UNIVERSITY OF CALIFORNIA

Los Angeles

The genomic basis of adaptation to climate across oak *(Quercus)* species and populations in California

A dissertation submitted in partial satisfaction

of the requirements for the degree

Doctor of Philosophy in Biology

by

Alayna Mead

2023

ABSTRACT OF THE DISSERTATION

The genomic basis of adaptation to climate across oak *(Quercus)* species and populations in California

by

Alayna Mead

Doctor of Philosophy in Biology

University of California, Los Angeles, 2023

Professor Victoria Sork, Chair

Characterizing the genomic basis of climate adaptation is essential both for understanding the process of evolution and for conserving populations under climate change. Here, I use several methods to investigate the genomic basis of adaptation across oak species and populations. In Chapter 1, we compared the gene expression response to a simulated drought stress across six oak species, two species from each of the three taxonomic sections of *Quercus* in California, which varied in their drought tolerance. We found that drought tolerant species had a less plastic response to leaf drying, suggesting that phenotypic traits and non-plastic patterns of gene expression contribute to their drought tolerance more than plasticity. We also found that the two deciduous trees, which were the most drought sensitive species, responded to drying with 22% of the same genes, indicating these responses had evolved in parallel across distantly related clades. In Chapter 2, we used whole-genome sequencing to characterize the rangewide genetic structure of a rare island endemic oak species, *Quercus tomentella*, from the California Channel Islands and Guadalupe Island in Mexico. We found evidence for widespread hybridization with a related oak species, isolation of the Guadalupe Island trees, and some gene flow among the California islands that is likely mediated by wind pollination. We also identified putatively adaptive SNPs that were associated with climate variables, compared the spatial patterns of neutral SNPs and candidate SNPs, and used this information to make recommendations for choosing seed sources for restoration projects in the light of climate change. In Chapter 3, we characterized the heat

stress response across species and populations. We performed a heat wave experiment and compared the gene expression responses among three oak species and two populations within each species, one from a southern, warmer site and one from a northern, cooler site. We found shared responses to heat stress among all species and populations, including the responses of individual genes as well as genes with related functions. We found limited evidence of differences in stress response among populations, suggesting a lack of local adaptation in the plastic heat stress response.

The dissertation of Alayna Mead is approved.

Kirk Edward Lohmueller

Lawren Sack

Karen Elizabeth Sears

Victoria Sork, Committee Chair

University of California, Los Angeles

2023

Contents

List of Figures

List of Tables

Acknowledgements

Chapter 1 is in preparation for publication with authors Alayna Mead, Camila D. Medeiros, Lawren Sack, and Victoria L. Sork. CDM performed physiological measurements and analysis, and all authors contributed to the experimental design.

Chapter 2 is in preparation for publication with authors Alayna Mead, John Knapp, Sorel Fitz-Gibbon, and Victoria L. Sork. JK and VLS contributed to experimental design and sampling, SFG performed variant calling, and VLS provided input on analyses.

Chapter 3 is in preparation for publication with authors Alayna Mead and Victoria L. Sork. VLS contributed to experimental design and analysis.

So many people have helped and supported me during my PhD.

I could not have become the scientist I am today without the support of my advisor, Victoria. Thank you for accepting me into your lab, for encouraging me, and for sharing your incredible knowledge of oaks and insight into how to do science. I admire you so much — both as a scientist and as a person.

Thank you to my committee members for your support, advice, and enthusiasm over the years. Kirk Lohmueller has provided me with me a deeper insight into population genetics, Lawren Sack has helped me think about what is happening at a physiological level when plants respond to their environment, and Karen Sears has given me a broader view of evolution. Others who have contributed to this dissertation and to my growth as a graduate student include Felipe Zapata, for his expertise on plant evolution, and Shane Campbell-Staton, for insight into parallel evolution and advice on designing gene expression experiments. Rachel Prunier also provided feedback on this work and helped with fieldwork, and more importantly, she has been a second mentor.

Thanks to all my fellow Sork labbies, past and present: Berenice Badillo, Krista Beckley, Luke Browne, Dylan Burge, Sorel Fitz-Gibbon, Claudia Henriquez, Yao Li, Zhizhong Li, Bran-

don MacDonald, Kari Merrill, Stephanie Steele, Scott O'Donnell, Marissa Ochoa, Heidi Yang, and Diego Zapata. This lab has been such a wonderful place to work, and I have learned so much from each of you. I'm thankful for all the lab meetings, field trips, pasta and taco dinners, and weekends spent measuring trees.

I am so thankful for my friends and fellow graduate students. So many people have helped me along the way with encouragement, advice, commiseration, and solidarity. Thank you for creating a supportive community and for always striving to make things better.

And lastly, of course, thanks to my family for always loving me and supporting me, throughout grad school and from the very beginning when I was a kid who loved reading and playing outside. Mom and Dad, thank you for giving me my love of nature and learning, and the endless support to pursue the life I wanted. To my partner Adam, for understanding my drive for learning because you share it, for coming along on collecting trips and helping me become better at coding, for knowing when to tell me to go write and when to tell me to take a break from writing, for helping me debug LaTeX while typesetting this dissertation(!) and for so much more, thank you.

VITA

Alayna Mead
Ph.D. Candidate in Biology

Education

MS Biology, 2017
The University of California, Los Angeles
Thesis: Water stress treatment in valley oak (*Quercus lobata*) seedlings reveals species–wide similarities and population–specific differences in ecophysiological and gene expression response

BS Organismal Biology, *magna cum laude,* 2015
Auburn University

Awards and Grants

2021: Vavra Research and Travel Award
2019: Vavra Research and Travel Award
2019: Best student poster presentation, Western Forest Genetics Association Conference
2019: Vavra Fellowship
2018: UCLA Alumni Fellowship

Publications

Mead AM, J Peñaloza Ramirez, MK Bartlett, JW Wright, L Sack, & VL Sork (2019). Seedling response to water stress in valley oak (*Quercus lobata*) is shaped by different gene networks across populations. Molecular Ecology, 28(24), 5248-5264.

Browne L, **AM Mead**, C Horn, K Chang, Z Celikkol., C Henriquez, F Ma, E Beraut., R Meyer. & VL Sork. (2020). Experimental DNA demethylation associates with changes in growth and gene expression of oak tree seedlings. G3: Genes, Genomes, Genetics, 10(3), 1019-1028.

Research Experience

2015–2023: Graduate Research with Dr. Victoria Sork, University of California, Los Angeles
2013–2015: Undergraduate Research with Dr. Kenneth M. Halanych, Auburn University
2014: Summer Research Experience for Undergraduates with Dr. Lena Hileman, University of Kansas

Conference Presentations

Mead AM, J Knapp, S Fitz-Gibbon, and VL Sork. August 2022. Using genomics to understand genetic variation and gene flow among populations of a rare island endemic oak, *Quercus tomentella*. Poster presentation, International Oak Society Conference.

Mead AM, J Knapp, S Fitz-Gibbon, and VL Sork. July 2022. Using genomics to understand genetic variation and gene flow among populations of a rare island endemic oak, *Quercus tomentella*. Poster presentation, Botany Conference.

Mead AM, C Medeiros, L Sack, VL Sork. June 2021. Climate niche, not phylogeny, explains differences among oak species in gene expression response to drought. Oral Presentation, Evolution Conference.

Mead AM. May 2021. Evaluating assisted gene flow as a strategy to conserve California tree species under climate change. Poster presentation, Forest Genetics Student & PostDoc Symposium.

Mead AM, C Medeiros, L Sack, VL Sork. July 2020. Oak drought tolerance traits and climatic niches determine their transcriptomic responses to leaf drying. Oral presentation, Botany Conference.

Mead AM, C Medeiros, L Sack, VL Sork. June 2019. Gene expression response to leaf drying in six California oak species reveals similarities among species occurring in similar climates. Poster presentation, Western Forest Genetics Association Conference.

Mead AM, C Medeiros, L Sack, VL Sork. May 2019. Gene expression response to leaf drying in six California oak species reveals similarities among species occurring in similar climates. Poster presentation, UCLA Ecology and Evolutionary Biology Research Symposium.

Mead AM, J Peñaloza Ramirez, M Bartlett, J Wright, L Sack, VL Sork. October 2018. Physiological and Gene Expression Response of Valley Oak (*Quercus lobata*) Seedlings to Water Stress. Poster Presentation, International Oak Society Conference.

Mead AM, J Peñaloza Ramirez, M Bartlett, J Wright, L Sack, VL Sork. June 2017. Variation in gene expression and ecophysiological response to water stress in valley oak seedling populations. Oral Presentation, Botany Conference.

Teaching Experience

Conservation Genetics	Winter 2022
Ecology and Behavior Lab	Fall 2015, Winter 2016, Fall 2018, Winter 2019, Spring 2019, Fall 2020, Winter 2021, Fall 2022
Ecology, Evolution, and Biodiversity Lab	Spring 2016, Fall 2017, Winter 2020
Life: Concepts and Issues	Fall 2016
Plant Physiology	Spring 2017, Spring 2022

Chapter 1

Gene expression response to drought shows both convergent and parallel evolution across oak species[1]

Abstract

The mechanisms underlying the evolution of adaptive phenotypes is a fundamental topic in evolutionary biology. Species from different lineages may evolve similar traits in response to the same climates through the same underlying genes (parallel evolution), or they may use different genes (convergent evolution). Here, we used comparative transcriptomics to test the extent to which repeated adaptations to the same climates found in different lineages of oaks are the result of parallel or convergent evolution. We considered six California oak species within three sections of the genus, including a pair of species within each section representing relatively mesic or xeric environments. Using individuals grown in a common environment, we measured gene expression within leaves in response to dehydration, and its association with functional traits. We tested whether species were more similar to another species within the same clade, indicating evolutionary limitations due to phylogenetic history, or to a distantly related congeneric species with a similar climate niche, indicating parallelism. We find that four of the six species (all with a tree growth form) had more drought-responsive genes in common with ecologically sim-

[1]This chapter will be submitted for publication with the following authors: Alayna Mead, Camila D. Medeiros, Lawren Sack, Victoria L. Sork

ilar species than within-clade species, consistent with parallel evolution. Yet, these ecologically-similar species responded to drought stress using different genes with similar functions, indicating an important role for convergent evolution as well. Species with greater drought sensitivity showed greater plasticity in gene expression during leaf dehydration. To understand the mechanisms associated with drought tolerance, we identified frontloaded genes with high constitutive expression in drought tolerant species and genes with strong divergence in expression levels among species and discovered multiple candidate genes that may underlie drought adaptation through non-plastic mechanisms. Together these results suggest that oak clades have adapted to a wide range of climates through both parallel and convergent evolution in their functional and physiological traits and the gene expression underlying those traits. The ability of oaks to adapt to climate may contribute to the high species richness and wide distribution of oaks across the northern hemisphere.

Introduction

Taxa occurring in similar environments often share phenotypes that are beneficial to that environment, exhibiting the ubiquitous influence of natural selection on evolution (Elmer and Meyer 2011). A key question in evolutionary biology is whether such similarities have evolved through parallel evolution, in which the same genomic mechanisms underlie the same phenotypes, or convergent evolution, in which different genomic mechanisms result in the same phenotype. Quantifying these two processes help us to understand how novel phenotypes will evolve: if evolution tends to occur repeatedly using the same genes each time, then adaptation may be more constrained and occur slowly, but if some gene functions are redundant so that mutations in different genes can achieve the same phenotype, taxa may be more flexible in adapting to new conditions (Losos 2011, Yeaman et al. 2016).

The repeated evolution of the same phenotypic trait may originate from repeated genomic changes at different levels of biological organization, such as the reuse of the same SNPs or amino

acid substitutions, mutations within the same gene but at different loci, and mutations in different genes that belong to the same pathway or have similar functions (Elmer and Meyer 2011). Because definitions of parallel and convergent evolution differ across studies (Arendt and Reznick 2008), we will use "parallelism" for cases when the same gene underlies an adaptive phenotype across species, "convergence" when different genes are used, and "repeated evolution" when the mechanism is unclear. Parallelism is more likely among closely related taxa, because they share more standing genetic variation that can be independently fixed under similar selection pressures (Conte et al. 2012, Ord and Summers 2015). Parallelism is also likely to occur when only a few genes underlie an adaptive phenotype, simply because there are limited mechanisms allowing convergence to take place (O'Quin et al. 2010). However, in many cases, whole-organism level characteristics are determined by multiple genes, allowing a similar phenotype to evolve through multiple mechanisms (Agrawal 2017). Therefore, in taxa that are more distantly diverged, and for traits that are polygenic, it is more likely that the evolution of similar phenotypes will occur through convergent changes at higher levels of biological organization (Bailey et al. 2015, Pfenninger et al. 2015, Renaut et al. 2014, Tenaillon et al. 2012).

Oaks (genus *Quercus*) are a diverse group of trees and shrubs that have adapted to a wide range of climates throughout their evolutionary history (Cavender-Bares 2018, Sork et al. 2022). Many North American oaks in different taxonomic sections of the genus have similar traits, and because species from different sections do not hybridize with each other (Nixon 2002), we can assume these traits arose independently within each section. Distantly related oak species often co-occur in the same community, and these co-occurring species tend to have similar traits that they may have independently evolved while adapting to the same climate (Cavender-Bares 2018, Cavender-Bares et al. 2004, Mohler 1990). Additionally, phylogenetic data show that two largest sections in North America, the red oaks (section *Lobatae*) and white oaks (section *Quercus*) have radiated into new climatic niches in parallel (Hipp et al. 2018), while species within *Protobalanus*, a small section that is geographically restricted to western North America, also inhabit different

climate niches (Nixon 2002). As the genus spread south from its paleo-arctic distribution over the last 50 million years, species have evolved traits that are beneficial for their new climate niches. One trait that is critical in niche adaptation in oak species is drought tolerance (Abrams 1990, Kaproth and Cavender-Bares 2016, Lobo et al. 2018, Ramírez-Valiente et al. 2018, Skelton et al. 2018). Given the polygenic nature of drought tolerance, it is unclear whether this trait has evolved through convergent or parallel evolution.

Evolutionary changes in gene expression are likely to be an important mechanism of adaptation (Gompel and Prud'homme 2009, Jones et al. 2012, King and Wilson 1975, Stern 2013, Whitehead and Crawford 2006). The amount of mRNA produced in a tissue under certain conditions can be considered a phenotype (Khaitovich et al. 2006) approximately corresponding to the amount of protein translated (Schwanhäusser et al. 2011). In contrast to traditional phenotypes, the measurement of gene expression makes it possible to identify specific genes. This allows us to measure the prevalence of parallelism in gene expression across species as the frequency at which species respond to a treatment by altering expression of the same genes. Here we focus specifically on genes which have altered levels of expression under leaf drying, as they may be important in the plant's plastic response to drought.

This study assesses the extent to which gene expression underlying drought tolerance reflects parallel or convergent evolution. We compared drought-related traits and gene expression responses to leaf drying across six oak species from three sections (clades within the genus *Quercus*), including two species from contrasting mesic or xeric climates within each section (Figure 1.1). We hypothesized that each species would have a response to drought stress that is more similar to an ecologically similar species than to a closely related species because the strength of natural selection exerted by drought would overcome phylogenetic constraints. We expect that these shared responses to drought will include both parallelism, in which separate species respond by altering expression of the same genes, and convergence, in which responses involve different genes with related functions. Parallelism in gene expression may be common

because regulatory regions can be evolutionary hotspots (Martin and Orgogozo 2013). However, we also expected that species will display convergence by responding to stress with different genes that have similar functions, because climate-adaptive traits are often polygenic (Lind et al. 2018, Savolainen et al. 2013, Yeaman et al. 2016). If adaptation to a given condition is able to evolve in multiple ways – either through multiple different phenotypes, or the same phenotype accomplished through different mechanisms – beneficial standing genetic variation is likely to be present in populations, which would facilitate adaptation to new environmental conditions (Barrett and Schluter 2008).

Methods

Sampling for functional traits

For this study we sampled six of the 21 species of California native oaks, with two representatives from each of the three *Quercus* sections found in North America, *Quercus*, *Lobatae*, and *Protobalanus* (Hipp et al. 2018). All sampled individuals were grown in a common garden at the California Botanic Garden (formerly Rancho Santa Ana) in Claremont, California (34.110738, −117.713913; 507 mm of rainfall per year; WorldClim, Hijmans et al. 2020). This common garden approach is ideal to test questions related to trait evolution across species, since it ensures that the observed differences are due to heritable interspecies variation and not to confounding environmental variables that may have elicited phenotypically plastic trait variation (Cavender-Bares et al. 2020, Cordell et al. 1998, Dunbar-Co et al. 2009, Fletcher et al. 2018, Givnish and Montgomery 2014, Mason and Donovan 2015).

Data were collected from three to six individuals per species, given availability in the common garden. For each individual, we used pole pruners to collect the most exposed to the sun and mature branch grown in the current year, with no signs of damage and herbivory. Branches were carried to the lab in dark bags with moist paper and rehydrated overnight under dark plastic

before harvesting stem sections and fully expanded leaves and stems for all subsequent analyses.

Functional trait measurements

Functional traits were measured for three sun leaves per individual, unless noted otherwise. Leaf area was measured using a flatbed scanner and analyzed using the software ImageJ (imagej.nih.gov/ij/). After scanning, leaves were oven-dried at 70° for 72 h and dry mass was measured using an analytical balance (0.01 mg; XS205; Mettler-Toledo, OH, USA). Leaf mass per area (LMA) was calculated as lamina dry mass divided by leaf area (Pérez-Harguindeguy et al. 2013).

The concentrations of leaf nitrogen per mass (N_{mass}, data provided in Table S6) was determined from oven-dried leaves by the Center for Stable Isotope Biogeochemistry from the University of California, Berkeley, by continuous flow dual isotope analysis (CHNOS Elemental Analyzer interfaced to an IsoPrime100 mass spectrometer). The concentration of leaf phosphorus per mass (P_{mass}) was determined from oven-dried leaves by the Baxter Lab in the Donald Danforth Plant Science Center, using a high throughput elemental composition (ionomics, approach from Salt et al. 2008, which uses Inductively Coupled Plasma-Mass Spectrometry (ICP-MS).

We estimated maximum rate of carboxylation per mass (V_{cmax}) and electron transport rate (J_{max}) from leaf N and P concentrations per mass (Domingues et al. 2010). Estimates of leaf lifetime integrated CO_2 assimilation rate ($\overline{A}_{mass}$) were derived from V_{cmax}, J_{max} and isotope composition data using the Farquhar, von Caemmerer and Berry model (Franks et al. 2009).

We measured the wood density (WD) from 5-cm branch segments after bark removal using the water displacement method (Pérez-Harguindeguy et al. 2013). The osmotic pressure at turgor loss point (π_{TLP}) was measured in two leaves from each of the 3-5 studied individuals. We used vapor-pressure osmometers (Vapro 5520 and 5600, Wescor, US) to obtain the osmotic concentration (π_o) of the leaves and used calibration equations to estimate π_{TLP} (Bartlett et al. 2012).

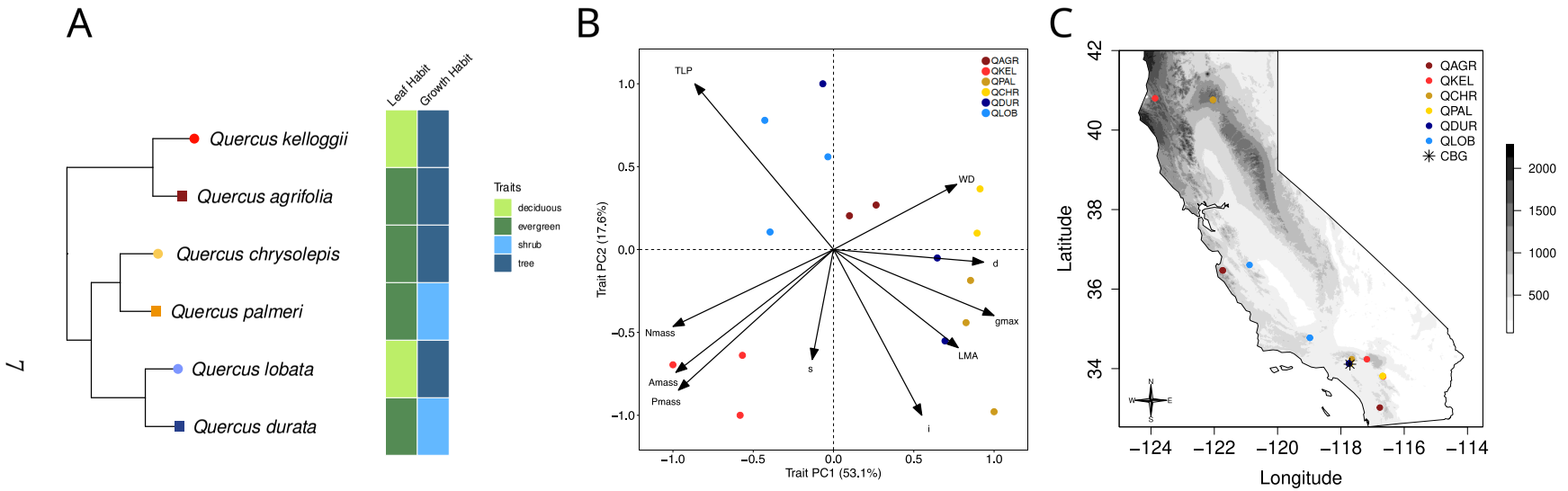


Figure 1.1. (A) Phylogeny of species, including their leaf habit and growth habit. Within each section, the more drought-tolerant species is indicated with a square, and the more drought-sensitive species with a circle. (B) PCA of mechanistic drought traits: leaf mass per area (LMA), wood density (WD), turgor loss point (TLP), leaf nitrogen and phosphorus content per mass (Nmass, Pmass), time-integrated photosynthetic rate per mass (Amass), stomatal density (d), stomatal initiation rate (i), and anatomical maximum stomatal conductance (gmax). (C) Location of provenance origins for each species and location of the common garden (California Botanic Garden, CBG).

Environmental variables for individual's provenances and species' native ranges

We obtained the coordinates for the location of sampling of acorns from the records of the California Botanic Garden and species occurrence data from the Global Biodiversity Information Facility (GBIF; references available in Table S2) and we used R software (version 3.4.4, R Core Team 2019 to extract and calculate the mean, range and standard deviation of environmental variables of the range of distribution of each species. Occurrence records were downloaded using the 'rgbif' and filtered to keep herbarium records only and remove incomplete (latitude or longitude missing) and duplicated records, non-natural occurrences (e.g., records from botanical gardens, planted urban trees) and to limit the temporal range to 1950-current (Riordan et al. 2015).

From open-access raster layers, we extracted nine environmental parameters: mean annual temperature (MAT), mean temperature of the warmest quarter (T_{max}), temperature seasonality (T_s), mean annual precipitation (MAP), precipitation of the driest quarter (P_{dry}), precipitation of the warmest quarter (P_{warm}; WorldClim, Hijmans et al. 2005), potential evapotranspiration (PET), aridity index (AI; CGIAR-CSI, NCAR-UCAR, Zomer et al. 2008) and soil water capacity ($soil_{WC}$; ISRIC Soilgrids Hengl et al. 2017) (Table S2). The raster layers with same resolution were stacked using the stack function from the 'raster' package (Hijmans et al. 2020) and the environmental parameters for each occurrence record were extracted using the extract function from the 'dismo' package (Hijmans et al. 2020).

Drying experiment

Branches were cut from the same adult individuals sampled for functional traits growing in a common environment at the California Botanic Garden, Claremont, California for use in a simulated leaf drying experiment. Due to space constraints, the experiment was run on two dates (11 May and 22 May 2018), with individuals from the same species and accession spread across the two experiments to avoid batch effects. All conditions were kept the same between the two experiments. After collection, branches were re-cut under water and rehydrated overnight, and

the treatment began the following day.

Two branchlets of approximately the same size (i.e., each having 8 leaves) were assigned to the control group or drought treatment. This allowed us to test the effects of drought on gene expression within the same individual and control for genetic variation in gene expression among individuals. Three replicates per treatment/species combination were used, including individuals from two accessions from a northern and southern part of the species range (except for *Q. durata*, which had one accession, and *Q. kelloggii*, which had three; Figure 1.1). Between 11 am and 12 pm, immediately after being cut, branches were placed on fans under high light to induce transpiration. Control branches were provided with water in a plastic bag wrapped around the base of the branch. This treatment continued for two hours, after which leaves were flash frozen in liquid nitrogen for RNAseq and stored at -80 °C (between 1 and 2 pm). Additionally, the water potential of leaves (Ψ_{leaf}) was measured for each branch before and after the experiment to ensure controls were well-hydrated and to compare the degree of stress among the treated branches (Figure S1). For all species, Ψ_{leaf} values were near zero for the control leaves (species averages ranging from -0.04 to -0.09), and Ψ_{leaf} values decreased under the drying treatment as expected (species averages ranging from -0.76 to -3.26).

RNA extraction and sequencing

To extract RNA from leaves, about 50 mg of leaf tissue was ground at room temperature using a Retsch mixer mill (model MM 301), placing grinding adapters and tubes in liquid nitrogen between rounds of grinding. For each extraction, tissue from three different leaves was used to reduce the effects of any variation among leaves within an individual branch. Polyphenolics and polysaccharides, which are common in oak leaves and interfere with downstream applications, were removed from leaves using a lithium chloride/urea-based pre-wash, described fully at: openwetware.org/wiki/Sork_Lab:Protocols#RNA_Extraction_for_Oak. Tissue was kept in the buffer overnight at 4 °C. The next day, mRNA was extracted using the Sigma-Aldrich Spectrum Plant

Total RNA Kit.

Sequencing libraries were prepared using the Illumina TruSeq RNA Library Prep Kit. RNA extractions were checked for quality using a NanoDrop instrument and Agilent TapeStation 2200, and final libraries were also quality checked on the Tapestation. Each library was diluted to 10 nM in EB buffer and 0.1% tween 20, based on DNA concentrations given by a Qubit Fluorometer, then samples were pooled by equal volume for sequencing. Samples were sequenced by the UCLA Broad Stem Cell Research Center's sequencing core on an Illumina 4000 machine, using 100 bp, paired-end sequencing. Samples were sequenced across 3 lanes (12 libraries per lane) with samples from the same species and treatment date distributed evenly across different lanes; control and treated samples of the same individual were run on the same lane.

RNAseq data processing

Sequence data files were converted from qseq to fastq format, demultiplexed (allowing for a one-base barcode mismatch), and reads failing the Illumina quality filter were removed using custom scripts (available at github.com/alaynamead/RNAseq_scripts). Adapters and low-quality reads (using a quality score cutoff of 27) were trimmed using Cutadapt 2.3. Reads were aligned to the *Quercus lobata* genome version 3.0 (Sork et al. 2022) using STAR 2.7 (Dobin et al. 2013) with default parameters. Only uniquely mapped reads (MAPQ score = 255) were retained. Sequencing platform duplicates (tagged 'DT:Z:SQ') were marked using the MarkDuplicates tool in Picard 2.18.23 and were removed. Read counts for each gene were obtained using the htseq-count command from HTSeq 0.11.2.

Gene expression analysis

To identify differentially expressed (DE) genes, we used DESeq2 (Love et al. 2014) because it has high power for experiments with low sample sizes (Schurch et al. 2016). Only genes from chromosomes 1-12 were included in the analysis, excluding those from unplaced scaffolds, leaving

a total of 38,510 genes used in the analysis. DE genes were identified using a species × treatment model, taking the paired design into account by including the individual as a factor. *P*-values were adjusted using the Benjamini-Hochberg procedure (Benjamini and Hochberg 1995) (FDR = 0.05).

We compared the gene expression responses across species to determine the factors shaping the evolution of the response to drought. To explore whether species were most similar to a closely related but ecologically different species, or to ecologically similar species from a different taxonomic section, we assigned species into ecologically-similar groups based on their drought tolerance, leaf habit, and growth habit. The deciduous trees, *Q. kelloggii* and *Q. lobata*, are the most drought-sensitive species. The other four species are all evergreen and relatively drought-tolerant, and we separated them into groups based on growth form: evergreen trees, *Q. agrifolia* and *Q. chrysolepis*; and evergreen shrubs, *Q. durata* and *Q. palmeri*. For each species, we compared its gene expression response to drying with the same-section species, and to the ecologically-similar species. The overlap in gene expression response to drying among species pairs was determined using the Jaccard index as in Bailey et al. (2015), calculated as the intersection (number of shared DE genes) divided by the union (total number of DE genes for both species). The same calculation was performed for the overlap among species of GO terms enriched within the DE genes.

To test whether patterns of gene expression were affected by alignment to the *Q. lobata* genome (because species more closely related to *Q. lobata* may align to the genome better than divergent species), we also aligned reads to the *Q. suber* genome (Ramos et al. 2018), which is an outgroup to all the species included in this study. Alignment and differential expression analysis were done using the same methods as in the primary analysis.

For analyses of gene expression variation among individuals (described below), DESeq2 was used to produce normalized gene expression values from raw transcript counts. First, genes that were not expressed in this dataset were removed, leaving 36,331 genes. The ‘varianceStabiliz-

ingTransformation' function with a parametric fit was used to transform count data, normalizing by library size across samples and stabilizing variance across genes (in particular, reducing the variance of lowly expressed genes) to better compare genes with different expression levels (Love et al. 2014).

GOseq (Young et al. 2010) was used to test whether the groups of upregulated and downregulated genes in each species were enriched for specific gene ontology (GO) terms. This method accounts for bias in differential expression results that are due to gene length, since longer genes will produce more reads during sequencing, resulting in greater statistical power for detecting expression differences. The default Wallenius distribution method was used to approximate the null distribution and calculate *P*-values. Each GO term is tested for both overrepresentation and underrepresentation, so *P*-values were adjusted separately for each group (using the Benjamini-Hochberg procedure) by determining whether a given GO term was more likely to be over- or underrepresented (based on *P*-values).

To test whether species differences were due to baseline gene expression in addition to gene expression plasticity, we identified genes that may be constitutively upregulated, or frontloaded (Barshis et al. 2013). Frontloaded genes may confer stress tolerance in some species by being highly expressed under all conditions, while only being expressed in response to stress in less tolerant species (Rivera et al. 2021). We identified the frontloaded genes for each species pair. Frontloaded genes in a given species (species 1) compared to another species (species 2) were defined as genes that were upregulated in response to drying in species 1, not differentially expressed in species 2, and having a higher expression under control conditions in species 2 than in species 1 (using a cutoff of 1.5, i.e. ≥1.5 times more highly expressed in species 2 under control conditions). To test the hypothesis that a given species had genes that were frontloaded compared to a second species, we tested whether frontloaded genes tended to be more highly expressed under control conditions in the second species, as might be expected if the gene is highly expressed at all times rather than plastically responding to a stimulus. For each species pair, the gene ex-

pression ratio under control conditions was calculated for the genes in the set of interest (genes that are upregulated in species 1 and non-DE in species 2), and the average ratio was compared to 10,000 replicates of a set of null genes (randomly sampled from all other genes). The *P*-value was calculated using the proportion of replicates in which the null set had a higher expression ratio than the test set.

We tested for relationships between gene expression response and drought tolerance traits in two ways. First, we tested for correlations between drought traits and the magnitude of the species gene expression response (the number of DE genes). Second, we tested for correlations between drought traits and average expression of genes sharing the same GO term, which should be functionally similar. The average expression response for a given GO term was calculated as the individual difference in drying and control expression for each gene, averaged across all genes annotated with a given GO term. This was correlated with species average trait values to test for a relationship between expression response of similar genes and species drought traits. This analysis was done for a subset of GO terms likely to respond to drought: those that were significantly enriched in the DE genes, as well as terms related to abscisic acid, response to water/osmotic stress, oxidative stress, and photosynthesis, which are known to be physiological drought responses (Pinheiro and Chaves 2011).

We used a partial redundancy analysis (RDA) in the R package vegan (Oksanen et al. 2019) to partition variance in gene expression response by the effects of phylogenetic relatedness and species climate. A redundancy analysis is a multivariate constrained ordination method that can be used to partition the variation in a data matrix into the proportion explained by two separate explanatory data matrices (Borcard et al. 2011). The gene expression response was measured as the difference in expression levels between drying and control branches within an individual. Pairwise branch distances were calculated for each species pair based on the oak phylogeny from Hipp et al. (2020) using the function 'cophenetic.phylo' from the R package ape, and the first and second axes of a PCoA on the branch distance matrix were used as variables measuring phyloge-

netic relatedness among species. The mean species climate was used as a measure of the climate a species has historically adapted to. We chose climate variables relevant to drought (precipitation of the driest quarter, mean temperature of the warmest quarter, potential evapotranspiration, soil water capacity, and temperature seasonality) and chose a subset of variables with low correlations with each other and that minimized variance inflation factors (tested using the function 'vif.cca'). The final model included three variables: precipitation of the driest quarter, mean temperature of the warmest quarter, and soil water capacity. We used a series of models to determine how much each group of variables explained variation in gene expression: both climate and phylogeny, each factor alone, and each factor controlling for the effects of the other factor. We used permutation tests (running the 'anova.cca' function with 1000 permutations) to determine whether each model explained more variation in gene expression than expected by chance. Models were run for two sets of genes: one set including all genes, and one set including the genes more highly ranked in the DE analysis. Because the number of significant genes varied widely among species, we did not choose these top genes by significance level to avoid species bias; instead, the top 500 genes from each species were selected as genes that are most likely to be plastic. Because of overlap in the top genes among species, this left 2078 "top" genes.

Evolutionary divergence in gene expression

To identify genes with highly divergent gene expression among species, we performed a phylogenetic ANOVA using the expression variance and evolution (EVE) model (Rohlfs and Nielsen 2015). This test uses the ratio of within- to between-species variance in gene expression (β), to identify genes that have high plasticity (high β) and genes that have high variation among species, but low variation within species (low β), consistent with directional selection on the expression levels of these genes. A likelihood ratio test (LRT) is used to determine whether a given gene likely has a different value of β than the rest of the genes. Only genes with no missing data were used (read count = 0 in one or more individuals), leaving 16,736 genes.

Significance was tested for each gene using 100 bootstrap replicates to calculate a null distribution of the likelihood ratio test statistic. This simulated the gene expression using the maximum likelihood distributions under the null hypothesis (that β is the same across all genes). *P*-values were calculated by testing how many bootstrap replicates had a LRT value lower than the actual value. Low *P*-values indicate that the actual LRT is lower than expected under the null hypothesis, consistent with directional selection on gene expression that diverged among species. To identify genes that have diverged among drought tolerant and drought sensitive species, we took the genes with the top 1% LRT (169 genes, all had *P*-values of 0 or 0.01, meaning 0 or 1 bootstrap replicates had LRT values lower than the actual value). For these top genes, we tested for a correlation between the species average gene expression and π_{TLP}.

Results

Functional traits

As expected from our experimental design, species within the same section did not necessarily share similar traits. Species do not cluster by section in a principal component analysis of species' traits (Figure 1.1B). PC1 clearly separates the two deciduous and most drought-sensitive species, *Q. kelloggii* (section *Lobatae*) and *Q. lobata* (section *Quercus*), from the other four evergreen species. Species from section *Protobalanus* (*Q. chrysolepis* and *Q. palmeri*), however, do cluster closely along PC1, perhaps because this clade is small and geographically restricted, consisting of only five evergreen species (Nixon 2002). The designated drought-tolerant species within each section had a more negative π_{TLP}, indicating greater resistance to wilting. However, the difference between the drought tolerance of same-section species pairs varied across sections, with *Q. lobata* and *Q. durata* in section *Quercus* having strongly contrasting drought tolerance traits, while *Q. chrysolepis* and *Q. palmeri* in section *Protobalanus* were both relatively drought tolerant (Figure 1.1B, Table S6).

Leaf drying experiment

Leaf water potential (Ψ_{leaf}), an indicator of osmotic stress, was near 0 for all individuals in the control treatment, indicating leaves were well-hydrated. Under the drying treatment, all individuals had a negative Ψ_{leaf}, indicating increased osmotic stress. Each species experienced similar levels of stress except for *Q. palmeri*, which had a Ψ_{leaf} significantly greater (less negative) than all other species (Figure S1), likely because it is the most drought tolerant species in this study.

Differential Expression

Of the 38,510 genes included in the analysis, 36,331 were expressed in the tissue. Species varied widely in the number of genes that were significantly differentially expressed between the control and drying treatments, with *Q. lobata* responding to drying with the most genes (5000, 13.8% of all genes) and *Q. palmeri* with the fewest (70, 0.2% of genes) (Table S1). The more drought-tolerant oak species responded to leaf dehydration with fewer genes (Figure 1.2) with the number of differentially expressed genes increasing exponentially in species with higher π_{TLP} values (P = 0.0002, R^2 = 0.97). The number of DE genes was not correlated with leaf water status (Ψ_{leaf}) (Figure S2). The results for the data aligned to the *Q. lobata* and *Q. suber* genomes were similar (Table S1), so further results reported here used the alignment to the chromosome-level *Q. lobata* genome rather than the scaffold-level *Q. suber* genome.

From the GO enrichment analysis of differentially expressed genes, we identified GO terms that were significantly enriched within the upregulated and downregulated genes for each species (Table S2) . Some GO terms were enriched within the upregulated genes for most species, except for one or both of the most drought-tolerant (*Q. palmeri* and *Q. chrysolepis*), including "calmodulin binding", "DNA-binding transcription factor activity", "protein ubiquitination", and "response to water" (Table S2). Significantly enriched downregulated GO terms were more species-specific; however, multiple species downregulated "ribosome" and "translation" functions.

We identified several GO terms whose overall expression was correlated with drought toler-

ance traits across species. Of the traits tested, the leaf potential at turgor loss point (π_{TLP}), an index of drought tolerance, had the greatest number of significant correlations with gene expression responses, followed by wood density (WD), leaf nitrogen content per mass (Nmass), time-integrated photosynthetic rate per mass (Amass), and leaf phosphorus content per mass (Pmass). For most of these GO terms, the species with the more drought-tolerant traits had little difference in expression between treatments, and the species with drought-sensitive traits more strongly upregulated or downregulated average GO expression in response to drying. More drought-sensitive species with higher π_{TLP} altered the expression of genes related to signaling (upregulated "double-stranded DNA binding" and "signal transduction", downregulated "ubiquitin-protein transferase activity" and "sequence-specific DNA binding"). They also upregulated genes related to lipid metabolic process, which may be related to the reorganization of cell membranes that may occur during drought (Gasulla et al. 2013, Hoekstra et al. 2001); and upregulated genes involved in carbon metabolism, which may be related to the accumulation of solutes as a stress response, or may be the result of metabolism changes due to decreased photosynthesis rates (Pinheiro and Chaves 2011) (e.g. processes involved in glycolysis, "beta-amylase activity", and "trehalose biosynthetic process"). Drought-sensitive species with lower WD generally downregulated expression of ribosomal and translation genes, while higher WD species had more constant levels of expression of these genes, suggesting drought-tolerant species maintained protein synthesis during the drying treatment, as found in drought-tolerant accessions of *Arabidopsis thaliana* (Des Marais et al. 2012) and populations of *Q. lobata* from warmer sites (Mead et al. 2019). Species with low WD also upregulated signaling-related genes, such as "MAP kinase activity" and "DNA binding". Species with higher Amass (photosynthetic rate) upregulated genes involved in "jasmonic acid biosynthetic process," which can be a stress response (Wang et al. 2020).

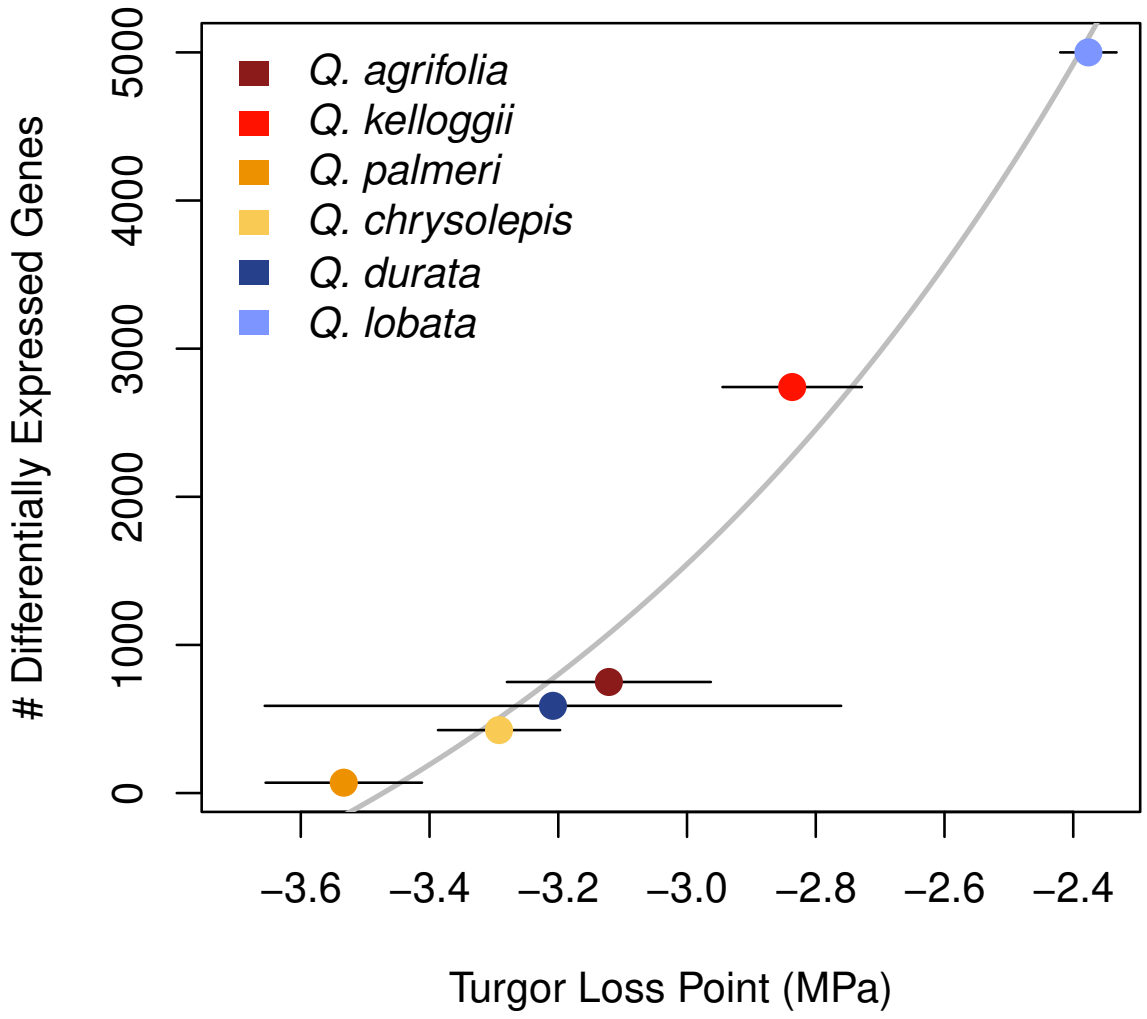


Figure 1.2. Relationship between species average turgor loss point (the water potential at which a leaf loses turgor; more negative values indicate stronger drought tolerance) and the number of differentially expressed genes. Drought tolerant species respond to drying with fewer genes. Error bars indicate standard error of turgor loss point. The number of differentially expressed genes is determined for each species rather than for individuals, so error bars are not shown. Grey line is an exponential fit (P = 0.0002, R^2 = 0.971).

.

Comparisons of the gene expression response across species

The proportion of shared genes responding to drying for a species pair ranged from 22% (*Q. kellogii* and *Q. lobata*) to just 1% (*Q. palmeri* compared to multiple species, Figure 1.2, Figure S1). We focused on testing whether species were more similar to the same-section species, as expected under neutral evolution, or to an ecologically similar species from a different section. At the level of individual genes, we found that four out of six species had gene expression responses that were more similar to the ecologically similar species than to the same-section species, consistent with parallel evolution of gene expression responses.

When comparing similarity in functional types of genes that were upregulated across species, similarities were generally higher; with the most similar species sharing 43% of their upregulated GO terms (Figure 1.3, Figure S1). Three species were more similar to the ecologically similar species than the same-section species, and of these species, similarity was higher for GO terms than for individual genes. However, similarity in the downregulated GO terms among species was lower; with the most similar species sharing only 11% of the downregulated GO terms. Four species had downregulated GO expression that was most similar to the ecologically similar species.

The similarity in gene expression response among two species was not related to the similarity in their average climate, and in fact some of the most similar species pairs came from contrasting climates (Figure S3). Species with similar responses did have relatively high overlap in their ranges (Figure S2); but the species with the highest overlap in ranges (*Q. agrifolia* and *Q. kelloggii*) were not particularly similar in their gene expression response (Figure S1). We used a redundancy analysis to further explore the relationship among functional traits, gene expression, climate, and phylogeny.

Figure 1.3. Testing whether a species is more similar to its same-section species, as expected under neutral evolution, or to its functionally similar species, as expected under parallel or convergent evolution in the response to drying stress. Similarity was measured using the Jaccard index and calculates the overlap between two species (the intersection divided by the union) for either the genes that were significantly differentially expressed or for the GO terms that were significantly enriched in the set of upregulated or downregulated genes. A '*' indicates that there is support for parallelism or convergence, because that species has a response that overlaps more with the functionally similar species than with the same-section species (i.e. 'same section' similarity < 'same functional group' similarity). Brackets on right indicate species pairs with same functional type that were compared to each other. Similarity values for all possible species pairs are shown in Figure S1.

Redundancy Analysis

The redundancy analysis partitioned the interspecific variation in functional traits and the gene expression response to drying into that explained by phylogeny alone, climate alone, or their joint influence (Table 1.1). We found that a species' climate explained equally as much or more variation as did phylogeny for all the phenotype sets tested: functional traits, the response to drying across all genes, and the response to drying for the top DE genes. The overall amount of variation explained (R^2 of the full model, Table 1.1) varied between the two gene expression sets, with climate and phylogeny explaining more variation in the response of the top DE genes than they did in the dataset containing all genes. The degree to which the influence of climate and phylogeny could be separated also varied among the functional traits and gene expression: the joint influence of climate and phylogeny explained little to none of the variation among species in gene expression, but explained 26% of the variation in functional traits.

Identification of frontloaded and evolutionarily diverged genes

We identified frontloaded genes as those that were upregulated in a given drought-sensitive species, non-DE in given drought-tolerant species, and more highly expressed in the drought-tolerant species under control conditions (Supplementary Table S3). This produced a set of genes for each species pair which were constitutively highly expressed in one species in comparison to the other species. For example, 149 genes were upregulated as a drought response in *Q. lobata* (drought-sensitive) and were more highly expressed in *Q. palmeri* (drought-tolerant) under the non-stressful control conditions. Generally, there was little overlap among the gene sets for each species pair; however, we identified 8 genes that were frontloaded in the most drought-tolerant species *Q. palmeri* when compared to the two least drought-tolerant species, *Q. kelloggii* and *Q. lobata*. These genes were annotated with GO terms that included protein kinase activity, DNA-binding transcription factor activity, and metal ion binding.

Using the expression variance and evolution (EVE) model (Rohlfs and Nielsen 2015), we

Table 1.1. Results of redundancy analysis partitioning the effects of climate (species mean) and phylogeny (phylogenetic distance) on the variance in drought traits and gene expression difference in the control and branch-drying treatment for an individual plant. The joint influence of phylogeny and climate explains the portion of the variation that cannot be disentangled, and is calculated by subtracting the variance explained by phylogeny alone and climate alone from the total variance explained by the full model; *P*-values are not given because it is not testable.

	Functional traits				**All genes**				**Top DE genes**			
	df	Adj. R^2	p		df	Adj. R^2	p		df	Adj. R^2	p	
full model	5	0.50	0.001	***	5	0.09	0.001	***	5	0.40	0.001	***
phylogeny alone	2	0.11	0.016	*	2	0.04	0.035	*	2	0.15	0.001	***
climate alone	3	0.12	0.030	*	3	0.06	0.008	**	3	0.24	0.001	***
phylogeny and climate	-	0.26	-	-	-	0	-	-	-	0.01	-	-

identified genes with strong divergence in expression levels across species, and then identified genes within this set that had expression levels correlated with drought tolerance across species (Figure 1.4, Supplementary Table S4). Of the 169 genes in the top 1%, 26 had expression levels that were significantly correlated with drought tolerance (as measured with π_{TLP}). Of these, 12 were more highly expressed in the drought-tolerant species, and 14 were more highly expressed in drought-sensitive species. Many of these correlations seemed to be driven by divergent expression in *Q. lobata*, the least drought-tolerant species; in Figure 1.4, we present correlations for which variation among all six species is more evident. This set of genes included several genes related to signaling had higher expression in the drought-tolerant species, including QL03p063179 (a gene annotated with the GO terms protein kinase activity; ATP binding; and protein phosphorylation), QL02p061158 (DNA-binding transcription factor activity and regulation of transcription, DNA-templated), and QL05p003384 (SAGA complex and transcription coregulator activity). These genes may enable drought tolerance through their higher baseline expression rather than a plastic responses to drought. QL05p055903 (microtubule motor activity; microtubule-based movement; ATP binding; and microtubule binding) was one of the genes with higher expression in drought-sensitive species.

Discussion

Parallelism and convergence in gene expression responses to drying

When comparing how species differed in their transcriptomic responses to drought, we compared ecologically similar species across different sections and identified instances of parallelism, in which these species responded to drying with the same genes; and convergence, in which they responded through genes with similar functions. We found evidence of parallelism in the transcriptomic response to drought in four out of six species: all four tree species shared more differentially expressed (DE) genes with their distantly-related, ecologically similar species than

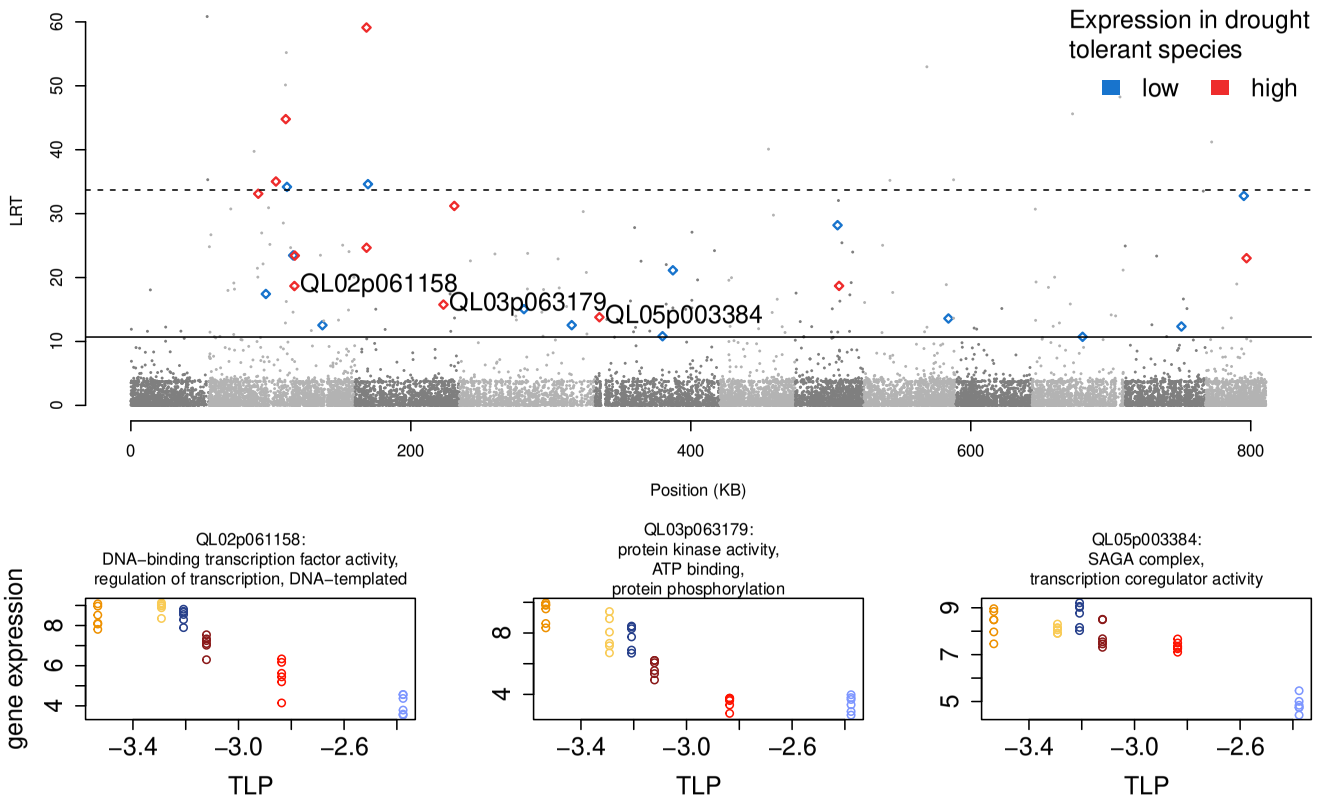


Figure 1.4. (A) Manhattan plot with the likelihood ratio on the y-axis and genomic position on the x-axis. Each point is a gene; dark and light grey colors represent different chromosomes. Genes with higher likelihood ration test (LRT) values have expression levels that are more divergent among species. Solid line represents the 99th percentile, dashed line represents the 99.9th percentile. Diamonds indicate genes that have LRT values in the top 1% and whose gene expression is correlated with drought tolerance (via turgor loss point). Blue diamonds have lower expression in more drought tolerant species and red diamonds have higher expression in more drought tolerant species. (B-D) Selected genes that are more highly expressed in drought-tolerant species. While some genes correlated with drought tolerance were more strongly differentiated among species (indicated by a higher LRT), many of these relationships were primarily driven by high or low expression in *Q. lobata* compared to all other species, so here we present genes for which variation among all six species is more evident. Titles give the gene name and associated GO terms.

with a closely related species (Figure 1.3). This finding supports the hypothesis that natural selection has acted similarly on gene expression responses as these four tree species independently adapted to similar conditions. However, the drought-tolerant shrub species *(Q. durata* and *Q. palmeri)* did not show evidence of a parallel response, and were more similar to a closely-related species than to each other. These species may have evolved different responses to drought, and *Q. durata* is a serpentine soil specialist (Nixon 2002), which may be a larger factor in its divergence from other species than drought tolerance.

In addition to responding with some of the same genes, ecologically-similar species often responded to drought stress using genes with similar functions, consistent with convergence. In fact, for GO terms enriched within upregulated genes, two of the three ecologically-similar species pairs were more similar at the level of GO terms than at the level of individual genes (Figure 1.3), supporting the hypothesis that repeated evolution of traits is more likely to occur through evolution at higher organizational levels (Agrawal 2017). Interestingly, the pair of deciduous tree species, which had the highest level of parallelism in DE genes, also showed the highest level of convergence compared to other species pairs. However, we also note that some species that were neither closely related nor functionally similar also showed high overlap in the types of genes that responded to drying, such as the similarity in upregulated GO terms between *Q. durata*, a drought-tolerant evergreen shrub, and *Q. kelloggii*, a drought-sensitive deciduous tree (Figure S1B), so patterns of gene expression are complex and are also determined by factors other than those tested here.

In contrast to the upregulated genes, genes that were downregulated in response to drying had fewer GO terms shared across species. In response to stress, plants typically upregulate drought-specific responses while downregulating their normal processes such as growth and photosynthesis (Chaves et al. 2003), so the lack of convergence in downregulated functions may be driven by species differences under non-stressful conditions. This hypothesis was supported by the GO enrichment analysis: multiple species upregulated GO terms related to drought responses,

including "response to water" and signaling-related terms such as "DNA-binding transcription factor activity" and "protein ubiquitination." In contrast, enriched GO terms within the down-regulated genes were usually species-specific and included terms related to photosynthesis and chloroplasts, transport, cell walls, and mitochondria (Table S5 and S6).

Parallel selection on gene expression appears to be present in oaks, but the majority of genes used in drought response are species-specific, and the level of parallelism in gene expression identified in this study is somewhat lower than that found in similar studies. When comparing the gene expression of two conifer species under a range of conditions, 74% of the orthologs with plastic responses showed similar expression patterns in both species (Yeaman et al. 2014). This higher level of parallelism may be because conifers generally have slower gene expression divergence than angiosperms (Bouzid et al. 2018, Walia et al. 2009, Yeaman et al. 2014). Overall these results suggest that oaks may evolve similar responses to stress through evolutionary changes in different genes. This convergence among species is consistent with previous research, such as population-level studies on oaks, which have found that populations with the same ecophysiological responses to drought used different gene networks in their drought response (Mead et al. 2019).

Climate shapes gene expression responses more than phylogeny

While seemingly parallel or convergent traits may be the result of neutral evolution, a relationship with climate can provide evidence that traits have evolved in response to similar selective pressures on multiple taxa (Losos 2011). We tested the extent to which climatic niche, phylogenetic history, and their joint influence statistically explain interspecific variation in traits and gene expression using a redundancy analysis. Under neutral evolution or adaptation constrained by phylogenetic history, phylogeny should explain more interspecific variation in the gene expression response than does climate. Conversely, our results showed that climate explained equally as much or more interspecific variation as phylogeny did (Table 1.1), consistent with closely re-

lated oak species adapting to contrasting climates by evolving traits beneficial to that climate. Additionally, climate and phylogeny explained more variation in gene expression in the dataset containing the most plastic genes than they did in the dataset containing all genes, consistent with the hypothesis that adaptations frequently arise through evolutionary changes in gene regulation (Jones et al. 2012, Stern 2013).

Evolutionary constraint may have a greater effect on phenotypic traits than on gene expression in these species. The joint influence of climate and phylogeny explained little to none of the variation among species in gene expression, but explained 26% of the variation in phenotypic traits (Table 1.1). The effects of climate and phylogeny on trait variation are difficult to disentangle in this dataset, which may mean that the evolution of traits is more limited by a species' evolutionary history than the evolution of gene expression. This result agrees with the results of previous studies on oak traits. The xylem hydraulic vulnerability of western North American oaks is phylogenetically conserved and species within clades tended to occupy similar climatic niches, perhaps because evolutionary changes in hydraulic traits require the coordinated evolution of a large suite of traits (Skelton et al. 2018). Medeiros (2021) identified groups of functionally related traits in California oak species, and found that including evolutionary history in their analysis resulted in stronger coordination among traits, suggesting that evolutionary history can have an important effect on the evolution of these traits. Our results suggest that gene expression is more evolutionarily labile than the evolution of functional traits, particularly in the most plastic genes. This accords with the hypothesis that plastic traits are more evolvable and thus more likely to evolve independently in separate lineages (Blomberg et al. 2003, Wund 2012).

While climate does explain some of the variation in the multivariate gene expression response to drought across species, it did not explain pairwise species similarities: the species pairs with the most similar responses did not come from similar climates (Figure S3), unlike another study which found that tree species with similar gene expression responses to drought tended to co-occur within a forest plot (Swenson et al. 2017). The fact that two oak species occur in similar

climates today does not mean that they will necessarily converge on the same traits; the relationship among the environment, traits, and fitness is complex and our results must be interpreted with an understanding of species' biology and evolutionary history. While species with shared responses to leaf drying did not typically occur in similar climates, they did often have similar functional traits. This observation underscores that drought adaptation is multifaceted; species' functional traits may be more important in determining their transcriptomic drought response than climate because traits can modulate the stress experienced at the tissue and cell level for a given climate, and there are multiple possible combinations of traits allowing adaptation to the same conditions (Sack and Buckley 2020). Additionally, species that currently co-occur do not necessarily have similar evolutionary histories, and historical patterns of range expansions or contractions (Gugger et al. 2013), habitat suitability (Ortego et al. 2015), or adaptational lag (Browne et al. 2019) can affect current genetic patterns. For example, *Q. agrifolia* and *Q. lobata* occur in similar climates, often at the same sites, but have contrasting traits and gene expression responses. *Q. lobata* has more drought-sensitive traits despite occurring in relatively hot, dry climates; this is probably because it is adapted to cooler conditions than it experiences in its current range (Browne et al. 2019), and it survives in these dry regions by colonizing microhabitats such as valleys where groundwater reserves can be accessed with its deep roots (Mahall et al. 2009). These characteristics of *Q. lobata* are consistent with drought avoidance, while *Q. agrifolia* may be better able to tolerate lower water potentials. Indeed, it is common for co-occurring oaks in western North America to have contrasting leaf traits, potentially allowing co-existence through complementary resource acquisition strategies (Cavender-Bares et al. 2018).

Drought-tolerant species have less-plastic gene expression responses to drying

Previous studies have found that plasticity in gene expression, i.e. the number of genes that respond to a treatment and the magnitude of their response, can be a major factor contributing to adaptation to environmental stresses (Kenkel and Matz 2016). However, here we find that drought

tolerance is associated with lower plasticity in gene expression. This finding suggests that the traits of more drought-tolerant species allow them to survive dehydration without significantly altering their gene expression, while drought-sensitive species must respond to drought through plasticity mediated by gene expression. This result also suggests a potential explanation for the patterns of parallelism and convergence we observed across species (see discussion above). The species that had the most similar gene expression response were the least drought-tolerant species pair, the deciduous trees; followed by the moderately drought-tolerant evergreen trees (Figure 1.3, Figure S1A). However, the drought-tolerant shrubs did not have similar responses. While in some cases, parallel and convergent evolution may have enabled different species to evolve similar drought responses, similar selection pressures on drought-tolerant species may not have acted on variation in gene expression plasticity if it is not a major contributor to drought tolerance.

Instead, drought tolerance in these oak species may have evolved primarily through other mechanisms, such as mutations affecting protein-coding regions, or changes in gene expression patterns that do not respond to the environment within the time frame tested here. We identified two groups of genes whose gene expression patterns may contribute to drought tolerance through mechanisms besides plasticity: genes that were frontloaded in drought-tolerant species and genes whose expression levels had evolutionarily diverged across species and were also associated with drought tolerance. Within these two sets of genes, there were several genes with functions related to signaling. Signaling genes could be beneficial as frontloaded and/or constitutively highly expressed genes in drought-tolerant species, as they may control networks of genes with related functions, resulting in a large overall phenotypic effect (Des Marais and Juenger 2010, Todaka et al. 2015; but see Des Marais et al. 2017). It is possible that the lack of plasticity in drought-tolerant species evolved from an ancestrally plastic state through canalization (Rivera et al. 2021), however, in this study we can not infer the ancestral state of gene expression to evaluate when shifts may have occurred.

The limited number of studies testing whether differential gene expression under stress

corresponds to stress tolerance across taxa do not follow a general trend (Bittner et al. 2021, Komoroske et al. 2021, Sandoval-Castillo et al. 2020). One study in house mice found that a desert population responded to dehydration with fewer differentially expressed genes compared to a non-desert population, and the gene expression of non-desert mice became more similar to desert mice under water stress, indicative of adaptive plasticity in response to dehydration (Bittner et al. 2021). However, several studies in fish have found that heat-tolerant species had higher gene expression plasticity in response to heat stress than heat-sensitive species (Komoroske et al. 2021, Sandoval-Castillo et al. 2020). While the degree of expression plasticity measured in different taxa was likely influenced by the specific experimental conditions, it may also indicate variation in the stress tolerance mechanisms evolved by the taxa. For increased plasticity to be beneficial, the organism must be able to accurately detect and respond to the environmental stress with minimal lag time, and there must be a cost to expressing the stress-induced phenotype under benign conditions (DeWitt et al. 1998). If these conditions are not met, stress tolerance is more likely to evolve through adaptations other than plasticity.

The importance of gene expression in adaptation

Understanding the mechanisms through which stress tolerance evolves is crucial to predicting evolutionary responses to climate change. Our results point to gene expression as a potentially important evolutionary target for selection on drought response in oaks. Oak species that have a strong plastic response to drought evolved similar responses through changes in both shared and different genes. While it is difficult to make conclusions about the overall "evolvability" of the drought response, our results showing the primarily species-specific responses to drought; the presence of parallelism among deeply diverged, but ecologically-similar species; and the increasing parallelism in functional groups of genes suggest that there are multiple mechanisms through which oaks may evolve drought responses. If the drought response is polygenic and similar responses can occur through changes in the expression of many different genes, as suggested

by our results, is is more likely that beneficial standing genetic variation in drought response is already present in populations, potentially allowing population persistence under climate change. However, short-term plasticity in gene expression does not seem to be the primary mechanism explaining drought tolerance among species, because the most drought-tolerant species responded with fewer genes than the drought-sensitive species. Instead, we find evidence that drought tolerance is associated with other evolutionary changes in the expression patterns of genes, such as frontloading and changes in baseline gene expression.

Acknowledgements

We thank the California Botanic Garden for allowing us to sample, and members of the Sork and Sack labs for helping with sampling and providing useful feedback on the manuscript. Genomic analyses and A.M. were supported by National Science Foundation Plant Genome Research Program (NSF IOS-#1444661). C.M. was supported by the Brazilian National Research Council (CNPq) through the Brazilian Science Without Borders Program (grant number: 202813/2014-2).

References

Abrams, M. D. (1990). Adaptations and responses to drought in *Quercus* species of North America. *Tree Physiology* 7(1-2-3-4):227–238.

Agrawal, A. A. (2017). Toward a predictive framework for convergent evolution: integrating natural history, genetic mechanisms, and consequences for the diversity of life. *The American Naturalist* 190(S1):S1–S12.

Arendt, J. and D. Reznick (2008). Convergence and parallelism reconsidered: what have we learned about the genetics of adaptation? *Trends in Ecology & Evolution* 23(1):26–32.

Bailey, S. F., N. Rodrigue, and R. Kassen (2015). The effect of selection environment on the probability of parallel evolution. *Molecular Biology and Evolution* 32(6):1436–1448.

Barrett, R. D. H. and D. Schluter (2008). Adaptation from standing genetic variation. *Trends in Ecology & Evolution* 23(1):38–44.

Barshis, D. J., J. T. Ladner, T. A. Oliver, F. O. Seneca, N. Traylor-Knowles, and S. R. Palumbi (2013). Genomic basis for coral resilience to climate change. *Proceedings of the National Academy of Sciences* 110(4):1387–1392.

Bartlett, M. K., C. Scoffoni, and L. Sack (2012). The determinants of leaf turgor loss point and prediction of drought tolerance of species and biomes: A global meta-analysis. *Ecology Letters* 15(5):393–405.

Benjamini, Y. and Y. Hochberg (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)* 57(1):289–300.

Bittner, N. K. J., K. L. Mack, and M. W. Nachman (2021). Gene expression plasticity and desert adaptation in house mice. *Evolution* 75(6):1477–1491.

Blomberg, S. P., T. Garland, and A. R. Ives (2003). Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution* 57(4):717–745.

Borcard, D., F. Gillet, and P. Legendre (2011). Canonical Ordination. *In* D. Borcard, F. Gillet, and P. Legendre (editors), Numerical Ecology with R, Use R, 153–225. Springer, New York, NY.

Bouzid, M., F. He, G. Schmitz, R. Haeusler, A. Weber, T. Mettler, and J. d. Meaux (2018). *Arabidopsis* species deploy distinct strategies to cope with drought stress. *bioRxiv* 341859.

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

Cavender-Bares, J., S. Kothari, J. E. Meireles, M. A. Kaproth, P. S. Manos, and A. L. Hipp (2018). The role of diversification in community assembly of the oaks (*Quercus* L.) across the continental U.S. *American Journal of Botany* 105(3):1–22.

Cavender-Bares, J. (2018). Diversification, adaptation, and community assembly of the American oaks (*Quercus*), a model clade for integrating ecology and evolution. *New Phytologist* 221(2):669–692.

Cavender-Bares, J., D. D. Ackerly, D. A. Baum, and F. A. Bazzaz (2004). Phylogenetic overdispersion in Floridian oak communities. *The American Naturalist* 163(6):823–843.

Cavender-Bares, J., C. G. Fontes, and J. Pinto-Ledezma (2020). Open questions in understanding the adaptive significance of plant functional trait variation within a single lineage. *New Phytologist* 227(3):659–663.

Chaves, M. M., J. P. Maroco, and J. S. Pereira (2003). Understanding plant responses to drought — from genes to the whole plant. *Functional Plant Biology* 30(3):239–264.

Conte, G. L., M. E. Arnegard, C. L. Peichel, and D. Schluter (2012). The probability of genetic parallelism and convergence in natural populations. *Proceedings of the Royal Society B: Biological Sciences* 279(1749):5039–5047.

Cordell, S., G. Goldstein, D. Mueller-Dombois, D. Webb, and P. M. Vitousek (1998). Physiological and morphological variation in *Metrosideros polymorpha*, a dominant Hawaiian tree species, along an altitudinal gradient: the role of phenotypic plasticity. *Oecologia* 113(2):188–196.

Des Marais, D. L., R. F. Guerrero, J. R. Lasky, and S. V. Scarpino (2017). Topological features of a gene co-expression network predict patterns of natural diversity in environmental response. *Proceedings of the Royal Society B: Biological Sciences* 284(1856):20170914.

Des Marais, D. L. and T. E. Juenger (2010). Pleiotropy, plasticity, and the evolution of plant abiotic stress tolerance. *Annals of the New York Academy of Sciences* 1206(1):56–79.

Des Marais, D. L., J. K. Mckay, J. H. Richards, S. Sen, T. Wayne, T. E. Juenger, D. L. D. Marais, J. K. Mckay, J. H. Richards, S. Sen, T. Wayne, and T. E. Juenger (2012). Physiological genomics of response to soil drying in diverse *Arabidopsis* accessions. *The Plant Cell* 24(March):893–914.

DeWitt, T. J., A. Sih, and D. S. Wilson (1998). Costs and limits of phenotypic plasticity. *Trends in Ecology & Evolution* 13(2):77–81.

Dobin, A., C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, and T. R. Gingeras (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29(1):15–21.

Domingues, T. F., P. Meir, T. R. Feldpausch, G. Saiz, E. M. Veenendaal, F. Schrodt, M. Bird, G. Djagbletey, F. Hien, H. Compaore, A. Diallo, J. Grace, and J. Lloyd (2010). Co-limitation of photosynthetic capacity by nitrogen and phosphorus in West Africa woodlands. *Plant, Cell & Environment* 33(6):959–980.

Dunbar-Co, S., M. Sporck, and L. Sack (2009). Leaf trait diversification and design in seven rare taxa of the Hawaiian *Plantago* radiation. *International Journal of Plant Sciences* 170(1):61–75.

Elmer, K. R. and A. Meyer (2011). Adaptation in the age of ecological genomics: Insights from parallelism and convergence. *Trends in Ecology and Evolution* 26(6):298–306.

Fletcher, L. R., H. Cui, H. Callahan, C. Scoffoni, G. P. John, M. K. Bartlett, D. O. Burge, and L. Sack (2018). Evolution of leaf structure and drought tolerance in species of Californian *Ceanothus*. *American Journal of Botany* 105(10):1672–1687.

Franks, P. J., P. L. Drake, and D. J. Beerling (2009). Plasticity in maximum stomatal conductance constrained by negative correlation between stomatal size and density: an analysis using *Eucalyptus globulus*. *Plant, Cell & Environment* 32(12):1737–1748.

Gasulla, F., K. vom Dorp, I. Dombrink, U. Zähringer, N. Gisch, P. Dörmann, and D. Bartels (2013). The role of lipid metabolism in the acquisition of desiccation tolerance in *Craterostigma plantagineum*: a comparative approach. *The Plant Journal* 75(5):726–741.

Givnish, T. J. and R. A. Montgomery (2014). Common-garden studies on adaptive radiation of photosynthetic physiology among Hawaiian lobeliads. *Proceedings of the Royal Society B: Biological Sciences* 281(1779):20132944.

Gompel, N. and B. Prud'homme (2009). The causes of repeated genetic evolution. *Developmental Biology* 332(1):36–47.

Gugger, P. F., M. Ikegami, and V. L. Sork (2013). Influence of late Quaternary climate change on present patterns of genetic variation in valley oak, *Quercus lobata* Née. *Molecular Ecology* 22(13):3598–3612.

Hengl, T., J. M. d. Jesus, G. B. M. Heuvelink, M. R. Gonzalez, M. Kilibarda, A. Blagotić, W. Shangguan, M. N. Wright, X. Geng, B. Bauer-Marschallinger, M. A. Guevara, R. Vargas, R. A. MacMillan, N. H. Batjes, J. G. B. Leenaars, E. Ribeiro, I. Wheeler, S. Mantel, and B. Kempen (2017). SoilGrids250m: Global gridded soil information based on machine learning. *PLOS ONE* 12(2):e0169748.

Hijmans, R. J., S. E. Cameron, J. L. Parra, P. G. Jones, and A. Jarvis (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* 25(15):1965–1978.

Hijmans, R. J., S. Phillips, J. Leathwick, and J. Elith (2020). dismo: Species Distribution Modeling. `https://CRAN.R-project.org/package=dismo`

Hipp, A. L., P. S. Manos, A. González-Rodríguez, M. Hahn, M. Kaproth, J. D. McVay, S. V. Avalos, and J. Cavender-Bares (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., P. S. Manos, M. Hahn, M. Avishai, C. Bodénès, J. Cavender-Bares, A. A. Crowl, M. Deng, T. Denk, S. Fitz-Gibbon, O. Gailing, M. S. González-Elizondo, A. González-Rodríguez, G. W. Grimm, X.-L. Jiang, A. Kremer, I. Lesur, J. D. McVay, C. Plomion, H. Rodríguez-Correa, E.-D. Schulze, M. C. Simeone, V. L. Sork, and S. Valencia-Avalos (2020). Genomic landscape of the global oak phylogeny. *New Phytologist* 226(4):1198–1212.

Hoekstra, F. A., E. A. Golovina, and J. Buitink (2001). Mechanisms of plant desiccation tolerance. *Trends in Plant Science* 6(9):431–438.

Jones, F. C., M. G. Grabherr, Y. F. Chan, P. Russell, E. Mauceli, J. Johnson, R. Swofford, M. Pirun, M. C. Zody, S. White, E. Birney, S. Searle, J. Schmutz, J. Grimwood, M. C. Dickson, R. M. Myers, C. T. Miller, B. R. Summers, A. K. Knecht, S. D. Brady, H. Zhang, A. A. Pollen, T. Howes, C. Amemiya, Broad Institute Genome Sequencing Platform & Whole Genome Assembly Team, E. S. Lander, F. Di Palma, K. Lindblad-Toh, and D. M. Kingsley (2012). The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* 484(7392):55–61.

Kaproth, M. and J. Cavender-Bares (2016). Drought tolerance and climatic distributions of the American oaks. *International Oaks* 27:49–60.

Kenkel, C. D. and M. V. Matz (2016). Gene expression plasticity as a mechanism of coral adaptation to a variable environment. *Nature Ecology & Evolution* 1(3):0014.

Khaitovich, P., W. Enard, M. Lachmann, and S. Pääbo (2006). Evolution of primate gene expression. *Nature Reviews Genetics* 7(9):693–702.

King, M.-C. and A. C. Wilson (1975). Evolution at Two Levels in Humans and Chimpanzees. *Science* 188(4184):107–116.

Komoroske, L. M., K. M. Jeffries, A. Whitehead, J. L. Roach, M. Britton, R. E. Connon, C. Verhille, S. M. Brander, and N. A. Fangue (2021). Transcriptional flexibility during thermal challenge corresponds with expanded thermal tolerance in an invasive compared to native fish. *Evolutionary Applications* 14(4):931–949.

Lind, B. M., M. Menon, C. E. Bolte, T. M. Faske, and A. J. Eckert (2018). The genomics of local adaptation in trees: are we out of the woods yet? *Tree Genetics & Genomes* 14(2):29.

Lobo, A., J. M. Torres-Ruiz, R. Burlett, C. Lemaire, C. Parise, C. Francioni, L. Truffaut, I. Tomášková, J. K. Hansen, E. D. Kjær, A. Kremer, and S. Delzon (2018). Assessing inter- and intraspecific variability of xylem vulnerability to embolism in oaks. *Forest Ecology and Management* 424:53–61.

Losos, J. B. (2011). Convergence, adaptation, and constraint. *Evolution* 65(7):1827–1840.

Love, M. I., W. Huber, and S. Anders (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* 15(12):1–21.

Mahall, B. E., C. M. Tyler, E. S. Cole, and C. Mata (2009). A comparative study of oak (*Quercus*, Fagaceae) seedling physiology during summer drought in southern California. *American Journal of Botany* 96(4):751–761.

Martin, A. and V. Orgogozo (2013). The loci of repeated evolution: A catalog of genetic hotspots of phenotypic variation. *Evolution* 67(5):1235–1250.

Mason, C. M. and L. A. Donovan (2015). Evolution of the leaf economics spectrum in herbs: Evidence from environmental divergences in leaf physiology across *Helianthus* (Asteraceae). *Evolution* 69(10):2705–2720.

Mead, A., J. Peñaloza Ramirez, M. K. Bartlett, J. W. Wright, L. Sack, and V. L. Sork (2019). Seedling response to water stress in valley oak (*Quercus lobata*) is shaped by different gene networks across populations. *Molecular Ecology* 28(24):5248–5264.

Medeiros, C. D. B. (2021). Chapter 3: Evolution of trait modules across California native oaks. Ph.D. thesis, UCLA.

Mohler, C. L. (1990). Co-occurrence of oak subgenera: implications for niche differentiation. *Bulletin of the Torrey Botanical Club* 117(3):247–255.

Nixon, K. C. (2002). The oak (*Quercus*) biodiversity of California and adjacent regions. Technical report, USDA Forest Service, San Diego, CA. `http://www.treesearch.fs.fed.us/pubs/26105`

Oksanen, J., F. G. Blanchet, M. Friendly, R. Kindt, P. Legendre, D. McGlinn, P. R. Minchin, R. B. O'Hara, G. L. Simpson, P. Solymos, M. H. H. Stevens, E. Szoecs, and H. Wagner (2019). vegan: Community Ecology Package. `https://CRAN.R-project.org/package=vegan`

O'Quin, K. E., C. M. Hofmann, H. A. Hofmann, and K. L. Carleton (2010). Parallel evolution of opsin gene expression in African cichlid fishes. *Molecular Biology and Evolution* 27(12):2839–2854.

Ord, T. J. and T. C. Summers (2015). Repeated evolution and the impact of evolutionary history on adaptation. *BMC Evolutionary Biology* 15(1):137.

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

Pfenninger, M., S. Patel, L. Arias-Rodriguez, B. Feldmeyer, R. Riesch, and M. Plath (2015). Unique evolutionary trajectories in repeated adaptation to hydrogen sulphide-toxic habitats of a neotropical fish (*Poecilia mexicana*). *Molecular Ecology* 24(21):5446–5459.

Pinheiro, C. and M. M. Chaves (2011). Photosynthesis and drought: Can we make metabolic connections from available data? *Journal of Experimental Botany* 62(3):869–882.

Pérez-Harguindeguy, N., S. Díaz, E. Garnier, S. Lavorel, H. Poorter, P. Jaureguiberry, M. S. Bret-Harte, W. K. Cornwell, J. M. Craine, D. E. Gurvich, C. Urcelay, E. J. Veneklaas, P. B. Reich, L. Poorter, I. J. Wright, P. Ray, L. Enrico, J. G. Pausas, A. C. D. Vos, N. Buchmann, G. Funes, F. Quétier, J. G. Hodgson, K. Thompson, H. D. Morgan, H. T. Steege, M. G. A. V. D. Heijden, L. Sack, B. Blonder, P. Poschlod, M. V. Vaieretti, G. Conti, A. C. Staver, S. Aquino, and J. H. C. Cornelissen (2013). New handbook for standardised measurement of plant functional traits worldwide. *Australian Journal of Botany* 61(3):167–234.

R Core Team (2019). R: A Language and Environment for Statistical Computing. `https://www.R-project.org/`

Ramos, A. M., A. Usié, P. Barbosa, P. M. Barros, T. Capote, I. Chaves, F. Simões, I. Abreu, I. Carrasquinho, C. Faro, J. B. Guimarães, D. Mendonça, F. Nóbrega, L. Rodrigues, N. J. M. Saibo, M. C. Varela, C. Egas, J. Matos, C. M. Miguel, M. M. Oliveira, C. P. Ricardo, and S. Gonçalves (2018). The draft genome sequence of cork oak. *Scientific Data* 5(1):180069.

Ramírez-Valiente, J. A., N. J. Deacon, J. Etterson, A. Center, J. P. Sparks, K. L. Sparks, T. Longwell, G. Pilz, and J. Cavender-Bares (2018). Natural selection and neutral evolutionary processes

contribute to genetic divergence in leaf traits across a precipitation gradient in the tropical oak *Quercus oleoides*. *Molecular Ecology* 27(9):2176–2192.

Renaut, S., G. L. Owens, and L. H. Rieseberg (2014). Shared selective pressure and local genomic landscape lead to repeatable patterns of genomic divergence in sunflowers. *Molecular Ecology* 23(2):311–324.

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

Rivera, H. E., H. E. Aichelman, J. E. Fifer, N. G. Kriefall, D. M. Wuitchik, S. J. S. Wuitchik, and S. W. Davies (2021). A framework for understanding gene expression plasticity and its influence on stress tolerance. *Molecular Ecology* 30(6):1381–1397.

Rohlfs, R. V. and R. Nielsen (2015). Phylogenetic ANOVA: The expression variance and evolution model for quantitative trait evolution. *Systematic Biology* 64(5):695–708.

Sack, L. and T. N. Buckley (2020). Trait Multi-Functionality in Plant Stress Response. *Integrative and Comparative Biology* 60(1):98–112.

Salt, D. E., I. Baxter, and B. Lahner (2008). Ionomics and the Study of the Plant Ionome. *Annual Review of Plant Biology* 59(1):709–733.

Sandoval-Castillo, J., K. Gates, C. J. Brauer, S. Smith, L. Bernatchez, and L. B. Beheregaray (2020). Adaptation of plasticity to projected maximum temperatures and across climatically defined bioregions. *Proceedings of the National Academy of Sciences* 117(29):17112–17121.

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

Schurch, N. J., P. Schofield, M. Gierliński, C. Cole, A. Sherstnev, V. Singh, N. Wrobel, K. Gharbi, G. G. Simpson, T. Owen-Hughes, M. Blaxter, and G. J. Barton (2016). How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use? *RNA* 22(10):1641–1641.

Schwanhäusser, B., D. Busse, N. Li, G. Dittmar, J. Schuchhardt, J. Wolf, W. Chen, and M. Selbach (2011). Global quantification of mammalian gene expression control. *Nature* 473(7347):337–342.

Skelton, R. P., T. E. Dawson, S. E. Thompson, Y. Shen, A. P. Weitz, and D. Ackerly (2018). Low vulnerability to xylem embolism in leaves and stems of North American oaks. *Plant Physiology* 177(3):1066–1077.

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

Stern, D. L. (2013). The genetic causes of convergent evolution. *Nature Reviews Genetics* 14(11):751–764.

Swenson, N. G., Y. Iida, R. Howe, A. Wolf, M. N. Umaña, K. Petprakob, B. L. Turner, and K. Ma (2017). Tree co-occurrence and transcriptomic response to drought. *Nature Communications* 8(1):1996.

Tenaillon, O., A. Rodríguez-verdugo, R. L. Gaut, P. Mcdonald, A. F. Bennett, A. D. Long, and B. S. Gaut (2012). The molecular diversity of adaptive convergence. *Science* 335(6067):457–461.

Todaka, D., K. Shinozaki, and K. Yamaguchi-Shinozaki (2015). Recent advances in the dissection of drought-stress regulatory networks and strategies for development of drought-tolerant transgenic rice plants. *Plant Biotechnology* 6:84.

Walia, H., C. Wilson, A. M. Ismail, T. J. Close, and X. Cui (2009). Comparing genomic expression patterns across plant species reveals highly diverged transcriptional dynamics in response to salt stress. *BMC Genomics* 10(1):398.

Wang, J., L. Song, X. Gong, J. Xu, and M. Li (2020). Functions of jasmonic acid in plant regulation and response to abiotic stress. *International Journal of Molecular Sciences* 21(4):1446.

Whitehead, A. and D. L. Crawford (2006). Variation within and among species in gene expression: Raw material for evolution. *Molecular Ecology* 15(5):1197–1211.

Wund, M. A. (2012). Assessing the impacts of phenotypic plasticity on evolution. *Integrative and Comparative Biology* 52(1):5–15.

Yeaman, S., K. A. Hodgins, K. E. Lotterhos, H. Suren, S. Nadeau, J. C. Degner, K. A. Nurkowski, P. Smets, T. Wang, L. K. Gray, K. J. Liepe, A. Hamann, J. A. Holliday, M. C. Whitlock, L. H. Rieseberg, and S. N. Aitken (2016). Convergent local adaptation to climate in distantly related conifers. *Science* 353(6306):1431–1433.

Yeaman, S., K. A. Hodgins, H. Suren, K. A. Nurkowski, L. H. Rieseberg, J. A. Holliday, and S. N. Aitken (2014). Conservation and divergence of gene expression plasticity following c. 140 million years of evolution in lodgepole pine (*Pinus contorta*) and interior spruce (*Picea glauca* × *Picea engelmannii*). *New Phytologist* 203(2):578–591.

Young, M. D., M. J. Wakefield, G. K. Smyth, and A. Oshlack (2010). Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biology* 11(2):R14.

Zomer, R. J., A. Trabucco, D. A. Bossio, and L. V. Verchot (2008). Climate change mitigation: A spatial analysis of global land suitability for clean development mechanism afforestation and reforestation. *Agriculture, Ecosystems & Environment* 126(1):67–80.

Chapter 2

Environment shapes adaptive genetic variation in an island endemic oak species: implications for restoration[1]

Abstract

Quercus tomentella, or island oak, is a rare oak restricted to six of the Channel Islands in California, USA and Baja California, Mexico. Like many island populations, its existence may be threatened by small population sizes and low genetic diversity. Previous work has shown that oaks on each island are genetically differentiated from each other, but less is known about the extent to which island populations are adapted to their local climate conditions. We performed whole-genome sequencing on island oak individuals from across the six islands it inhabits to characterize the genetic variation across its range and assess geographic patterns of climate-associated genetic variation. In addition, to assess the influence of hybridization and introgression, we sequenced mainland and island samples of *Q. chrysolepis*, a closely related species that hybridizes with island oak. The overall goal of this study is to assess whether the underlying genetic composition of each island should be taken into account when developing restoration strategies for *Q. tomentella.* Our analysis of the genetic structure across the landscape for both neutral and putatively adaptive climate-associated alleles revealed evidence of gene flow among some islands but also some

[1]This chapter will be submitted for publication with the following authors: Alayna Mead, John Knapp, Sorel Fitz-Gibbon, and Victoria Sork

genetic isolation among islands, particularly for Guadalupe Island, which is the most geographically isolated. We identify climate-associated SNPs and find that the genetic turnover across space differs between adaptive and neutral alleles in some regions, suggesting that local adaptation to climate has occurred both within and between islands. Using these associations, we predict that populations on Catalina and eastern Santa Cruz islands will be most maladapted to future climates. These results can be used to make specific recommendations on the location of populations to be used as seed sources for restoration projects to maximize adaptation to future climates at restoration sites.

Introduction

The evolutionary history of a species and the populations within it are shaped by their environment, yet intraspecific variation is rarely considered when making decisions about restoration. Geographic patterns of environmental variation and rates of gene flow across the landscape may enable or constrain the ability of populations to adapt to their local conditions. When gene flow occurs more frequently among individuals that are geographically close to each other, a genetic pattern of isolation by distance (IBD) can result, in which genetic distance increases with geographic distance. Conversely, gene flow may be highest among individuals from similar environments, regardless of their geographic distance from each other. This pattern of isolation by environment (IBE) may result from a range of mechanisms, including natural selection against individuals from different environments or against hybrid local/immigrant offspring and biased dispersal in which individuals preferentially disperse to environments that are similar to those of their parents (Sexton et al. 2014, Wang and Bradburd 2014).

In reality, the environmental distance and geographic distance between populations is often strongly correlated, making it difficult to disentangle their contributions to genetic structure. However, landscape genomics methods that sample across many localities and many loci can be used to maximize the range of geographic and environmental distances and minimize their

correlations (Rellstab et al. 2015, Sork et al. 2013, Wang and Bradburd 2014). It is also possible to identify genetic variants that may contribute to local adaptation using genotype-environment association (GEA) analyses, which associate genetic variants with climate and can account for the effects of demographic history on population structure (Sork et al. 2013). Using multivariate GEA analyses, many loci can be analyzed together to better identify weak, polygenic selection across many loci (Forester et al. 2018). Additionally, associations between alleles and environmental gradients can be modeled to predict how allele frequencies vary across the landscape in association with environmental variables (genomic turnover, Fitzpatrick and Keller 2015) and to predict the genomic composition necessary for a population to be well-adapted to future climate (Fitzpatrick et al. 2021, Fitzpatrick and Keller 2015, Ingvarsson and Bernhardsson 2020, Rellstab et al. 2021, 2016). The difference between the current composition and that needed for future climates, termed the genetic offset, is a metric of how much evolutionary change would be necessary to maintain adaptation to the environment under climate change, and can be used to identify the regions of a species range that are at the greatest risk of maladaptation. Landscape genomics can help us understand the geographic and climatic factors that determine genetic structure within a species, and determine the extent of local adaptation, providing insight into the evolutionary history of a species and how it will respond to rapid environmental shifts under climate change.

Forest trees often experience high levels of gene flow across the species range, facilitated by a predominantly outcrossing mating system and long-distance pollen dispersal by wind (Petit and Hampe 2006). Despite this lack of genetic structure, local adaptation is common within forest trees, likely due to the high genetic diversity of a large panmictic species combined with high seedling mortality introducing strong selection against non-locally adapted alleles (Sork 2016, 2017, Sork et al. 2013). Consistent with this, landscape genomic studies on trees have found that geographic distance strongly affects genetic structure, but that environmental variables also play a role, and that candidate genetic variants associated with climate are present (Fitzpatrick and Keller 2015, Gugger et al. 2021, Jia et al. 2020, Martins et al. 2018). However, few studies

have tested for local adaptation in tree species with restricted ranges and isolated populations. In such species, small population sizes and low levels of gene flow among populations may result in lower genetic diversity and a decreased ability to adapt to changing climatic conditions. Studies on tree species with island populations have found genetic differentiation across disjunct populations indicating limited gene flow (Di Santo et al. 2022, Gugger et al. 2018), and island endemic plant species may have lower genetic diversity than mainland congeners (Hamabata et al. 2019). However, the genetic diversity of island species depends on the characteristics and demographic history of the species and varies widely; in fact, in many species, island populations have greater diversity than mainland populations (García-Verdugo et al. 2015).

Despite relatively small population sizes, studies on island species have found signs of local adaptation. In Hawaiian koa, precipitation was an important factor structuring genetic variation across its range (Gugger et al. 2018). In Torrey pine, disjunct island and mainland populations show morphological differences in a common garden that are consistent with the climate differences (Hamilton et al. 2017). These studies show that local adaptation may be possible across species with fragmented ranges, even when low levels of gene flow reduce genetic diversity. Similarly, local adaptation may be present at small scales, such as within islands. In koa populations within the island of Hawaii, genetic variation was associated with precipitation (Gugger et al. 2018). Fine-scale local adaptation has also been described in bird species (Gamboa et al. 2022, Langin et al. 2015). For example, Island Scrub Jays, which are endemic to a single California Channel Island, have bill shapes that are locally adapted to different vegetation types (Cheek et al. 2022, Langin et al. 2015). Ultimately the presence of local adaptation in a given species will depend on many evolutionary and environmental factors, and understanding patterns of neutral and adaptive genetic variation will facilitate conservation efforts for rare or fragmented species.

In this study, we investigate the factors that shape spatial patterns of genetic variation in an island relictual oak, *Quercus tomentella*, which is present on six of the Channel Islands off the coast of California and Mexico and then use these patterns to address the question of how best

to restore island oak to Santa Rosa Island where the distribution has been reduced dramatically by humans. This species is categorized as endangered by the IUCN (Beckman and Jerome 2017), and its threats include grazing from non-native herbivores, erosion, changes in hydrology, and climate change. Previous work using microsatellite markers found that that *Q. tomentella* is genetically differentiated across islands and has lower genetic diversity than its more widespread relative *Q. chrysolepis* (Ashley et al. 2018). Low genetic diversity and geographically distant island populations may constrain the ability of this species to adapt to local conditions or respond to a changing climate. However, adaptive alleles may show different spatial patterns than neutral alleles (Fitzpatrick and Keller 2015, Gugger et al. 2021, Martins et al. 2018). Additionally, this species may have historically had sufficiently large population sizes to avoid low genetic diversity and fixation of deleterious alleles. While its present range is small and fragmented, *Q. tomentella* once had a larger range throughout mainland California. Fossil evidence shows it was present in the western Mojave Desert during the Miocene (25-5.3 Ma) (Axelrod 1939) and the central coast of California during the Pliocene (5.3-2.6 Ma) (Axelrod 1944a,b). It is thought that *Q. tomentella* gradually became restricted to the coast and eventually to the Channel Islands as the mainland climate became less temperate (Axelrod 1967, Muller 1965). During recent glacial periods with lower sea levels, all islands were once closer to the mainland, and the northern Channel Islands were part of the same landmass as recently as 11,000 years ago (Kennett et al. 2008, Reeder-Myers et al. 2015). During this time, there may have been more gene flow among island populations, as well as between island and mainland populations, if they coexisted. Hybridization may provide another source of genetic variation: *Q. tomentella* hybridizes with *Q. chrysolepis*, a widespread species on mainland California that is also present on the islands, and genomic and fossil data support a history of co-occurrence and ancient introgression between the two species (Axelrod 1944a,b, Ortego et al. 2018).

In this study, we use whole-genome resequencing data to analyze the processes structuring genetic variation, including local adaptation, in *Q. tomentella* and provide a foundation for

developing conservation strategies for maintaining and restoring populations on the Channel Islands. Our specific goal is to characterize the genetic structure of *Q. tomentella* at both neutral and putatively adaptive climate-associated alleles, and make predictions about maladaptation under future climate change. Based on these analyses, we will identify which islands are at greater risk from climate change, and which populations would be best used as seed sources for a proposed restoration project on Santa Rosa Island. This study will illustrate of how landscape genomic analyses can inform conservation strategies.

Methods

Collections

Leaf samples from *Q. tomentella* individuals were collected across the six Channel Islands encompassing the species range. Because *Q. tomentella* hybridizes with *Q. chrysolepis*, we also collected co-occurring *Q. chrysolepis* and putative hybrid individuals from the islands, and *Q. chrysolepis* individuals from across mainland California. As *Q. tomentella* can reproduce clonally, we attempted to avoid collecting multiple samples from the same clone by collecting from only one stem within the same grove or cluster. Leaf samples were either dried on silica or were placed on ice while transported to the Sork lab, then were frozen at -80° C. In total, we sequenced and analyzed data for 107 *Q. tomentella* individuals, 17 *Q. chrysolepis* individuals (8 from mainland California and 9 from the islands), and 3 putative hybrids.

DNA extraction and sequencing

Leaves were hand-ground on liquid nitrogen using a mortar and pestle, and approximately 50 mg of tissue was used for DNA extraction. DNA was extracted from leaves using a modified version of the Qiagen DNeasy Plant Mini Kit protocol. First, to remove polyphenols, a prewash step was performed twice. 1 mL of prewash buffer was added to ground leaf tissue, ground in

a bead mill for 20 seconds, centrifuged for 10 minutes at 10,000 RPM, and the supernatant was discarded. Prewash buffer consisted of (per sample) 100 ul Tris, 100 ul EDTA, 200 ul 5 M NaCl, 600 ul molecular grade water, and 0.01 g PVP. Following the prewash, the Qiagen protocol was followed. For the silica-dried leaf samples (from Guadalupe and San Clemente islands) we were unable to extract sufficient amounts of DNA, so extraction was performed by the CCGP mini-core using the Macherey-Nagel NucleoMag Plant kit, with the following modifications: PVP and Proteinase K were added during the digest, and an additional ethanol wash was added before elution. Extracted DNA was sent to UC Davis DNA Technologies and Expression Analysis Cores for library preparation using a custom SeqWell kit. Whole-genome sequencing was performed on a NovaSeq 6000 using 150 bp, paired-end sequencing.

Filtering and variant calling

Adapters were trimmed from raw reads using Trim Galore, and reads with a length less than 20 bp were removed. Reads were not trimmed based on quality scores during this step. Reads were aligned to the *Q. lobata* genome (Sork et al. 2022) using BWA-MEM, with 'markShorterSplits' and 'readGroupHeaderLine' options enabled. Duplicate reads were marked and removed using GATK MarkDuplicates. Variants were called using GATK HaplotypeCaller with the 'emit-ref-confidence' option set to 'GVCF'. Variants were hard-filtered using GATK VariantFiltration, with SNPs and indels filtered separately. For SNPs, we removed variants with quality by depth (QD) <2, quality (QUAL) <30, mapping quality (MQ) < 40, phred-scaled strand bias (FS) >60, symmetric odds ratio strand bias (SOR) >3, mapping quality rank sum (MQRankSum) < -12.5, and read position rank sum (ReadPosRankSum) < -8. We removed indels with QD<2, FS>200, QUAL<30, and ReadPosRankSum< -20. Repetitive regions of the genome were removed using vcftools based on the reference genome.

From this set of high-quality variants, we selected biallelic SNPs with high coverage across all samples for further analysis. Using bcftools (version 1.15.1, Danecek et al. 2021), we selected

only biallelic SNPs, and removed SNPs with a mean depth across all samples <5, set individual genotypes with depth <5 to missing, removed SNPs with a minor allele frequency <0.01, and removed SNPs with ≥90% missingness across all individuals. The resulting filtered VCF file was converted to BED file format using PLINK (version 1.90b6.26, Chang et al. 2015), and variants in linkage disequilibirum (LD) were pruned using a window size of 50 variants, a window shift value of 10 variants, and an R^2 threshold of 0.1, which identifies variant pairs with a correlation >0.1 within the given window and prunes them until no correlated pairs remain. Unless otherwise noted, analyses were run on this filtered and LD-pruned dataset of 585,298 SNPs. For analyses requiring no missing data, SNPs were imputed by assigning missing individuals the most common SNP (total missingness in the filtered dataset was 5%).

Climate data

Climate variables were extracted for each locality from WorldClim (version 2.1, historic climate data for 1970-2000) at 30-second resolution. Of the five California Channel Islands with island oaks present, the northern islands (Santa Rosa, Santa Cruz, and Anacapa) are cooler and receive more precipitation than the southern islands (Catalina and San Clemente). Precipitation seasonality varies across an east/west gradient in the northern Channel Islands, with the western part of the range experiencing less precipitation seasonality and more summer rainfall. Guadalupe Island, off the coast of Baja California, Mexico, has warmer winter temperatures and less precipitation, but has less yearly variation in temperature and precipitation, and receives more precipitation during the dry season than the California islands. We also extracted future predictions of climate at our sample sites using downscaled CMIP5 climate models from WorldClim, projected to 2050 under the RCP 8.5 (rising emissions) scenario. We choose two climate models with different outcomes for Southern California, CNRM-CM5 (a warmer/wetter scenario) and MIROC-ESM (a hotter/drier scenario) (Underwood et al. 2019).

Genetic Structure

Divergence among islands was calculated using Weir and Cockerham's F_{ST} in the R package heirfstat 0.5-11 (Goudet and Jombart 2022). To perform a principal components analysis (PCA) on the imputed SNP set, we used the R package vegan (version 2.6-2, Oksanen et al. 2019). Because Guadalupe Island trees were divergent from the other populations (see results), we also ran a PCA with those samples excluded. ADMIXTURE (version 1.3.0, Alexander et al. 2009) was used to estimate the ancestry of individuals and characterize genetic structure across the species range. For the entire dataset, we ran ADMIXTURE for K values 1-10, where K is the number of hypothetical ancestral populations. We also ran ADMIXTURE on subsets of the data to investigate fine-scale genetic structure for the following sets: all samples except Guadalupe Island (K=1-10) and individuals from Santa Cruz and Santa Rosa alone (K=1-5). Because Guadalupe Island populations appear deeply diverged from the other island oak populations, we excluded those individuals from further analyses of adaptive genetic variation.

To assess the contribution of geography and climate, we used a partial redundancy analysis (RDA) in the vegan package, which partitioned the genetic variance within the California islands samples into proportions that could be statistically explained by climate, by geographical location (via latitude and longitude), by each of these factors alone while controlling for the other, and by their joint influence. We tested the significance of each explanatory factor using the anova.cca function with 99 permutations.

A redundancy analysis was also used to identify candidate SNPs that are associated with climate variables. We reduced the climate and environmental variables to a subset with pairwise correlations below 0.75: BIO5 (Max Temperature of Warmest Month), BIO6 (Min Temperature of Coldest Month), BIO15 (Precipitation Seasonality), BIO18 (Precipitation of Warmest Quarter), BIO19 (Precipitation of Coldest Quarter), and elevation. A preliminary analysis found that results were strongly influenced by the five individuals from Anacapa, which are from one small grove and had high genetic similarity to each other. Many of the resulting candidate SNPs were associ-

ated with BIO5, which is lowest at the Anacapa site. Since these trees are relatively isolated, it is difficult to determine whether this association is the result of natural selection on temperature or due to inbreeding within this population, so we removed the Anacapa individuals from the analysis. Following Forester et al. (2018), we identified candidate SNPs as those that were strongly associated with the multivariate climate space. We considered candidate SNPs to be those that were outliers (more than 4 standard deviations) on the first three PC axes. We did not correct for population structure because the partial RDA showed a minor influence of geography on genetic variation after controlling for climate, and because correcting for population structure in an RDA results in reduced power and increased false positive rates in systems with low population structure (Forester et al. 2018).

Gradient Forest (version 0.1-32, Ellis et al. 2012) was used to identify nonlinear relationships between SNPs and environmental factors, and to predict genetic turnover across the range of island oak. We used the imputed SNP dataset and ran the analysis on the set of all SNPs, as well as on the set of candidate SNPs identified from the redundancy analysis. Geographic location cannot be included as a predictor in Gradient Forest analysis, so PCNM axes were calculated as a proxy for location using the pcnm function in the vegan package. A truncation distance of 150 km was used in calculating PCNM axes to maintain connections between populations on the northern and southern islands and avoid statistical artifacts resulting from our clustered sampling design. As recommended by Fitzpatrick and Keller (2015), we included the first half of the axes with positive eigenvectors (denoting a positive spatial correlation) in the analysis, in this case 5 PCNM axes. In addition to the PCNM axes, the same climate variables that were used in the redundancy analysis were included as explanatory variables. The gradientForest command was run with the following parameters: number of trees = 500, correlation threshold = 0.5, and max level = $\log_2(0.368 \times \text{number of samples}/2)$.

From the relationships between SNP frequency and climate gradients identified by Gradient Forest analysis, we can predict the genetic composition and turnover across the range, in-

cluding non-sampled locations. The first three predicted axes were mapped to RGB color values which were combined for each cell and mapped, as in Fitzpatrick and Keller (2015). We did this color assignment for both sets of SNPs, and also mapped the difference in genetic turnover between the two sets using the Procrustes residuals. Using the Gradient Forest predictions of genomic turnover, we calculated the predicted change in genomic composition of the candidate SNPs that would be necessary for populations to be well-adapted to future climate conditions at their present location under two climate models, or the genetic offset.

Results

Genetic Structure

F_{ST} values among island pairs and mainland *Q. chrysolepis* ranged from near 0 to 0.05 (Table 2.1). Within the island populations, the population from Guadalupe Island in Mexico was the most divergent from other populations (average pairwise F_{ST} = 0.032), followed by Santa Rosa and Anacapa (average pairwise F_{ST} = 0.021 and 0.020, respectively). In fact, the average divergence between Guadalupe populations and the California populations was similar to that of the divergence between mainland *Q. chrysolepis* and the island populations primarily consisting of *Q. tomentella.* PCA and ADMIXTURE plots also supported strong divergence of Guadalupe populations from other *Q. tomentella* samples in California (Figure 2.1, Figure 2.2).

We found evidence of widespread introgression between the two species. Mainland *Q. chrysolepis* samples clustered separately from the island individuals in the PCA, but island individuals that were identified in the field as *Q. chrysolepis* or possible hybrids primarily cluster within the island samples rather than with mainland *Q. chrysolepis* (Figure 2.1). ADMIXTURE plots show that *Q. chrysolepis* or hybrid individuals on the northern islands have mixed ancestry, with portions from mainland *Q. chrysolepis* and from the northern *Q. tomentella* groups (Figure 2.2). Trees from the southern California islands had more *Q. chrysolepis* ancestry than those from the

Table 2.1. Pairwise F_{ST} (Weir and Cockerham's) across all island samples and mainland *Q. chrysolepis* samples. Average divergences for each group were calculated using each pairwise F_{ST} value for a location, both with and without with the mainland *Q. chrysolepis* individuals included.

	Mainland (Qchr)	Santa Rosa	Santa Cruz	Anacapa	Catalina	San Clemente	Guadalupe	Average	Average, mainland excluded
Mainland (Qchr)	-	0.049	0.026	0.044	0.012	0.02	0.028	0.030	
Santa Rosa	0.049	-	0.009	0.023	0.019	0.011	0.042	0.026	0.021
Santa Cruz	0.026	0.009	-	0.01	0.009	0.001	0.027	0.014	0.011
Anacapa	0.044	0.023	0.01	-	0.011	0.013	0.045	0.024	0.020
Catalina	0.012	0.019	0.009	0.011	-	0.003	0.026	0.013	0.014
San Clemente	0.02	0.011	0.001	0.013	0.003	-	0.021	0.012	0.010
Guadalupe	0.028	0.042	0.027	0.045	0.026	0.021	-	0.032	0.032

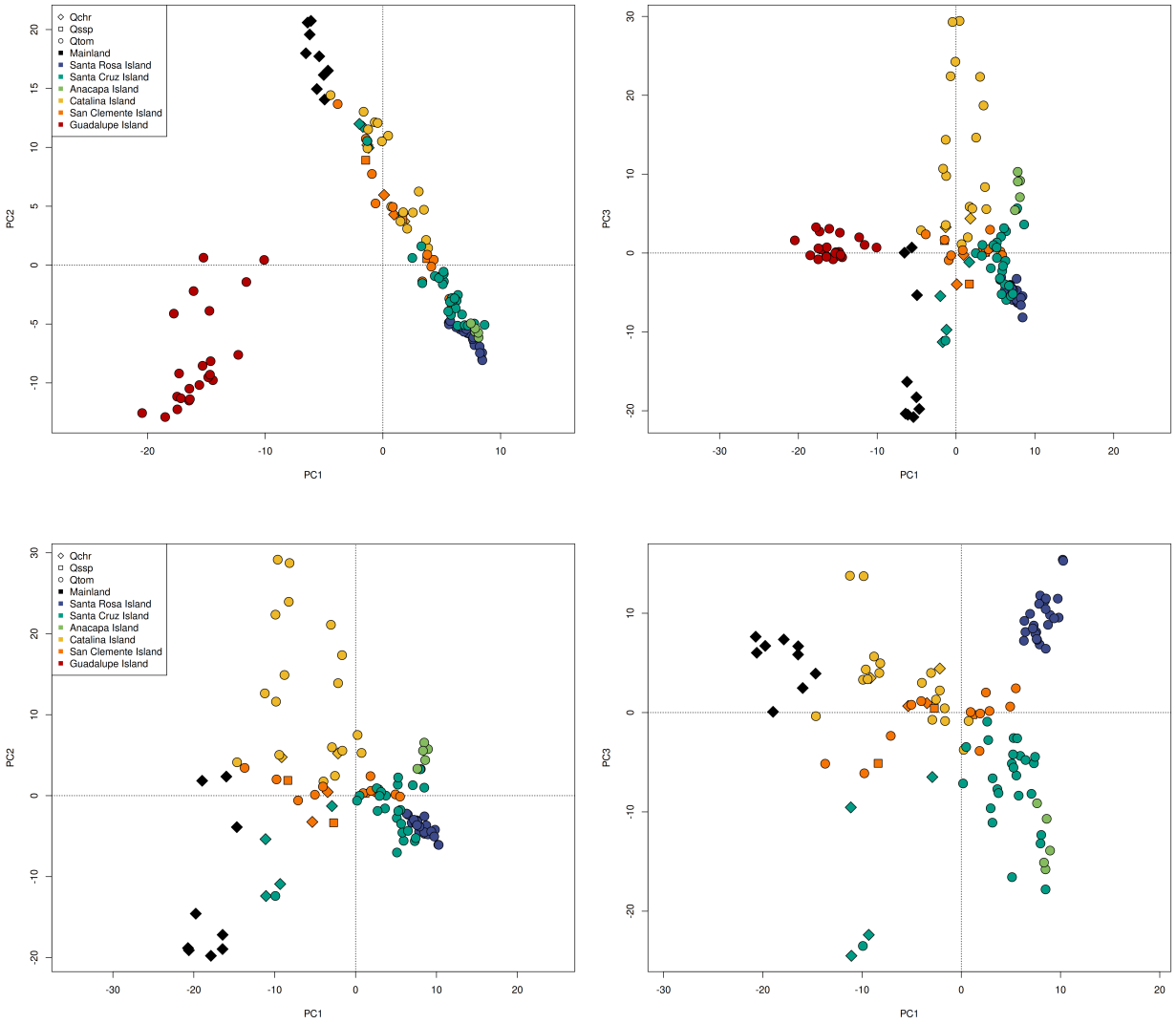


Figure 2.1. PCAs of genetic variation across SNPs for axes 1-3 for all samples (top) and with Guadalupe Island excluded (bottom). Color of points indicates their collection location, and shape indicates the species as identified in the field.

Figure 2.2. ADMIXTURE results when the number of ancestral groups (K) was set to values 2-5. Each color indicates a group, and each bar shows the ancestry proportions of each group for an individual tree.

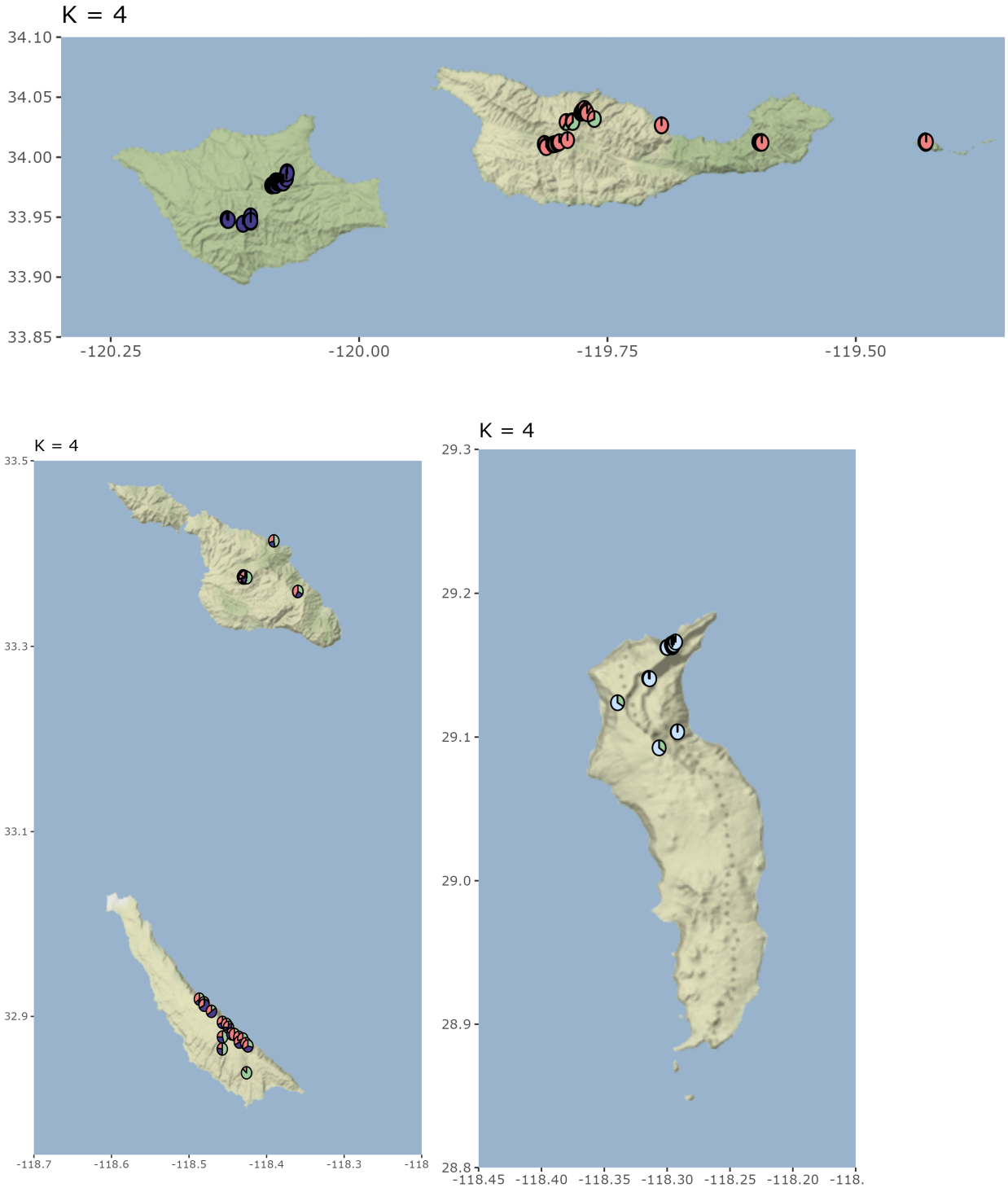


Figure 2.3. Map of ADMIXTURE ancestry proportions for each individual when K=4. Top shows the northern California islands, bottom left shows the southern California islands, and bottom right shows Guadalupe Island in Mexico.

northern islands (Figure 2.2). Because these results suggest a history of introgression between the two species on the islands rather than occasional F1 offspring, and because hybridization can be a source of genetic variation for natural selection (Suarez-Gonzalez et al. 2018), we decided to include the hybrid individuals and those that morphologically appeared to be *Q. chrysolepis* from the islands in further analyses of putatively adaptive genetic variation.

ADMIXTURE results for the full dataset, for the dataset with Guadalupe sample excluded, and for the dataset with only the northern California islands (Santa Cruz and Santa Rosa) all had the lowest cross-validation error for K=1, meaning that the best model includes all samples within a single population. We plotted admixture proportions for multiple K values to visualize genetic structure (Figure 2). The plot for K = 2 separates mainland *Q. chrysolepis* and Guadalupe Island as one ancestry group and the northern California islands as another, with admixture between the two groups occurring in hybrid individuals and in the southern California islands. For K = 3, Guadalupe Island forms a separate ancestry group. Across multiple K values, trees from the southern California islands were more admixed, with ancestry proportions from both mainland *Q. chrysolepis* and the northern California island populations. Running ADMIXTURE on a subset dataset including only two northern Channel Islands, Santa Rosa and Santa Cruz, revealed fine-scale genetic structure. For example, trees on the southern ridge of Santa Cruz shared some ancestry with Santa Rosa samples, suggesting that gene flow has occurred from Santa Rosa to Santa Cruz populations.

Adaptive genetic variation

Multiple redundancy analysis models were tested in order to partition genetic variance into proportions explained by climate, geography, their joint influence, and each factor alone (controlling for the effects of the other factor). All models significantly explained genetic variation (P = 0.01 for all), indicating that both geography and climate are factors shaping genetic variation (Table 2.2). However, climate explained a greater proportion of variance than geography: in the

Table 2.2. Partitioning of genetic variance among California island samples when climate and geography (latitude and longitude) are both included, and the contributions of each factor when controlling for the other. First column includes the total amount of variation explained, and second column is as a proportion of variance explained by the full model (0.038). *P*-values were calculated from 99 permutations, except for the joint influence, which is not testable.

	Adjusted R^2	Proportion of explainable variance	*P*	
Full model (climate + geography)	0.038	1	0.01	**
Climate alone	0.023	0.616	0.01	**
Geography alone	0.009	0.225	0.01	**
Joint influence of climate and geography	0.006	0.159		

full model, geography and climate together explained 3.8% of the genetic variation; with climate alone explaining 2.3%, geography alone explaining 0.9%, and their joint influence explaining 0.6%. As a proportion of total explainable variance, climate alone explained 61.6%, geography alone explained 22.5%, and their joint influence explained explained 15.9%.

The redundancy analysis identified 560 candidate climate-associated SNPs. For each candidate SNP, we determined which climate variable was most strongly correlated, and found 174 SNPs correlated most strongly with BIO19, 155 with elevation, 88 with BIO5, 73 with BIO18, 47 with BIO15, and 23 with BIO6.

When Gradient Forest was used to analyse all SNPs, the five PCNM axes describing geographic distance had greater importance values than the climate variables (Figure 2.4), suggesting a spatial influence on genetic variation. However, for the candidate SNPs, important variables included both PCNM axes and environmental variables (in order of importance: PCNM4, BIO6, PCNM2, elevation, PCNM1). As expected, overall importance values were also higher when considering only the candidate SNPs, indicating a stronger associations. Genetic turnover was mapped onto the California islands to visualize predicted spatial patterns of neutral and climate-associated genetic variation (Figure 2.5). Results for both candidate and neutral SNPs showed genetic differentiation across all islands, but southern California islands were more similar to each other, and northern California islands showed variation across an east-west gradient associated with precipitation seasonality (BIO15) and precipitation of the coldest quarter (BIO19). Eastern

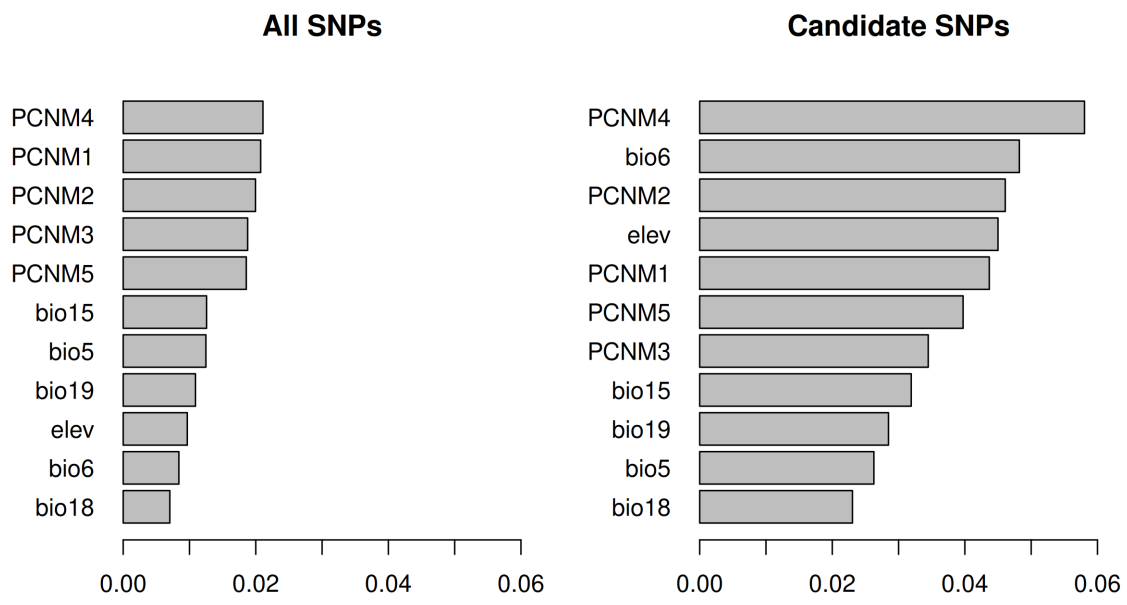


Figure 2.4. Gradient forest R^2 weighted importance values of environmental variables (BioClim and elevation) and distance variables (PCNM) in explaining genetic variation, for all SNPs and the for only the set of 560 candidate climate SNPs. BioClim variables abbreviations are: bio5 (Max Temperature of Warmest Month), bio6 (Min Temperature of Coldest Month), bio15 (Precipitation Seasonality), bio18 (Precipitation of Warmest Quarter), and bio19 (Precipitation of Coldest Quarter).

Santa Cruz Island and the coastline of San Clemente Island were the regions with the most the most prominent difference in allele turnover between the set of all SNPs and the candidate SNPs (Figure 2.6).

Maps of the genetic offset were similar for both models of future climate, and predicted that populations on Catalina Island will be most maladapted to future conditions, followed by Anacapa and eastern Santa Cruz (Figure 2.7). We also tested whether the historic climate at some sites may match the future climate at other sites, potentially providing seed sources pre-adapted to future climate conditions. We visualized shifts in climate by plotting the historic annual mean temperature (BIO1) and annual precipitation (BIO12) for current and future conditions for all sample locations (Figure 2.8). The results vary depending on the climate model used, but in general there are few sites with historic climate that match both future temperature and precipitation

values at other sites. However, under the CNRM model, the future climate of San Clemente sites is similar to the present climate at some Catalina sites.

Discussion

Genetic structure

Our results show varying levels of isolation among islands. Overall F_{ST} values are low, indicating weak population structure (Table 2.1). The Guadalupe Island oaks, which are the most geographically distant from the other populations, are strongly isolated from the California islands. The California island populations are isolated to various degrees: overall, Santa Rosa appears to be the most genetically distinct, while other islands are less differentiated from each other and show signs of admixture (Table 2.1, Figure 2.2). The patterns of genetic structure found here may result from asymmetric gene flow shaped by typical wind patterns during spring flowering. In the California islands, winds are primarily from the northwest, so trees on Santa Rosa Island are unlikely to receive pollen from other islands, while Santa Cruz and Anacapa may receive pollen from Santa Rosa, and San Clemente and Catalina may receive pollen from all three northern islands, explaining the increased admixture observed in these individuals (Figure 2.2). Similar to our results, results from microsatellite data found isolation of Guadalupe and greater admixture in the Catalina and San Clemente populations (Ashley et al. 2018), but our ADMIXTURE results show more distinct ancestries between Santa Rosa and the islands to its east.

Introgression with *Q. chrysolepis* appears to be widespread and is likely to be another important factor shaping genetic structure in island oak. Admixture between the two species is particularly evident in the southern Channel Islands, but also present within some individuals in the northern Channel Islands. Island trees that were identified as *Q. chrysolepis* based on morphology clustered more closely with the island *Q. tomentella* individuals than with the mainland *Q. chrysolepis* individuals, suggesting a long history of introgression between the two species on

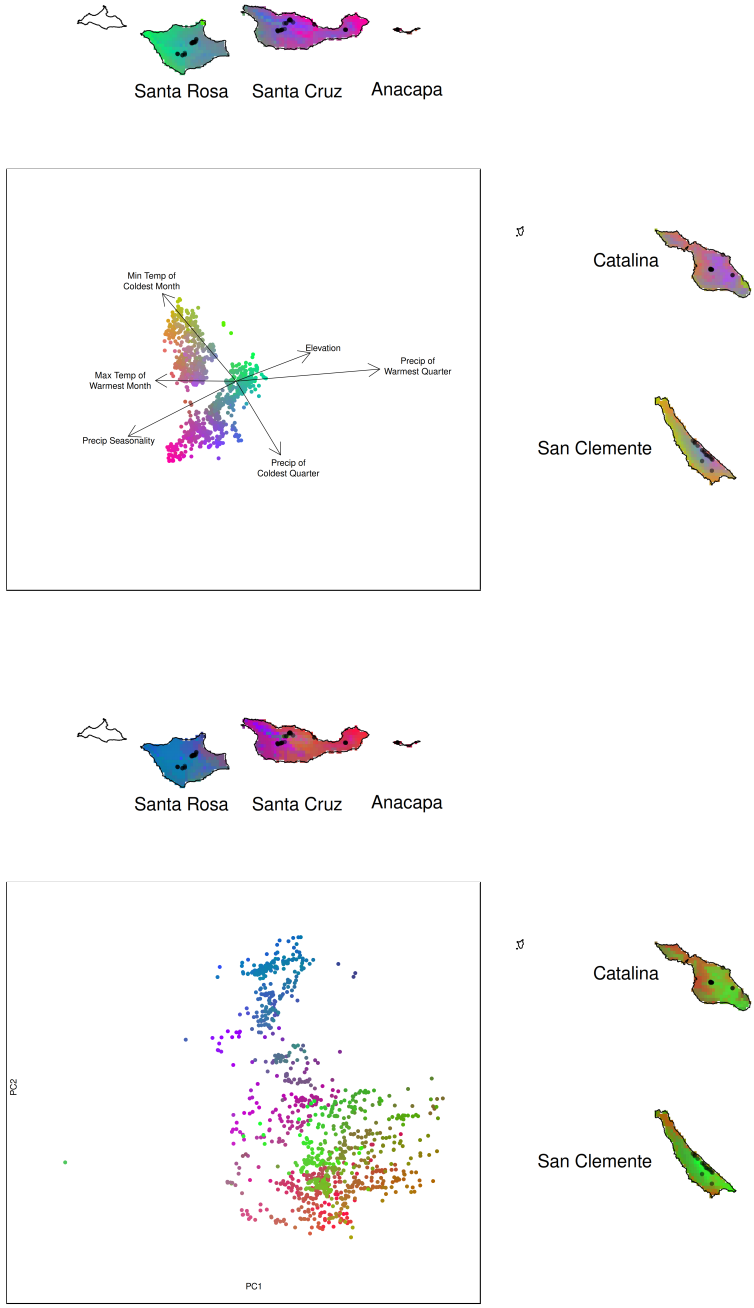


Figure 2.5. Maps of genetic turnover predicted by gradient forest for all SNPs (top) and candidate SNPs (bottom). Regions that are predicted to be genetically similar based on SNP-climate associations have similar colors. Inset shows the first two axes of a PCA that explains variation in climate across all cells included in the map. Loadings are not shown for the candidate SNPs for greater visibility. Points indicate sample locations.

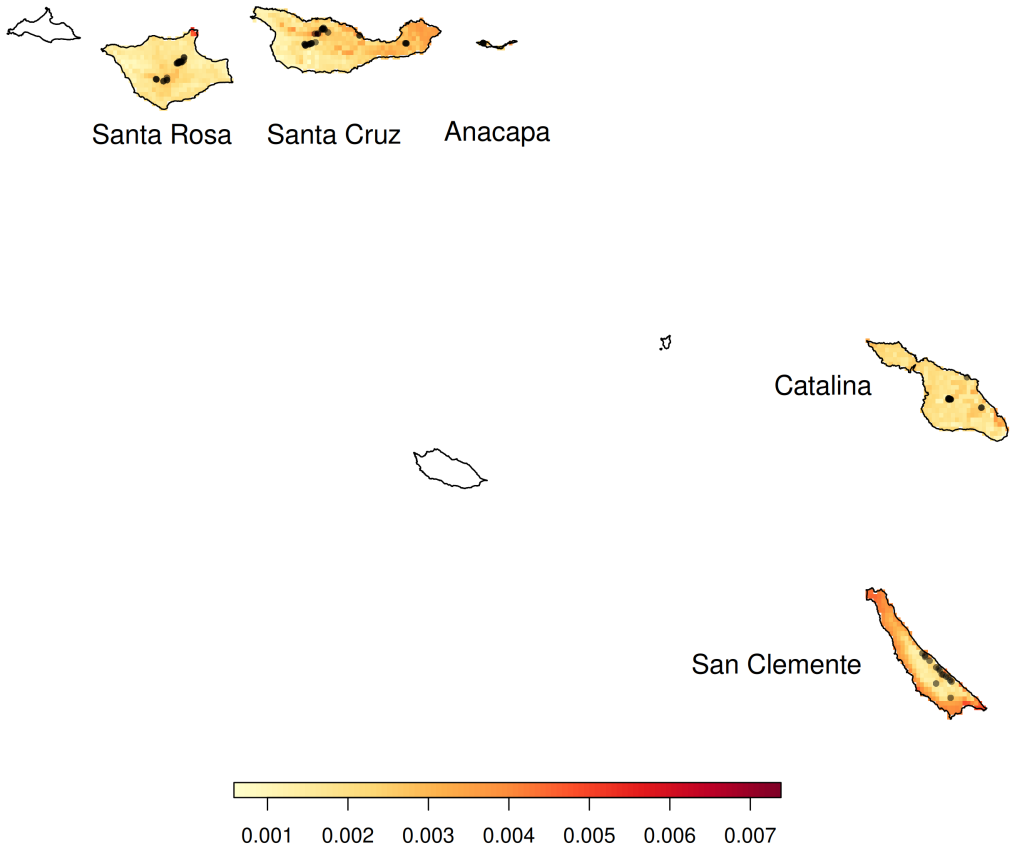


Figure 2.6. Difference in the genomic turnover patterns between the candidate SNPs and all SNPs, as shown in Figure 2.5. Darker red indicates that the two sets of SNPs produce different predictions of genetic composition for that region.

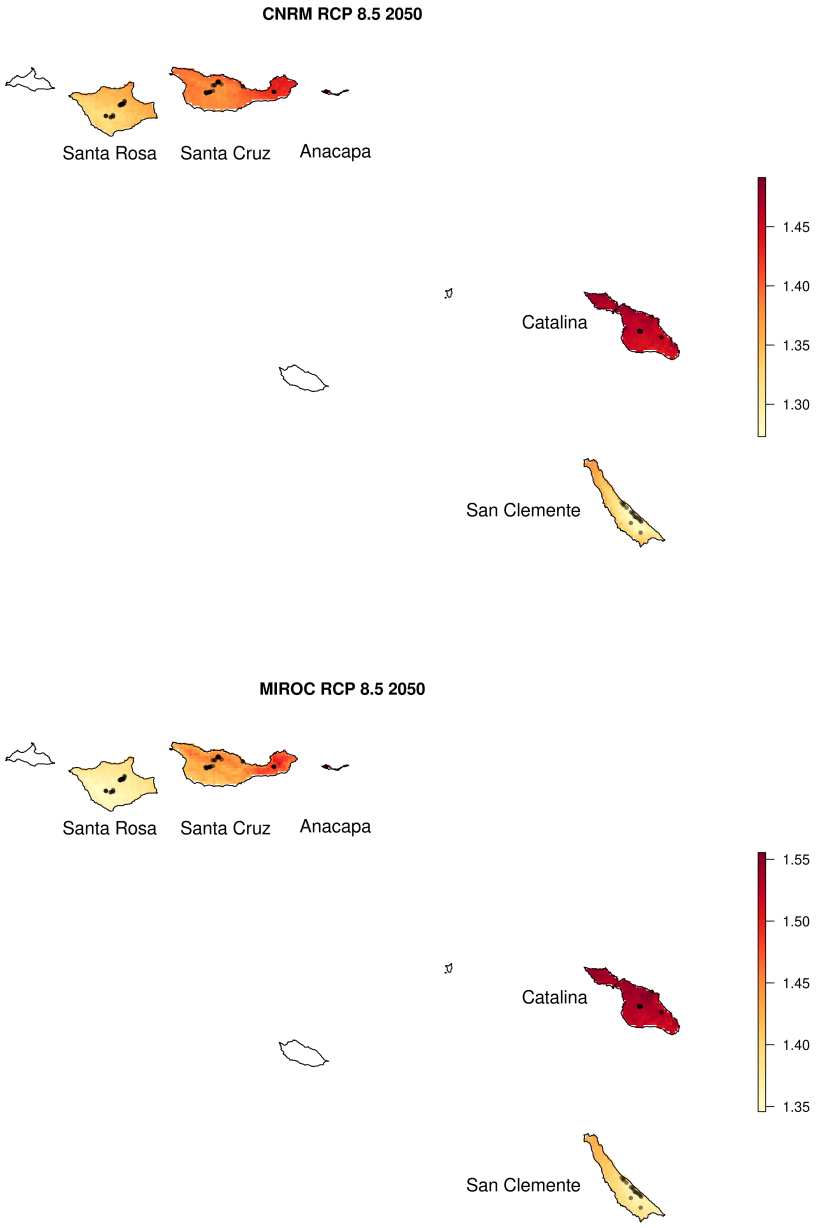


Figure 2.7. Map of the genetic offset, or predicted difference in the current genetic composition and that which would be ideal under future climate conditions. Darker red indicates a greater likelihood of future maladaptation. Results are shown for two different climate models, CNRM (upper) and MIROC (lower), for the year 2050 under the RCP 8.5 (rising emissions) scenario.

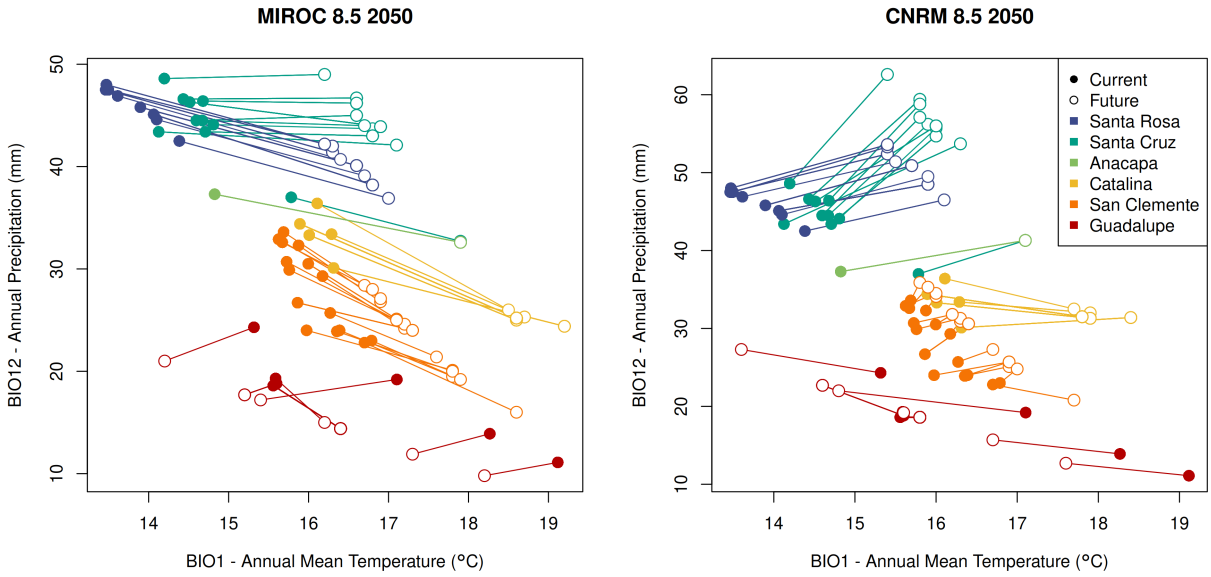


Figure 2.8. Predicted shifts in climate at each location based on annual mean temperature (BIO1) and annual precipitation (BIO12) under two future climate models. Closed circles indicate current climate for a given location, and they are connected to open circles that indicate the projected future climate for the same location.

the islands (Figure 2.1). The isolated Guadalupe Island population, however, appears to be genetically similar to the mainland *Q. chrysolepis;* having similar values along PC1 (Figure 2.1) and forming the same ancestry group for K=2 (Figure 2.2). Previous work has hinted at a complex evolutionary history in these two species. Ortego et al. (2018) suggested that *Q. chrysolepis* is a non-monophyletic species split into a northern and southern lineage, and that *Q. tomentella* split from the southern lineage of *Q. chrysolepis* in the late Pliocene or early Pleistocene. If *Q. tomentella* originated in southwest North America, Guadalupe Island may have been colonized early in the species' evolutionary history, before its range moved northward. Colonization of the California islands may have occurred later, as California shifted toward a more Mediterranean climate and *Q. tomentella* became restricted to the coast. Additionally, it is possible that the Guadalupe population has experienced introgression with *Q. cedrocensis*, which we were unable to sample but which is present on Cedros Island to the southeast of Guadalupe Island and on the mainland in San Diego County and northern Baja California. In the future, demographic model-

ing that includes the Guadalupe Island population could be used to disentangle the evolutionary history of these species.

Local adaptation

Our analyses revealed that climate accounted for most of the explainable genetic variation among the California island populations when controlling for the effects of geography (Table 2.2). Thus, natural selection is having a major impact on the evolution of these island populations in comparison to neutral differences resulting from genetic drift and l imited gene flow within or among islands. We identified 560 candidate SNPs associated with climate across the islands that may contribute to local adaptation. SNPs were most frequently associated with precipitation of the coldest quarter (BIO19) and elevation (which was also correlated with temperature annual range, BIO17, in our dataset). In the Gradient Forest analysis, precipitation seasonality (BIO15) and maximum temperature of the warmest month (BIO5) were the most important climate variables. Together these results suggest that both temperature and precipitation variables are important in local adaptation, and there is not a strong influence of any one factor, unlike some tree species for which genetic turnover is most associated with precipitation variables (Gugger et al. 2018, Martins et al. 2018). In our results, the PCNM axes had greater importance values than climate in explaining genetic turnover. This pattern could result from spatial location, unmeasured environmental factors, or both factors shaping genetic turnover more strongly than the measured climate variables (Martins et al. 2018).

This study also revealed evidence that populations on the same island may be locally adapted to different conditions. On Santa Cruz Island, climate-associated genetic variation differed across an east/west gradient (Figure 2.5). Candidate SNPs and climate-associated SNPs produced different predictions of allelic turnover on eastern Santa Cruz Island, suggesting that gene flow occurs between eastern and western Santa Cruz populations, but that selection results in differentiation of climate-associated alleles between the regions. Local adaptation could occur

as a result of the precipitation gradient in the northern Channel Islands, in which eastern regions have less summer precipitation and greater precipitation seasonality. This adds to the existing evidence that local adaptation may be possible at small scales and in the presence of gene flow, even within relatively small island populations (Cheek et al. 2022, Gamboa et al. 2022, Gugger et al. 2018, Hamilton et al. 2017, Langin et al. 2015). In contrast, populations on Catalina and San Clemente show little turnover of adaptive alleles within islands.

Genetic offset

Genetic offset is a measure of vulnerability to climate change, based on associations between genomic structure and future climates. Our results predict that populations on Catalina will be most maladapted to future climates in comparison to the other California Island populations, followed by eastern Santa Cruz populations (Figure 2.7). If gene flow occurs from the northern to southern islands, as suggested by the genetic structure, both Catalina and San Clemente may receive alleles that are maladaptive at hotter sites, particularly as climate change progresses. Catalina is projected to experience a greater increase in annual mean temperature than San Clemente (Figure 2.8), resulting in a larger genetic offset. While all populations are likely to experience some degree of maladaptation to future climates, populations with a greater genetic offset could be prioritized in conservation efforts.

Seed transfer guidelines

For restoration projects currently taking place on several islands, landscape genomic findings can inform decisions about what seed sources should be used to maximize success for plantings. However, different strategies may be preferred depending on management priorities, and conservation decisions must take values into account as well as data. If the priority is to maintain "natural" patterns of genetic variation, seeds should only be introduced among populations that have historically experienced gene flow: for example, seeds from Santa Rosa could be introduced

to other islands, but other populations should not be introduced to Santa Rosa since it seems to be more isolated. However, if the priority is to ensure that species can persist in their current ranges under climate change, it may be necessary to introduce alleles that are "pre-adapted" to future climates through assisted gene flow (Aitken and Bemmels 2016, Aitken and Whitlock 2013). Since all populations are expected to be maladapted to future conditions (Figure 2.7), maintaining historic patterns of gene flow may not effectively conserve populations. For example, while Santa Rosa populations are relatively isolated, they have also historically experienced the coolest temperatures and may benefit from the introduction of alleles that are beneficial under warmer conditions.

Future temperatures on the three northern California islands are expected to become more like historic temperatures on San Clemente and Catalina islands (Figure 2.8), so introducing seeds from these warmer islands may benefit the northern populations. However, it may be difficult to identify seed sources that match the precise conditions projected in the future. When considering both mean temperature and mean precipitation, few populations are likely to be pre-adapted to future climates at other sites, with the possible exception of seeds from San Clemente matching future conditions on Catalina under one climate model (Figure 2.8). Additionally, future climate predictions vary by model and scenario, and predictions about precipitation changes have greater uncertainty than temperature changes (IPCC 2014, Trenberth 2011). In this case, transferring seeds northward would result in increased precipitation, which may not be deleterious for this species.

This study demonstrates the importance of using genomic information when determining the best management strategies to conserve rare species. Our recommendations depend on the assumption that populations are locally adapted to modern climate conditions. Because *Q. tomentella* is a paleoendemic that experienced range contraction and likely survives on the Channel Islands because of their more moderate climates, it is possible that populations are already maladapted to present climate conditions, as found for California populations of valley oak (Browne

et al. 2019). Setting up reciprocal transplant experiments that can be monitored over 5-10 years would improve our predictions of the populations that are ideal seed sources for a restoration site, in comparison to this landscape genomic study alone. Nonetheless, without that information, our findings demonstrate that island populations are likely to be maladapted to future climates. Therefore, selecting genotypes from populations that will be best adapted to future conditons is likely to be more beneficial for habitat restoration than using local seeds.

Acknowledgements

This work took place on the ancestral lands of the Chumash (Northern Channel Islands) and the Tongva (Southern Channel Islands), and we acknowledge them as the traditional caretakers of the land.

We thank those who helped us with fieldwork and collecting: Elizabeth Becker, James Cabrera, Lauren Dennhardt, Adam Fontenot, Robert Kaaret, Seth Kauppinen, John Knapp, Reena Lam, Betty Lee, Zhizhong Li, Luciana Luna, Sergio A. Luvianos-Colín, Scott O'Donnell, John Randall, Joseph Retelskyj, and Kristi Smith. Kari Merrill and Dan Oliveira helped perform DNA extractions. Sequencing was performed by the DNA Technologies and Expression Analysis Cores at the UC Davis Genome Center, supported by NIH Shared Instrumentation Grant 1S10OD010786-01. Sorel Fitz-Gibbon performed variant calling and provided advice on genomic 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]; and by The Nature Conservancy.

References

Aitken, S. N. and J. B. Bemmels (2016). Time to get moving: assisted gene flow of forest trees. *Evolutionary Applications* 9(1):271–290.

Aitken, S. N. and M. C. Whitlock (2013). Assisted gene flow to facilitate local adaptation to climate change. *Annu. Rev. Ecol. Evol. Syst* 44:367–88.

Alexander, D. H., J. Novembre, and K. Lange (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research* 19(9):1655–1664.

Ashley, M. V., J. R. Backs, L. Kindsvater, and S. T. Abraham (2018). Genetic variation and structure in an endemic island oak, *Quercus tomentella*, and mainland canyon oak, *Quercus chrysolepis*. *International Journal of Plant Sciences* 179(2):151–161.

Axelrod, D. I. (1939). A miocene flora from the western border of the Mohave desert. Contributions to paleontology. The Carnegie Institution of Washington, Washington, D. C.

Axelrod, D. I. (1944a). The Mulholland Flora. *In* Pliocene floras of California and Oregon, [Carnegie Institution of Washington] Contributions to paleontology. The Carnegie Institution of Washington, Washington, D. C.

Axelrod, D. I. (1944b). The Sonoma Flora. *In* Pliocene floras of California and Oregon, [Carnegie Institution of Washington] Contributions to paleontology. The Carnegie Institution of Washington, Washington, D. C.

Axelrod, D. I. (1967). Geologic history of the Californian insular flora. *In* R. N. Philbrick (editor), Proceedings of the Symposium on the Biology of the California Islands. Santa Barbara Botanic Garden.

Beckman, E. and D. Jerome (2017). *Quercus tomentella*. `https://www.iucnredlist.org/species/30959/2799049`

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

Chang, C. C., C. C. Chow, L. C. Tellier, S. Vattikuti, S. M. Purcell, and J. J. Lee (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* 4(1):s13742–015–0047–8.

Cheek, R. G., B. R. Forester, P. E. Salerno, D. R. Trumbo, K. M. Langin, N. Chen, T. Scott Sillett, S. A. Morrison, C. K. Ghalambor, and W. Chris Funk (2022). Habitat-linked genetic variation supports microgeographic adaptive divergence in an island-endemic bird species. *Molecular Ecology* 31(10):2830–2846.

Danecek, P., J. K. Bonfield, J. Liddle, J. Marshall, V. Ohan, M. O. Pollard, A. Whitwham, T. Keane, S. A. McCarthy, R. M. Davies, and H. Li (2021). Twelve years of SAMtools and BCFtools. *GigaScience* 10(2):giab008.

Di Santo, L. N., S. Hoban, T. L. Parchman, J. W. Wright, and J. A. Hamilton (2022). Reduced representation sequencing to understand the evolutionary history of Torrey pine (*Pinus torreyana* parry) with implications for rare species conservation. *Molecular Ecology* 31(18):4622–4639.

Ellis, N., S. J. Smith, and C. R. Pitcher (2012). Gradient Forests: calculating importance gradients on physical predictors. *Ecology* 93:156–168.

Fitzpatrick, M. C., V. E. Chhatre, R. Y. Soolanayakanahally, and S. R. Keller (2021). Experimental support for genomic prediction of climate maladaptation using the machine learning approach Gradient Forests. *Molecular Ecology Resources* 21(8):2749–2765.

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

Forester, B. R., J. R. Lasky, H. H. Wagner, and D. L. Urban (2018). Comparing methods for detecting multilocus adaptation with multivariate genotype–environment associations. *Molecular Ecology* 27(9):2215–2233.

Gamboa, M. P., C. K. Ghalambor, T. Scott Sillett, S. A. Morrison, and W. Chris Funk (2022). Adaptive divergence in bill morphology and other thermoregulatory traits is facilitated by restricted gene flow in song sparrows on the California Channel Islands. *Molecular Ecology* 31(2):603–619.

García-Verdugo, C., M. Sajeva, T. La Mantia, C. Harrouni, F. Msanda, and J. Caujapé-Castells (2015). Do island plant populations really have lower genetic variation than mainland populations? Effects of selection and distribution range on genetic diversity estimates. *Molecular Ecology* 24(4):726–741.

Goudet, J. and T. Jombart (2022). hierfstat: Estimation and Tests of Hierarchical F-Statistics. `https://CRAN.R-project.org/package=hierfstat`

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

Gugger, P. F., C. T. Liang, V. L. Sork, P. Hodgskiss, and J. W. Wright (2018). Applying landscape genomic tools to forest management and restoration of Hawaiian koa (*Acacia koa*) in a changing environment. *Evolutionary Applications* 11(2):231–242.

Hamabata, T., G. Kinoshita, K. Kurita, P.-L. Cao, M. Ito, J. Murata, Y. Komaki, Y. Isagi, and T. Makino (2019). Endangered island endemic plants have vulnerable genomes. *Communications Biology* 2(1):1–10.

Hamilton, J. A., R. Royauté, J. W. Wright, P. Hodgskiss, and F. T. Ledig (2017). Genetic conservation and management of the California endemic, Torrey pine (*Pinus torreyana* Parry): Implications of genetic rescue in a genetically depauperate species. *Ecology and Evolution* 7(18):7370–7381.

Ingvarsson, P. K. and C. Bernhardsson (2020). Genome-wide signatures of environmental adaptation in European aspen (*Populus tremula*) under current and future climate conditions. *Evolutionary Applications* 13(1):132–142.

IPCC (2014). Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Technical report, Intergovernmental Panel on Climate Change. `https://www.ipcc.ch/report/ar5/syr/`

Jia, K.-H., W. Zhao, P. A. Maier, X.-G. Hu, Y. Jin, S.-S. Zhou, S.-Q. Jiao, Y. A. El-Kassaby, T. Wang, X.-R. Wang, and J.-F. Mao (2020). Landscape genomics predicts climate change-related genetic offset for the widespread *Platycladus orientalis* (Cupressaceae). *Evolutionary Applications* 13(4):665–676.

Kennett, D. J., J. P. Kennett, G. J. West, J. M. Erlandson, J. R. Johnson, I. L. Hendy, A. West, B. J. Culleton, T. L. Jones, and T. W. Stafford (2008). Wildfire and abrupt ecosystem disruption on California's Northern Channel Islands at the Ållerød–Younger Dryas boundary (13.0–12.9ka). *Quaternary Science Reviews* 27(27):2530–2545.

Langin, K. M., T. S. Sillett, W. C. Funk, S. A. Morrison, M. A. Desrosiers, and C. K. Ghalambor (2015). Islands within an island: Repeated adaptive divergence in a single population. *Evolution* 69(3):653–665.

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

Muller, C. H. (1965). Relictual origins of insular endemics in *Quercus*. *In* 1st Symposium on the Biology of the California Islands. National Park Service.

Oksanen, J., F. G. Blanchet, M. Friendly, R. Kindt, P. Legendre, D. McGlinn, P. R. Minchin, R. B. O'Hara, G. L. Simpson, P. Solymos, M. H. H. Stevens, E. Szoecs, and H. Wagner (2019). vegan: Community Ecology Package. https://CRAN.R-project.org/package=vegan

Ortego, J., P. F. Gugger, and V. L. Sork (2018). Genomic data reveal cryptic lineage diversification and introgression in Californian golden cup oaks (section *Protobalanus*). *New Phytologist* 218(2):804–818.

Petit, R. J. and A. Hampe (2006). Some evolutionary consequences of being a tree. *Annual Review of Ecology, Evolution, and Systematics* 37:187–214.

Reeder-Myers, L., J. M. Erlandson, D. R. Muhs, and T. C. Rick (2015). Sea level, paleogeography, and archeology on California's Northern Channel islands. *Quaternary Research* 83(2):263–272.

Rellstab, C., B. Dauphin, and M. Exposito-Alonso (2021). Prospects and limitations of genomic offset in conservation management. *Evolutionary Applications* 14(5):1202–1212.

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

Rellstab, C., S. Zoller, L. Walthert, I. Lesur, A. R. Pluess, R. Graf, C. Bodénès, C. Sperisen, A. Kremer, and F. Gugerli (2016). Signatures of local adaptation in candidate genes of oaks (*Quercus* spp.) with respect to present and future climatic conditions. *Molecular Ecology* 25(23):5907–5924.

Sexton, J. P., S. B. Hangartner, and A. A. Hoffmann (2014). Genetic isolation by environment or distance: Which pattern of gene flow is most common? *Evolution* 68(1):1–15.

Sork, V. L. (2016). Gene flow and natural selection shape spatial patterns of genes in tree populations: Implications for evolutionary processes and applications. *Evolutionary Applications* 9(1):291–310.

Sork, V. L. (2017). Genomic studies of local adaptation in natural plant populations. *Journal of Heredity* 109(1):3–15.

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

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

Suarez-Gonzalez, A., C. Lexer, and Q. C. B. Cronk (2018). Adaptive introgression: a plant perspective. *Biology Letters* 14(3):20170688.

Trenberth, K. E. (2011). Changes in precipitation with climate change. *Climate Research* 47:123–138.

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

Wang, I. J. and G. S. Bradburd (2014). Isolation by environment. *Molecular Ecology* 23(23):5649–5662.

Chapter 3

Shared responses to heat stress and limited local adaptation across three California oak species[1]

Abstract

The frequency and intensity of heat waves are increasing from those historically experienced by tree populations. Previous work has shown that oak species and populations can differ in their response to drought, but few studies have examined response to increased temperatures. Comparing stress responses among species and populations can clarify how traits and genotypes contribute to differences in stress response and tolerance among taxa. Here, we characterize the gene expression response to heat stress in three California oak species, *Quercus douglasii*, *Q. kelloggii*, and *Q. lobata*. We also sampled acorns from a warmer and cooler site within each species range to test whether populations varied in their heat responses, consistent with local adaptation, and whether population differences were similar across all three species. We exposed two-year-old seedlings to a heat wave treatment with a maximum temperature of 35 °C for 14 days and compared them to seedlings in a control treatment with a maximum temperature of 25 °C. We found that leaf temperatures were higher in seedlings exposed to the heat treatment. All species and populations shared some core responses to this elevated temperature, including upregula-

[1]This chapter will be submitted for publication with the following authors: Alayna Mead and Victoria Sork

tion of genes related to heat response, such as heat shock proteins and chaperonins. Within each species, populations showed similarities in their responses, suggesting that these oak species are either well-adapted to tolerate the heat wave they experienced through our experiment, or that that the response to heat stress was similar among populations. These findings, in combination with earlier studies on drought stress, imply that oaks may be able to withstand increased temperatures if they do not experience drought stress concurrently.

Introduction

The frequency and intensity of extreme heat events is expected to continue to increase (IPCC 2014), and these heat waves can be a major cause of mortality in trees, particularly when heat stress is combined with drought or biotic stresses (Allen et al. 2010, Teskey et al. 2015). Previous studies have suggested that temperature is a major factor shaping plant evolution, including trait, genetic, epigenetic, and gene expression variation (Gugger et al. 2016a,b, Mead et al. 2019, Moles et al. 2014). However, despite its importance, few studies have assessed the response to multi-day heat stress in trees, with short-term "heat shock" experiments being more common (Teskey et al. 2015). Existing studies have found that heat stress caused decreased photosynthesis rates, negative effects on photochemistry, decreased growth, and lower biomass accumulation, while other studies have found decreases in growth only when heat is combined with drought (Ameye et al. 2012, Arend et al. 2013, Bauweraerts et al. 2013, Daas et al. 2008, Guha et al. 2018).

Gene expression can be used to quantify fine-scale physiological responses to stress as well as their underlying genetic basis, and it can also be an important mode of adaptation (Carroll 2005, Jones et al. 2012, Kenkel and Matz 2016). Transcriptomic responses to heat stress are well-characterized in model species and crops (Ohama et al. 2017, Shinozaki and Yamaguchi-Shinozaki 2022) and gene expression differences may help explain variation in stress tolerance (Halter et al. 2017). Comparisons of heat-responsive genes among species and populations could enable a better understanding of the genes that contribute to differences in drought tolerance. Previous

work has found differences in the response to drought stress among species (see Chapter 1) and populations (Gugger et al. 2016b, Mead et al. 2019) that may contribute to differential tolerance, suggesting that responses to heat may vary similarly.

Some stress responses are likely to be common across many species, such as increased production of heat shock proteins (HSPs), a conserved protein family present in all domains of life which has undergone multiple diversification events in plants (Chen et al. 2018, Waters and Vierling 2020). Other responses may vary among species and contribute to differences in heat tolerance, such as the temperature response curves of photochemistry and transpiration, both of which ultimately determine growth and fitness (Arend et al. 2013, Daas et al. 2008, Guha et al. 2018, Hamerlynck and Knapp 1996). Populations may also differ in their responses to heat stress due to adaptation to varying local conditions (Bigras 2000, Teskey et al. 2015, Weston et al. 2008). If local adaptation is widespread, populations that typically experience higher temperatures should be more heat-tolerant than those at cooler sites. For example, in many tree species, populations from more xeric sites have thicker leaves, higher stomatal density, and higher photosynthesis rates during drought than mesic populations (Abrams 1994). In trees, there is some evidence of population differentiation in responses to heat stress (Bigras 2000, Konôpková et al. 2018, Teskey et al. 2015, Weston and Bauerle 2007), but the genetic basis of these responses is relatively unknown.

In this study, we compared the gene expression responses to heat stress across three California oak species with different climatic niches, and two populations within each species from sites in northern and southern California which experience contrasting temperatures. In general, oaks are relatively heat tolerant (Arend et al. 2013, Guha et al. 2018), but species vary widely in their morphology and climatic niches (Cavender-Bares 2018). Additionally, populations within species are often adapted to local conditions (Gugger et al. 2016a, 2021), and oak populations differ in their responses to drought stress (Gugger et al. 2016b, Mead et al. 2019). If the selective pressures from heat waves vary across a species range, responses to heat may differ among

populations, similar to that observed in responses to drought. Conversely, if populations have shared responses to heat, this may indicate similar selective pressure across all sites, or that local adaptation has occurred through mechanisms besides plasticity. This study had two primary objectives: 1) Identify responses to heat stress that are shared among the three species and which may be important across oaks, as well as responses that differ across species and may contribute to their contrasting climate tolerances. 2) Test whether populations within species differed in their responses, consistent with local adaptation to contrasting climates, and whether there were shared patterns of local adaptation across the three species.

Methods

Study species and sample design

Four species were included in the heat stress experiment, and three of these were sequenced for their gene expression responses. *Quercus agrifolia* (QAG) is an evergreen tree (Tucker 2014a), while the other three species are deciduous. *Q. douglasii* (QDO) is a small tree within the California scrub oak clade (Hipp et al. 2018, Sork et al. 2016), occurs on dry slopes and woodlands, and is drought deciduous (Abrams 1990, Tucker 2014b). *Q. lobata* (QLO) is a large tree that occurs in savannas, primarily in valleys with groundwater access (Cannon 1914, Tucker 2014d). *Q. kelloggii* (QKE) occurs at higher elevations and therefore typically experiences lower temperatures (Tucker 2014c). Of these species, *Q. douglasii* and *Q. agrifolia* are more resistant to embolism (indicating adaptation to xeric habitats) than *Q. kelloggii* and *Q. lobata* (Skelton et al. 2018).

For each species we sampled two populations from contrasting sites, a warmer and a cooler site. At Blue Oak Ranch Reserve (BO), a cooler and wetter site in Northern California, all four species co-occur and were sampled. At Sedgwick Reserve (SW), a warmer and drier site in Southern California, all species except *Q. kelloggii* were sampled. For a warmer site with *Q. kelloggii* we sampled from the James San Jacinto Mountains Reserve (JA), also in Southern California. In

Fall 2019, we collected acorns from 7-11 mother trees of each species at each site. After collection, acorns from *Q. agrifolia*, *Q. douglasii*, and *Q. lobata* were stored at 4 °C until they were planted in December. Because *Q. kelloggii* occurs at higher elevations and requires a longer stratification period than the other species (Bonner and Karrfalt 2008), they were stored at 4 °C until the end of January, when they were planted. Prior to planting, acorns that floated in water were discarded and remaining acorns were scrubbed with 10% bleach to prevent mold growth. They were planted in Stuewe & Sons D40 pots (with a depth of 10 inches and diameter of 2.5 inches) in SunGrow Sunshine Mix #4, and grown in the Plant Growth Center at UCLA.

Heat stress experiment

The heat stress experiment took place in July 2021, during the seedlings' second summer. The experiment was designed to test oak responses to a heat wave that is hotter and longer than historically experienced at the sample sites, similar to heat events that are will occur more frequently as climate change progresses. To design the experiment, we analyzed historic weather data collected by weather stations at Sedgwick and Blue Oak Ranch reserves. We downloaded data weather data from Dendra (Dendra Science 2022), and analyzed air temperatures taken at a height of two meters every ten minutes across all days of July from 2011-2020 (Figure 3.1). The experiment was run for 14 days, from July 7 – July 20, which is a longer period than previous July heat waves at or above that temperature since 2011 (SW had 11 consecutive days and BO had 7 consecutive days at 35 °C or higher).

Seedlings were placed in either a control or heat treatment growth chamber, and temperatures were set to mimic natural conditions by shifting gradually over the course of a day, with the highest temperatures occurring in the afternoon. Temperature changes are summarized in Figure 3.1. For both treatments, the growth chambers were set to 15 °C at midnight. At 6:00 am, the temperature was increased linearly, reaching its maximum at noon, and was held at the maximum temperature for two hours, then linearly decreased from 2:00 pm until it reached 15 °C at

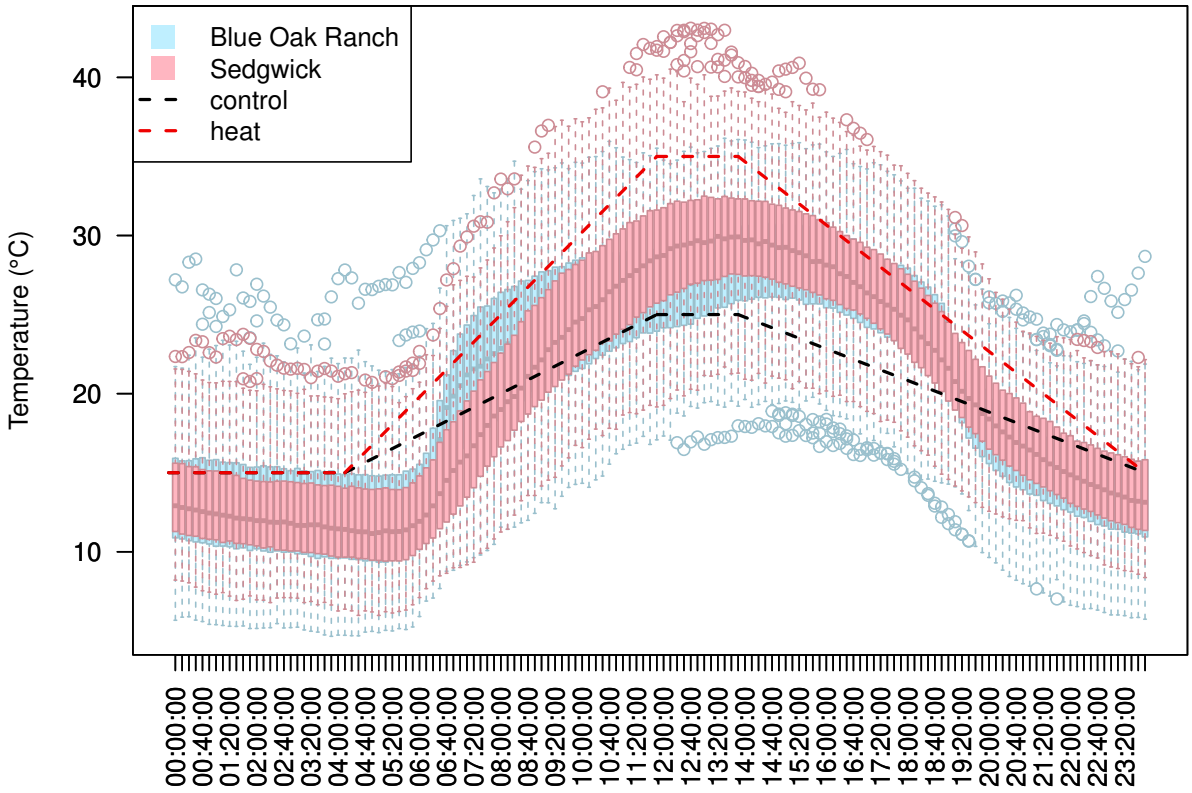


Figure 3.1. Comparison of historic July temperatures at Blue Oak Ranch and Sedgwick reserves with temperatures used in the control and heat stress treatments in this experiment. Boxplots show the air temperature for 10-minute increments across all days in July with recorded weather data from 2011-2020. Dotted lines show the temperatures in the control and heated growth chambers.

midnight (Figure 3.1). Growth chamber lights were on for 14 hours from 6:00 am to 8:00 pm. For the control treatment, the maximum temperature was 25 °C, and for the heat stress treatment the maximum temperature was 35 °C. Relative humidity was set to 50%. Plants were well-watered throughout the experiment to avoid the compounding effects of heat and drought stress.

From each population, 8-12 seedlings were included in each treatment. When enough seedlings were available we used a paired design, in which two seedlings from the same mother tree (likely half-siblings) were included, one in each treatment group, to better account for the effect of relatedness on gene expression patterns. We measured stem height and the number of fully ex-

panded leaves at the beginning and end of the treatment (July 2 and July 19). Relative growth rate (RGR) was calculated using stem heights at these two time points, using the following formula:

$$RGR = \frac{(\ln H_2 - \ln H_1)}{(t_2 - t_1)}$$

where H_1 and H_2 are the height at the beginning and end of the experiment, respectively, and $t_2 - t_1$ is the duration of the experiment in days.

Leaf thickness was measured with calipers before the experiment began (July 2). We also measured leaf temperature using a thermocouple at two time points in the experiment: the first full day (July 7) and the day before the end of the experiment (July 19). Leaf temperature was measured during the hottest part of the day, from about noon to 2:00 pm, alternating measurements between the control and treatment chambers throughout. On the last day of the experiment, leaves were collected from seedlings after the hottest part of the day (from 2:00 pm to 4:30 pm), flash frozen in liquid nitrogen, and stored at -80 °C. We collected fully expanded leaves that had not developed during the experiment to control for developmental effects across samples.

RNA extractions

We extracted RNA from frozen leaf tissue for 5 replicate individuals per group for a total of 60 samples (2 treatments x 3 species x 2 populations x 5 replicates). Approximately 50 mg of leaf tissue was ground in a tube for one minute at 30 Hz using a Retsch mixer mill (model MM 301), repeated twice, re-freezing tubes in liquid nitrogen between rounds of grinding. Polyphenolics and polysaccharides, which are common in oak leaves and interfere with downstream applications, were removed from leaves using a lithium chloride/urea-based pre-wash, made from 18.75 mL 8 M LiCl solution, 24 g Urea, 8 mL 11% PVP K-60 solution, and 0.5 mL 1 M Dithiothreitol (DTT). After grinding, 1.8 mL of the pre-wash extraction buffer was added to the tube, ground for 10 seconds at 30 Hz to resuspend tissue in buffer, and centrifuged for 10 minutes, 1000 RCF at 4 °C. 1.4 mL of supernatant was transferred to a new tube and placed on ice at 4 °C overnight.

The next day, the supernatant was spun in a centrifuge at 4 °C for 30 minutes at 20,000 RCF. Supernatant was discarded, and the resulting tissue pellet was washed by adding 500 ul 70% ethanol then spinning for 5 minutes at 5,000 RCF, repeated twice, then allowing ethanol to evaporate by air drying for 10 minutes. After this prewash procedure, RNA was extracted from the remaining pellet using either the Sigma-Aldrich Spectrum Plant Total RNA kit or Qiagen RNeasy Plant Mini Kit. RNA quality was checked using a NanoDrop and Qubit Fluorometer.

Extracted RNA was sent to the Genomic Sequencing and Analysis Facility at the University of Texas at Austin for library prep and sequencing. We used TagSeq, a method which sequences only the 3' end of the RNA strand rather than the entire transcript, which reduces the coverage necessary for differential expression analysis (Lohman et al. 2016, Marx et al. 2020). The library prep protocol is provided by the University of Texas at Austin (Aglyamova et al. 2019). Libraries were sequenced on a NovaSeq 6000 using single-end 100 basepair reads, and each library was run on two lanes to avoid lane effects.

RNAseq read processing

Read files were processed using the scripts available from UT Austin (Matz 2021). We removed duplicated reads with the 'tagseq_clipper' script. Adapters and poly-A tails were trimmed using Cutadapt (version 2.3, Martin 2011). Reads were trimmed for quality using a cutoff of 15, and reads with a length <20 bp were removed. Reads were aligned to the *Q. lobata* genome version 3.2 (Sork et al. 2022) using STAR (version 2.7.10b, Dobin et al. 2013). Reads that were not uniquely mapped (with a MAPQ score <255) were removed. The number of reads present for each sample was counted using htseq-count from the python package HTseq (version 0.13.5), with setting stranded=yes.

We obtained a list of pairwise orthologs between the *Q. lobata* and *Arabidopsis thaliana* reference genomes from the OMA database (Altenhoff et al. 2021). We also obtained annotations, including gene ontology (GO) terms for the *Arabidopsis* genome from the Arabidopsis Information

Resource (TAIR). We selected the orthologs with one-to-one matches between the two species and applied the *Arabidopsis* annotations to the oak genes.

Analysis

Growth rates and leaf temperatures were analyzed with a linear model ('lm' function in R) to detect effects of the provenance (source site), treatment, and provenance × treatment interaction, with family (nested within provenance and species) included in the model to account for similarity due to high relatedness.

We tested for differentially expressed (DE) genes using the R package limma (version 3.52.4, Ritchie et al. 2015), as well as some functions from edgeR (version 3.38.4, Robinson et al. 2010). First we filtered out genes with low read counts using 'filterByExpr', removing genes that did not have a minimum of five reads in at least some samples, and a minimum of 30 reads in total across all samples, which reduced the number of genes from 18,209 to 10,562. We tested for genes that were differentially expressed between the control and heat stress experiments for each provenance within a species, including a term for family nested within species and provenance to control for the effects of relatedness on gene expression.

We tested for GO terms that were significantly enriched in each set up upregulated or downregulated genes using GOseq 1.48.0 (Young et al. 2010). GOseq takes a list of significant and non-significant DE genes and weights them by gene length, because longer genes have more reads and are more likely to be DE. However, since TagSeq sequences only a short portion of 3' end of the transcript for all genes, gene length bias is not expected, and we gave all genes equal weighting. Genes without assigned GO terms were ignored in the analysis. GOseq returns results for tests of whether a given GO term is overrepresented and underrepresented. We focused on the overrepresented terms, and adjusted *P*-values using the Benjamini-Hochberg procedure for genes that were more likely to be over-represented than under-represented. To visualize overall differences in the expression of functional groups across species and provenances, we took the

set of enriched GO terms and calculated the average log fold change between heat and control seedlings per group for all genes associated with that GO term.

Gene networks

We used weighted gene co-expression network analysis (WGCNA) to explore broader patterns of gene expression responses among populations and species (Langfelder and Horvath 2008). In this analysis, genes are grouped into clusters based on variation in their expression across samples, and the expression of the group of genes (or eigengene) can be correlated with traits and compared across treatment groups. We created a network for each species (20 individuals each) using the same parameters. A signed network was used, meaning gene were only grouped together if they were positively correlated. The minimum module size was 50 genes, and modules with a correlation >0.8 were merged together (cutHeight = 0.2). Modules are assigned names of colors. We identified modules with expression values that were associated with treatment, provenance, and their interaction using the 'lm' function in R.

Results

Heat stress experiment

The heat stress experiment caused seedlings to have higher leaf temperatures in comparison to the control seedlings both at the beginning and end of the experiment (P <2e-16 for both, Figure 3.2). Leaf temperatures in the heat treatment were slightly lower than the air temperature (35 °C), indicating that there may have been some transpirational cooling occurring. In contrast, seedlings in the control treatment had temperatures slightly higher than air temperature. The effect of the heat stress treatment on leaf temperature was consistent across all species and populations: there was not a significant treatment × species or treatment × provenance effect. However, there was a significant effect of species on leaf temperature for the measurements taken at the begin-

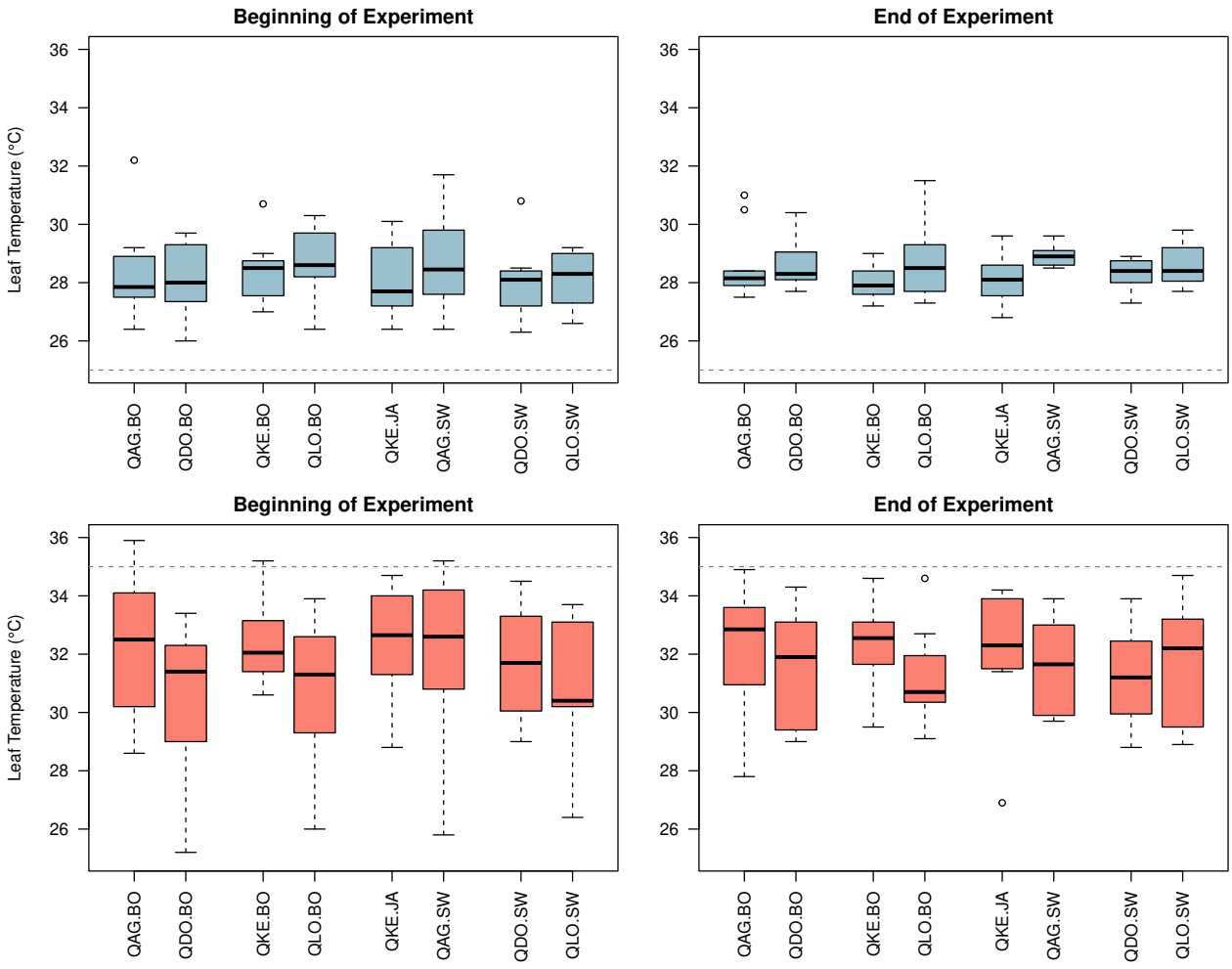


Figure 3.2. Leaf temperature across populations measured after the hottest part of the day in the control group (top) and heat treatment (bottom) at the beginning and end of the experiment. Dotted lines indicate the air temperatures for the control growth chamber (25 °C) and heat stress growth chamber (35 °C) for comparison.

ning of the experiment (P = 0.05). The difference in leaf temperature between the two time points was small and did not vary among treatment or population groups, indicating the transpirational response was constant over the course of the experiment. Heat treatment did not affect the relative growth rates of seedling height (P = 0.23, Figure 3.3), suggesting that their fitness did not decrease when exposed to higher temperatures. However, growth rates did vary among species (P = 0.05) and family (P = 0.02), indicating a genetic effect.

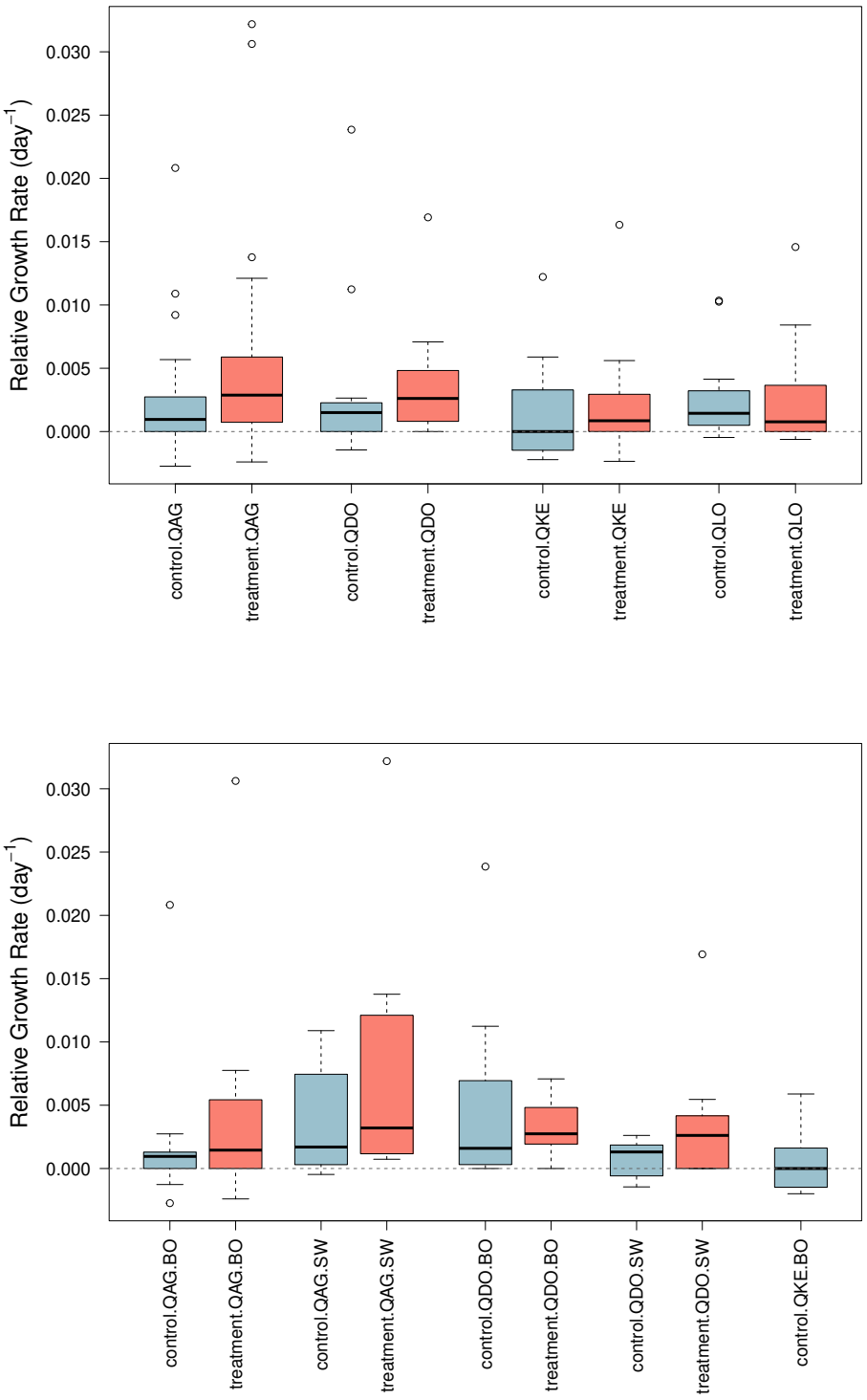


Figure 3.3. Relative growth rates of each species (top) and population (bottom) over the course of the experiment compared across treatments.

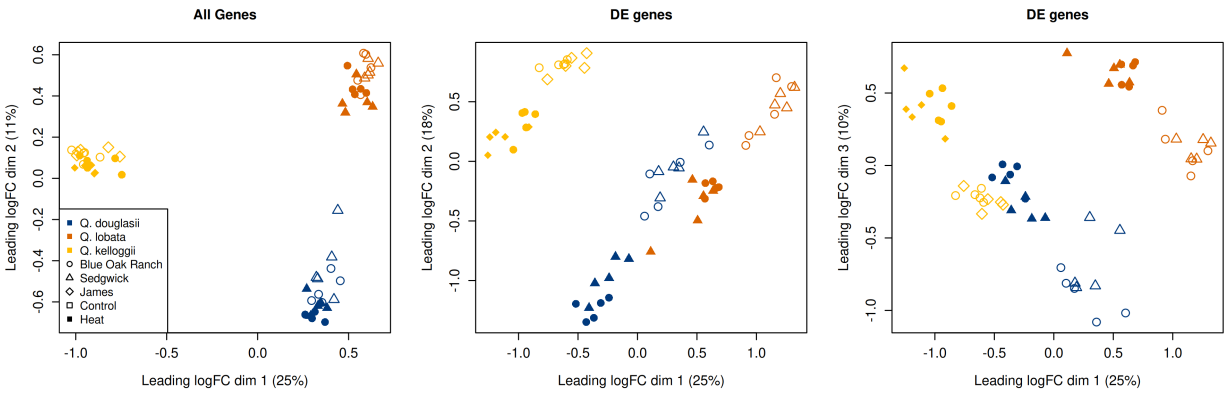


Figure 3.4. MDS plots indicating approximate log fold change between individual seedlings. Open symbols indicate seedlings from the heat treatment and closed circles indicate seedlings from the control.

Differential expression

Species strongly differed in the expression patterns of the entire set of genes (Figure 3.4), while populations within species did not cluster together. Treatment also affected clustering, even within the set of all genes. When only genes that were significantly differentially expressed in at least one population, differences in treatment were more dramatic. Populations did not appear to cluster separately within a treatment, suggesting responses were similar across populations.

The number of differentially expressed (DE) genes responding to heat stress in a population ranged from 13 to 158, with a total of 421 genes that were DE in at least on population (Table 1). The majority of these genes were specific to one population, suggesting few shared responses across species or populations. However, most of these genes had the responded in the same direction (either upregulated or downregulated) across groups, even if they were not significant (Figure 3.5), which may indicate that the lack of shared genes is due to low power rather than a true biological pattern.

Genes that were upregulated in response to the heat wave included heat shock proteins and chaperonins, including four genes that were significantly upregulated for all six populations: LOC115949676, LOC115958111, LOC115968877, and LOC115984301. The sets of upregulated gene sets across all populations was enriched for "induction of programmed cell death." As with in-

Table 3.1. Number of significantly differentially expressed genes in response to heat treatment in each population.

	Qdo.SW	Qdo.BO	Qke.JA	Qke.BO	Qlo.SW	Qlo.BO
Downregulated	41	76	34	60	50	11
Upregulated	45	82	44	40	40	2
Total	86	158	78	100	90	13

dividual genes, few GO terms were significantly enriched across all populations, but most responded in the same direction across all populations, including upregulation of typical heat stress responses, such as "heat acclimation," "chaperone binding," and "response to light intensity."

Some responses did differ among populations. For example, "secondary metabolite biosynthetic process" was enriched in downregulated genes for Blue Oak Ranch *Q. douglasii* populations, and was more highly expressed in Blue Oak Ranch *Q. kelloggii* populations. The Blue Oak Ranch *Q. kelloggii* populations also upregulated expression of "secondary metabolite biosynthetic process," "growth," and "response to salicylic acid" in response to heat, and Blue Oak Ranch *Q. lobata* populations upregulated chloroplast and thylakoid related genes.

Gene networks

For *Q. douglasii*, 35 modules were identified, of which 9 were associated with treatment, and none were associated with provenance. For *Q. kelloggii*, 32 modules were identified, of which 8 were associated with treatment and 2 with provenance. For *Q. lobata*, 30 modules were identified, of which 8 were associated with treatment and 2 with provenance. There were no modules with a significant treatment × provenance interaction in any species, indicating that populations did not differ in their plastic response to treatment. This was supported by the clustering of samples by module expression for the treatment-associated modules (Figure 3.7): samples from the control and heat treatment formed separate clusters with contrasting module expression values, with no clear differences between populations from warm and cool sites.

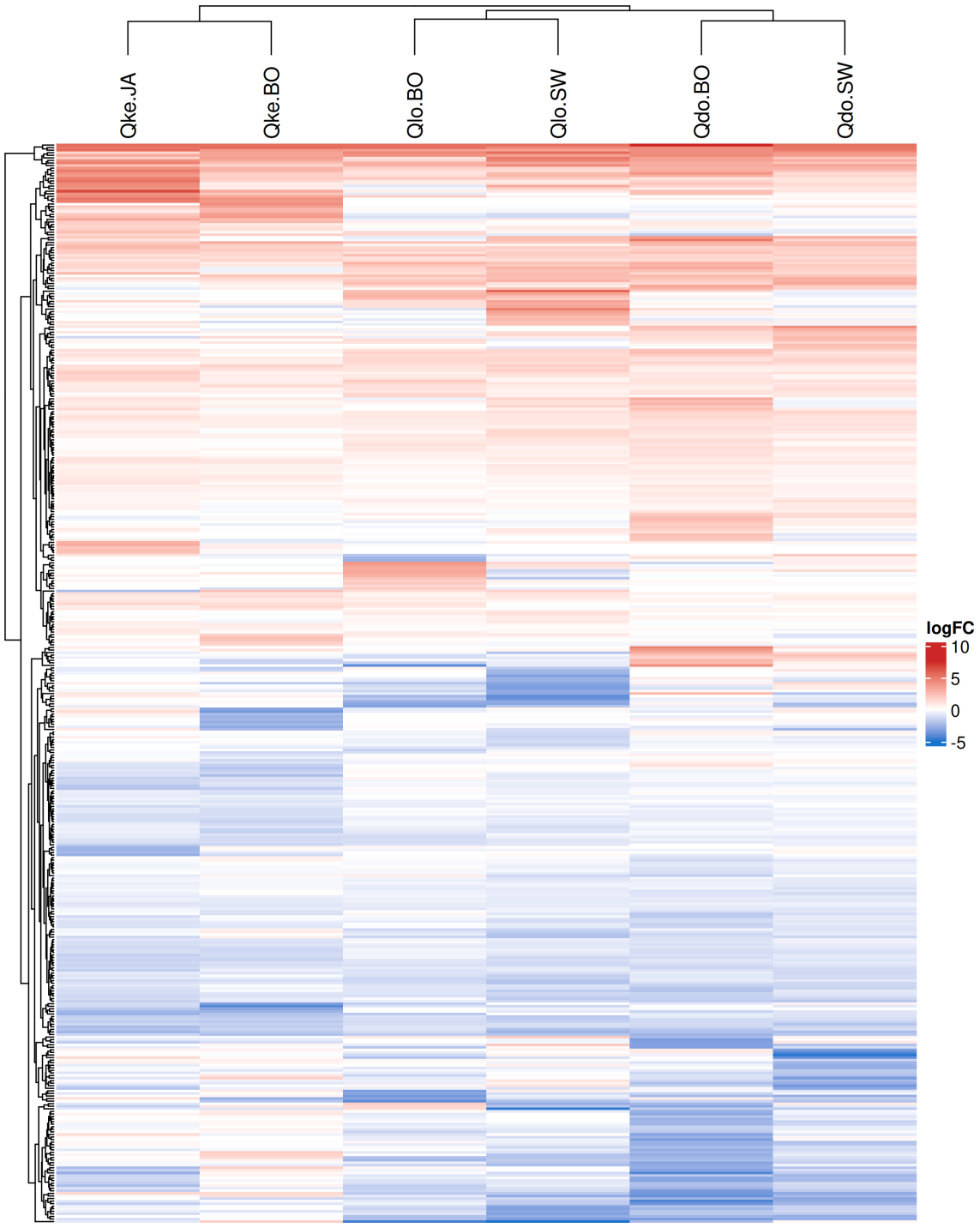


Figure 3.5. Log fold change across populations for all genes that are DE in at least one population. Dendrograms indicate hierarchical clustering of populations (columns) and genes (rows). Positive logFC values (red) indicate higher expression under the heat treatment and negative values (blue) indicate lower expression under the heat treatment.

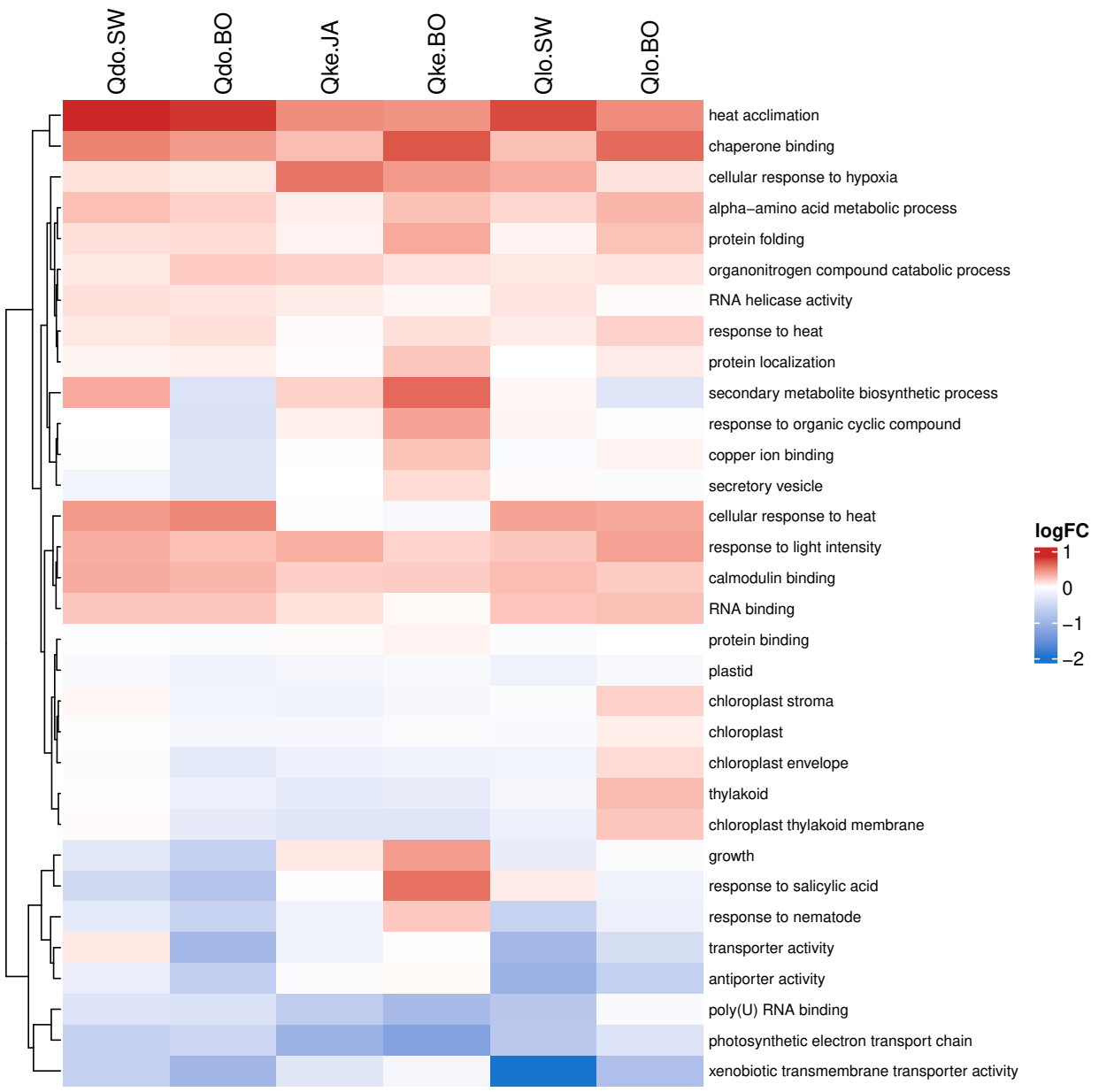
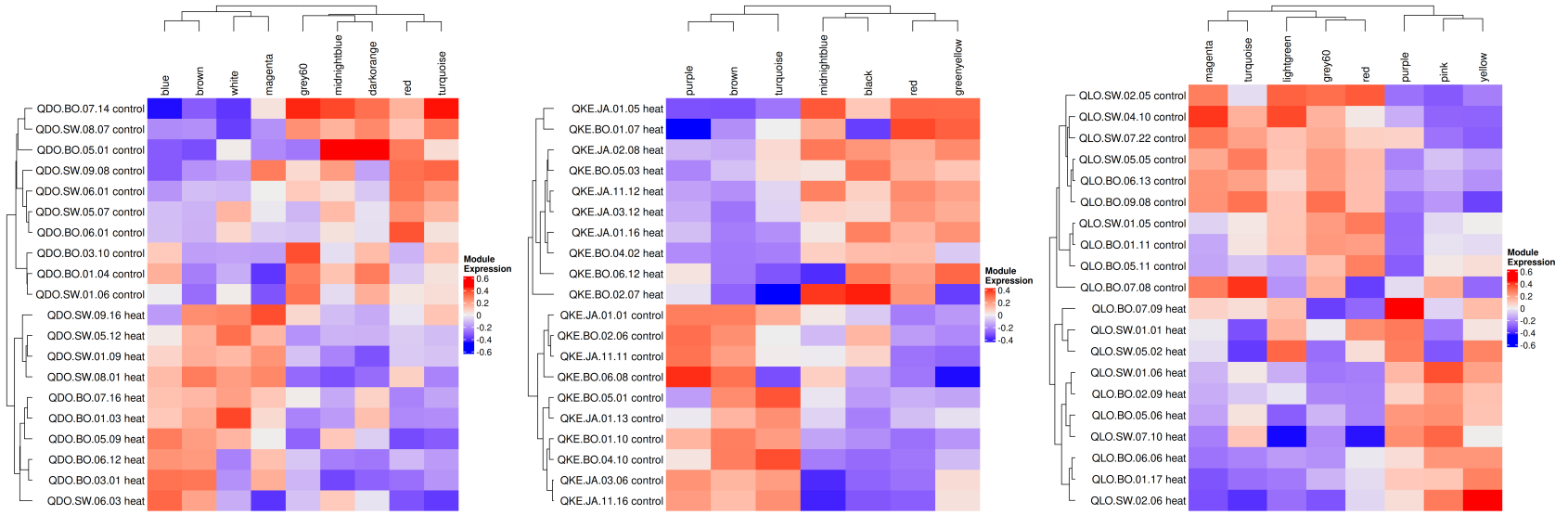


Figure 3.6. Log fold change between the heat and control treatments averaged across all genes associated with a GO term (including genes that are not significantly DE). GO terms that were significantly enriched within differentially expressed genes for at least one population are included. GO terms that were only associated with one gene were excluded for clarity.

Figure 3.7. Clustering of gene expression modules identified by WGCNA that differed between control and heat treated seedlings for each species. Individuals (rows) and modules (columns) are clustered by similarity. Individual IDs indicate the species, provenance, family, and individual number. The module expression is the eigenvalue of the expression of all genes included in the module and shows the overall expression of that module. Modules are assigned the names of colors to distinguish between them; module names are not comparable among species.



Discussion

Shared responses to heat stress across oak species

Four genes were significantly differentially expressed across all populations and species, indicating their importance across two of the major oak sections. Two of these genes were annotated as heat shock proteins (18.1 kDa class I heat shock protein-like and 17.3 kDa class II heat shock protein-like) and one was annotated as a chaperone regulator (BAG family molecular chaperone regulator 6). In addition to these shared genes, species were broadly similar in the types of genes that responded to heat stress, with heat response genes such as heat shock proteins and chaperonins being upregulated (Figure 3.6). These shared responses may contribute to the heat tolerance of the oak clade, or they may have originated earlier, as HSPs are essential across all land plants (Waters and Vierling 2020); regardless, they demonstrate the importance of these genes in the heat stress response.

Species differences in heat response

Variation in overall responses among species were consistent with their evolutionary history, with *Q. douglasii* and *Q. lobata*, both from section *Quercus*, being more similar to each other than to *Q. kelloggii* from section *Lobatae* (Figure 3.5). Additionally, some species-specific responses to heat stress may be associated with adaptation to their niches. While there is no prior research comparing physiological heat tolerance among these three species, their drought tolerance may provide insights into their responses to heat stress because of the relationship between transpiration and water availability. *Q. douglasii*, a relatively drought-tolerant species (Griffin 1976, Skelton et al. 2018), showed the strongest upregulation of heat acclimation genes when compared to the other two species (Figure 3.6). Genes that were upregulated only in *Q. douglasii* included genes coding for a small chloroplastic heat shock protein (LOC115971695), a DnaJ-like chaperone protein (LOC115989601), and a sorbitol dehydrogenase (LOC115960266) which is associated with

stress tolerance (Jia et al. 2015). These responses may enable it to survive at warm, sunny sites without consistent access to groundwater (Cannon 1914). *Q. kelloggii*, which occurs at high elevations and is likely to be less heat-tolerant, did not upregulate "cellular response to heat" genes, unlike the other two species, which could indicate a lack of an adaptive response. *Q. kelloggii* also downregulated TAP38 (LOC115960227, an ortholog of AT4G27800), which is involved in regulating state transitions to maintain redox balance of the photosynthetic electron transport chain under varying light conditions (Longoni and Goldschmidt-Clermont 2021).

Populations had few differences in their response to heat

We found limited evidence for local adaptation in the response to heat stress for all three oak species. Leaf temperatures did not differ among populations, suggesting similar transpirational responses to heat. In general, treatment had a larger effect on the gene expression profile of an individual than its source population (Figure 3.4). However, populations from warmer and cooler sites did differ in some of their gene expression responses. For example, *Q. kelloggii* from Blue Oak Ranch site upregulated growth and salicylic acid genes (Figure 3.6), suggesting they may sustain growth through warm periods, although it is unclear whether this would be beneficial at a cooler site. *Q. lobata* from Blue Oak Ranch upregulated chloroplast and thylakoid genes, which may be a response to maintain chloroplast membrane function under higher temperatures (Sharkey 2005). Overall, though, population differences were limited. None of the gene networks differed in the plastic response across treatments (provenance × treatment effect, Figure 3.7), indicating that responses were shared across populations.

Previous studies testing whether tree populations differ in their stress responses and ecophysiological traits in trees have found mixed results. Some found evidence of different responses among populations, consistent with the expectation that populations are locally adapted (Konôpková et al. 2018). Conversely, some studies found more variation among individuals than among populations (Hess et al. 2016, Skelton et al. 2019). High gene flow and an outcrossing mating

system combined with a lack of selection at the population level could explain this pattern, but previous studies in oaks have found evidence of population differences in gene expression in response to drought stress (Mead et al. 2019), so it seems unlikely gene flow is high enough to prevent local adaptation. A lack of population differences could also suggest canalization of traits driven by strong selection across the entire species (Skelton et al. 2019). Frontloading, or constitutive upregulation, may be more important in adaptation to heat than plasticity (Barshis et al. 2013, Collins et al. 2021).

It is also possible that the level of heat stress experienced by the seedlings was not sufficient to induce population-specific responses. Seedlings in the heat treatment had growth rates similar to control seedlings (Figure 3.3). In fact, the average relative growth rate was slightly higher in heat-treated seedlings for *Q. agrifolia*, *Q. douglasii*, and *Q. kelloggii*. This difference was not significant and only reflects changes in height, which may differ from biomass accumulation. However, this result may hint that higher temperatures could be beneficial to some oak species, as found previously (Arend et al. 2013). Additionally, heat stress alone may not be as harmful as the combination of heat and drought stress that is typical of California summers, because water access enables plants to regulate their temperature through transpiration (Bauweraerts et al. 2013, Chaves et al. 2016). When water is not available, actual temperatures experienced at the tissue level may be much higher. Seedlings in the heat treatment were well-watered in our experiment, and leaf temperatures were lower than air temperatures. While seedlings did respond with heat-related genes such as heat shock proteins, these responses may have been sufficient to avoid reductions in fitness.

The minor differences among populations that we observed did not vary between the warm and cool sites in the same way for each species. Instead, population differences in gene expression were specific to each species (Figures 5 and 6). Unlike the results from Chapter 1, the magnitude of the gene expression response was not associated with the source location. In *Q. douglasii* and *Q. kelloggii*, the populations from cooler sites responded with more genes, while for *Q. lobata*

the population from the warmer site responded with more genes. Previous studies have found that different species adapted to warmer sites by evolving the same gene expression responses to heat, but this was hypothesized to result from selection on standing variation from before the two species diverged (Zhao et al. 2015). Since parallelism is more likely in recently diverged taxa (Bohutínská et al. 2021, Conte et al. 2012), these oak species may not have shared standing variation recently enough for local adaptation to act similarly.

Conclusions

We found responses to increased temperature that were shared across three oak species, including both individual genes and overall patterns in the types of genes. Species differed in some ways that may explain their habitat preferences. Populations from warmer and cooler sites had limited differences in their response, potentially indicating a lack of local adaptation. Additionally, observed population differences were not consistent across species, indicating that, if local adaptation is present, selection has targeted different responses in each species. Overall, seedlings grew equally well in the control and heat treatments, indicating that these shared responses may contribute to their overall heat tolerance. These oak species may be well-adapted to moderate heat stress when they are also well-watered. Future increases in the frequency and severity of heat waves may be more likely to reduce fitness of oaks when combined with increased drought stress, and differences in tolerance among species and populations may be more evident under such conditions.

Acknowledgements

We acknowledge the Native Peoples of California as the traditional caretakers of the oak ecosystems sampled for this study. We thank the University of California Natural Reserve System for sampling permission. Samples were collected from Blue Oak Ranch Reserve, Sedgwick Reserve,

and James San Jacinto Mountain Reserve. Aldo De La Mora Rodriguez, Zac Harlow, Jennifer Hunter, and Shane Waddell provided help with collection logistics. Julia Lindner helped with planting and collecting growth and germination data, and Weimin Deng helped with planting and taking care of the seedlings. We also thank the Genomic Sequencing and Analysis Facility at The University of Texas at Austin for library prep and sequencing services.

References

Abrams, M. D. (1990). Adaptations and responses to drought in *Quercus* species of North America. *Tree Physiology* 7(1-2-3-4):227–238.

Abrams, M. D. (1994). Genotypic and phenotypic variation as stress adaptations in temperate tree species: a review of several case studies. *Tree Physiology* 14(7-8-9):833–842.

Aglyamova, G., J. Podnar, and M. Matz (2019). TagSeq library preparation. `https://github.com/z0on/tag-based_RNAseq/blob/master/TagSeq_sample_prep_june2019.docx`

Allen, C. D., A. K. Macalady, H. Chenchouni, D. Bachelet, N. Mcdowell, M. Vennetier, T. Kitzberger, A. Rigling, D. D. Breshears, E. H. T. Hogg, P. Gonzalez, R. Fensham, Z. Zhang, J. Castro, N. Demidova, J.-h. Lim, G. Allard, S. W. Running, A. Semerci, and N. Cobb (2010). A global overview of drought and heat-induced tree mortality reveals emerging climate change risks for forests. *Forest Ecology and Management* 259:660–684.

Altenhoff, A. M., C.-M. Train, K. J. Gilbert, I. Mediratta, T. Mendes de Farias, D. Moi, Y. Nevers, H.-S. Radoykova, V. Rossier, A. Warwick Vesztrocy, N. M. Glover, and C. Dessimoz (2021). OMA orthology in 2021: website overhaul, conserved isoforms, ancestral gene order and more. *Nucleic Acids Research* 49(D1):D373–D379.

Ameye, M., T. M. Wertin, I. Bauweraerts, M. A. McGuire, R. O. Teskey, and K. Steppe (2012). The effect of induced heat waves on *Pinus taeda* and *Quercus rubra* seedlings in ambient and elevated CO2 atmospheres. *New Phytologist* 196(2):448–461.

Arend, M., A. Brem, T. M. Kuster, and M. S. Günthardt-Goerg (2013). Seasonal photosynthetic responses of European oaks to drought and elevated daytime temperature. *Plant Biology* 15(s1):169–176.

Barshis, D. J., J. T. Ladner, T. A. Oliver, F. O. Seneca, N. Traylor-Knowles, and S. R. Palumbi (2013). Genomic basis for coral resilience to climate change. *Proceedings of the National Academy of Sciences* 110(4):1387–1392.

Bauweraerts, I., T. M. Wertin, M. Ameye, M. A. McGuire, R. O. Teskey, and K. Steppe (2013). The effect of heat waves, elevated [CO2] and low soil water availability on northern red oak (*Quercus rubra* L.) seedlings. *Global Change Biology* 19(2):517–528.

Bigras, F. J. (2000). Selection of white spruce families in the context of climate change: heat tolerance. *Tree Physiology* 20(18):1227–1234.

Bohutínská, M., J. Vlček, S. Yair, B. Laenen, V. Konečná, M. Fracassetti, T. Slotte, and F. Kolář (2021). Genomic basis of parallel adaptation varies with divergence in *Arabidopsis* and its relatives. *Proceedings of the National Academy of Sciences* 118(21).

Bonner, F. T. and R. P. Karrfalt (2008). The woody plant seed manual. *Agric. Handbook No. 727. Washington, DC. U.S. Department of Agriculture, Forest Service. 1223 p.* 727.

Cannon, W. A. (1914). Specialization in vegetation and in environment in California. *The Plant World* 17(8):223–237.

Carroll, S. B. (2005). Evolution at two levels: On genes and form. *PLoS Biology* 3(7):1159–1166.

Cavender-Bares, J. (2018). Diversification, adaptation, and community assembly of the American oaks (*Quercus*), a model clade for integrating ecology and evolution. *New Phytologist* 221(2):669–692.

Chaves, M. M., J. M. Costa, O. Zarrouk, C. Pinheiro, C. M. Lopes, and J. S. Pereira (2016). Controlling stomatal aperture in semi-arid regions—The dilemma of saving water or being cool? *Plant Science* 251:54–64.

Chen, B., M. E. Feder, and L. Kang (2018). Evolution of heat-shock protein expression underlying adaptive responses to environmental stress. *Molecular Ecology* 27(15):3040–3054.

Collins, M., M. S. Clark, J. I. Spicer, and M. Truebano (2021). Transcriptional frontloading contributes to cross-tolerance between stressors. *Evolutionary Applications* 14(2):577–587.

Conte, G. L., M. E. Arnegard, C. L. Peichel, and D. Schluter (2012). The probability of genetic parallelism and convergence in natural populations. *Proceedings of the Royal Society B: Biological Sciences* 279(1749):5039–5047.

Daas, C., P. Montpied, B. Hanchi, and E. Dreyer (2008). Responses of photosynthesis to high temperatures in oak saplings assessed by chlorophyll-a fluorescence: inter-specific diversity and temperature-induced plasticity. *Annals of Forest Science* 65(3):305.

Dendra Science (2022). UC Natural Reserve System. `https://dendra.science/orgs/ucnrs`

Dobin, A., C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, and T. R. Gingeras (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29(1):15–21.

Griffin, J. R. (1976). Regeneration in *Quercus lobata* savannas, Santa Lucia Mountains, California. *The American Midland Naturalist* 95(2):422–435.

Gugger, P. F., S. J. Cokus, and V. L. Sork (2016a). Association of transcriptome-wide sequence variation with climate gradients in valley oak (*Quercus lobata*). *Tree Genetics & Genomes* 12(2):1–14.

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

Gugger, P. F., J. M. Peñaloza-Ramírez, J. W. Wright, and V. L. Sork (2016b). Whole-transcriptome response to water stress in a California endemic oak, *Quercus lobata*. *Tree Physiology* 00:1–13.

Guha, A., J. Han, C. Cummings, D. A. McLennan, and J. M. Warren (2018). Differential ecophysiological responses and resilience to heat wave events in four co-occurring temperate tree species. *Environmental Research Letters* 13(6):065008.

Halter, G., N. Simonetti, C. Suguitan, K. Helm, J. Soroksky, and E. R. Waters (2017). Patterns of thermotolerance, chlorophyll fluorescence, and heat shock gene expression vary among four *Boechera* species and *Arabidopsis thaliana*. *Botany* 95(1):9–27.

Hamerlynck, E. and A. K. Knapp (1996). Photosynthetic and stomatal responses to high temperature and light in two oaks at the western limit of their range. *Tree Physiology* 16(6):557–565.

Hess, M., H. Wildhagen, L. V. Junker, and I. Ensminger (2016). Transcriptome responses to temperature, water availability and photoperiod are conserved among mature trees of two divergent Douglas-fir provenances from a coastal and an interior habitat. *BMC Genomics* 17(1):1–18.

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

IPCC (2014). Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Technical report, Intergovernmental Panel on Climate Change. `https://www.ipcc.ch/report/ar5/syr/`

Jia, Y., D. C. Wong, C. Sweetman, J. B. Bruning, and C. M. Ford (2015). New insights into the evolutionary history of plant sorbitol dehydrogenase. *BMC Plant Biology* 15(1):101.

Jones, F. C., M. G. Grabherr, Y. F. Chan, P. Russell, E. Mauceli, J. Johnson, R. Swofford, M. Pirun, M. C. Zody, S. White, E. Birney, S. Searle, J. Schmutz, J. Grimwood, M. C. Dickson, R. M. Myers, C. T. Miller, B. R. Summers, A. K. Knecht, S. D. Brady, H. Zhang, A. A. Pollen, T. Howes, C. Amemiya, Broad Institute Genome Sequencing Platform & Whole Genome Assembly Team,

E. S. Lander, F. Di Palma, K. Lindblad-Toh, and D. M. Kingsley (2012). The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* 484(7392):55–61.

Kenkel, C. D. and M. V. Matz (2016). Gene expression plasticity as a mechanism of coral adaptation to a variable environment. *Nature Ecology & Evolution* 1(3):0014.

Konôpková, A., D. Kurjak, J. Kmeť, R. Klumpp, R. Longauer, Ľ. Ditmarová, and D. Gömöry (2018). Differences in photochemistry and response to heat stress between silver fir (*Abies alba* Mill.) provenances. *Trees* 32(1):73–86.

Langfelder, P. and S. Horvath (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 9(559).

Lohman, B. K., J. N. Weber, and D. I. Bolnick (2016). Evaluation of TagSeq, a reliable low-cost alternative for RNAseq. *Molecular Ecology Resources* 16(6):1315–1321.

Longoni, F. P. and M. Goldschmidt-Clermont (2021). Thylakoid protein phosphorylation in chloroplasts. *Plant and Cell Physiology* 62(7):1094–1107.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 17(1):10–12.

Marx, H. E., S. Scheidt, M. S. Barker, and K. M. Dlugosch (2020). TagSeq for gene expression in non-model plants: A pilot study at the Santa Rita Experimental Range NEON core site. *Applications in Plant Sciences* 8(11):e11398.

Matz, M. (2021). Genome-wide gene expression profiling with tag-based RNA-seq (TagSeq). `https://github.com/z0on/tag-based_RNAseq/`

Mead, A., J. Peñaloza Ramirez, M. K. Bartlett, J. W. Wright, L. Sack, and V. L. Sork (2019). Seedling response to water stress in valley oak (*Quercus lobata*) is shaped by different gene networks across populations. *Molecular Ecology* 28(24):5248–5264.

Moles, A. T., S. E. Perkins, S. W. Laffan, H. Flores-Moreno, M. Awasthy, M. L. Tindall, L. Sack, A. Pitman, J. Kattge, L. W. Aarssen, M. Anand, M. Bahn, B. Blonder, J. Cavender-Bares, J. H. C. Cornelissen, W. K. Cornwell, S. Díaz, J. B. Dickie, G. T. Freschet, J. G. Griffiths, A. G. Gutierrez, F. A. Hemmings, T. Hickler, T. D. Hitchcock, M. Keighery, M. Kleyer, H. Kurokawa, M. R. Leishman, K. Liu, Ü. Niinemets, V. Onipchenko, Y. Onoda, J. Penuelas, V. D. Pillar, P. B. Reich, S. Shiodera, A. Siefert, E. E. Sosinski, N. A. Soudzilovskaia, E. K. Swaine, N. G. Swenson, P. M. v. Bodegom, L. Warman, E. Weiher, I. J. Wright, H. Zhang, M. Zobel, and S. P. Bonser (2014). Which is a better predictor of plant traits: temperature or precipitation? *Journal of Vegetation Science* 25(5):1167–1180.

Ohama, N., H. Sato, K. Shinozaki, and K. Yamaguchi-Shinozaki (2017). Transcriptional regulatory network of plant heat stress response. *Trends in Plant Science* 22(1):53–65.

Ritchie, M. E., B. Phipson, D. Wu, Y. Hu, C. W. Law, W. Shi, and G. K. Smyth (2015). *limma* powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research* 43(7).

Robinson, M. D., D. J. McCarthy, and G. K. Smyth (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26(1):139–140.

Sharkey, T. D. (2005). Effects of moderate heat stress on photosynthesis: importance of thylakoid reactions, rubisco deactivation, reactive oxygen species, and thermotolerance provided by isoprene. *Plant, Cell & Environment* 28(3):269–277.

Shinozaki, K. and K. Yamaguchi-Shinozaki (2022). Functional genomics in plant abiotic stress responses and tolerance: From gene discovery to complex regulatory networks and their application in breeding. *Proceedings of the Japan Academy, Series B* 98(8):470–492.

Skelton, R. P., L. D. L. Anderegg, P. Papper, E. Reich, T. E. Dawson, M. Kling, S. E. Thompson, J. Diaz, and D. D. Ackerly (2019). No local adaptation in leaf or stem xylem vulnerability to

embolism, but consistent vulnerability segmentation in a North American oak. *New Phytologist* 223(3):1296–1306.

Skelton, R. P., T. E. Dawson, S. E. Thompson, Y. Shen, A. P. Weitz, and D. Ackerly (2018). Low vulnerability to xylem embolism in leaves and stems of North American oaks. *Plant Physiology* 177(3):1066–1077.

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

Sork, V. L., Erin Riordan, Paul F. Gugger, Sorel Fitz-Gibbon, Xinzeng Wei, and Joaquín Ortego (2016). Phylogeny and introgression of California scrub white oaks (*Quercus* section *Quercus*). *International Oaks* 27.

Teskey, R., T. Wertin, I. Bauweraerts, M. Ameye, M. A. Mcguire, and K. Steppe (2015). Responses of tree species to heat waves and extreme heat events. *Plant, Cell & Environment* 38(9):1699–1712.

Tucker, T. J. R., John M. (2014a). *Quercus agrifolia*. *In* Jepson Flora Project, Jepson eFlora. The Jepson Herbarium, revision 2 edition.

Tucker, T. J. R., John M. (2014b). *Quercus douglasii*. *In* Jepson Flora Project, Jepson eFlora. The Jepson Herbarium, revision 2 edition.

Tucker, T. J. R., John M. (2014c). *Quercus kelloggii*. *In* Jepson Flora Project, Jepson eFlora. The Jepson Herbarium, revision 2 edition.

Tucker, T. J. R., John M. (2014d). *Quercus lobata*. *In* Jepson Flora Project, Jepson eFlora. The Jepson Herbarium, revision 2 edition.

Waters, E. R. and E. Vierling (2020). Plant small heat shock proteins – evolutionary and functional diversity. *New Phytologist* 227(1):24–37.

Weston, D. J. and W. L. Bauerle (2007). Inhibition and acclimation of C3 photosynthesis to moderate heat: a perspective from thermally contrasting genotypes of *Acer rubrum* (red maple). *Tree Physiology* 27(8):1083–1092.

Weston, D. J., L. E. Gunter, A. Rogers, and S. D. Wullschleger (2008). Connecting genes, coexpression modules, and molecular signatures to environmental stress phenotypes in plants. *BMC Systems Biology* 2(16).

Young, M. D., M. J. Wakefield, G. K. Smyth, and A. Oshlack (2010). Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biology* 11(2):R14.

Zhao, L., J. Wit, N. Svetec, and D. J. Begun (2015). Parallel gene expression differences between low and high latitude populations of *Drosophila melanogaster* and *D. simulans*. *PLoS Genetics* 11(5):1–25.