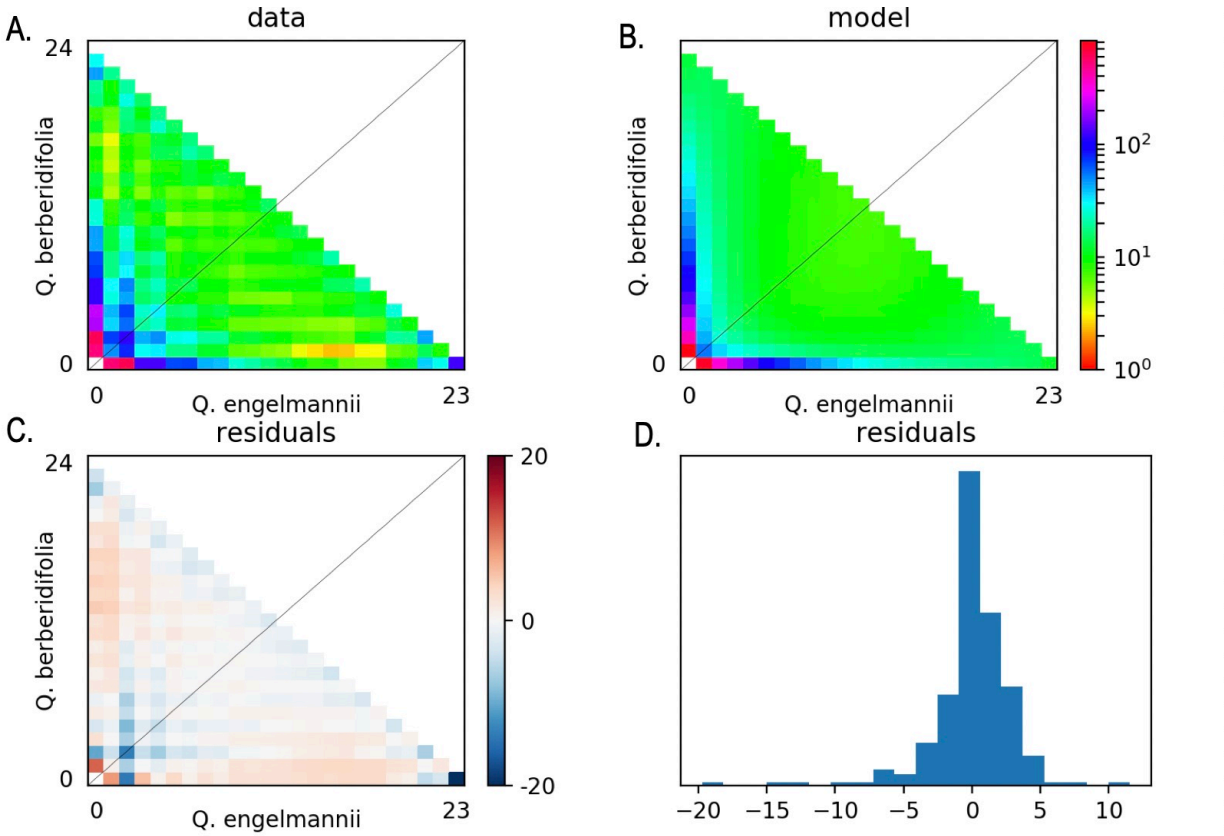

Figure 1.3: *DaDi* results for the site frequency spectrum of the most likely demographic model. The most likely demographic model (PreBotAsymEAM) includes a bottleneck in *Q. engelmannii* that coincided with the resumption of asymmetric gene flow between the two taxa. (**A**) Heatmap using the observed site frequency spectrum generated from sequence data of both species. (**B**) Heatmap of the site frequency spectrum of the most likely demographic model. Heatmaps of site frequency spectra (SFS) indicate which alleles are private to one species (boxes found on the x and y axes for *Q. engelmannii* and *Q. berberidifolia* respectively) compared to alleles that are shared to some extent between species (boxes that do not fall on the axes). The axes are labeled with the number of individuals for each species. Colors of boxes indicate the density or number of alleles in each given point of the heatmap. The residuals between the model and the data are presented as both a heatmap (**C**) and bar plot (**D**).

Table 1.1: Results of the six most likely demographic scenarios. The best fit model was PreBotAsymEAM, which included a bottleneck in *Quercus engelmannii* that occurred at the same time as asymmetric gene flow between both species. (See Appendix 1 for description of the other scenarios).

Model	Model Description	Num. Params.	LL	AIC	ΔAIC
PreBotAsymEAM	**Asymmetric gene flow during bottleneck in QE**	**9**	**1851.70**	**3723.4**	**0**
SexpAM	Population expansion with asymmetric gene flow	6	-1873.43	3758.8576	35.4576
BotAsymB	Asymmetric gene flow during bottleneck in QB	9	-1878.71	3773.4152	50.0152
P2ExpSM	Population growth in QB with symmetric gene flow	5	-1890.73	3791.45	68.05
ABotAsymEAM	Asymmetric gene flow after bottleneck in QE	9	-1895.32	3810.6346	87.2346

The best fit model when comparing the observed site frequency spectrum to the site frequency spectrum of a series of hypothetical models included an ancient split between the ancestors of *Q. engelmannii* and *Q. berberidifolia* followed by the initiation of gene flow during a bottleneck in *Q. engelmannii,* which persisted into contemporary times (Figure 1.3). Based on the most likely demographic model and an estimated generation time of 5 years, the split between the ancestors of these species occurred approximately 45mya (+/- 240,751 years) followed by a bottleneck in *Q. engelmannii* at ~10mya (+/- 151,000 years) that coincided with asymmetric gene flow that was greater from *Q. berberidifolia* into *Q. engelmannii* (m_{21} = 1.1489 +/- 0.0078). Gene flow continued throughout the bottleneck and into present day. *Quercus engelmannii* recovered to its current effective population size at ~8mya (+/- 215, 000 years) (Figure 1.4). Timing estimates were also calculated using 10 and 20 year generation times, but

divergence between the two species exceeded estimates for the crown age of the genus (Hipp et al., 2019) and are not reported here (see Appendix 1, Table S3).

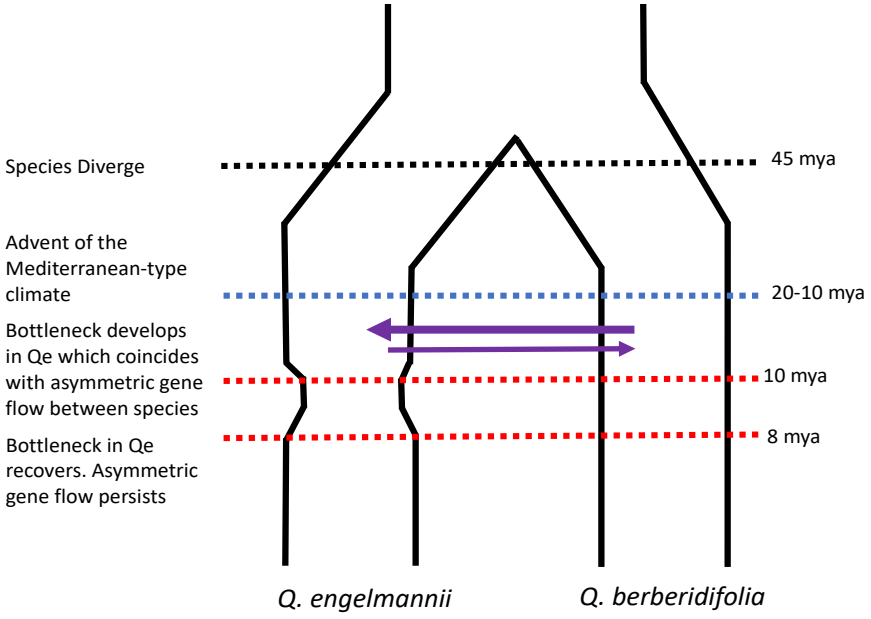

Figure 1.4: Visualization of the most likely demographic model with estimates of the timing of each specific demographic event. Purple arrows denote gene flow between species and thickness of the arrow indicates the rate of gene flow. Times were estimated using a mutation rate of 1.01×10^{-8} and a generation time of 5 years.

Discussion

This study provides evidence that ancient introgression contributed to the genetic composition of southern California populations of both species, particularly *Quercus engelmannii*. As predicted, our hypothesis of asymmetric gene flow into the more recently arriving oak was supported. The most likely demographic scenario indicates that introgression between the two species is asymmetric due to higher rates of gene flow from *Q. berberidifolia* into *Q. engelmannii*. This pattern is similar to recent evidence found in other oak species where rates and direction of gene flow between species are variable (Lepais et al., 2013) and shaped by

environmental context (Dodd & Afzal-Rafii, 2004). Similar patterns are also seen as a result of large-scale environmental disturbances such as glaciation (Leroy et al., 2019) or human-mediated changes in fire regime (Ortego et al., 2017) in which gene flow has also been shown to be asymmetric.

The timing of the initiation of gene flow between these two species also points to a period of large-scale climatic shifts during the transition to a Mediterranean-type climate in California. According to the most likely demographic model, *Q. berberidifolia* exerted greater influence on the genome of *Q. engelmannii* in terms of overall amounts of gene flow. While it was an ongoing exchange over the last ~9 million years, the initiation of this exchange seems to be associated with the emergence of a Mediterranean type climate in the California Floristic Province, which has been estimated to have occurred between 10-20mya (Rundel et al., 2016). This climatic shift is associated with a reduction in size of populations of *Q. engelmannii* that occurs simultaneously with gene flow between species. Even when extending the length of generation times, the initiation of gene flow during secondary contact and the formation of a genetic bottleneck in *Q. engelmannii* still occur during the transition to a Mediterranean-type climate in California. This type of association of changes in demographic parameters with large-scale climatic shifts has been seen in European oaks in response to glaciation cycles (Leroy et al., 2017). In light of contemporary population decline and predictions of further loss of suitable habitat throughout its current range as climatic conditions in California are predicted to be hotter and drier (Riordan et al., 2015), it is likely that *Q. engelmannii* is poorly adapted to survive outside of its narrow niche of low to mid-elevation plateaus in southern California and will be under greater threat of population declines in the future. It is possible that ancient introgression allowed this species to persist and current hybridization may continue to play that role.

Our findings revealed that the most likely demographic model did not include significant reductions in population size in *Q. berberidifolia* as were seen in *Q. engelmannii,* which implies that *Q. berberidifolia* may have been able to avoid similar reductions in population size even under large-scale climatic shifts. Perhaps *Q. berberidifolia* was not as strongly affected by such a shift and was able to rapidly adapt to changes in climate. Under this demographic scenario it is possible to infer that *Q. berberidifolia* had higher fitness during this transitionary period, which may explain the higher rates of gene flow from *Q. berberidifolia* into *Q. engelmannii*. Two alternative explanations are possible. The larger population sizes of *Q. berberidifolia* populations could have led to differential pollen swamping of *Q. engelmannii*, a process that has been hypothesized in European oak taxa (Lepais et al., 2009). In addition, the asymmetrical gene flow could also be promoted by other pre-zygotic barriers, such as asymmetric overlap in flowering time or differential pollen competition between species (Lepais et al., 2013). Gene flow between species could have altered phenological responses in either species, as has been shown in monkey flowers (*Mimulus* spp.) (Brandvain et al., 2014), which could further reinforce asymmetric gene flow between species. Finally, evidence of asymmetric introgression could be due to the transfer of adaptive genetic material across species boundaries (Barton, 2001). Given that the ancient asymmetrical gene flow coincides with the transition to a Mediterranean-type climate in the initiation of gene flow between these species, the evidence supports the adaptive introgression hypothesis, but does not preclude the others.

This study provides circumstantial evidence that asymmetrical introgression may reflect the exchange of adaptive genetic variation from the California endemic oaks into the more distantly related Mexican clade. The next step would be to examine whole genome sequences to identify and classify regions of introgression throughout the genome. For example, recent studies

of humans and multiple other extinct hominid species found evidence of introgression throughout the human genome (Green et al., 2010; Sankararaman et al., 2014; Vernot et al., 2016). Using statistical approaches such as *D*-statistics (Green et al., 2010) and *S** (Vernot et al., 2016) in the analysis of whole-genome data, it is possible to identify regions of the genome that are more similar to other species than expected by chance alone. By mapping these regions to a well annotated reference genome, these studies were able to show that functional genes were involved in introgression. The two distantly related species examined here provide an excellent opportunity to test adaptive introgression in oaks and identify whether genes associated with climate were involved.

Literature Cited

Abbott, R. J., Barton, N. H., & Good, J. M. (2016). Genomics of hybridization and its evolutionary consequences. *Molecular Ecology*, *25*(11), 2325-2332. https://doi.org/10.1111/mec.13685

Anderson, E. (1949). Introgressive hybridization. *Chapman and Hall, London*.

Anderson, E., & Stebbins, G. (1954). Hybridization as an evolutionary stimulus. *Evolution*, *8*(4), 378-388.

Arnold, M. L. (1997). *Natural hybridization and evolution*. Oxford University Press on Demand.

Arnold, M. L., Buckner, C. M., & Robinson, J. J. (1991). Pollen-mediated introgression and hybrid speciation in Louisiana irises. *Proceedings of the National Academy of Sciences*, *88*(4), 1398-1402.

Barton, N. (2001). The role of hybridization in evolution. *Molecular Ecology*, *10*(3), 551-568.

Bay, R. A., & Ruegg, K. (2017). Genomic islands of divergence or opportunities for introgression? *Proceedings of the Royal Society B-Biological Sciences*, *284*(1850), Article 20162414. https://doi.org/10.1098/rspb.2016.2414

Becker, M., Gruenheit, N., Steel, M., Voelckel, C., Deusch, O., Heenan, P. B., McLenachan, P. A., Kardailsky, O., Leigh, J. W., & Lockhart, P. J. (2013). Hybridization may facilitate in situ survival of endemic species through periods of climate change. *Nature Climate Change*, *3*(12), 1039-1043. https://doi.org/10.1038/nclimate2027

Brandvain, Y., Kenney, A. M., Flagel, L., Coop, G., & Sweigart, A. L. (2014). Speciation and Introgression between *Mimulus nasutus* and *Mimulus guttatus*. *PLoS Genetics*, *10*(6), Article e1004410. https://doi.org/10.1371/journal.pgen.1004410

Christe, C., Stölting, K. N., Bresadola, L., Fussi, B., Heinze, B., Wegmann, D., & Lexer, C. (2016). Selection against recombinant hybrids maintains reproductive isolation in hybridizing *Populus* species despite F1 fertility and recurrent gene flow. *Molecular Ecology*, *25*(11), 2482-2498.

Christe, C., Stölting, K. N., Paris, M., Fraïsse, C., Bierne, N., & Lexer, C. (2017). Adaptive evolution and segregating load contribute to the genomic landscape of divergence in two tree species connected by episodic gene flow. *Molecular Ecology*, *26*(1), 59-76.

Coffman, A. J., Hsieh, P. H., Gravel, S., & Gutenkunst, R. N. (2015). Computationally efficient composite likelihood statistics for demographic inference. *Molecular Biology and Evolution*, *33*(2), 591-593.

Crawford, J. E., Riehle, M. M., Guelbeogo, W. M., Gneme, A., Sagnon, N., Vernick, K. D., Nielsen, R., & Lazzaro, B. P. (2015). Reticulate speciation and barriers to introgression in the *Anopheles gambiae* species complex. *Genome Biology and Evolution*, *7*(11), 3116-3131. https://doi.org/10.1093/gbe/evv203

Dasmahapatra, K. K., Walters, J. R., Briscoe, A. D., Davey, J. W., Whibley, A., Nadeau, N. J., Zimin, A. V., Hughes, D. S., Ferguson, L. C., & Martin, S. H. (2012). Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature*, *487*(7405), 94.

De La Torre, A. R., Wang, T., Jaquish, B., & Aitken, S. N. (2014). Adaptation and exogenous selection in a *Picea glauca* × *Picea engelmannii* hybrid zone: implications for forest management under climate change. *New Phytologist*, *201*(2), 687-699.

Dodd, R. S., & Afzal-Rafii, Z. (2004). Selection and dispersal in a multispecies oak hybrid zone. *Evolution*, *58*(2), 261-269.

Edelman, N. B., Frandsen, P., Miyagi, M., Clavijo, B. J., Davey, J., Dikow, R., Accinelli, G. G., Van Belleghem, S., Patterson, N. J., & Neafsey, D. E. (2018). Genomic architecture and introgression shape a butterfly radiation. *bioRxiv*, 466292.

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PloS one*, *6*(5), e19379.

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, *14*(8), 2611-2620.

Fick SE, & Hijmans, R. (2017). Worldclim 2: new 1km spatial resolution climate surfaces for global land areas.

Fitz-Gibbon, S., Hipp, A. L., Pham, K. K., Manos, P. S., & Sork, V. L. (2017). Phylogenomic inferences from reference-mapped and de novo assembled short-read sequence data using RADseq sequencing of California white oaks (*Quercus* section *Quercus*). *Genome*, *60*(9), 743-755.

Green, R. E., Krause, J., Briggs, A. W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., Li, H., Zhai, W., & Fritz, M. H.-Y. (2010). A draft sequence of the Neandertal genome. *Science*, *328*(5979), 710-722.

Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H., & Bustamante, C. D. (2009). Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genetics*, *5*(10), e1000695.

Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H., & Bustamante, C. D. (2010). Diffusion approximations for demographic inference: DaDi. *Nature Precedings*, *1*(1).

Hamilton, J. A., Lexer, C., & Aitken, S. N. (2013). Genomic and phenotypic architecture of a spruce hybrid zone (*Picea sitchensis* x *P. glauca*). *Molecular Ecology*, *22*(3), 827-841. https://doi.org/10.1111/mec.12007

Harrison, R. G., & Larson, E. L. (2016). Heterogeneous genome divergence, differential introgression, and the origin and structure of hybrid zones. *Molecular Ecology*, *25*(11), 2454-2466.

Hipp, A. L., Manos, P. S., González-Rodríguez, A., Hahn, M., Kaproth, M., McVay, J. D., Avalos, S. V., & Cavender-Bares, J. (2018). Sympatric parallel diversification of major oak clades in the Americas and the origins of Mexican species diversity. *New Phytologist*, *217*(1), 439-452.

Hipp, A. L., Manos, P. S., Hahn, M., Avishai, M., Bodénès, C., Cavender-Bares, J., Crowl, A. A., Deng, M., Denk, T., & Fitz-Gibbon, S. (2019). Genomic landscape of the global oak phylogeny. *New Phytologist*, *226*(4), 1198-1212.

Howard, D. J., Preszler, R. W., Williams, J., Fenchel, S., & Boecklen, W. J. (1997). How discrete are oak species? Insights from a hybrid zone between *Quercus grisea* and *Quercus gambelii*. *Evolution*, *51*(3), 747-755.

Kim, B. Y., Wei, X., Fitz-Gibbon, S., Lohmueller, K. E., Ortego, J., Gugger, P. F., & Sork, V. L. (2018). RADseq data reveal ancient, but not pervasive, introgression between Californian tree and scrub oak species (*Quercus* sect. *Quercus*: *Fagaceae*). *Molecular Ecology*, *27*(22), 4556-4571.

Kremer, A. (2016). Microevolution of European temperate oaks in response to environmental changes. *Comptes rendus biologies*, *339*(7), 263-267.

Lathrop, E. W., Osborne, C., Rochester, A., Yeung, K., Soret, S., & Hopper, R. (1991). *Size class distribution of Quercus engelmannii (Engelmann Oak) on the Santa Rosa Plateau, Riverside County, California* The Symposium on Oak Woodlands and Hardwood Rangeland Management, Davis, California.

Lepais, O., Petit, R. J., Guichoux, E., Lavabre, J. E., Alberto, F., Kremer, A., & Gerber, S. (2009). Species relative abundance and direction of introgression in oaks. *Molecular Ecology*, *18*(10), 2228-2242. https://doi.org/10.1111/j.1365-294X.2009.04137.x

Lepais, O., Roussel, G., Hubert, F., Kremer, A., & Gerber, S. (2013). Strength and variability of postmating reproductive isolating barriers between four European white oak species [Article]. *Tree Genetics & Genomes*, *9*(3), 841-853. https://doi.org/10.1007/s11295-013-0602-3

Leroy, T., Rougemont, Q., Dupouey, J. L., Bodenes, C., Lalanne, C., Belser, C., Labadie, K., Le Provost, G., Aury, J. M., & Kremer, A. (2019). Massive postglacial gene flow between European white oaks uncovered genes underlying species barriers. *New Phytologist*, *226*(4), 1183-1197.

Leroy, T., Roux, C., Villate, L., Bodenes, C., Romiguier, J., Paiva, J. A. P., Dossat, C., Aury, J. M., Plomion, C., & Kremer, A. (2017). Extensive recent secondary contacts between four European white oak species. *New Phytologist*, *214*(2), 865-878. https://doi.org/10.1111/nph.14413

Mallet, J. (2005). Hybridization as an invasion of the genome. *Trends in Ecology & Evolution*, *20*(5), 229-237.

Martin, N. H., Bouck, A. C., & Arnold, M. L. (2006). Detecting Adaptive Trait Introgression between *Iris fulva* and *Iris brevicaulis* in highly-selective field conditions. *Genetics*, *172*(4), 2481-2489.

Ortego, J., Gugger, P. F., Riordan, E. C., & Sork, V. L. (2014). Influence of climatic niche suitability and geographical overlap on hybridization patterns among southern Californian oaks. *Journal of biogeography*, *41*(10), 1895-1908.

Ortego, J., Gugger, P. F., & Sork, V. L. (2015). Climatically stable landscapes predict patterns of genetic structure and admixture in the Californian canyon live oak. *Journal of biogeography*, *42*(2), 328-338.

Ortego, J., Gugger, P. F., & Sork, V. L. (2017). Impacts of human-induced environmental disturbances on hybridization between two ecologically differentiated Californian oak species. *New Phytologist*, *213*(2), 942-955.

Ortego, J., Riordan, E. C., Gugger, P. F., & Sork, V. L. (2012). Influence of environmental heterogeneity on genetic diversity and structure in an endemic southern Californian oak. *Molecular Ecology*, *21*(13), 3210-3223.

Pearse, I. S., & Hipp, A. L. (2009). Phylogenetic and trait similarity to a native species predict herbivory on non-native oaks. *Proceedings of the National Academy of Sciences*, *106*(43), 18097-18102.

Petit, R. J., Bialozyt, R., Garnier-Gere, P., & Hampe, A. (2004). Ecology and genetics of tree invasions: from recent introductions to Quaternary migrations. *Forest Ecology and Management*, *197*(1-3), 117-137. https://doi.org/10.1016/j.foreco.2004.05.009

Petit, R. J., Bodénès, C., Ducousso, A., Roussel, G., & Kremer, A. (2004). Hybridization as a mechanism of invasion in oaks. *New Phytologist*, *161*(1), 151-164.

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*(2), 945-959.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., Maller, J., Sklar, P., De Bakker, P. I., & Daly, M. J. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, *81*(3), 559-575.

Rieseberg, L. H., Raymond, O., Rosenthal, D. M., Lai, Z., Livingstone, K., Nakazato, T., Durphy, J. L., Schwarzbach, A. E., Donovan, L. A., & Lexer, C. (2003). Major ecological transitions in wild sunflowers facilitated by hybridization. *Science*, *301*(5637), 1211-1216.

Rieseberg, L. H., & Wendel, J. F. (1993). Introgression and its consequences in plants. *Hybrid zones and the evolutionary process*, 70-109.

Rieseberg, L. H., Whitton, J., & Gardner, K. (1999). Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species. *Genetics*, *152*(2), 713-727.

Riordan, E. C., Gillespie, T. W., Pitcher, L., Pincetl, S. S., Jenerette, G. D., & Pataki, D. E. (2015). Threats of future climate change and land use to vulnerable tree species native to Southern California. *Environmental Conservation*, *42*(2), 127-138.

Riordan, E. C., Gugger, P. F., Ortego, J., Smith, C., Gaddis, K., Thompson, P., & Sork, V. L. (2016). Association of genetic and phenotypic variability with geography and climate in three southern California oaks. *American Journal of Botany*, *103*(1), 73-85.

Roberts, F. M. (1995). *Illustrated guide to the oaks of the Southern Californian floristic province*. FM Roberts Publications.

Ruegg, K., Anderson, E. C., Boone, J., Pouls, J., & Smith, T. B. (2014). A role for migration-linked genes and genomic islands in divergence of a songbird. *Molecular Ecology*, *23*(19), 4757-4769. https://doi.org/10.1111/mec.12842

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

Sankararaman, S., Mallick, S., Dannemann, M., Prüfer, K., Kelso, J., Pääbo, S., Patterson, N., & Reich, D. (2014). The genomic landscape of Neanderthal ancestry in present-day humans. *Nature*, *507*(7492), 354-357.

Sork, V. L., Cokus, S. J., Fitz-Gibbon, S. T., Zimin, A. V., Puiu, D., Garcia, J. A., Gugger, P. F., Henriquez, C. L., Zhen, Y., & Lohmueller, K. E. (2021). High-quality genome and methylomes illustrate features underlying evolutionary success of oaks. *bioRxiv*.

Suarez-Gonzalez, A., Hefer, C. A., Christe, C., Corea, O., Lexer, C., Cronk, Q. C., & Douglas, C. J. (2016). Genomic and functional approaches reveal a case of adaptive introgression from *Populus balsamifera* (balsam poplar) in *P. trichocarpa* (black cottonwood). *Molecular Ecology*, *25*(11), 2427-2442.

Suarez-Gonzalez, A., Hefer, C. A., Lexer, C., Cronk, Q. C., & Douglas, C. J. (2018). Scale and direction of adaptive introgression between black cottonwood (*Populus trichocarpa*) and balsam poplar (*P. balsamifera*). *Molecular Ecology*, *27*(7), 1667-1680.

Vallejo-Marin, M., Cooley, A. M., Lee, M. Y., Folmer, M., McKain, M. R., & Puzey, J. R. (2016). Strongly asymmetric hybridization barriers shape the origin of a new polyploid species and its hybrid ancestor. *American Journal of Botany*, *103*(7), 1272-1288. https://doi.org/10.3732/ajb.1500471

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

Vernot, B., Tucci, S., Kelso, J., Schraiber, J. G., Wolf, A. B., Gittelman, R. M., Dannemann, M., Grote, S., McCoy, R. C., & Norton, H. (2016). Excavating Neandertal and Denisovan DNA from the genomes of Melanesian individuals. *Science*, *352*(6282), 235-239.

vonHoldt, B. M., Kays, R., Pollinger, J. P., & Wayne, R. K. (2016). Admixture mapping identifies introgressed genomic regions in North American canids. *Molecular Ecology*, *25*(11), 2443-2453.

Whitney, K. D., Broman, K. W., Kane, N. C., Hovick, S. M., Randell, R. A., & Rieseberg, L. H. (2015). Quantitative trait locus mapping identifies candidate alleles involved in adaptive introgression and range expansion in a wild sunflower. *Molecular Ecology*, *24*(9), 2194-2211. https://doi.org/10.1111/mec.13044

Whittemore, A. T., & Schaal, B. A. (1991). Interspecific gene flow in sympatric oaks. *Proceedings of the National Academy of Sciences*, *88*(6), 2540-2544.

Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., & Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*, *28*(24), 3326-3328.

Chapter 2

Environmental association analysis reveals candidate loci involved in local adaptation and future maladaptation in two hybridizing white oak (*Quercus* sect. *Quercus*)

Abstract

Landscape genomic techniques have been used to identify patterns of local adaptation and are a vital tool in developing successful conservation plans for many plant and animal taxa. Comparing patterns of local adaptation in species that share similar geographic distributions can reveal how selection influences the evolution of different taxa within similar geographic regions. Of increasing concern in conservation biology is the effect that selection plays on hybridization and introgression between compatible taxa. This study aims to use this framework to compare two hybridizing oak taxa; *Quercus engelmannii* and *Q. berberidifolia*, to identify which climatic variables may be shaping patterns of local adaptation in each species and whether selection may favor the exchange of adaptive alleles through hybridization and introgression. Using a landscape genomic method, Gradient Forest, which maps the whole genome sequence data of 92 *Q. engelmannii* individuals and 79 *Q. berberidifolia* individuals onto geographic and climate gradients, we found that the most important climatic factor was precipitation seasonality for both species within the region of their distribution that overlaps. When comparing current predicted

geographic patterns of genomic diversity to predicted patterns of genomic diversity under future climatic scenarios, our results indicated a pattern of potential maladaptation under future climate scenarios in eastern populations of *Q. engelmannii*, but not *Q. berberidifolia.* We conclude that further climate warming is likely to result in maladaptation in *Q. engelmannii*, but hybridization with the more drought tolerant shrub oak may provide a source of adaptive variation that will mitigate against maladaptation.

Introduction

Generating accurate assessments of the drivers of local adaptation and species distributions is one of the most critical priorities as we attempt to develop and improve conservation efforts in light of rapidly changing climate regimes (Manel & Holderegger, 2013; Shaffer et al., 2022). This priority is especially true for sessile organisms such as plants, which often are unable to migrate fast enough to avoid rapid climatic or environmental upheaval, yet still show signatures of selection throughout their geographic distribution (Browne et al., 2019; Gugger et al., 2021; Sork et al., 2013). By analyzing the correlation between spatial patterns of genetic diversity in plant taxa and environmental factors using landscape genomic techniques, such as environmental association analyses (EAAs), it is possible to identify the association between climate and evolutionary processes, including genetic drift, the development of potential abiotic barriers to gene flow between populations, and demographic history correlated with environmental factors (De La Torre et al., 2014; Holliday et al., 2017; Sork et al., 2013; Storfer et al., 2018). When selection results in a given phenotype having increased fitness in a local environment over phenotypes from other environments, then the frequencies of the alleles underlying that phenotype will follow the environmental gradient that is generating that selective pressure (Savolainen et al., 2013; Savolainen et al., 2007; Storfer et al., 2007). Thus, it is likely that climatic factors can influence the distribution of genetic and genomic diversity within the distribution of a species (Savolainen et al., 2013). Increases in the availability and application of whole genome sequencing techniques have led to an expansion of landscape genomic methods, which can identify potentially adaptive loci underlying local adaptation (Fitzpatrick & Keller, 2015; Frichot et al., 2013; Holderegger & Wagner, 2008; Rellstab et al., 2015; Van Strien et al., 2012). Furthermore, by utilizing advanced climatic modeling data, predicted future

environmental conditions can also be used to predict future genetic and genomic turnover in response to changing environmental conditions (Shryock et al., 2021). Therefore, a landscape genomic approach can offer insight into how current and future climatic conditions affect the evolution of species in their native range.

Hybridization is one evolutionary process that is likely strongly associated with local climatic gradients and may also serve as a source of local adaptation through the exchange of alleles that improve fitness across species boundaries (Abbott et al., 2016; Anderson & Stebbins, 1954; Arnold, 1997; Barton & Hewitt, 1985). Adaptive introgression has recently become recognized as a much more important influence on the adaptation and persistence of many plant taxa with examples in poplar (*Populus* spp.) (Christe et al., 2017; Suarez-Gonzalez et al., 2017), spruce (*Picea* spp.) (Hamilton et al., 2013) and oaks (*Quercus* spp.) (Leroy et al., 2020). However, previous studies that have identified signals of adaptive introgression have not focused on the environmental or climatic factors that may be driving local adaptation and thus patterns of introgression throughout a species distribution and its overlap with any potential reproductively compatible conspecific individuals (Abbott et al., 2016). An important first step in identifying adaptive introgression should be the characterization of the environmental gradients driving local adaptation and the alleles that are potentially associated with those gradients.

Oaks (*Quercus* spp.) represent a candidate system to study the effects of climate on the distribution of neutral versus candidate loci for local adaptation. Landscape genetic and genomic analyses have shown evidence of local adaptation in valley oak (*Q. lobata*) (Gugger et al., 2021; Sork et al., 2013), netleaf oak (*Q. rugosa*) (Martins et al., 2018), and cork oak (*Q. suber*) (Vanhove et al., 2021). By utilizing comparisons of background genomic variation to candidate loci for local adaptation, it was also possible to identify regions of the distribution of each of

these oak species that are sensitive to selective pressures associated with climatic variables and which of these variables are most important in driving local adaptation (Gugger et al., 2021; Martins et al., 2018; Vanhove et al., 2021). Oaks are also known to hybridize within phylogenetic sections and evidence indicates that patterns of hybridization are potentially associated with ecological factors (Burge et al., 2019; Howard et al., 1997; Van Valen, 1976). The development of hybrid zones between oak taxa has been associated with climatic niche and environmental stability, indicating that hybridization is a dynamic process in oaks that is strongly associated with local environmental conditions (Ortego et al., 2014; Ortego et al., 2015). Utilizing landscape genomic methods for observed allele frequencies in contemporary oak species has shown evidence of adaptive introgression associated with climate gradients in European oaks (Leroy et al., 2020). Additionally, the extent and direction of hybridization and introgression has been associated elevational gradients (Burge et al., 2019) and fire disturbance regime (Ortego et al., 2017), which points to a significant role of selection in shaping patterns of introgression between oak taxa.

The overall goal of this chapter is to identify and compare the distribution of genomic variation of two hybridizing species, especially genomic variation that is associated with climate variables, to understand the extent to which the environment may favor hybridization and introgression. The two hybridizing oak species, Engelmann oak (*Q. engelmannii*) and California scrub oak (*Q. berberidifolia*), have overlapping distributions and environmental niches, but divergent ecological preferences along fine-scale geographical gradients (Ortego et al., 2014; Riordan et al., 2016). The dynamics of such fine-scale geographic gradients allow an extensive opportunity for hybridization and introgression to occur between these species (Ortego et al., 2014). Species distribution modelling efforts have revealed strong associations between the

distribution of both species and similar environmental factors such as temperature seasonality and precipitation seasonality due to their broadly overlapping distributions (Riordan et al., 2016). EAAs have revealed a relationship between two climatic factors, precipitation seasonality and winter precipitation, with genetic gradients in *Q. engelmannii* using microsatellites. However, similar trends could not be drawn for *Q. berberidifolia* (Riordan et al., 2016). Using whole genome sequence data, it is more feasible possible to simultaneously detect both neutral and putative adaptive genetic variation to determine if patterns of local adaptation are driven by different environmental factors in each species (Fitzpatrick & Keller, 2015). Furthermore, by utilizing future climate models, it will be possible to determine if there are portions of the range of each species that will be susceptible to future shifts in climate regimes and whether there is overlap in the geographic distribution of hybrid tension zones and predicted future patterns of adaptive genomic diversity.

This chapter will utilize whole-genome resequencing data to (i) identify the climatic variables that have the strongest influence on the distribution of neutral genomic diversity in each species; (ii) use latent factor mixed modeling to identify outlier loci associated with climatic variables to develop a set of potentially adaptive loci; (iii) map neutral and putatively adaptive genomic diversity of each species to identify the areas of each species distribution are exhibiting the strongest signals of genomic offset due to climate-associated selection and highlight populations of conservation concern, and (iv) compare the current distribution of climate-associated outlier loci to future predicted distributions under two potential future climate regimes. This framework will allow us to assess the influence of climatic factors in local adaptation and how patterns of local adaptation are distributed across the landscape. If divergent patterns of local adaptation under similar selection regimes occur in this system under changing

climate scenarios, then the potential for adaptative introgression may increase over time as well. By understanding the potential for adaptive introgression future research can then be extended to what alleles are being exchanged and what impacts on the fitness of individuals they may have.

Methods

Study system

The two study species, Engelmann oak (*Quercus engelmannii)* and California scrub oak (*Q. berberidifolia),* have broadly overlapping distributions in southern California. *Quercus engelmannii* is a tree oak with a restricted distribution south of the San Gabriel Mountains in southern CA and Baja California, Mexico. Sampling for this project was restricted to the two disjunct populations in southern California. (Figure 2.1). It is a medium-large sized tree with a restrictive habitat preference of mesas with well-developed soil between 500-1500m in elevation (Scott, 1991). *Q. berberidifolia* is a shrub growing up to 3m in height and has a widespread, coastal distribution in California. *Q. berberidifolia* is strongly associated with mixed chaparral ecosystems in which it can be a locally dominant species throughout its range. In southern California, these species share broadly similar habitat niches and geographic distributions, but their habitat preferences result in ecological separation across fine-scale geographic distances throughout the area in which their distributions overlap (Ortego et al., 2014; Riordan et al., 2016).

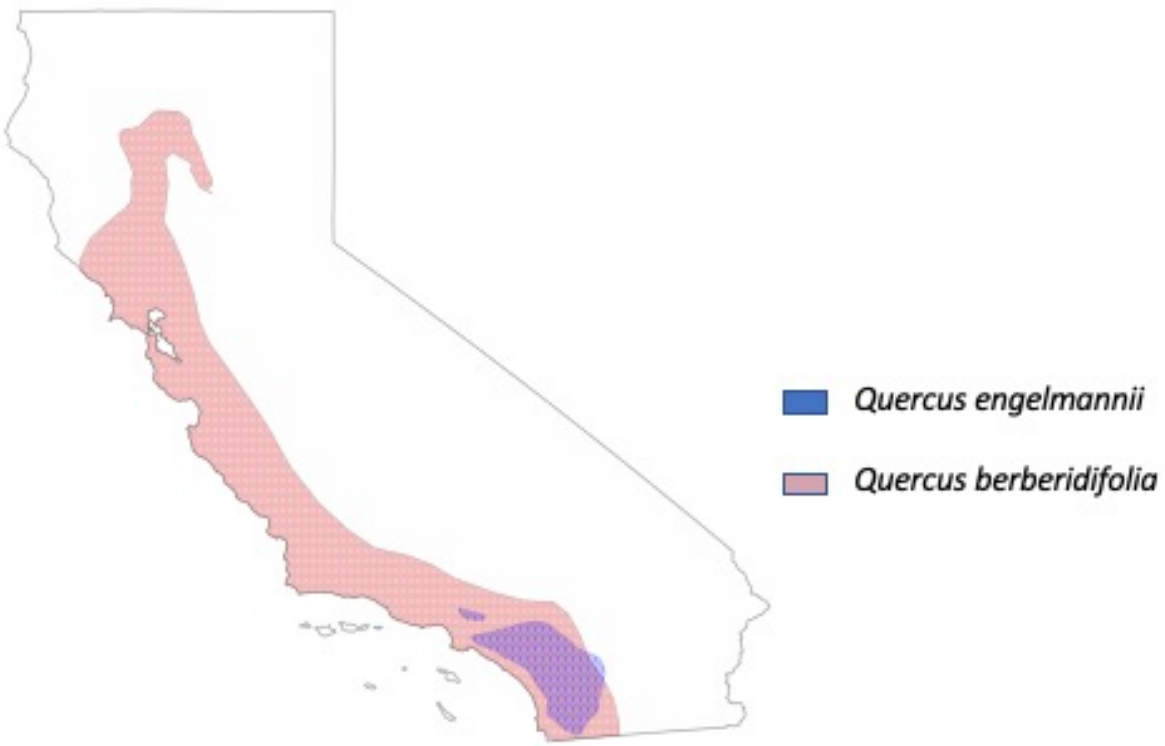

Figure 2.1: A distribution map of *Quercus berberidifolia* (pink) and *Q. engelmannii* (blue) across California, USA excluding portions of their distribution in Baja California, MX.

Sampling and DNA Extraction

Range-wide sampling was conducted for each species: 98 *Quercus engelmannii* individuals and 100 *Q. berberidifolia* (Fig. 2.2). Fresh leaf tissue was collected from each individual and store on ice until transferred to long-term storage at -80C. DNA was extracted using a modified Qiagen DNEasy mini plant kit, which utilized liquid nitrogen to flash-freeze the leaf tissue and manual grinding using a mortar and pestle and two additional repetitions of the prewash steps.

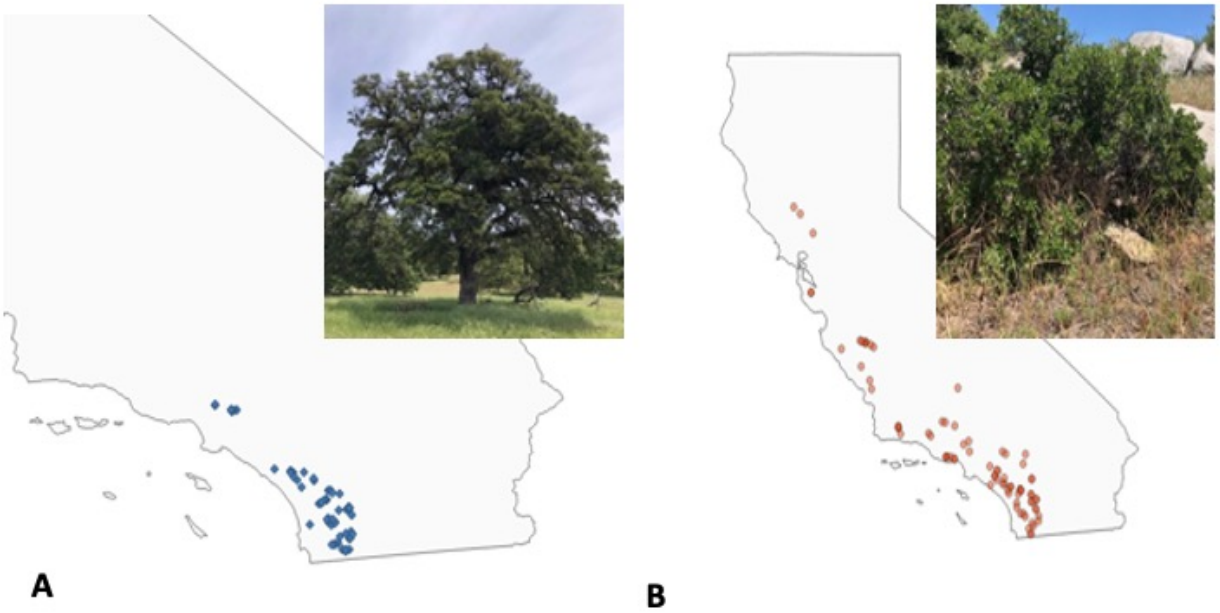

Figure 2.2: Sampling sites for *Quercus engelmannii* (A) and *Q. berberidifolia* (B).

Sequencing and Variant Calling

Extracted samples were sent to the genomic sequencing core at University of California, Davis for library preparation and sequencing. Libraries were prepared using a custom SeqWell kit designed for ½ reaction volume due to low DNA quantities per sample. Sequences were mapped to the reference genomes for *Quercus engelmannii* and *Q. berberidifolia* respectively following methods previously used for other oak species (Fitz-Gibbon et al., 2017). Variants were called using the default parameters in the package ForestQC (Li et al., 2019). Samples that did not meet the threshold of 7x total mapped coverage were discarded from downstream analysis resulting in a total of 79 *Quercus berberidifolia* individuals and 92 *Q. engelmannii* individuals respectively. For each sample, biallelic markers were retained and filtered for linkage disequilibrium using sliding windows of 50bp moving 1bp each step and an r^2 value of 0.5 with a minor allele frequency of 0.065 using Plink v. 2.0 (Purcell et al., 2007).

LFMM Methods

Outlier SNPs associated with climatic variables were identified using latent factor mixed modeling (LFMM) (Frichot et al., 2013) in R 4.2.1 (Team, 2022) using the package "LEA" (Frichot & François, 2015), following the methods previously described for landscape genomic analysis in other oak taxa (Gugger et al., 2021; Martins et al., 2018). Briefly, to control for underlying population structure, the discrete number of genetic clusters (K) and admixture coefficients were identified using SNMF. The most likely number of clusters were chosen where cross-entropy plateaus for each species. The most likely number of genetic clusters for each species were then used as latent factors and missing SNP data was imputed using commands in the package "LEA" as recommended by the authors (Frichot & François, 2015; Frichot et al., 2013).

Climate variables were downloaded from the worldclim database at a resolution of 10 seconds (Fick & Hijmans, 2017). Climate variables were filtered for autocorrelation and the remaining variables used for this analysis were Bio 3; isothermality, Bio 6; mean temperature of driest quarter, Bio 13; precipitation of wettest month, Bio 15; precipitation seasonality, Bio 14; precipitation of driest month, and elevation. LFMM was run to determine the association of SNPs with environmental factors using 10 repetitions, a burn-in of 5,000 steps with 10,000 iterations. SNPs were ranked according to probability of association with climate variables based on their z-score value. Resulting z-scores were converted to q-values using the "qvalue" 2.6.0 package in R to determine the statistical significance of z-score values. SNPs with a q-value < 0.01 and a z-score of >5 were chosen to select loci that had a significant association with climatic gradients that also were of large effect size. These steps were repeated for a subset of

Quercus berberidifolia individuals that comprised only the southern portion of the range of the species to compare the distribution of genetic offset between current climate-associated outliers and their predicted distributions under multiple future climate scenarios.

Gradient Forest Methods

We ran Gradient Forest using the "LEA" package in R for both genomic diversity in all contexts as well as candidate outlier loci identified by LFMM. Gradient Forest utilizes a machine learning approach to identify non-linear relationships between genomic data and climatic or environmental variables across the landscape (Fitzpatrick & Keller, 2015). To compare and reinforce the associations between climate and outlier alleles identified using LFMM, Gradient Forest was run twice for each species: once using a subset of 10,000 random SNPs and once using outliers identified in LFMM. The climatic variables used were the same variables described in the LFMM methods (see above). Because Gradient Forest does not explicitly test for associations between allelic variables and geographic variables, a set of principal coordinates of neighborhood matrices (PCNM), otherwise known as Moran's eigenvector maps (MEM) were developed using geospatial coordinates of the sampled individuals using the "pcnm" function in the R package "vegan" (Oksanen et al., 2013). Gradient Forest was run using the default parameters as recommended for each SNP dataset (Fitzpatrick & Keller, 2015).

Gradient Forest results were mapped using the "predict" function, which acts to predict genetic variation across all grid cells. The principal components of the predicted genetic diversity were used to generate a red-blue-green color scale for the first three PC axes for each iteration of GF according to the methods outlined in Fitzpatrick and Keller (2015). This function also provides the importance of each variable to the prediction of the distribution of genomic

diversity. It is important to note that this does not quantify the predictive power of each variable, but only the weighted relative importance of individual predictive values (Figure 2.4). We identified the top 5 most important predictor variables and compared them both across species using all SNPs and only outlier SNPs. To identify genomic offset between all SNPs vs. outlier SNPs, the GF results for each category were compared independently for each species using a Procrustes superimposition in "vegan". The resulting Procrustes residuals were mapped to highlight the absolute distance in genetic composition between outliers and all SNPs and identify regions of each species distribution where genomic offset was highest.

To model potential vulnerability of both species under future climate regimes we measured genetic offset, which is proportional to F_{ST} between current and future populations. Genomic offset can be used to represent the expected difference in allele frequencies between current and future climate conditions (Fitzpatrick & Keller, 2015). The same steps were repeated for two predicted climate scenarios in California that represent a warmer wetter (MIROC-ESM6-1) and hotter, drier (CNRM-ES6-1) climate scenario. Outlier SNPs identified using LFMM for *Q. engelmannii* and *Q. berberidifolia* were used to predict future genetic offset under both climate scenarios. To compare local adaptation in both species under the same geospatial extent and the area of secondary contact between them, individuals of *Q. berberidifolia* were restricted to populations from south of the Transverse Ranges in southern California. Outliers identified by LFMM were used in Gradient Forest analyses to predict future genetic turnover under both scenarios. Genetic offset was identified by comparing the predicted genetic turnover of outlier SNPs under each future climate scenario to current predicted genetic turnover. Procrustes residuals of these comparisons were mapped to identify predicted geographic regions of high genetic offset under each scenario.

Results

LFMM

SNMF results for *Quercus engelmannii* indicate that the most likely number of genetic clusters in this sample is 4 (K=4). *Q. berberidifolia* had a most likely number of genetic clusters of 2 (K=2) for the range-wide dataset. To account for any additional hidden substructure in southern populations, K= 3 was used when identifying outliers in only southern populations (Fitzpatrick & Keller, 2015). For LFMM analyses, a K=4 was used as a latent factor for *Quercus engelmannii*. For *Q. berberidifolia*, a K=2 and a K=3 were used as latent factors for range-wide and southern populations respectively. After filtering for q and z scores, a total of 2,502 outlier alleles significantly associated with climate gradients were identified using the range-wide dataset for *Q. berberidifolia* and a total of 780 climate-associated SNPs were identified in the southern populations. A total of 55 climate-associated outlier alleles passed filtering thresholds for *Q. engelmannii*. These outliers were then used in downstream Gradient Forest analyses.

Gradient Forest Results

Gradient Forest analyses indicate that overall genomic turnover in both *Quercus engelmannii* and *Quercus berberidifolia* are strongly correlated with numerous but differing climatic variables (Figure 2.3). In both species, the most important variables determining the distribution of genomic diversity were related to "hidden" geospatial variables, but precipitation seasonality and minimum temperature of coldest month were among the top five most important variables in shaping neutral genomic diversity in *Q. engelmannii* and *Q. berberidifolia* respectively (Figure 2.3A and 2.3C. respectively). This indicates that differing climatic variables are key in shaping the range-wide patterns of genomic diversity in each species. However, when

only including individuals from southern populations of *Q. berberidifolia*, GF analyses indicate that the most important factor in shaping genomic diversity was precipitation seasonality with less of an impact of geospatial variables. Precipitation seasonality was also important in shaping the species distribution models of both species (Riordan et al., 2016).

Genetic offset under current climate conditions

Genetic offset between overall genomic diversity and candidate adaptive loci in *Q. engelmannii* reveals stronger divergence in genomic turnover between neutral and potentially adaptive genomic variation along the eastern edge of its distribution (Figure 2.4C). Additionally, there are signals of genetic offset found in the disjunct northern population, which may be associated with population substructure. The core distribution of the species shows little genomic offset between overall genomic turnover and turnover at outlier loci identified using LFMM, which points to a strong association between the west-east gradient in precipitation seasonality and signals of selection for outlier loci.

Genetic offset between neutral and outlier candidate loci revealed differences in distribution of neutral genetic variation and outlier loci in multiple areas of the range of *Q. berberidifolia* (Figure 2.4F). Genomic offset was strongest in the Tehachapi Mountains and along the southeastern edge of its distribution in southern California. When including only southern populations, the major areas of genomic offset were along the eastern edge of its distribution, which directly overlaps with high areas of genetic turnover of outlier loci in *Q. engelmannii* and points to a similar relationship with precipitation seasonality gradients in this species. However, the magnitude of genetic offset is lower in *Q. engelmannii* (Figure 2.4C).

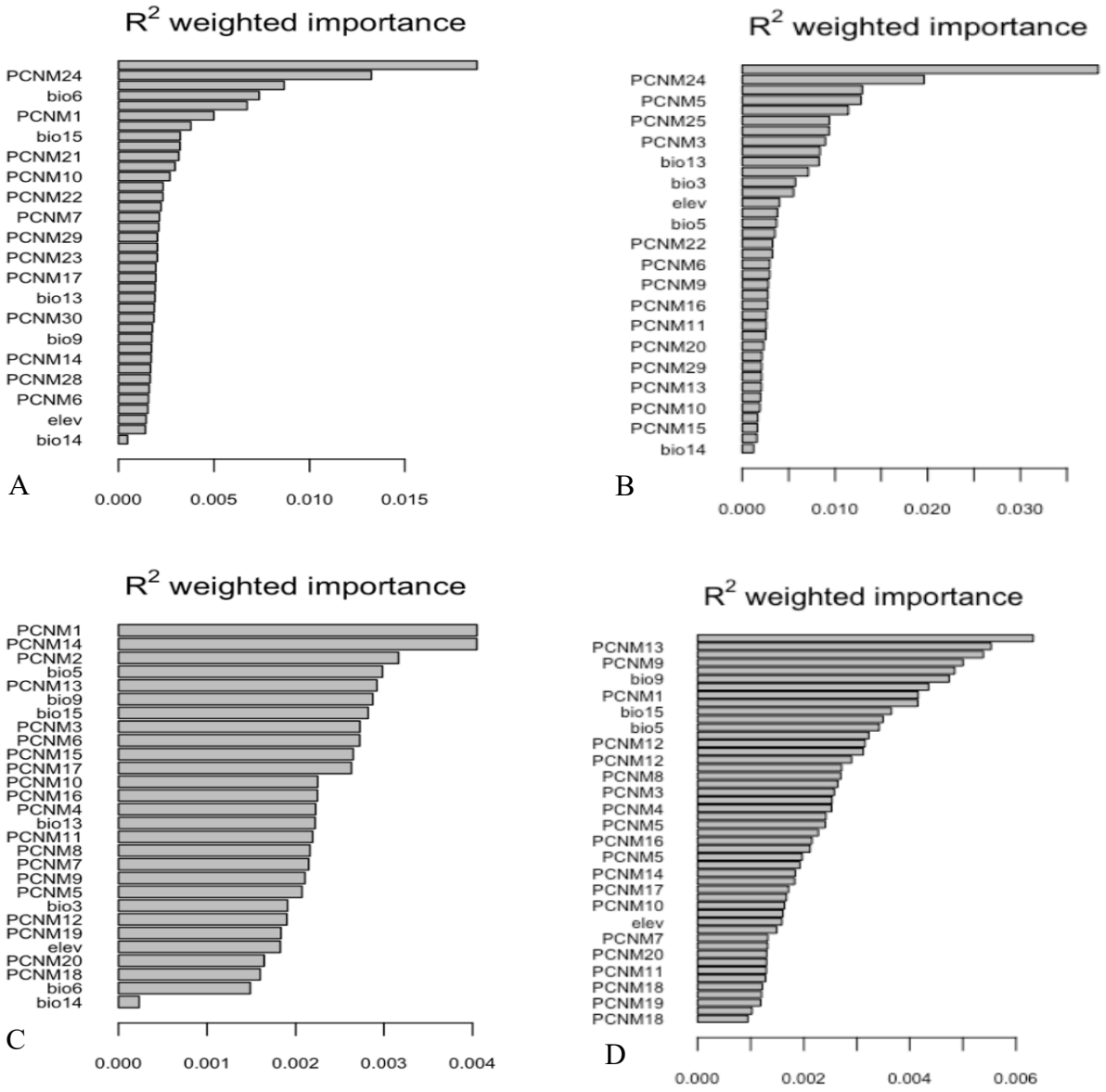

Figure 2.3**:** Bar plots of the relative weighted importance (y-axis) of predictor variables in predicting patterns of genomic diversity using Gradient Forest. The top row represents the importance values for *Q. engelmannii* using all loci (A) and outlier loci only (B). The bottom row represents the most important predictor variables for all and outlier loci in *Q. berberidifolia* (C and D respectively). PCNM variables represent "hidden" geographic variables while "bio" variables represent bioclim climatic variables.

Genetic offset between neutral and outlier candidate loci revealed differences in distribution of neutral genetic variation and outlier loci in multiple areas of the range of *Q. berberidifolia* (Fig. 2.5). Genomic offset was strongest in the Tehachapi Mountains and along the southeastern edge of its distribution in southern California. When including only southern populations, the major areas of genomic offset were along the eastern edge of its distribution, which directly overlaps with high areas of genetic turnover of outlier loci in *Q. engelmannii* and points to a similar relationship with precipitation seasonality gradients in this species. However, the magnitude of genetic offset is lower in *Q. engelmannii* (Figure 2.4C).

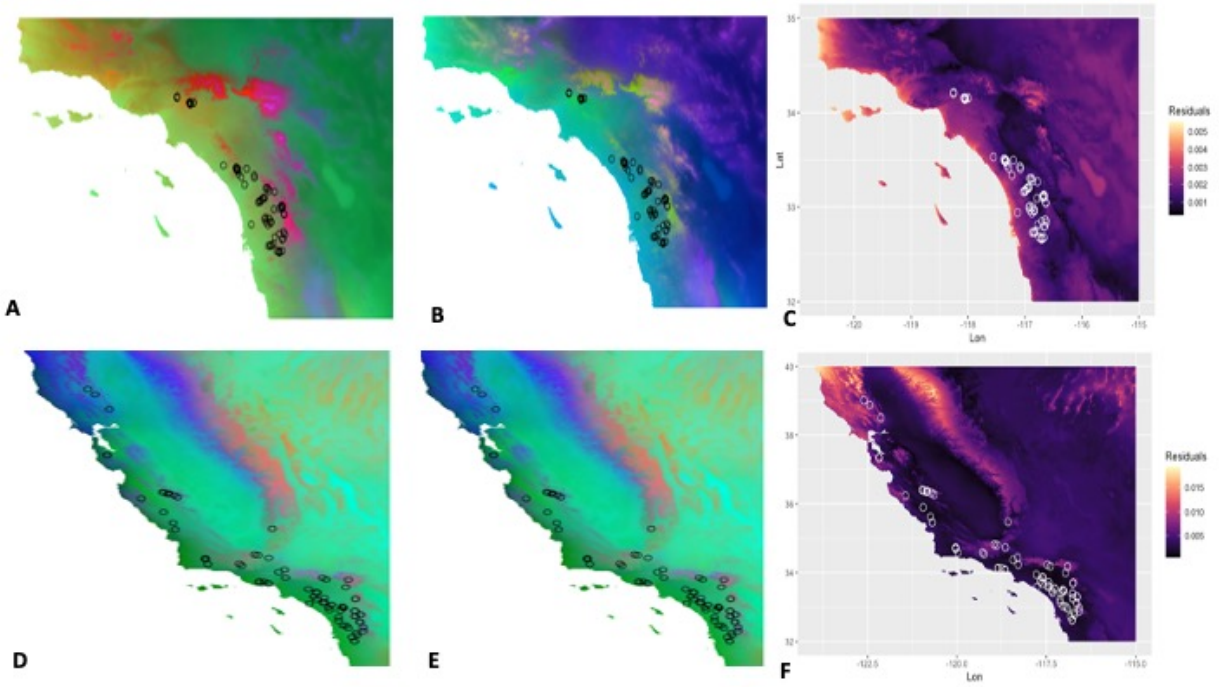

Figure 2.4: Gradient Forest results for neutral genomic turnover (A, D), outlier loci (B, E) and genomic offset (C, F) due to environmental factors for *Q. engelmannii* (A:C) and *Q. berberidifolia* (D:F). The circles represent sampling sites for each species respectively. Similar colors on GF maps (A, B, D, E) represent similar genomic clusters while different colors represent predictions of unique genomic clusters associated with environmental variables. Greater genomic offset in figures C and F is represented by darker colors with red representing the strongest signals of genomic offset.

Genetic offset under future climate scenarios

Predicted future genomic offset of *Q. engelmannii* increases in magnitude under both warmer and wetter and hotter and drier future climate models, but genomic offset in *Q. berberidifolia* remains stable in both scenarios. Similar to patterns of current genetic offset, the regions with greatest genetic offset are found along the eastern edge of the distribution of *Q. engelmannii* and are similar in magnitude under both potentially warmer, wetter, and hotter, drier future climate scenarios (Fig 2.5A and 2.5B). Under the same future climate scenarios, the genetic offset between current outliers and their predicted future allele frequencies does not show a significant change in either scenario (Fig 2.5C and 2.5D).

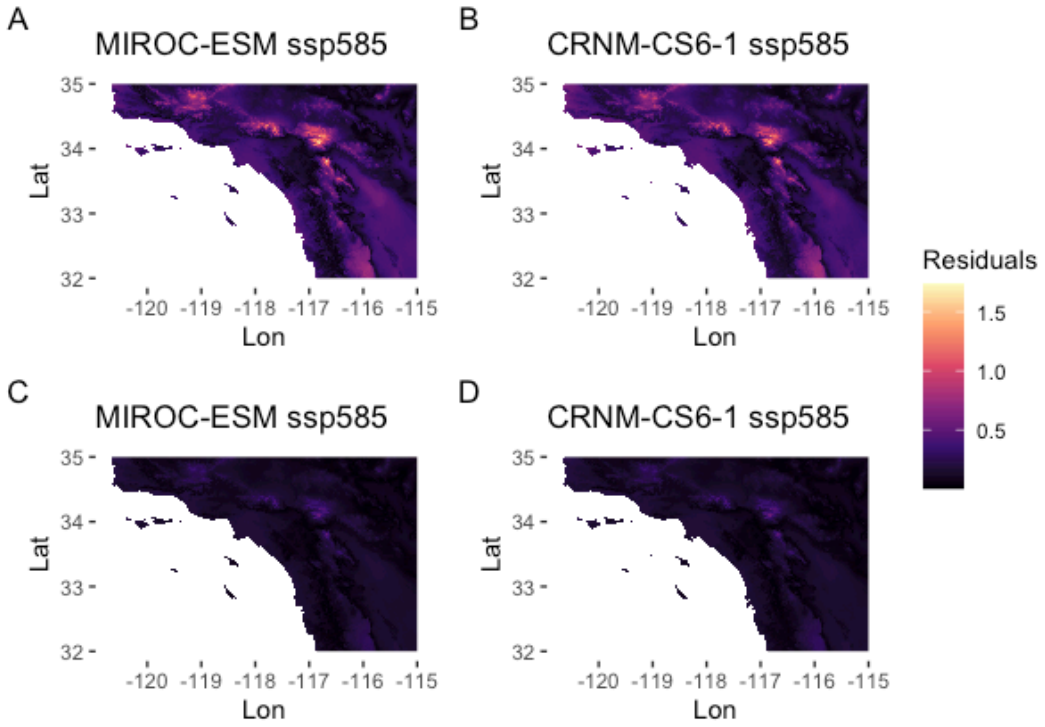

Figure 2.5: Plots of predicted future genomic offset under warmer, wetter (MIROC-ESM; A and C) and hotter, drier (CRNM-CS6-1; B and D) climate conditions for *Q. engelmannii* (top row) and *Q. berberidifolia* (bottom row). Lighter colors indicate stronger signals of genomic offset.

Discussion

Our results indicate that *Quercus engelmannii* is strongly adapted to current climatic conditions and more susceptible than *Q. berberidifolia* to changing future climatic regimes. In *Quercus engelmannii* the most important variables shaping the distribution of genomic diversity were geospatial variables, but precipitation seasonality was the most important climate variable in driving the distribution of neutral and adaptive genomic diversity. Precipitation seasonality was also shown to be important in previous landscape genetics analyses using neutral genetic markers as well as in defining the distribution of the species (Riordan et al., 2016). This agreement indicates that this climatic variable is the driving force behind adaptation, population diversification, and species distribution and is likely a strong selective pressure on the species. Range-wide GF results indicate that the most important variables shaping the distribution of genomic diversity in *Q. berberidifolia* are primarily geospatial, but minimum temperature of coldest month was the most important climatic variable. SDMs predicted that the most important environmental variables shaping the distribution of *Q. berberidifolia* included temperature seasonality and precipitation seasonality (Riordan et al., 2016). However, precipitation seasonality was most important in predicting the distribution of genomic diversity when controlling for only southern populations of *Q. berberidifolia*. It is possible that while similar variables may be important in shaping the overall patterns of genomic diversity in both species where their ranges overlap, the genomic offset between overall genomic diversity and putatively outlier alleles are higher in *Q. engelmannii* than in populations of *Q. berberidifolia* along similar portions of their distributions. Evidence of this genomic offset grows stronger under both wetter and drier future climate scenarios in *Q. engelmannii*, which may show evidence of populations of future conservation concern for the species. When combining evidence of potential

maladaptation in *Q. engelmannii* throughout its range with the predicted stability of populations of *Q. berberidifolia*, the potential for introgression may exist between these two species.

Gradient Forest results for *Q. engelmannii* indicate a significant relationship between geospatial variables and patterns of genomic diversity, but precipitation seasonality is also significantly associated with shaping landscape genomic distributions. This variable was also found to be significant in shaping neutral genetic variation and with overall species distribution models. It is likely that the narrow habitat preference of *Q. engelmannii* is strongly associated with local patterns of precipitation, which serves as a strong selective pressure limiting the overall distribution of the species (Ortego et al., 2014; Scott, 1991). Even though neither genetic nor genomic population substructure is found between populations found at the eastern edge of its distribution and the core distribution of the species, genomic offset shows divergence of outlier alleles in this area. Such a divergence in outlier alleles provides evidence of selective pressures due to variables effecting both the distribution of the species and patterns of genomic diversity of adaptive alleles, which has been seen in other oak species and may mark these populations as important to conservation efforts (Gugger et al., 2021; Martins et al., 2018). By comparing current distributions of outlier alleles to their predicted distributions under both climate scenarios, this predicted genetic offset increased in magnitude and may represent a signal of maladaptation to future climate scenarios in those populations (Capblancq et al., 2020). Similar results were found in popular (*Populus* spp.) (Fitzpatrick & Keller, 2015) and desert plantain (*Plantago ovata*) (Shryock et al., 2021), and the findings for current and future climate models were used to identify populations for conservation priority. When combined with previous work that predicts a significant loss of suitable habitat for this species under future climate change scenarios, our findings highlight that Engelman oak, a major tree oak in the

southern part of California, should be one of significant conservation concern (Riordan et al., 2015).

When controlling for populations found in southern California, GF results indicate that the most important climatic variables in shaping the distribution of genomic diversity in *Q. berberidifolia* are the same as those found to be important in *Q. engelmannii*. Range-wide results for *Q. berberidifolia* show minimum temperature in the coldest month is the most important climatic variable associated with the distribution of neutral and adaptive genetic variation in the species. However, here their populations overlap, similar climatic variables seem to be shaping the genomic diversity of both species along similar climatic gradients. Similarly, to *Q. engelmannii*, the strongest signals of genomic offset between neutral and adaptive candidate alleles occur along the southeastern edge of its range as well as in the Tehachapi Mountains in southern California. Genomic offset is not strong throughout the rest of its range, even when comparing northern and southern populations separated by the Transverse Ranges, which have been identified as a significant barrier to gene flow in this species (Sork et al., 2016). Regions of high genomic offset are also associated with areas that are either known to show unique patterns of genomic diversity in other oak species (i.e., *Q. lobata* in the Tehachapi Mountains; (Gugger et al., 2021)) or possibly represent the extreme edge of the species distribution near the transition to desert habitats in southeastern California. Genetic offset under future climate models does not indicate a significant increase in genomic offset of outlier loci under either warmer, wetter, or colder, drier climate scenarios. A weaker signal of genomic offset likely indicates that *Q. berberidifolia* is under less threat than eastern populations of *Q. engelmannii* under predicted future climatic conditions (Capblancq et al., 2020). However, a lack of maladaptation to future climate conditions in *Q. berberidifolia* combined with an overlap in significant climate variables

potentially underlying selection in both species can provide the opportunity for adaptive introgression between these species.

The hybrid system of *Q. engelmannii* and *Q. berberidifolia* possesses many of the key characteristics that increase the potential of adaptive introgression. Since *Q. berberidifolia* shows less of a genomic offset under both current and future climate scenarios, it is possible that it is better adapted to both current and future climatic conditions. Due to the degree of geographic overlap and often close geographic proximity of the distributions of these species, hybridization and introgression can potentially lead to the asymmetric exchange and assimilation of adaptive alleles into the less well-adapted *Q. engelmannii* (Abbott et al., 2016; Barton, 2001). Potential evidence of adaptive introgression has been shown in in European white oaks that associated adaptive introgression to cycles of continental glaciation (Leroy et al., 2020; Leroy et al., 2019). Similar patterns of asymmetric, potentially adaptive introgression have been shown in other hybrid systems that include *Q. berberidifolia* in which the direction and extent of introgression changed in response to fire regime in the local environment (Ortego et al., 2017). Evidence of ancient introgression between *Q. engelmannii* and *Q. berberidifolia* showed similar patterns of asymmetric gene flow from *Q. berberidifolia* into *Q. engelmannii,* which may have been associated with changing climate (O'Donnell et al., 2021). The ability of this system to freely hybridize when living in close physical proximity can lead to the exchange of alleles across species boundaries while experiencing divergent selection regimes, increasing the potential for adaptive introgression (Abbott et al., 2016). This is an especially critical process to examine with the increased predicted genetic offset under future climate conditions in eastern populations of the already restricted distribution of *Q. engelmannii*. These populations are already exhibiting a stronger signal of genomic offset than *Q. berberidifolia* populations in similar regions of their

distributions and that is only predicted to increase, making these populations of a particular concern under most conservation genomic philosophies (Shaffer et al., 2022). It may be that introgression allows *Q. engelmannii* to persist in regions in which it is maladapted to future environmental conditions by exchanging alleles with a species that is more fit under most predicted future scenarios (Rieseberg & Burke, 2001; Rieseberg & Wendel, 1993).

In conclusion, this landscape genomic analysis reveals the vulnerability of Engelman oak to climate change, but the fact that *Q. berberidifolia* shows less vulnerability highlights a potential benefit of hybridization and introgression that has been observed between these two species. Previous research (O'Donnell et al., 2021) has demonstrated that interspecific gene flow is asymmetric between these two species favoring gene flow into *Q. engelmannii*. Asymmetric gene flow between these species could lead to an increase in fitness and potentially the persistence of *Q. engelmannii* in regions where it is predicted to be vulnerable under predicted future climate regimes. Future genomic research can test this hypothesis by determining whether potentially adaptive alleles such as climate-associated outlier loci are being exchanged (Savolainen et al., 2013). An important caveat of the current analysis is the inability to determine whether climate, geography, or a combination of both types of variables has the greatest predictive power in mapping patterns of genomic diversity. Additional analyses such as redundancy analyses (RDAs) could be used to quantify whether climate variables are better predictors of the distribution of genomic diversity than geospatial variables and to identify the strength of the interaction between them. Findings from this study demonstrate the opportunity and the climatic conditions that would make adaptive introgression possible and future study of the genes associated with introgression and their potential evolutionary impact will show whether adaptive introgression is occurring between these two oak species.

Literature Cited

Abbott, R. J., Barton, N. H., & Good, J. M. (2016). Genomics of hybridization and its evolutionary consequences. *Molecular Ecology*, *25*(11), 2325-2332. https://doi.org/10.1111/mec.13685

Anderson, E., & Stebbins, G. (1954). Hybridization as an evolutionary stimulus. *Evolution*, *8*(4), 378-388.

Arnold, M. L. (1997). *Natural hybridization and evolution*. Oxford University Press on Demand.

Barton, N. (2001). The role of hybridization in evolution. *Molecular Ecology*, *10*(3), 551-568.

Barton, N. H., & Hewitt, G. M. (1985). Analysis of hybrid zones. *Annual review of Ecology and Systematics*, *16*(1), 113-148.

Browne, L., Wright, J. W., Fitz-Gibbon, S., Gugger, P. F., & Sork, V. L. (2019). Adaptational lag to temperature in valley oak (Quercus lobata) can be mitigated by genome-informed assisted gene flow. *Proceedings of the National Academy of Sciences*, *116*(50), 25179-25185.

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

Capblancq, T., Fitzpatrick, M. C., Bay, R. A., Exposito-Alonso, M., & Keller, S. R. (2020). Genomic prediction of (mal) adaptation across current and future climatic landscapes. *Annual Review of Ecology, Evolution, and Systematics*, *51*(1).

Christe, C., Stölting, K. N., Paris, M., Fraïsse, C., Bierne, N., & Lexer, C. (2017). Adaptive evolution and segregating load contribute to the genomic landscape of divergence in two tree species connected by episodic gene flow. *Molecular ecology*, *26*(1), 59-76.

De La Torre, A. R., Wang, T., Jaquish, B., & Aitken, S. N. (2014). Adaptation and exogenous selection in a *Picea glauca* × *Picea engelmannii* hybrid zone: implications for forest management under climate change. *New Phytologist*, *201*(2), 687-699.

Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International journal of climatology*, *37*(12), 4302-4315.

Fitz-Gibbon, S., Hipp, A. L., Pham, K. K., Manos, P. S., & Sork, V. L. (2017). Phylogenomic inferences from reference-mapped and de novo assembled short-read sequence data using RADseq sequencing of California white oaks (*Quercus* section *Quercus*). *Genome*, *60*(9), 743-755.

Fitzpatrick, M. C., & Keller, S. R. (2015). Ecological genomics meets community-level modelling of biodiversity: Mapping the genomic landscape of current and future environmental adaptation. *Ecology Letters*, *18*(1), 1-16.

Frichot, E., & François, O. (2015). LEA: An R package for landscape and ecological association studies. *Methods in Ecology and Evolution*, *6*(8), 925-929.

Frichot, E., Schoville, S. D., Bouchard, G., & François, O. (2013). Testing for associations between loci and environmental gradients using latent factor mixed models. *Molecular Biology and Evolution*, *30*(7), 1687-1699.

Gugger, P. F., Fitz-Gibbon, S. T., Albarrán-Lara, A., Wright, J. W., & Sork, V. L. (2021). Landscape genomics of Quercus lobata reveals genes involved in local climate adaptation at multiple spatial scales. *Molecular Ecology*, *30*(2), 406-423.

Hamilton, J. A., Lexer, C., & Aitken, S. N. (2013). Genomic and phenotypic architecture of a spruce hybrid zone (*Picea sitchensis* x *P. glauca*). *Molecular Ecology*, *22*(3), 827-841. https://doi.org/10.1111/mec.12007

Holderegger, R., & Wagner, H. H. (2008). Landscape genetics. *Bioscience*, *58*(3), 199-207.

Holliday, J. A., Aitken, S. N., Cooke, J. E., Fady, B., González-Martínez, S. C., Heuertz, M., Jaramillo-Correa, J. P., Lexer, C., Staton, M., & Whetten, R. W. (2017). Advances in ecological genomics in forest trees and applications to genetic resources conservation and breeding. *Molecular ecology*, *26*(3), 706-717.

Howard, D. J., Preszler, R. W., Williams, J., Fenchel, S., & Boecklen, W. J. (1997). How discrete are oak species? Insights from a hybrid zone between *Quercus grisea* and *Quercus gambelii*. *Evolution*, *51*(3), 747-755.

Leroy, T., Louvet, J. M., Lalanne, C., Le Provost, G., Labadie, K., Aury, J. M., Delzon, S., Plomion, C., & Kremer, A. (2020). Adaptive introgression as a driver of local adaptation to climate in European white oaks. *New Phytologist*, *226*(4), 1171-1182.

Leroy, T., Rougemont, Q., Dupouey, J. L., Bodenes, C., Lalanne, C., Belser, C., Labadie, K., Le Provost, G., Aury, J. M., & Kremer, A. (2019). Massive postglacial gene flow between European white oaks uncovered genes underlying species barriers. *New Phytologist*, *226*(4), 1183-1197.

Li, J., Jew, B., Zhan, L., Hwang, S., Coppola, G., Freimer, N. B., & Sul, J. H. (2019). ForestQC: Quality control on genetic variants from next-generation sequencing data using random forest. *PLoS computational biology*, *15*(12), e1007556.

Manel, S., & Holderegger, R. (2013). Ten years of landscape genetics. *Trends in ecology & evolution*, *28*(10), 614-621.

Martins, K., Gugger, P. F., Llanderal-Mendoza, J., González-Rodríguez, A., Fitz-Gibbon, S. T., Zhao, J. L., Rodríguez-Correa, H., Oyama, K., & Sork, V. L. (2018). Landscape genomics provides evidence of climate-associated genetic variation in Mexican populations of Quercus rugosa. *Evolutionary Applications*, *11*(10), 1842-1858.

O'Donnell, S. T., Fitz-Gibbon, S. T., & Sork, V. L. (2021). Ancient Introgression Between Distantly Related White Oaks (Quercus sect. Quercus) Shows Evidence of Climate-Associated Asymmetric Gene Exchange. *Journal of Heredity*, *112*(7), 663-670.

Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'hara, R., Simpson, G. L., Solymos, P., Stevens, M. H. H., & Wagner, H. (2013). Package 'vegan'. *Community ecology package, version*, *2*(9), 1-295.

Ortego, J., Gugger, P. F., Riordan, E. C., & Sork, V. L. (2014). Influence of climatic niche suitability and geographical overlap on hybridization patterns among southern

Californian oaks [Article]. *Journal of Biogeography*, *41*(10), 1895-1908. https://doi.org/10.1111/jbi.12334

Ortego, J., Gugger, P. F., & Sork, V. L. (2015). Climatically stable landscapes predict patterns of genetic structure and admixture in the Californian canyon live oak. *Journal of Biogeography*, *42*(2), 328-338.

Ortego, J., Gugger, P. F., & Sork, V. L. (2017). Impacts of human-induced environmental disturbances on hybridization between two ecologically differentiated Californian oak species. *New Phytologist*, *213*(2), 942-955.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., Maller, J., Sklar, P., De Bakker, P. I., & Daly, M. J. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, *81*(3), 559-575.

Rellstab, C., Gugerli, F., Eckert, A. J., Hancock, A. M., & Holderegger, R. (2015). A practical guide to environmental association analysis in landscape genomics. *Molecular Ecology*, *24*(17), 4348-4370.

Rieseberg, L. H., & Burke, J. M. (2001). The biological reality of species: gene flow, selection, and collective evolution. *Taxon*, 47-67.

Rieseberg, L. H., & Wendel, J. F. (1993). Introgression and its consequences in plants. *Hybrid zones and the evolutionary process*, 70-109.

Riordan, E. C., Gillespie, T. W., Pitcher, L., Pincetl, S. S., Jenerette, G. D., & Pataki, D. E. (2015). Threats of future climate change and land use to vulnerable tree species native to Southern California. *Environmental Conservation*, *42*(2), 127-138.

Riordan, E. C., Gugger, P. F., Ortego, J., Smith, C., Gaddis, K., Thompson, P., & Sork, V. L. (2016). Association of genetic and phenotypic variability with geography and climate in three southern California oaks. *American Journal of Botany*, *103*(1), 73-85.

Savolainen, O., Lascoux, M., & Merilä, J. (2013). Ecological genomics of local adaptation. *Nature Reviews Genetics*, *14*(11), 807-820.

Savolainen, O., Pyhäjärvi, T., & Knürr, T. (2007). Gene flow and local adaptation in trees. *Annual Review of Ecology, Evolution, and Systematics*, 595-619.

Scott, T. A. (1991). The distribution of Engelmann oak (*Quercus engelmannii*) in California. In: Standiford, Richard B., tech. coord. 1991. Proceedings of the symposium on oak woodlands and hardwood rangeland management; October 31-November 2, 1990; Davis, California. Gen. Tech. Rep. PSW-GTR-126. Berkeley, CA: Pacific Southwest Research Station, Forest Service, US Department of Agriculture; p. 351-359,

Shaffer, H. B., Toffelmier, E., Corbett-Detig, R. B., Escalona, M., Erickson, B., Fiedler, P., Gold, M., Harrigan, R. J., Hodges, S., & Luckau, T. K. (2022). Landscape genomics to enable conservation actions: the California Conservation Genomics Project. *Journal of Heredity*.

Shryock, D. F., Washburn, L. K., DeFalco, L. A., & Esque, T. C. (2021). Harnessing landscape genomics to identify future climate resilient genotypes in a desert annual. *Molecular Ecology*, *30*(3), 698-717.

Sork, V., Aitken, S., Dyer, R., Eckert, A., Legendre, P., & Neale, D. (2013). Putting the landscape into the genomics of trees: approaches for understanding local adaptation and population responses to changing climate. *Tree Genetics & Genomes*, *9*(4), 901-911.

Sork, V. L., Gugger, P. F., Chen, J.-M., & Werth, S. (2016). Evolutionary lessons from California plant phylogeography. *Proceedings of the National Academy of Sciences*, *113*(29), 8064-8071.

Storfer, A., Murphy, M., Evans, J., Goldberg, C., Robinson, S., Spear, S., Dezzani, R., Delmelle, E., Vierling, L., & Waits, L. (2007). Putting the 'landscape'in landscape genetics. *Heredity*, *98*(3), 128-142.

Storfer, A., Patton, A., & Fraik, A. K. (2018). Navigating the interface between landscape genetics and landscape genomics. *Frontiers in Genetics*, *9*, 68.

Suarez-Gonzalez, A., Hefer, C. A., Lexer, C., Cronk, Q. C., & Douglas, C. J. (2017). *Introgression underlies adaptively significant variation and range boundaries in forest trees* University of British Columbia].

Team, R. C. (2022). *R: A language and environment for statistical computing*. In R Core Development Team, Vienna, Austria. https://www.R-project.org/.

Van Strien, M. J., Keller, D., & Holderegger, R. (2012). A new analytical approach to landscape genetic modelling: Least-cost transect analysis and linear mixed models. *Molecular Ecology*, *21*(16), 4010-4023.

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

Vanhove, M., Pina-Martins, F., Coelho, A. C., Branquinho, C., Costa, A., Batista, D., Príncipe, A., Sousa, P., Henriques, A., & Marques, I. (2021). Using gradient Forest to predict climate response and adaptation in Cork oak. *Journal of Evolutionary Biology*, *34*(6), 910-923.

Chapter 3

Genome-wide introgression analysis reveals asymmetric gene flow of stress response genes in two hybridizing white oaks (*Quercus* sect. *Quercus*)

Abstract

Hybridization and exchange of alleles through introgression is hypothesized to play an important role in local adaptation and the ability of plant species to rapidly adapt to changing environmental conditions. Adaptive introgression between hybridizing taxa has been proposed since early work on hybridization, but the availability of whole genome sequences now make it possible find evidence. This study tests the adaptive introgression hypothesis using two white oak (*Quercus* sect. *Quercus*) species; Engelmann oak (*Q. engelmannii*) and California shrub oak (*Q. berberidifolia*) that hybridize where their distributions overlap. By analyzing whole genome sequence data for 33 *Q. berberidifolia* and 31 *Q. engelmannii* individuals using the python package **sstar***,* evidence of significant introgression was found to be asymmetric with significant introgression discovered only in *Q. engelmannii.* To detect evidence of adaptive introgression, previously identified climate associated outliers from landscape genomic analyses of *Q. berberidifolia* were mapped to regions of increased introgression in *Q. engelmannii*. We identified 8 SNPs associated with climate in *Q. berberidifolia* that were also found in introgressed regions in *Q. engelmannii* individuals. Additionally, 31 characterized functional genes were found within regions of increased introgression in *Q. engelmannii,* including genes associated with stress response. These findings indicate the potential exchange of adaptive alleles of functional genes that may be directly associated with phenotypic expression and local

adaptation. Such results suggest that the introgressed genes may have enhanced the ability of *Q. engelmannii* to adapt to the Mediterranean climate of southern California.

Introduction

Hybridization leading to the exchange of alleles across species boundaries has been recognized as a potential mechanism for rapid local adaptation to changing environmental conditions or to novel environments as species expand their distribution into new habitats (Abbott et al., 2016; Anderson, 1949; Anderson & Stebbins, 1954; Barton & Hewitt, 1985). It is possible that processes such as hybridization provide an increase in overall genetic diversity by introducing novel alleles across species boundaries. If the alleles exchanged across species boundaries increase fitness in local populations, then they may spread throughout a species much like expectations for novel adaptive mutations yet potentially at much shorter time scales than mutations alone through a process otherwise known as adaptive introgression (Abbott et al., 2016; Barton, 2001; Barton & Hewitt, 1989; Wu, 2001). It is possible that such adaptive introgression is a mechanism through which species may rapidly evolve to changing climatic conditions through the exchange of alleles with species that may be more likely to be adapted to future climatic conditions (Abbott et al., 2016). Analyzing patterns of ancient introgression can provide evidence of this hypothesis by identifying signals of introgression throughout the genome that have persisted through long periods of evolutionary time and may be evidence of adaptive introgression (Barton, 2001). For example, studies of monkeyflowers (*Mimulus* spp.) have shown evidence for introgression of alleles associated with genes that control flowering time along a latitudinal gradient as species distributions shift under climate change (Vallejo-Marin & Hiscock, 2016). Analysis of introgression in poplar (*Populus* spp.) showed signals of the exchange of functional genes through introgression associated with cold hardiness and drought tolerance (Suarez-Gonzalez et al., 2017, 2018). Sunflowers have also shown the ability to exchange disease resistance traits through introgression in response to local pest infestations

leading to range expansion (Whitney et al., 2015). By utilizing this framework and testing for ancient adaptive introgression, it is possible to determine the influence of introgression on the long-term evolution of species involved in a hybrid system.

Oaks represent an ideal system to test for adaptive introgression due to their ability to freely hybridize with closely related species as well as their strong association between local ecological conditions and rates of hybridization (Van Valen, 1976). Evidence has indicated that while there may be intrinsic pre-zygotic barriers to gene flow between species that may affect the directionality and rate of gene flow between species (Lepais et al., 2013), several studies of hybrid and early-generation backcrossed individuals indicate local adaptation may be a driver of hybridization and introgression between species (Dodd & Afzal-Rafii, 2004; Lepais et al., 2009; Leroy et al., 2020). Additionally, patterns of ancient introgression also coincide with large-scale climate fluctuations in both California as well as in Europe, which also point to adaptation to rapidly changing climatic conditions strongly influencing the exchange of alleles across species boundaries (Leroy et al., 2017; O'Donnell et al., 2021). Previous work in European oaks has shown evidence of introgression of SNPs associated with genotype-environment and genotype-phenotype interactions that occurred adjacent to or within functional genes that may indicate signals of adaptive introgression (Leroy et al., 2020). However, patterns of glaciation in Europe resulted in extensive secondary contact leading to hybridization and introgression being a more recent evolutionary phenomenon in those taxa (Leroy et al., 2017) potentially making the identification of signals of adaptive introgression harder to identify (Shchur et al., 2020). Comparatively, the evolution of oaks in California in western North America has been predominantly influenced by the transition to a Mediterranean climate, which is predicted to have begun approximately 10-15mya (Rundel et al., 2016). Evidence in Californian white oaks

has shown that secondary contact between many of the species was ancient and signals of hybridization and introgression can be traced back to the transition to a Mediterranean climate (Kim et al., 2018; O'Donnell et al., 2021). By testing for introgression using species that underwent persistent secondary contact at much deeper time scales, it may be possible to identify introgression at functional genes associated with adaptation and long-term persistence under shifting climate conditions.

This study aims to test for adaptive introgression in two white oak (*Quercus* sect. *Quercus*) species; the more xeric adapted shrub species, *Q. berberidifolia,* and the more mesic tree species *Q. engelmannii*. These species differ in their broad habitat preferences yet live in close enough proximity that hybridization and introgression is possible, making this an ideal system to test for adaptive introgression (Ortego et al., 2014; Riordan et al., 2016). Their ability to exist in close geographic proximity, but within different ecological niches points to the influence of climatic variables on patterns of hybridization and introgression. It may be possible to identify genes associated with adaptation to climate by analyzing the exchange of alleles strongly associated with climatic gradients. It is possible that gene flow with the more drought adapted *Q. berberidifolia* may have helped *Q. engelmannii* expand into or persist within its current native range through adaptive introgression. Conversely, it is also possible that gene flow from *Q. engelmannii* into *Q. berberidifolia* helps aid in the range expansion of *Q. berberidifolia* into regions currently inhabited by *Q. engelmannii*. By identifying climate outliers and determining if they are associated with introgression, it is possible to test these hypotheses. Thus, this study attempts to answer three main questions: (i) Using whole genome sequence data, do these two species share genes through ancient introgression with each other? If so, is it symmetric or asymmetric? (ii) Is there evidence of specific regions of the genome that are

experiencing significant levels of gene flow between species? and (iii) Are functional genes or candidate loci associated with climate found within regions of increased introgression between these two species? The presence of climate-associated outlier alleles within regions of introgression would be strong evidence of adaptive introgression. If the direction of gene flow is asymmetric from the Californian endemic shrub oak into the Mexican-originating tree oak, it would suggest that the introgressed genes may have improved the adaptation of *Q. engelmannii* to southern California Mediterranean climate and provide evidence of the current adaptive capacity of the donor and recipient species (Barton, 2001). By characterizing the genes and alleles involved in gene flow across species boundaries it is possible to predict the importance of introgression on the evolution of both species under future climate conditions.

Methods

Study System

This two study species are the tree oak *Quercus. engelmannii* and the shrub oak *Q. berberidifolia. Q. engelmannii* has a restricted distribution to mesas with well-developed soils between 500-1500m in elevation south of the San Gabriel Mountains of southern California with a core distribution in the Southern Coast Ranges where local precipitation levels are higher than surrounding areas along fine-scale habitat gradients (Figure 2.A) (Scott, 1991). *Q. engelmannii* belongs to a sub-clade of white oaks that includes species found in the southwestern U.S. and northern Mexico (Hipp et al., 2020). It is the only species in its clade found in a Mediterranean-type climate with precipitation primarily occurring in the winter rather than Summer monsoonal precipitation regimes typically found in the American Southwest and northern Mexico. *Q. berberidifolia* is a widespread California endemic shrub oak with a coastal distribution (Fig 2.1).

It is typically found within mixed chaparral ecosystems in the more arid local environments where those species assemblages are present (Figure 2.1B). While these species are highly divergent ecologically, the geographic heterogeneity of southern California leads to frequent transitions between divergent ecosystems at the microhabitat scale. These species often exist at close physical proximity throughout the area where their distributions and provides the opportunity for hybridization (Ortego et al., 2014; Riordan et al., 2016), which may explain ancient introgression between these species (Kim et al., 2018). Based on predictions of current and future habitat loss under multiple future climate scenarios, *Q. engelmannii* is more likely to experience population decline than *Q. berberidifolia* under predicted changes to climatic conditions throughout its range in the future (see results of previous chapter) (Riordan et al., 2015). The combination of prior opportunity to hybridize and the divergence between the ecological preferences of these species increases the opportunity for adaptive introgression.

Sampling, Sequencing and Variant Calling

We selected 32 *Q. berberidifolia* and 34 *Q. engelmannii* whole genome sequences (64 total), which were sampled from the extent of range overlap between the two species in southern California from the overall range-wide sampling described in Chapter 2. Sequences were mapped to the *Quercus lobata* reference genome (v 3.1) (Sork et al., 2021) using previously described methods for oaks (Fitz-Gibbon et al., 2017). Variants were called using the default parameters present in the package ForestQC (Li et al., 2019). Individuals were chosen that exceeded 7x average mapped coverage based on previous analysis (Chapter 2). To ensure that signals of introgression were preserved, the steps involving filtering for linkage disequilibrium described in the previous chapter were skipped for subsequent downstream analysis.

ADMIXTURE Analysis and Detecting Introgression

In order to accurately test for signals of ancient adaptive introgression, early generation backcrossed individuals are excluded to increase the likelihood signals of introgression of potentially adaptive loci or within genes is due to adaptation and not the exchange of neutral genetic variation (Abbott et al., 2016; Barton, 2001). ADMIXTURE analysis was performed using 50,000 burn-in steps and 10,000 replications testing across a likely number of genetic clusters (K) of 1-5 to determine the most likely number of genetic clusters. An early-backcrossed individual was considered any individual that did not have a proportion of ancestry of at least 85% belonging to one of the two genetic clusters associated with either species based on previous methods used for oaks (Lepais et al. 2013). Individuals with a proportion of ancestry less than 85% were removed from downstream analysis.

Areas of the genome experiencing significant levels of introgression were identified using a machine learning approach developed in the python package **sstar** (Huang et al., 2022). This introgression analysis method is based on the summary statistic *S**, which was designed to detect sequence similarity between species that is greater than expected due to chance alone through the identification of patterns of similar mutations predicted to have accumulated in an archaic population, otherwise known as congruent sites (Plagnol & Wall, 2006; Wall, 2000). This analysis is applied in the package **sstar** by defining the hypothesized demographic model, the test or archaic population, and the target population in which you are attempting to detect introgression. In this framework, **sstar** was used to test for introgression between the two species and utilizing each species as a test and target population in replicate analyses. Previous analyses have shown that *Q. berberidifolia* has higher genetic diversity and a greater effective population

size, so the neutral demographic model was defined with a time to most recent common ancestor of 45 million years, a starting effective population size of 1,000,000 and 100,000 individuals for *Q. berberidifolia* and *Q. engelmannii* respectively (O'Donnell et al., 2021), a mutation rate of 1.2 x 10^{-8} (Sork et al., 2022), a recombination rate of 1.01 x 10^{-8} based on recombination rates in eucalyptus (Gion et al., 2016), and no gene flow since the time of divergence between the two lineages. ***S**** analysis was performed using sliding windows of 50kbp and steps of 1kbp. To identify the threshold for significance of *S** scores, the bootstrapping method in the "quantiles" command in **sstar** was used. This command uses the program *ms* (Hudson, 2002) to simulate sequence data to establish significance thresholds through large-scale replication of data. This test was done for 20,000 replicates of 50kb sequencing windows using the above values for effective population size, mutation, and recombination rates. Simulated *S** scores were used to generate the 99th percentile values for observed *S** scores and any window exceeding that threshold was considered an introgressed window.

Detecting Evidence of Adaptive Introgression

The expectation is that most introgression will likely be selectively neutral, but exchanging alleles found within or near functional genes have the potential of changing phenotypic expression associated with those genes and altering the fitness of hybridizing species (Barton & Hewitt, 1985). Based on this framework, windows that showed evidence of introgression were mapped to the latest annotation of the *Quercus lobata* reference genome (v 3.1) and visualized using the R package chromomap (Anand & Rodriguez Lopez, 2022) to determine the physical location of those windows within the genome. Introgressed windows were then scanned for annotated functional genes that may be associated with functional

responses, which may provide a source of adaptive introgression. Additionally, to determine if there are any signals of climate-associated adaptive introgression, 720 putative candidate loci of large effect size significantly associated with climate gradients in the southern portion of the distribution of *Q. berberidifolia* identified in chapter 2 (see LFMM methods) were mapped to the genome and analyzed for overlap with introgressed windows. Any climate associated outlier SNPs found within introgressed windows are potentially a source of adaptive introgression involved in increasing local adaptation to climate variables and were further annotated to determine if they fall within functional genes within those windows. To test whether climate associated outlier SNPs fell within introgressed windows due to chance alone, a permutation test was conducted using the R package regioneR (Gel et al., 2016).

Results

ADMIXTURE

ADMIXTURE results indicated a most likely K=2 with genetic clusters separating along species boundaries and strong evidence of asymmetric gene flow between species (Figure 3.1). Of the 32 *Quercus berberidifolia* and 34 *Q. engelmannii* individuals, two individuals did not meet proportion of ancestry thresholds and were excluded from further analysis resulting in a total of 31 *Q. berberidifolia* and 33 *Q. engelmannii* individuals (64 total individuals) used in *S** analyses. Of the remaining individuals, several *Q. engelmannii* individuals showed evidence of mixed ancestry associated with later-generation backcross events, but little evidence of mixed ancestry in the 31 remaining *Q. berberidifolia* individuals, potentially indicating asymmetric gene flow between these species (Figure 3.1).

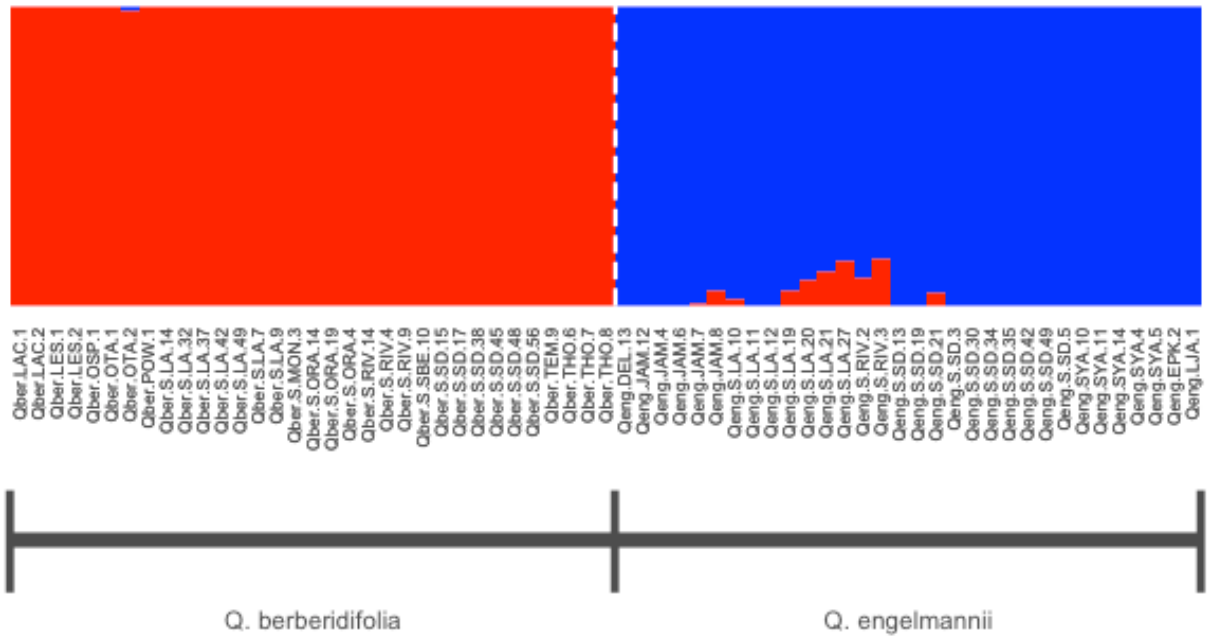

Figure 3.1: ADMIXTURE results plot depicting predicted proportion of ancestry under the most likely scenario of K=2. Quercus berberidifolia individuals are on the left and associated with the red genetic cluster while Q. engelmannii individuals are on the right and associated with the blue genetic cluster

Identification of Introgressed Regions

Introgressed sequences were distributed throughout the genomes, but they were strongly asymmetric with windows crossing the significance threshold for *S** scores only being found being exchanged from the drought-adapted shrub *Quercus berberidifolia* into the tree oak *Q. engelmannii*. Reciprocal tests using each species as the test or archaic population and target population show no evidence of significant introgression from *Q. engelmannii* into *Q. berberidifolia*. In contrast, we found strong evidence of *Q. berberidifolia* introgression into *Q. engelmannii* found on 5 out of 12 total chromosomes. Out of all overlapping windows tested, 125 potentially overlapping windows had *S** scores that crossed the significance threshold resulting in 29 unique regions of the genome that showed significant levels of introgression

(Figure 3.2). The greatest number of windows were found on chromosome 4 while chromosomes 2 and 8 had comparatively few significant windows, which indicates introgression is likely associated with specific regions of the genome of *Q. engelmannii*. Conversely, the lack of significant windows of introgression from any *Q. berberidifolia* individual indicates that even background rates of hybridization and introgression are low from *Q. engelmannii* into *Q. berberidifolia*, which indicates a lack of adaptive introgression occurring into *Q. berberidifolia.*

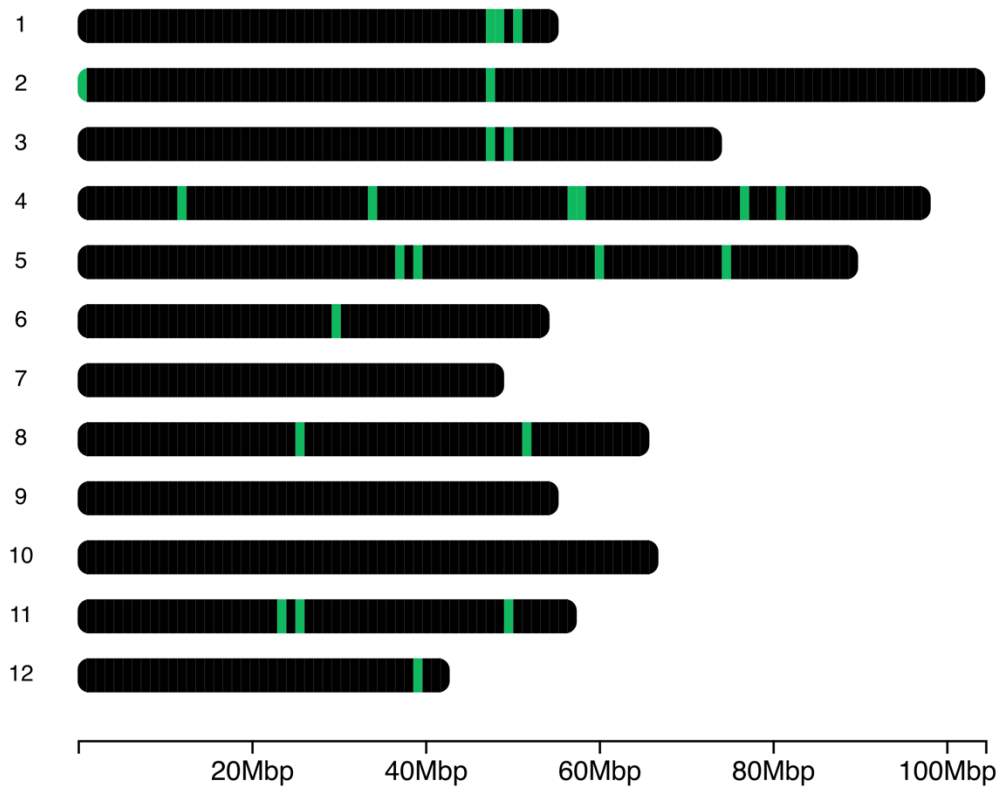

Figure 3.2: Map of the 12 oak chromosomes including genomic candidate introgressed regions in *Q. engelmannii*, based on significant S* scores (green bars).

Functional Genes Found in Introgressed Windows

Across the 125 introgressed windows and 24 unique regions of overlapping windows found in *Quercus engelmannii*, a total of 32 functional genes with detailed annotations were discovered, including several genes known to be associated with stress responses in plants (Table 3.1). Additionally, numerous unclassified genes were also discovered in these regions which lacked annotations, so it is not possible to assess whether they are adaptive. Functional genes were found within all identified introgressed regions except for one 50kbp window found on chromosome 6, which did not contain a functional gene. Two genes from gene families associated with stress response were found within introgressed windows: a zinc finger family

protein and a hypersensitive-induced response protein. Proteins from the zinc finger family are typically associated with binding of transcription factors and have been implicated in repression of transcription and are involved in defense and acclimation response in plants (Ciftci-Yilmaz & Mittler, 2008; Kiełbowicz-Matuk, 2012; Liu et al., 2022). Hypersensitive-induced response genes are associated with immune response in *Arabidopsis thaliana* in both species (Choi et al., 2011; Qi et al., 2011). Introgression of these functional stress response genes in addition to other functional genes found within introgressed windows implies that adaptive introgression may be associated with these windows.

Table 3.1: Identified annotated genes and their chromosome locations. Genes of known association with biotic or abiotic stress response highlighted in bold.

Chrom.	Gene
1	Transcription factor RF2a
1	ER membrane protein complex sub-unit
1	Putative RNA binding protein Luc 7-like
1	Methylene blue sensitivity 1
1	**Zinc finger CCH Domain-containing protein ZFN-like**
1	IN080 complex subunit D-like
1	Fatty alcohol caffeyol-COA acetyltransferase
2	LRR receptor-like serine/theonine-protein kinase
2	High mobility protein group B protein 7-like
2	Helicase-like transcription factor CHR28
2	Phytochrom-associated serine threonine-protein phosphate
2	Protease 2-like
2	**hypersensitive-induced response protein 1-like**
2	**AP2-like ethylene-responsive transcription factor BBM1**
3	Glutathione S-transferase T3-like
3	RAB11-binding protein RELCH homolog
3	receptor protein-kinase-like protein ZAR1
4	Putative disease-resistant protein At3g14460
4	Myosin heavy chain, striated muscle-like
4	F-box/Kelch-repeat protein At3g61590
4	Anthocyanidin 3-0-glucosyltransferase 7-like
5	**WRKY transcription factor 28-like**
5	Tryptophan synthase beta chain 1-like
5	COBRA-like protein 10
8	ras-related protein Rab11 D-like
8	Histone deactylase complex subunit SAP18
8	Polygalacturonate 4-alpha galactronosyltransferase
8	Gibberellin 20 oxidase-like
11	**MADS-box trasncription factor 23**
12	XIAP-associated factor 1
12	Pentatricopeptide repeat-containing protein At1g0990

Introgression of climate associated outlier loci

When we mapped climate-associated outlier SNPs identified through a landscape genomic study (Chapter 2) to introgressed windows in *Q. engelmannii*, several SNPs were found within or near genes, providing evidence of potential adaptive introgression. Specifically, of the 720 outlier loci identified in southern populations of *Q. berberidifolia*, 8 SNPs were found to be near or within introgressed windows. Permutation tests indicated that the expected number of SNPs expected to be randomly found within introgressed windows was ~2, indicating that the number of outlier SNPs was significantly greater than expected. Of the 8 climate outlier SNPs, three of them were found in proximity to, or within, annotated functional genes found within introgressed windows and may be associated with adaptive introgression.

We found several SNPs located close to functional genes. SNPs were found within 10kb of both WRKY transcription factor 53 as well as a beta-glucosidase 46-like. WRKY transcription factors are associated with plant immune response and may represent adaptative genetic variation through their regulation of multiple pathways of plant immune response in addition to their response to drought or cold stress making them critical in plant responses to biotic and abiotic stressors (Chen et al., 2017; Rushton et al., 2010). Beta-glucosidase is not involved in any specific plant responses to environmental stress. However, it has been shown to be involved in signaling processes for the natural insect predators of herbivorous insects during periods of intense predation by herbivores in other plant taxa (Mattiacci et al., 1995), but it has not been studied in oaks. One climate associated outlier SNP was found within a MADS-box transcription factor. The MADS-box gene family is highly conserved within plants and is directly associated with root, shoot and floral development in flowering plants (Becker & Theißen, 2003; Ng & Yanofsky, 2001). The association of climate-associated, putatively adaptive SNPs found through

a separate study and found here within introgressed windows and within or near functional genes provide evidence of adaptive introgression of traits associated with adaptation to climate.

Discussion

Our results point to evidence of ancient asymmetric adaptive introgression of climate-associated outlier alleles as well as introgression within proximity of other functional genes associated with stress and pathogen response in plant taxa. These results are direct evidence of potential adaptive introgression associated with functional genes and show a distinct correlation with adaptation to abiotic factors such as response to changing climate regimes. Additionally, introgression was heavily asymmetric with no evidence of persistent introgression from *Quercus engelmannii* into *Q. berberidifolia*. The timing of secondary contact leading to introgression between these species is predicted to have begun during the transition to a Mediterranean climate in southern California (O'Donnell et al., 2021). This ancient asymmetric adaptive introgression may have allowed the adaptation and persistence of *Q. engelmannii* under climatic conditions that are unique to coastal California and Baja California in North America.

Evidence of asymmetric gene flow and introgression supports the hypothesis of increased introgression from the wide-spread, more drought adapted *Quercus berberidifolia* that may have enhanced the successful expansion of *Q. engelmannii* into southern California. The closest relatives of *Q. engelmannii* are currently found in Arizona and New Mexico in the southwestern U.S. and northern Mexico (Hipp et al., 2020). It is likely that the shift to a Mediterranean-type climate resulted in significantly different seasonal precipitation patterns than those experienced by its closest relatives and resulted in strong selective pressure on *Q. engelmannii* as has been shown in other plant genera in the California Floristic Province (Rundel et al., 2016). Evidence

of adaptive introgression of genes involved in stress response from this study suggests that gene flow between species likely played a key role in the adaptation of *Q. engelmannii* to the Mediterranean-type climate. These findings are consistent with previous research that showed asymmetric ancient introgression that favored gene flow from *Q. berberidifolia* rather than vice-versa (Kim et al., 2018; O'Donnell et al., 2021). Similar patterns of asymmetric adaptive introgression have been found in other tree species including poplar (*Populus* spp.) (Suarez-Gonzalez et al., 2017) and European white oaks (Leroy et al., 2020). Specifically, evidence in both systems implicated more recent adaptive introgression in the range expansion of taxa into novel habitats based on climatic gradients and local environmental conditions. Current evidence in this study supports a similar pattern of asymmetric gene flow and adaptive introgression in this system during the more distant evolutionary past that also may have assisted the colonization and persistence of a species to a novel environment in *Q. engelmannii*.

Models suggest that *Q. engelmannii* is likely to become increasingly maladapted under future climate scenarios, further reinforcing the potential of adaptive introgression increasing the fitness of this species (Chapter 2). Evidence of contemporary introgression between these species may also support the adaptive introgression hypothesis. Transcriptome analysis of alleles associated with drought response genes showed also supports asymmetric gene flow from *Q. berberidifolia* into *Q. engelmannii* (Oney-Birol et al., 2018). Previous landscape genetics work in this system has shown that contemporary introgression is strongly associated abiotic factors such as temperature and precipitation seasonality, which also supports the adaptive introgression hypothesis in this system (Ortego et al., 2014; Riordan et al., 2016). In many cases, critical abiotic factors, such as temperature and precipitation regimes, are predicted to exceed current tolerance thresholds of many species under predicted future climate scenarios making rapid

evolution in response to these changing conditions critical for the survival of many taxa globally (Franks & Hoffmann, 2012; Shaw & Etterson, 2012). However, the rates of change of local climate conditions are likely to exceed the evolutionary rates of many species, especially long-lived species such as trees (Browne et al., 2019; Lexer et al., 2007). Evidence of ancient introgression showed introgressed windows contained functional genes associated with drought, temperature, and disease resistance and response, which will likely be critical functional processes as southern California is predicted to suffer increasing temperatures and duration and severity of droughts under future climate scenarios (Underwood et al., 2019). The evidence showing ancient adaptive introgression may indicate that persistent hybridization and introgression between these species may result in further exchange of adaptive alleles because of changing climate.

Evidence of ancient adaptive introgression of functional genes involved in stress response indicates that these genes were likely crucial to the persistence of *Q. engelmannii* to a Mediterranean-type climate in the past and may be significantly involved with future adaptation. This study was designed to test whether climate -associated outlier SNPs were found near functional genes. However, it is unknown whether these outlier SNPs are non-randomly distributed throughout the genome of these species and current results may reflect this biased distribution. Future work to annotate all identified climate outlier SNPs can reveal the extent of this bias and whether significant results found using permutation tests that did not account for non-randomly distributed SNPs are supported. Since most plant responses to environmental stress are the result of highly complex interactions between genes and gene networks, it is likely that larger networks of genes than the genes found in this study are involved (Becker & Theißen, 2003; Choi et al., 2011). This study focused on identifying signals of adaptive introgression of

loci of large effective size, meaning the genes containing or adjacent to climate-associated outliers in *Q. berberidifolia* likely played a large role in adaptation to past climatic regimes. It may be possible to detect additional genes associated with adaptive introgression when reducing the threshold of filtering procedures to include more loci of small effective size. Additionally, the functional role that many genes involved plant responses to environmental stress is uncharacterized and it may be the case that genes play multiple functional roles in separate plant response pathways (Chaves et al., 2003). It may be the case in this system that more of the 32 identified functional genes are associated with plant response pathways and may also be involved in adaptive introgression.

In conclusion, this study finds compelling evidence of ancient asymmetric adaptive introgression for genes involved in phenotypic response to climate stress from the endemic species *Quercus berberidifolia* into the non-endemic species *Quercus engelmannii*. It is possible that adaptive introgression enhanced the ability of *Q. engelmannii* to expand its range into California and persist under a transition to a Mediterranean-type climate. Previous research has shown evidence of asymmetric gene flow between these species from *Q. berberidifolia* into *Q. engelmannii*, but these findings represent the first evidence connecting this gene flow to functional genes and potentially to adaptive introgression of phenotypic traits. Evidence from this analysis supports the hypothesis of adaptive introgression. Future research into the fitness consequences of introgressed individuals will determine importance the importance of contemporary introgression in adaptation to future climatic regimes.

Literature Cited

Abbott, R. J., Barton, N. H., & Good, J. M. (2016). Genomics of hybridization and its evolutionary consequences. *Molecular Ecology*, *25*(11), 2325-2332. https://doi.org/10.1111/mec.13685

Anand, L., & Rodriguez Lopez, C. M. (2022). ChromoMap: an R package for interactive visualization of multi-omics data and annotation of chromosomes. *BMC bioinformatics*, *23*(1), 33.

Anderson, E. (1949). Introgressive hybridization. *Chapman and Hall, London*.

Anderson, E., & Stebbins, G. (1954). Hybridization as an evolutionary stimulus. *Evolution*, *8*(4), 378-388.

Barton, N. (2001). The role of hybridization in evolution. *Molecular Ecology*, *10*(3), 551-568.

Barton, N., & Hewitt, G. M. (1989). Adaptation, speciation and hybrid zones. *Nature*, *341*(6242), 497-503.

Barton, N. H., & Hewitt, G. M. (1985). Analysis of hybrid zones. *Annual review of Ecology and Systematics*, *16*(1), 113-148.

Becker, A., & Theißen, G. (2003). The major clades of MADS-box genes and their role in the development and evolution of flowering plants. *Molecular Phylogenetics and Evolution*, *29*(3), 464-489.

Browne, L., Wright, J. W., Fitz-Gibbon, S., Gugger, P. F., & Sork, V. L. (2019). Adaptational lag to temperature in valley oak (Quercus lobata) can be mitigated by genome-informed

assisted gene flow. *Proceedings of the National Academy of Sciences*, *116*(50), 25179-25185.

Chaves, M. M., Maroco, J. P., & Pereira, J. S. (2003). Understanding plant responses to drought—from genes to the whole plant. *Functional plant biology*, *30*(3), 239-264.

Chen, F., Hu, Y., Vannozzi, A., Wu, K., Cai, H., Qin, Y., Mullis, A., Lin, Z., & Zhang, L. (2017). The WRKY transcription factor family in model plants and crops. *Critical reviews in plant sciences*, *36*(5-6), 311-335.

Choi, H. W., Kim, Y. J., & Hwang, B. K. (2011). The hypersensitive induced reaction and leucine-rich repeat proteins regulate plant cell death associated with disease and plant immunity. *Molecular plant-microbe interactions*, *24*(1), 68-78.

Ciftci-Yilmaz, S., & Mittler, R. (2008). The zinc finger network of plants. *Cellular and Molecular Life Sciences*, *65*, 1150-1160.

Dodd, R. S., & Afzal-Rafii, Z. (2004). Selection and dispersal in a multispecies oak hybrid zone. *Evolution*, *58*(2), 261-269.

Fitz-Gibbon, S., Hipp, A. L., Pham, K. K., Manos, P. S., & Sork, V. L. (2017). Phylogenomic inferences from reference-mapped and de novo assembled short-read sequence data using RADseq sequencing of California white oaks (*Quercus* section *Quercus*). *Genome*, *60*(9), 743-755.

Franks, S. J., & Hoffmann, A. A. (2012). Genetics of climate change adaptation. *Annual review of genetics*, *46*, 185-208.

Gel, B., Díez-Villanueva, A., Serra, E., Buschbeck, M., Peinado, M. A., & Malinverni, R. (2016). regioneR: an R/Bioconductor package for the association analysis of genomic regions based on permutation tests. *Bioinformatics*, *32*(2), 289-291.

Gion, J.-M., Hudson, C. J., Lesur, I., Vaillancourt, R. E., Potts, B. M., & Freeman, J. S. (2016). Genome-wide variation in recombination rate in Eucalyptus. *BMC genomics*, *17*(1), 1-12.

Hipp, A. L., Manos, P. S., Hahn, M., Avishai, M., Bodénès, C., Cavender-Bares, J., Crowl, A. A., Deng, M., Denk, T., & Fitz-Gibbon, S. (2020). Genomic landscape of the global oak phylogeny. *New Phytologist*, *226*(4), 1198-1212.

Huang, X., Kruisz, P., & Kuhlwilm, M. (2022). sstar: A Python package for detecting archaic introgression from population genetic data with S. *Molecular Biology and Evolution*, *39*(11), msac212.

Hudson, R. R. (2002). Generating samples under a Wright–Fisher neutral model of genetic variation. *Bioinformatics*, *18*(2), 337-338.

Kiełbowicz-Matuk, A. (2012). Involvement of plant C2H2-type zinc finger transcription factors in stress responses. *Plant Science*, *185*, 78-85.

Kim, B. Y., Wei, X., Fitz-Gibbon, S., Lohmueller, K. E., Ortego, J., Gugger, P. F., & Sork, V. L. (2018). RADseq data reveal ancient, but not pervasive, introgression between Californian tree and scrub oak species (*Quercus* sect. *Quercus*: *Fagaceae*). *Molecular Ecology*, *27*(22), 4556-4571.

Lepais, O., Petit, R. J., Guichoux, E., Lavabre, J. E., Alberto, F., Kremer, A., & Gerber, S. (2009). Species relative abundance and direction of introgression in oaks. *Molecular Ecology*, *18*(10), 2228-2242. https://doi.org/10.1111/j.1365-294X.2009.04137.x

Lepais, O., Roussel, G., Hubert, F., Kremer, A., & Gerber, S. (2013). Strength and variability of postmating reproductive isolating barriers between four European white oak species [Article]. *Tree Genetics & Genomes*, *9*(3), 841-853. https://doi.org/10.1007/s11295-013-0602-3

Leroy, T., Louvet, J. M., Lalanne, C., Le Provost, G., Labadie, K., Aury, J. M., Delzon, S., Plomion, C., & Kremer, A. (2020). Adaptive introgression as a driver of local adaptation to climate in European white oaks. *New Phytologist*, *226*(4), 1171-1182.

Leroy, T., Roux, C., Villate, L., Bodenes, C., Romiguier, J., Paiva, J. A. P., Dossat, C., Aury, J. M., Plomion, C., & Kremer, A. (2017). Extensive recent secondary contacts between four European white oak species. *New Phytologist*, *214*(2), 865-878. https://doi.org/10.1111/nph.14413

Lexer, C., Van Loo, M., & Barbara, T. (2007). Towards forest community and ecosystem genomics. *New Phytologist*, *173*(4), 673-676. <Go to ISI>://WOS:000244060200005

Li, J., Jew, B., Zhan, L., Hwang, S., Coppola, G., Freimer, N. B., & Sul, J. H. (2019). ForestQC: Quality control on genetic variants from next-generation sequencing data using random forest. *PLoS computational biology*, *15*(12), e1007556.

Liu, Y., Khan, A. R., & Gan, Y. (2022). C2H2 zinc finger proteins response to abiotic stress in plants. *International Journal of Molecular Sciences*, *23*(5), 2730.

Mattiacci, L., Dicke, M., & Posthumus, M. A. (1995). beta-Glucosidase: an elicitor of herbivore-induced plant odor that attracts host-searching parasitic wasps. *Proceedings of the National Academy of Sciences*, *92*(6), 2036-2040.

Ng, M., & Yanofsky, M. F. (2001). Function and evolution of the plant MADS-box gene family. *Nature Reviews Genetics*, *2*(3), 186-195.

O'Donnell, S. T., Fitz-Gibbon, S. T., & Sork, V. L. (2021). Ancient Introgression Between Distantly Related White Oaks (Quercus sect. Quercus) Shows Evidence of Climate-Associated Asymmetric Gene Exchange. *Journal of Heredity*, *112*(7), 663-670.

Oney-Birol, S., Fitz-Gibbon, S., Chen, J.-M., Gugger, P. F., & Sork, V. L. (2018). Assessment of shared alleles in drought-associated candidate genes among southern California white oak species (Quercus sect. Quercus). *BMC genetics*, *19*(1), 88.

Ortego, J., Gugger, P. F., Riordan, E. C., & Sork, V. L. (2014). Influence of climatic niche suitability and geographical overlap on hybridization patterns among southern Californian oaks. *Journal of biogeography*, *41*(10), 1895-1908.

Plagnol, V., & Wall, J. D. (2006). Possible ancestral structure in human populations. *PLoS genetics*, *2*(7), e105.

Qi, Y., Tsuda, K., Nguyen, L. V., Wang, X., Lin, J., Murphy, A. S., Glazebrook, J., Thordal-Christensen, H., & Katagiri, F. (2011). Physical association of Arabidopsis hypersensitive induced reaction proteins (HIRs) with the immune receptor RPS2. *Journal of Biological Chemistry*, *286*(36), 31297-31307.

Riordan, E. C., Gillespie, T. W., Pitcher, L., Pincetl, S. S., Jenerette, G. D., & Pataki, D. E. (2015). Threats of future climate change and land use to vulnerable tree species native to Southern California. *Environmental Conservation*, *42*(2), 127-138.

Riordan, E. C., Gugger, P. F., Ortego, J., Smith, C., Gaddis, K., Thompson, P., & Sork, V. L. (2016). Association of genetic and phenotypic variability with geography and climate in three southern California oaks. *American Journal of Botany*, *103*(1), 73-85.

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

Rushton, P. J., Somssich, I. E., Ringler, P., & Shen, Q. J. (2010). WRKY transcription factors. *Trends in Plant Science*, *15*(5), 247-258.

Scott, T. A. (1991). The distribution of Engelmann oak (*Quercus engelmannii*) in California. In: Standiford, Richard B., tech. coord. 1991. Proceedings of the symposium on oak woodlands and hardwood rangeland management; October 31-November 2, 1990; Davis, California. Gen. Tech. Rep. PSW-GTR-126. Berkeley, CA: Pacific Southwest Research Station, Forest Service, US Department of Agriculture; p. 351-359,

Shaw, R. G., & Etterson, J. R. (2012). Rapid climate change and the rate of adaptation: insight from experimental quantitative genetics. *New Phytologist*, *195*(4), 752-765.

Shchur, V., Svedberg, J., Medina, P., Corbett-Detig, R., & Nielsen, R. (2020). On the distribution of tract lengths during adaptive introgression. *G3: Genes, Genomes, Genetics*, *10*(10), 3663-3673.

Sork, V. L., Cokus, S. J., Fitz-Gibbon, S. T., Zimin, A. V., Puiu, D., Garcia, J. A., Gugger, P. F., Henriquez, C. L., Zhen, Y., & Lohmueller, K. E. (2021). High-quality genome and methylomes illustrate features underlying evolutionary success of oaks. *bioRxiv*.

Sork, V. L., Cokus, S. J., Fitz-Gibbon, S. T., Zimin, A. V., Puiu, D., Garcia, J. A., Gugger, P. F., Henriquez, C. L., Zhen, Y., & Lohmueller, K. E. (2022). High-quality genome and methylomes illustrate features underlying evolutionary success of oaks. *Nature communications*, *13*(1), 2047.

Suarez-Gonzalez, A., Hefer, C. A., Lexer, C., Cronk, Q. C., & Douglas, C. J. (2017). *Introgression underlies adaptively significant variation and range boundaries in forest trees* University of British Columbia].

Suarez-Gonzalez, A., Hefer, C. A., Lexer, C., Cronk, Q. C., & Douglas, C. J. (2018). *Scale and direction of adaptive introgression between black cottonwood (Populus trichocarpa) and balsam poplar (P. balsamifera)* University of British Columbia].

Underwood, E. C., Hollander, A. D., Safford, H. D., Kim, J. B., Srivastava, L., & Drapek, R. J. (2019). The impacts of climate change on ecosystem services in southern California. *Ecosystem Services*, *39*, 101008.

Vallejo-Marin, M., & Hiscock, S. J. (2016). Hybridization and hybrid speciation under global change. *New Phytologist*, *211*(4), 1170-1187. https://doi.org/10.1111/nph.14004

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

Wall, J. D. (2000). Detecting ancient admixture in humans using sequence polymorphism data. *Genetics*, *154*(3), 1271-1279.

Whitney, K. D., Broman, K. W., Kane, N. C., Hovick, S. M., Randell, R. A., & Rieseberg, L. H. (2015). Quantitative trait locus mapping identifies candidate alleles involved in adaptive introgression and range expansion in a wild sunflower. *Molecular Ecology*, *24*(9), 2194-2211. https://doi.org/10.1111/mec.13044

Wu, C. I. (2001). The genic view of the process of speciation. *Journal of Evolutionary Biology*, *14*(6), 851-865.

Appendix 1.1

Model Descriptions

Models were tested in a nested fashion in which the simplest model is a specialized version of more complex models. Models were compared in tiers according to the number of free parameters. The most likely model from each scenario was then used to develop more complex models. For example, when comparing models that have two free parameters, the best performing model was one that included symmetric gene flow after secondary contact between the study species. Gene flow between species was then considered in each subsequent tier of models (either symmetric or asymmetric gene flow depending on the number of free parameters in each model). Below are descriptions of each model and the free parameters used to describe them grouped into tiers based on the number of free parameters. Abbreviations in parentheses of each model description correspond to their label in table S1.

3 Parameter Models

Species Divergence with No Gene Flow (Iso)- Species diverged at time T_{split} and remained at the same effective population size (v_1 and v_2 for *Quercus engelmannii* and *Quercus berberidifolia* respectively) as the ancestral populations and did not experience intraspecific gene flow since the time of divergence. (Parameters: v_1, v_2, T_{split})

4 Parameter Models

Population expansion in Quercus engelmannii *with no gene flow* (P1Exp)- Species diverged at time T_{split} with ancestral effective population sizes of v_1 and v_2 (for *Quercus engelmannii* and *Quercus berberidifolia* respectively) and did not come into secondary contact and experience secondary gene flow. At time T_{split}, *Quercus engelmannii* experienced an increase in effective population size at rate (s) while the effective population size of *Quercus berberidifolia* did not change in size.

(Parameters: v_1, v_2, s, T_{split})

Population expansion in Quercus berberidifolia *with no gene flow* (P2Exp)- Species diverged at time T_{split} with ancestral effective population sizes of v_1 and v_2 (for *Quercus engelmannii* and *Quercus berberidifolia* respectively) and did not come into secondary contact and experience secondary gene flow. At time T_{split}, *Quercus berberidifolia* experienced a continuous increase in effective population size at rate (s) while the effective population size of *Quercus engelmannii* did not change in size.

(Parameters: v_1, v_2, s, T_{split})

Species divergence with continuous symmetric gene flow (Smig)- Species diverged at time T_{split} with ancestral effective population sizes of v_1 and v_2 and experienced intermittent, but continuous symmetric gene flow at rate (m) beginning at time T_{split} and continuing until present day. (Parameters: v_1, v_2, m, T_{split})

Symmetric population expansion after divergence (SExp)- Species diverged at time T_{split} with ancestral effective population sizes of v_1 and v_2 (for *Quercus engelmannii* and *Quercus berberidifolia* respectively) and did not come into secondary contact and experience secondary

gene flow. Beginning at time T_{split} both species experienced effective population growth at rate (s). (Parameters: v_1, v_2, s, T_{split})

5 Parameter Models

Continuous asymmetric gene flow (Amig)- Species diverged at time T_{split} and experienced intermittent, but continuous asymmetric gene flow at rates m_{12} (gene flow **from** *Q. engelmannii* **into** *Q. berberidifolia*) and m_{21} (gene flow **from** *Q. berberidifolia* **into** *Q. engelmannii*) beginning at time T_{split} and continuing until present day. (Parameters: $v_1, v_2, m_{12}, m_{21}, T_{split}$)

Symmetric population expansion with symmetric gene flow (SexpSM)- Species diverged at time T_{split} with ancestral effective population sizes of v_1 and v_2 and experienced intermittent, but continuous symmetric gene flow at rate (m) beginning at time T_{split} and continuing until present day. Concurrent with this symmetric gene flow, both species also experienced symmetric growth in effective population at rate (s) that persisted from divergence into contemporary populations. (Parameters: $v_1, v_2, s, m, T_{split}$)

Asymmetric effective population size growth in Quercus engelmannii *with continuous symmetric gene flow* (P1ExpSM)- Species diverged at time T_{split} with ancestral effective population sizes of v_1 and v_2 and experienced intermittent, but continuous symmetric gene flow at rate (m) beginning at time T_{split} and continuing until present day. Concurrent with this symmetric gene flow, *Quercus engelmannii* also experienced growth in effective population at rate (s) that persisted from divergence into contemporary populations while the effective population size of *Quercus beberidifolia* remained the same. (Parameters: $v_1, v_2, s, m, T_{split}$)

Asymmetric effective population size growth in Quercus berberidifolia *with continuous symmetric gene flow* (P2ExpSM)- Species diverged at time T_{split} with ancestral effective

population sizes of v_1 and v_2 and experienced intermittent, but continuous symmetric gene flow at rate (m) beginning at time T_{split} and continuing until present day. Concurrent with this symmetric gene flow, *Quercus berberidifolia* also experienced growth in effective population at rate (s) that persisted from divergence into contemporary populations while the effective population size of *Quercus engelmannii* remained the same.
(Parameters: $v_1, v_2,$ s, m, T_{split})

Species Divergence with symmetric gene flow after secondary contact (SplitSM)- Species diverged at time T_{split} with ancestral effective population sizes of v_1 and v_2 and remained isolated until time T_{mig} when the species came into secondary contact. Upon secondary contact, symmetric gene flow between species occurred at rate (m). (Parameters: $v_1, v_2,$ m, T_{split}, T_{Mig})

6 Parameter models

Symmetric effective population size growth with continuous asymmetric gene flow (SexpAM)- Species diverged at time T_{split} with ancestral effective population sizes of v_1 and v_2 and experienced intermittent, but continuous asymmetric gene flow at rates m_{12} (gene flow **from** *Q. engelmannii* **into** *Q. berberidifolia*) and m_{21} (gene flow **from** *Q. berberidifolia* **into** *Q. engelmannii*) beginning at time T_{split} and continuing until present day. Concurrent with this gene flow, both species also experienced growth in effective population size at rate (s). (Parameters: v_1, v_2, m_{12}, m_{21}, s, T_{split})

Species Divergence with asymmetric gene flow after secondary contact (SplitAM)- Species diverged at time T_{split} with ancestral effective population sizes of v_1 and v_2 and remained isolated until time T_{mig} when the species came into secondary contact. Upon secondary contact, asymmetric gene flow between species occurred at rates m_{12} (gene flow **from** *Q. engelmannii*

into *Q. berberidifolia*) and m_{21} (gene flow **from** *Q. berberidifolia* **into** *Q. engelmannii*).

(Parameters: $v_1, v_2, m_{12}, m_{21}, T_{split}, T_{Mig}$)

8 Parameter Models

Population bottleneck Q. engelmannii *with continuous symmetric gene flow after secondary contact* (BotSM)- Species diverged at time T_{split} with ancestral effective population sizes of v_1 and v_2 and remained isolated until time T_{Bot} at which *Q. engelmannii* underwent a population bottleneck of depth v_B and recovered to effective population size v_F at time T_F. At the point in which the bottleneck occurs, these species also come into secondary contact and begin symmetric gene flow at rate (m).

(Parameters: v_1, v_2, v_B, v_F , m, T_{split}, T_{Bot}, T_F)

9 Parameter Models

Bottleneck in Q. engelmannii *with asymmetric gene flow during the bottleneck event* (PreBotAsymEAM)- Species diverged at time T_{split} with ancestral effective population sizes of v_1 and v_2 and remained isolated until time T_{Bot} at which *Q. engelmannii* underwent a population bottleneck of depth v_B and recovered to effective population size v_F at time T_F. Secondary contact between these species occurs at time T_B (at the beginning of the bottleneck event) and asymmetric gene flow occurs at rate m_{12} (gene flow **from** *Q. engelmannii* **into** *Q. berberidifolia*) and m_{21} (gene flow **from** *Q. berberidifolia* **into** *Q. engelmannii*).

(Parameters: v_1, v_2, v_B, v_F , $m_{12}, m_{21}, T_{split}, T_{Bot}, T_F$)

Bottleneck in Q. engelmannii *with asymmetric gene flow after the bottleneck event* (ABotAsymEAM)- Species diverged at time T_{split} with ancestral effective population sizes of v_1

and v_2 and remained isolated until time T_{Bot} at which *Q. engelmannii* underwent a population bottleneck of depth v_B and recovered to effective population size v_F at time T_F. Secondary contact between these species occurs at time T_B (at the beginning of the bottleneck event) and asymmetric gene flow occurs at rate m_{12} (gene flow **from** *Q. engelmannii* **into** *Q. berberidifolia*) and m_{21} (gene flow **from** *Q. berberidifolia* **into** *Q. engelmannii*).

(Parameters: v_1, v_2, v_B, v_F , m_{12}, m_{21}, T_{split}, T_{Bot}, T_F)

Table S1: All tested models ranked by ΔAIC scores in comparison to the most likely model according to log likelihood scores. Score highlighted in bold represents the model with the highest log likelihood.

Model	LL	AIC	ΔAIC
PreBotAsymEAM	**1851.7**	**3723.4**	**0**
AexpSM	-1873.4288	3758.8576	35.4576
Bot Asym B	-1878.7076	3773.4152	50.0152
P2ExpSM	-1890.725	3791.45	68.05
ABotAsymEAM	-1895.3173	3810.6346	87.2346
P1ExpSM	-1910.3164	3830.6328	107.2328
Amig	-2027.1024	4064.2048	340.8048
SplitSM	-2135.5278	4271.0556	547.6556
Smig	-2355.9259	4719.8518	995.6
SplitAM	-2367.2398	4746.4796	1023.0796
BotSym	-2816.7769	5649.5538	1926.1538
BotAsymBE	-2819.22	5654.44	1931.04
P2Exp	-4087.1723	8182.3446	4458.9446
SExp	-4818.656	9645.312	5921.912
Iso	-5014.309	10038.618	96315.22
P1Exp	-5522.1018	11052.2036	7328.804

Table S2: Estimated parameters for each tested model. Models include ancestral population size (n), effective population size before and after bottleneck (n_B, n_F) time of divergence (T_s), time of migration (T_m), time of bottleneck (T_b) and migration rate (m, m_{12}, m_{21}). Parameter values represent estimates from the most likely model for each scenario.

Model	**Ts**	**Tm**	**Tb**	**s**	**Nu1**	**Nu2**	**NuB**	**NuF**	**m**	**m12**	**m21**
Iso	0.6234	x	x	x	0.8195	0.50365	x	x	x	x	x
SExp	0.505	x	x	0.5799	1.275	1.5955	x	x	x	x	x
Smig	0.063	x	x	x	0.2086	0.178	x	x	0.0247	x	x
P1Exp	0.5001	x	x	0.2076	0.6217	1.0655	x	x	x	x	x
P2Exp	0.5004	x	x	0.2254	0.714	0.7362	x	x	x	x	x
P1ExpSM	1.8401	x	x	0.907	0.4628	0.338	x	x	1.5767	x	x
P2ExpSM	1.4548	x	x	0.6106	1.1721	0.6756	x	x	0.5083	x	x
Amig	1.883	x	x		0.3891	0.76939	x	x	x	0.99375	0.36834
AexpSM	1.6673	x	x	0.4942	0.2029	0.1809	x	x	x	2.9062	2.605
SplitSM	2.8823	0.9279	x	x	0.6946	0.6791	x	x	0.5916	x	x
SplitAM	2.963	1.2661	x	x	0.7989	0.6807	x	x	x	0.4837	0.42133
Bot_Sym	1.1405	1.5803	0.6995	x	0.3866	1.0023	0.0221	0.6613	0.7274	x	x
Bot Asym B	0.909	0.1513	0.101	x	0.1445	1.0911	2.215	0.0218	x	0.3411	2.695
Bot Asym BE	1.9996	1.02933	0.053851	x	0.6143	0.59582	0.0878	0.5826	x	0.54916	0.73863
ABotAsymEAM	1.4232	0.64569	0.499679	x	0.7594	0.80568	0.1934	0.6664	x	1.05991	0.88959
PreBotAsymEAM	**1.9933**	**0.45099**	**0.34155**	**x**	**0.6143**	**0.59582**	**0.0878**	**0.5826**	**x**	**0.86777**	**1.14819**

Table S3: Estimated timing of events of the most likely demographic model using an estimated generation time of 5, 10 and 20 years. It is worth noting that recent time calibrated phylogenies of the genus *Quercus* estimate the age of the genus to be ~60mya and divergence between *Quercus engelmannii* and *Q. berberidifolia* to be ~40mya (Hipp et al., 2018; Hipp et al., 2019).

Generation Time	Time of Divergence	Time of Bottleneck	Timing of Gene Flow	Timing of Bottleneck Recovery
5 years	45mya	10mya	10mya	8mya
10 years	90mya	20mya	20mya	16mya
20 years	180mya	40mya	40mya	32mya

Literature Cited

Hipp, A. L., Manos, P. S., González-Rodríguez, A., Hahn, M., Kaproth, M., McVay, J. D., Avalos, S. V., & Cavender-Bares, J. (2018). Sympatric parallel diversification of major oak clades in the Americas and the origins of Mexican species diversity. *New Phytologist*, *217*(1), 439-452.

Hipp, A. L., Manos, P. S., Hahn, M., Avishai, M., Bodénès, C., Cavender-Bares, J., Crowl, A. A., Deng, M., Denk, T., & Fitz-Gibbon, S. (2019). Genomic landscape of the global oak phylogeny. *New Phytologist*, *226*(4), 1198-1212.