UNIVERSITY OF CALIFORNIA

Los Angeles

# Comparing phenotypic and genetic variation in California walnuts (*Juglans californica* and *J. hindsii*) to resolve species identities and scan for climate adaptations

A thesis submitted in partial satisfaction

of the requirements for the degree

Master of Science in Biology

by

Diego Julian Zapata

2024

# ABSTRACT OF THE THESIS

## Comparing phenotypic and genetic variation in California walnuts (*Juglans californica* and *J. hindsii*) to resolve species identities and scan for climate adaptations

by

Diego Julian Zapata

Master of Science in Biology

University of California, Los Angeles, 2024

Professor Victoria Sork, Chair

*Juglans californica* and *J. hindsii* are sister taxa endemic to California with significant economic, ecological, and cultural relevance to the state. Each taxon is a foundation species for a rare and threatened plant community, the walnut woodland, in their respective regions. Both species are distributed throughout much of California, but their species identities still need clarification due to some authorities' treatment of both taxa as subspecies of *Juglans californica*. The future of either taxon will need to confront climate change within a highly urbanized and fragmented landscape, where the usual options for long-living tree species are to adapt, disperse, or die, but landscape genomic-based predictions cannot occur without a clear understanding of how each taxon is genetically and ecologically distinct. Using herbarium specimens and whole-genome sequences for 104 field-identified *Juglans californica* sampled throughout California, we describe spatially explicit patterns of leaf morphology, genetic structure, and genetic diversity in both taxa and leverage genetic-environment associations to describe distinct patterns of local adaptation that corroborate that these taxa are should be treated as different species. Genetic structure analyses suggest that *J. californica* is restricted to the Los Angeles Metropolitan area and the Transverse Ranges. *J. hindsii* occurs throughout much of Northern California and extends into the southmost portion of the species range in San Diego County. Hybridization with a third cluster occurs throughout both of the species. Leaf trait values, including abaxial vein axil hair area, leaflet number, and petiole length, significantly differ between species and corroborate genetic structure results. *J. hindsii* has higher genetic diversity than *J. californica* (nucleotide diversity, biallelic richness, and observed heterozygosity) and local adaptation patterns strongly influenced by winter bioclimatic variables based on GEA analyses. *J. californica* has local adaptation patterns that are strongly influenced by summer temperature and water availability. These results suggest that each species may respond differently to climate change projections and require species-specific protections against ongoing habitat loss and climate change.

The thesis of Diego Julian Zapata is approved.

Felipe Zapata

Kirk Edward Lohmueller

Victoria Sork, Committee Chair

University of California, Los Angeles

2024

**Table of Contents**

## LIST OF FIGURES

# LIST OF TABLES

## 1. Introduction

The endemic black walnut (*Juglans* spp.) trees of California are a unique and important species in ecosystems in the state but continue to remain cryptic among conservation practitioners and scholars regarding their species identity and conservation risk. Molecular evidence supports the classification of two distinct *Juglans* species, *Juglans californica* S. Watson (southern California black walnut) and *Juglans hindsii* Jepson (Northern California black walnut) (Stone et al., 2009), but other authorities, like the California Native Plant Society and formerly the Jepson Manual, describe these species as varieties of the same species (*Juglans californica var. californica* and *Juglans californica var. hindsii*) (Anderson, 2002). The result is an age-old debate between lumpers and splitters that has only held back the pursuit and relevance of the limited research on either taxon despite a narrowing window of opportunity for the conservation of each species under the threat of ongoing habitat loss and climate change.

Ecologically, both trees serve as foundation species, providing significant structural heterogeneity, habitat, and essential food sources for associated species of wildlife and people (Quinn, 1989). Southern California black walnuts (SCBW) and Northern California black walnut (NCBW) trees dominate one of California's most charismatic and threatened ecosystems, the California walnut forests and woodland alliances, within their respective regions (CNPS, 2021). Southern California black walnut trees are low-growing, winter-deciduous hardwood trees that grow to a height of 15 meters. Their northern counterparts are morphologically and ecologically very similar but can be distinguished by their generally loftier height and canopy size and by subtle differences in their leaf morphology (Baldwin et al., 2012). Both California black walnut species occur on north- and east-facing slopes, within ravines and canyons, or along riparian corridors with deep alluvial soils throughout the state; their appropriate habitat is consequently specialized to sites that remain at least partially moist throughout much of the year (Longcore & Noujdina, 2022). Both species belong to the black walnut section (*Juglans* sect. *Rhysocaryon*). Historically, both taxa were widely used by indigenous peoples of California; the Tongva, Chumash, and Kumeyaay consumed SCBW nuts regularly when abundant, utilized the husks to produce a black dye used in basketry, and modified the nut shells into dice to play a game with significant social and political implications (Timbrook & Chapman, 2007; Wilken-Robertson, 2018). Many of these traditions persist amongst California's First Nations peoples as an expression of their culture and resistance. NCBW also has significant economic relevance for the state as a widely used species for locally disease-resistant rootstock for the state's walnut industry and to produce the "Paradox" (English walnut *J. regia* x NCBW *J. hindsii*) hybrid, a tree widely used to make fine lumber or as a street tree in northern California. The extinction of either taxon in their respective regions would accordingly incur irreparable losses to California's agricultural industry, natural resources, biodiversity, and cultural identity.

Despite their shared ecology, both taxa have been considered conservation priorities in California for distinct reasons. The Northern California black walnut was only recently considered to be a rare and threatened subspecies by the California Department of Fish and Wildlife and the California Native Plant Society due to a widely held belief that the subspecies was restricted to only three or four sites within Contra Costa, Sacramento, and Napa counties before European settlement in the mid-19th century led to widespread planting and naturalization outside of this purported pre-colonial range (Smith, 1909). NCBW's frequent outcrossing with English walnuts (*Juglans regia*) by the California walnut industry to develop the desirable rootstock "Paradox" hybrid further galvanized the perceived threat of genetic swamping for these pre-colonial sites until recent studies utilizing microsatellite markers revealed that the majority of putatively ‘wild’ J. hindsii individuals represented genetically pure members of the species (Potter et al., 2018). While finding no evidence of *Juglans regia* introgression within NCBW, this study did describe an appreciable amount of hybridization between *Juglans hindsii* and *Juglans californica,* given that there were several spontaneous and likely human-introduced populations of genetically pure and SCBW-hybridized NCBW trees co-occurring within SCBW’s range.

Southern CA black walnuts are currently listed under IUCN’s Red List of Threatened Species as ‘Near Threatened’ due to significant habitat loss leading to discrete and isolated populations vulnerable to ongoing land use change, fire, grazing, competitive exclusion by invasive species, and climate change (Stritch & Barstow, 2019). The southern CA black walnut’s main population center largely co-occurs with the Los Angeles metropolitan area, persisting in islands of hillsides that were previously too steep and suboptimal for urban development (Longcore & Noujdina, 2022). According to a species distribution model including land use change, the southern California black walnut has already lost 31% of its suitable habitat due to urbanization alone, distinguishing it as a significant conservation priority at odds with ongoing urban and economic growth in the second-largest metropolitan area of the United States (Riordan et al., 2015). Despite receiving ‘Protected Tree’ designations by the California Department of Fish and Wildlife, the California Native Plant Society, and under a City of Los Angeles local ordinance (Protected Tree Ordinance, 2006), a permit to remove a mature *Juglans californica* tree in the Los Angeles area is issued at a rate of every 7.2 days (CNPS, 2017; Longcore & Noujdina, 2022). Under future land use and climate change, the species is projected to lose 64–70% of existing and undeveloped suitable habitat, and the future of this species will depend on its capacity to disperse northward and out of highly fragmented urban areas (Riordan et al., 2015).

Regardless of their current conservation status, long-lived trees, like walnuts, are often considered among the most vulnerable taxa to climate change, given the high likelihood of such species incurring an adaptational lag. Adaptational lags occur because species cannot immediately adjust ecological or evolutionary adaptations to respond to rapid environmental changes, leading to a period during which their fitness and survival are compromised (Browne et al., 2019; Bisbing et al., 2021, Aitken et al., 2008). Because trees take much longer to reach reproductive maturity, have no mobility other than to disperse seeds, and usually rely on symbiosis or phenologic synchronicity for successful reproduction, an adaptational lag is likely to occur between where a tree currently exists and where its favorable habitat has been displaced (Sork et al., 2010, Browne et al., 2019). Unlike outright mortality, adaptational lags can manifest as a subtle reduction in fitness that may persist for over a century in trees (Vellend et al., 2006), silencing the "conservation alarm" by causing the time-delayed but deterministic extinction or extirpation for a given population whether or not perturbations continue to act upon a species. How well a tree species can migrate or adapt to these perturbations will ultimately dictate whether the species is a case of what Janzen has referred to as the “living dead”—“evolutionarily dead” individuals or entire metapopulations of a species that “continue to live out their physiological life” but which are incapable of “serving as a reproductive member of that species” (2001). Adaptational lags are already widely speculated to have been incurred across both CBW taxa due to a series of observations regarding age gap structures and seedling recruitment. Stark gaps in SCBW age classes between two study sites in Los Angeles County were hypothesized to be caused by increased seedling mortality and seed production failure following longer drought years and greater moisture variation between sites (Keeley, 1990). Others have noted the complete absence of natural California black walnut recruitment for both species in some portions of their current distributions (Tibor, 2001; CNPS, 2017). Because California black walnut seeds are recalcitrant and do not possess long-term dormancy mechanisms to await more favorable conditions, successful California black walnut reproduction will be contingently restricted as the increasing magnitude and length of drier and warmer conditions limit regeneration opportunities for the species (Longcore & Noujdina, 2022). In a climate scenario where increased intervals of aridity prevent the reproduction and migration of California black walnut trees, the probability that a fatal adaptational lag will be incurred markedly increases.

Leveraging landscape genomic approaches and whole-genome sequence data for these species provides the opportunity to compare patterns in genetic structure and to scan for the loci involved with local climate adaptations, which will resolve uncertainties regarding the genomic differentiation between both walnut taxa and can further identify how these taxa are ecologically distinct from one another based on gene-environment associations. Gene-environment association studies (GEA) are powerful tools used to identify genetic variations associated with specific environmental factors. These studies can provide crucial insights into how species, such as long-lived trees, adapt to local climatic conditions and help address the challenges of adaptational lags across the landscape scale (Sork et al., 2013). Many studies have used a landscape genomic approach to identify vulnerable regions for tree species under climate change (Gugger et al., 2020; Jia et al., 2019; Martins et al., 2018; Steane et al, 2014). Given that an adaptive phenotype is the product of an entire genome and may be a result of many loci of varying influence (Allendorf et al., 2010), landscape genomic inquiries that use scans of whole-genome sequences provide a wholly integrated assessment of adaptive fitness for any species under a reasonable timeline and cost, regardless of the extent to which a species has been well-studied (Joost et al., 2007; Shryock et al., 2020). Understanding how adaptive traits can be augmented based on patterns of existing genotype-environment associations may provide a narrow window for genetic intervention via assisted gene flow or migration in the case of an adaptational lag (Aitken & Bemmels, 2016; Aitken et al., 2008). This approach represents an alternative to more time-consuming and costly methods for identifying adaptive traits, such as long common garden experiments, that cannot be afforded to taxa like California black walnuts with a narrow window of opportunity for their conservation.

In this study, I will use herbarium specimens and whole-genome sequences for field-identified *Juglans californica* individuals sampled throughout California to determine if northern and southern occurrences of the putative taxon represent two distinct species based on differences in leaf morphology, genetic structure, and genetic differentiation and diversity. Following the determination of species identities across the state, I will also describe spatial patterns of hybridization occurring in each species' genomes to provide insight into the risk of introgression. Finally, I will leverage gene-environment associations to explore how climate has shaped local patterns of genomic variation in each California black walnut taxon and to isolate the environmental variables likely to serve as important selection pressures on each species under climate change. This thesis will thereby address the following questions: 1) Are *Juglans californica* and *Juglans hindsii* different species based on phenotypic and genomic divergence and genetic structure? 2) Does either taxon exhibit different patterns in local adaptation across a climate gradient based on gene-environment associations? This thesis will thereby serve as the first inquiry to leverage whole-genome variation to resolve the demographic and conservation uncertainties for one of California's most threatened and emblematic sister species to provide a clearer understanding of the current and future state of California walnuts based on patterns of genetic structure, genetic diversity, and gene-environment associations.

## Methods

### *Sampling Design*

Georeference coordinates for *Juglans californica* individuals were obtained using herbarium records from the Consortium of California Herbaria. Given the increased likelihood of missing trees for older records, we filtered for records with collection dates after the year 2000. Redundant records occurring within the same 1 km$^2$ grid cell were excluded from our sample list to encapsulate as much of a climate gradient for the species as possible. Leaf tissue was sampled for 120 *Juglans californica* individuals throughout the species' northern and southern putative distributions for DNA extraction. A new herbarium voucher was derived for each tree sampled for later leaf trait analyses.

### *DNA Extraction and Sequencing*

Approximately 50 mg of leaf tissue was frozen in liquid nitrogen and processed into a fine powder using a homogenizer. DNA extraction procedures followed a modified version of the Qiagen DNEasy Plant Mini Kit extraction protocol. A prewash step was performed twice to remove polyphenols. 1 mL of prewash buffer, consisting of 100 ul Tris, 100 ul EDTA, 200 ul 5 M NaCl, 600 ul molecular grade water, and 0.01 g PVP, was added to the ground leaf tissue of each sample. The ground leaves and buffer were ground in a bead mill for 20 seconds, centrifuged for 10 minutes at 10,000 RPM, and the supernatant was discarded. Thereafter, the Qiagen protocol was followed. 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

### *Variant Calling & Filtering*

Filtering and variant calling . Adapters were trimmed from raw reads using Trim Galore, and reads with a length of less than 20 bp were removed. Reads were not trimmed based on quality scores during this step. Reads were aligned to the *J. californica* reference genome (Fitz-Gibbon et al., 2023) using BWA-MEM (Li, 2013), with 'markShorterSplits' and 'readGroupHeaderLine' options enabled. Duplicate reads were marked and removed using GATK MarkDuplicates (Van der Auwera & O'Connor, 2020). 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. 104 individuals out of 120 samples were retained in the whole dataset after removing individuals with a coverage of less than 7. Using bcftools (version 1.15.1, Danecek et al., 2021), we selected only biallelic SNPs across all samples from this set of high-quality variants for further analysis.

Using bcftools, the vcf was subsetted by genetic cluster according to PCA and ADMIXTURE results, SNPs with a mean depth across all samples <5 were removed, individual genotypes with depth <5 were set to missing, SNPs with a minor allele frequency <0.01 were removed, and SNPs with ≥90% missingness across all individuals were removed. 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 disequilibrium (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 only independent pairs remain. Genetic structure and gene-environment association analyses were run on these filtered and LD-pruned datasets (Table S1). SNPs were imputed for analyses requiring no missing data by assigning missing individuals the most common allele.

### *Genetic Structure & Diversity*

Individual ancestry coefficients for each sample were derived using ADMIXTURE 1.3 (Alexander, Novembre, & Lange, 2009). Ten replicates for each number of ancestral populations (k) varying between 1 and 10 were run on the complete genetic dataset and then on each subsetted vcf based on genetic clusters. The best k was chosen based on cross-validation errors. A PCA on the complete genetic dataset was derived using PLINK (version 1.90b6.26, (Chang et al., 2015)) to corroborate ADMIXTURE results. The PCA was colored by genetic cluster assignments for each sample based on ADMIXTURE results.

Genomic diversity metrics, observed heterozygosity, biallelic richness, and nucleotide diversity ($\boldsymbol{\pi}$), were calculated and plotted spatially using the function window_gd(), which calculates each metric via sliding window calculations with rarefication, in the R package 'wingen' (Bishop et al., 2023) using each VCF subsetted by genetic cluster. Genomic diversity metrics for unsampled areas were derived for each metric using kriging methods with the function krig_gd(). To calculate the genetic differentiation in the whole genomes of individuals from different genetic clusters (n = 8,777,656), we used vcftools with the "--weir-fst-pop" option, which calculates the mean and Weir and Cockerham's weighted $F_{\mathrm{st}}$ statistic across loci. To calculate $F_{\mathrm{st}}$ statistics, we only used individuals with ancestry coefficients greater than 0.9 for each genetic cluster.

### ***Deriving Environmental Data and Variable Selection***

Bioclimatic variables for each set of coordinates wherein the samples were collected were derived from WorldClim 2 (Fick & Hijmans, 2017) using the R package ‘raster’ (Hijmans et al., 2015) at a resolution of 10 (minutes of a degree). To identify the environmental variables that explain the greatest proportion of genetic variation within each cluster across a climate gradient, a Gradient Forest model was run on the complete dataset of SNPs for each major genetic cluster identified by genetic structure results using the R package, ‘gradientForest’ with 500 regression trees and default parameters (Ellis et al., 2012). Gradient Forest utilizes an unsupervised machine learning approach to identify non-linear relationships between genomes and the environment by evaluating how much each environmental variable contributes to reducing error, considering interactions with other features in genomic data. (Fitzpatrick & Keller, 2015). To include how geography might underlie genetic variation, a set of principal coordinates of neighborhood matrices (PCNM), or Moran’s eigenvector maps (MEM), were imputed using the set of coordinates for each sample using the function “pcnm" in the R package “vegan” (Oksanen et al., 2013). The PCNM is derived by truncating and deriving a spatial weights matrix from the Euclidean distances between samples to represent spatial relationships between all samples. Based on the $R^2$ weighted importance outcomes for each gradientForest run, environmental variables with the overall greatest predictive power in the model were selected stepwise from highest to lowest. Environmental variables that were highly correlated based on correlation matrices derived using the function ‘pairs.panels’ in the R package, ‘psych’ (version 2.4.3; Revelle & Revelle, 2015) were removed to reduce variance inflation of RDA downstream (Figure S2). The resulting environmental dataset contained five poor to moderately correlated variables with the highest explanatory power on genomic variation (Table 1).

### ***Gene-environment Associations***

Two gene-environment associations were employed to explore how the environment shapes genomic variation and to scan for climate adaptive loci: latent factor mixed models (LFMM) and redundancy analysis (RDA). Latent factor mixed models are a univariate approach that accounts for confounding variables, such as population structure, while identifying loci associated across a climate gradient to significantly reduce the risk of false positives (Frichot et al., 2013). In landscape genomics, RDA is a technique that uses multivariate regression to find combinations of environmental data that explain combinations of genetic markers (Capblancq & Forester, 2021). RDAs can thereby identify which genetic loci are strongly associated with a set of environmental predictors. Both of these approaches assume outlier loci based on gene-environment associations are putatively adaptive. Only subsetted VCFs were used in both RDA and LFMM runs.

To identify outlier loci using LFMM, we used the R package ‘LEA’ version 2.4 (Frichot & François, 2015). Missing SNPs were imputed using the function ‘impute(),’ following an sNMF run on each subsetted VCF. K=1 was the number of ancestral lineages with the lowest cross-entropy values for both clusters included in this study, so it was used as the argument for imputing missing SNPs. The lfmm2() in the LEA package was used to run an LFMM for each environmental variable (five times in total on each set of genetic markers) with K=1 to account for population structure. Using lfmm2.test(), P-values and genomic inflation factors (GIF) were determined for each LFMM run with default parameters (Frichot & François, 2015). Because of multiple hypothesis testing, we used the function qvalue() in the R package "qvalue" (Storey et al., 2015) to convert p-values into q-values. Loci with a q-value < 0.1 (FDR <10%) were retained as candidate adaptive loci (Sandercock et al., 2023).

To identify outlier loci using RDA, we used the function rda() in the 'vegan' package (Oksanen et al., 2013) by setting the bioclimate data as independent variables and the genomic data as dependent variables for each genetic cluster. To test if the relationship between climate and genomic data is significant, we used the function anova.cca(). We conducted partial redundancy analyses (pRDA) by including a conditioned matrix in each pRDA model to partition the variance explained by climate, geography, and the interactions between the two. For instance, we included geography as a condition in a pRDA and used the function anova.cca() on the output to test if climate variables were still significant in explaining variance when including the effects of geography.  We used the inertia values of each constrained matrix from both pRDAs conditioned on either climate or geography to derive the proportion of variance explained by each respective independent variable on the full model (climate + geography + collinearity). Joint effects were calculated using the remaining inertia of the entire model, which was not explained by the sum of climate and geography. Finally, we defined outlier SNPs as loci with loadings outside of 3 standard deviations from the mean (two-tailed p-value = 0.0027) (Ferchaud et al., 2022).

### *Leaf Morphology Comparative Analysis*

To confirm species identity and corroborate genetic structure results, we assessed whether leaf morphology differed significantly between samples of different genomic clusters.

We randomly selected a leaf from each herbarium sample collected at the time of leaf tissue sampling for whole-genome sequencing from the set of leaves that were largely undamaged (missing a portion of leaves, folded, or otherwise damaged by handling). For each sample, we measured leaf thickness using digital calipers and avoided any major leaf veins, measured the dry mass of each whole leaf in grams, and assessed the presence or absence of abaxial vein axil hairs (Figure 1). The absence of abaxial vein axil hairs in California walnut leaves is unique to southern California walnuts (*Juglans californica*) and is thereby a diagnostic phenotype used in field-based identifications (Baldwin et al., 2012). Using the digitized leaves included in the herbarium vouchers representing each sample, we measured leaf area, leaf mass area (LMA), specific leaf area (SLA), leaf perimeter, petiole length, leaf minor length (width), and leaf major length (length) using ImageJ (Abràmoff et al., 2004). We performed a MANOVA using the base function manova() in R to assess whether leaf traits significantly differ between genetic clusters. To identify which traits differed between genetic clusters, we performed an ANOVA for each leaf trait in relation to genetic structure assignments. Finally, we generated a PCA of all the scaled leaf traits using the function prcomp() in the R package ‘stats’ (Team R.C, 2018).

## Results

### *Genetic structure aligns with speculative species ranges of* **Juglans californica** *and* **Juglans hindsii**

Using the entire set of 104 individual trees possessing a minimum coverage of 7, ADMIXTURE predicts K=3 was the best model, given the lowest ten-fold cross-validation error of 0.25687, from K selections 1-10 using default point estimation and bootstrapping procedures (Figure 1). 45 individuals derive most of their ancestry to a genetic cluster that only occurs in southern California, encompassing Ventura, Los Angeles, Riverside, San Bernardino, and

Orange Counties; 33 other samples, including all samples from the southernmost sampling locations in San Diego County, derive most of their ancestry from a genetic cluster that occurs broadly throughout Northern and Central California. The remaining 26 samples derived at least some of their ancestry from a third genetic cluster that exhibits no clear geographic pattern and occurs across California within the two other clusters. Three of these 26 samples derived most of their ancestry from this third cluster and occur in Los Angeles, San Bernardino, and Santa Clara County. A PCA of this genetic data largely corroborates the existence of three genetic clusters, with PC1 largely partitioning samples based on their ancestry between the southern and northern clusters and PC2 largely explaining any hybridization with the third unknown cluster (Figure 2).

Because two of these genetic clusters appeared to overlap well with the putative ranges of *Juglans californica* and *Juglans hindsii*, we decided to include all individuals that derive their ancestry from these two genetic clusters in gene-environment analyses. Given this dataset's many hybrids, we assigned samples to one of these putative species based on the relative proportion of their ancestry coefficients. We excluded samples with ancestry from the third unknown cluster from downstream analyses. Individuals grouped within the putative *J. californica* and *J. hindsii* genetic clusters do not exhibit any population substructure according to ADMIXTURE analyses using default point estimation and bootstrapping procedures, with K=1 being the most likely model for each cluster (ten-fold CV error value of 0.31029 and 0.21568 respectively). Finally, we excluded the southernmost *J. hindsii* samples from downstream analyses, as including these would result in significant gaps in sampling across a climate gradient and because it is unclear if these individuals were introduced to this region of California.

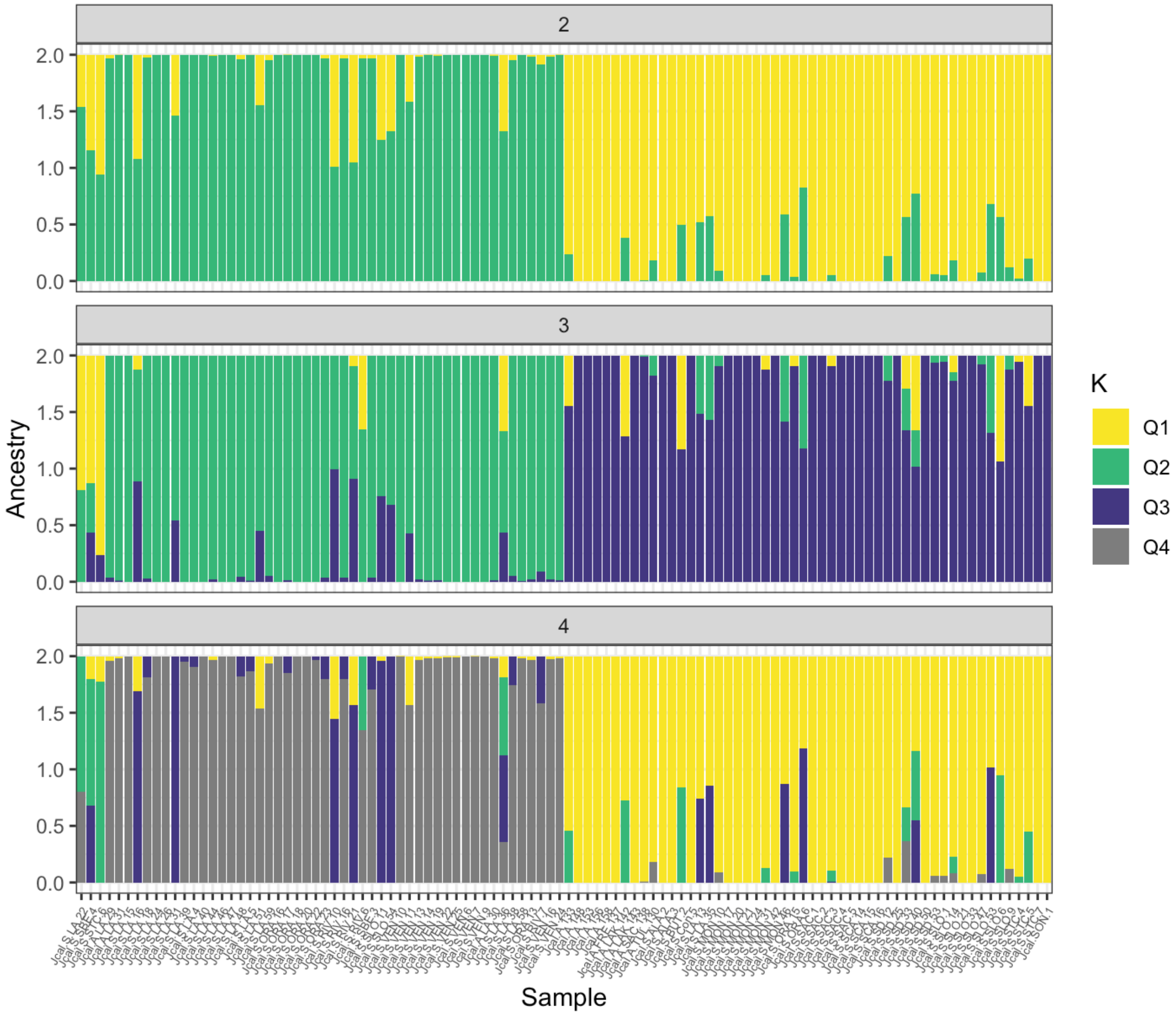


**Figure 1. ADMIXTURE bar plot illustrating the ancestry proportions of 104 samples under the best fit model of K=3 genetic clusters.** Each bar represents an individual sample, with colors denoting the proportion of ancestry derived from each of the three inferred genetic clusters. ADMIXTURE uses a maximum likelihood approach to estimate these proportions based on genotype data. The optimal number of clusters (K=3) was determined by the lowest ten-fold cross-validation error (CV error = 0.25687). A high degree of hybridization is observed across the majority of samples, indicating significant admixture among the ancestral lineages.

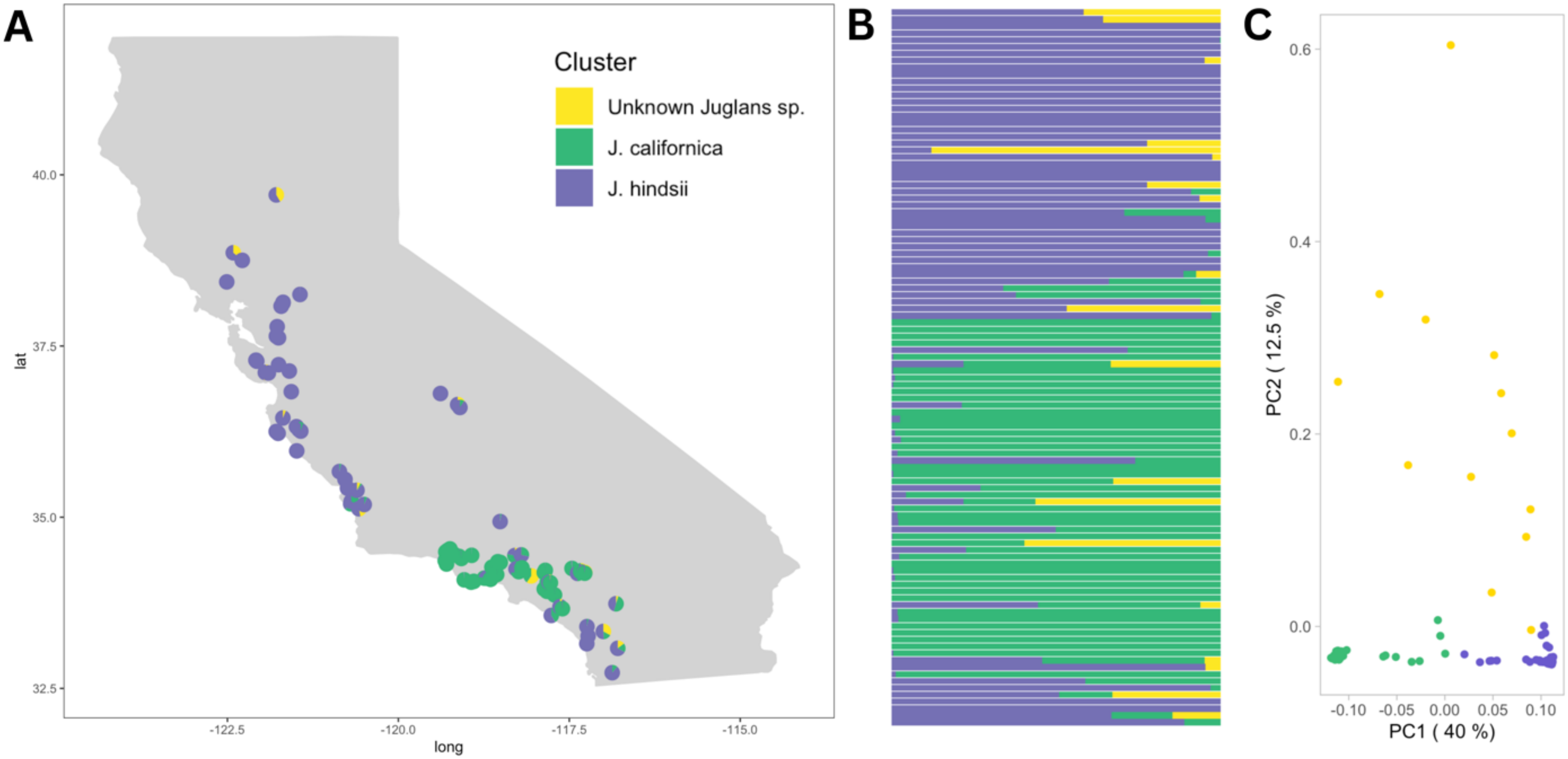


**Figure 2. Genetic structure patterns define new species ranges for California walnut species. (A)** Geographic distribution of genetic structure indicates putative species ranges for *Juglans californica* and *Juglans hindsii*. Despite initial identification as *J. californica*, genetic patterns suggest that northern samples are actually *J. hindsii*. Additionally, the southernmost samples cluster with northernmost samples, suggesting a more limited range for *J. californica*. **(B)** Bar plot displaying relative ancestry proportions for each genetic cluster, organized by latitude. **(C)** Principal Component Analysis (PCA) of genomic variation supports the ADMIXTURE results. PC1 differentiates between *J. californica* (green) and *J. hindsii* (blue), while samples with signatures of a third, unknown *Juglans* taxon (yellow) align along PC2.

***Leaf trait differentiation corroborates species identities and genetic structure results***

Although all of our samples were identified as *Juglans californica* by Consortium of California Herbaria contributors, the clear geographic split between northern and southern samples prompted us to assess whether the northern samples of *J. californica* were misidentified by comparing leaf trait differentiation between *J. californica* and *J. hindsii*. We found that the northern samples were strongly associated with the presence of abaxial vein axil hairs ($X^2$-square test, p-value = 2.056e-12). Given that this diagnostic phenotype is associated with *J. hindsii* (Figure 1; Baldwin et al., 2012) and that none of the samples that derived all of their ancestry from the southern cluster exhibited this trait, we conclude these two genetic clusters represent two distinct taxa: *Juglans californica* and *Juglans hindsii*.

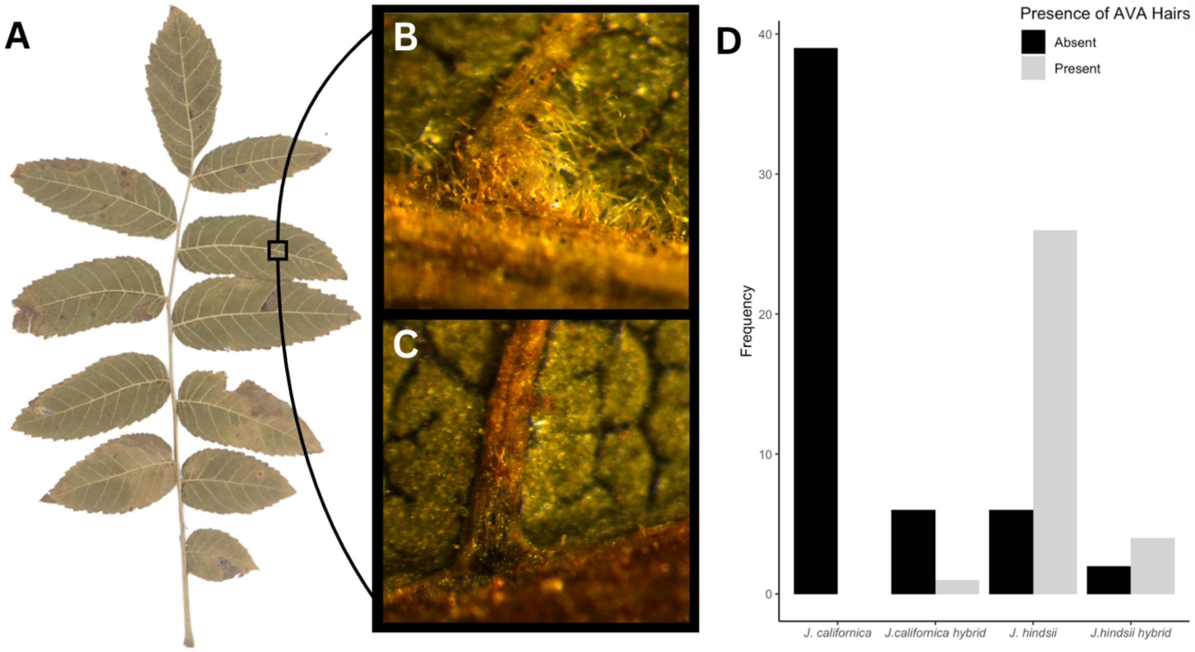


**Figure 3**. **Abaxial vein axil (AVA) hairs are a diagnostic phenotype to differentiate species identity between *Juglans californica* and *Juglans hindsii*. (A)** A typical *J. californica* leaf. A singular abaxial vein axil is highlighted. **(B)** An abaxial vein axil of a putative pure *J. hindsii* individual sampled in San Bernardino County. Notice the abundant hairs typical of this taxon. **(C)** An abaxial vein axil of a pure *J. californica* individual sampled in Riverside County. Notice the complete absence of hairs. **(D)** Bar plot showing the frequency of AVA hairs in each species. Individuals were assigned to each species according to ADMIXTURE and PCA genetic structure results.

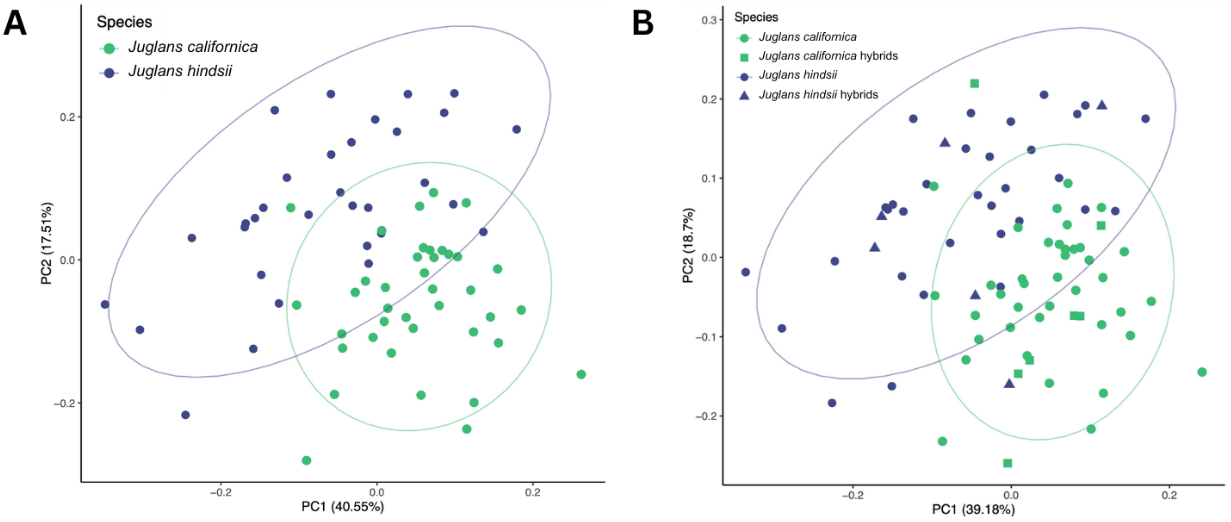


**Figure 4. Leaf trait variation clusters each species and their hybrids.** A PCA of eight leaf traits (abaxial vein axil hair presence, leaflet number, petiole length, leaf thickness, leaf mass, leaf perimeter, SLA, and leaf width/length ratio) shows similar trait convergence between **(A)** pure individuals and **(B)** pure individuals and their hybrids. Most hybrids cluster with pure individuals of the species from which they derive most of their ancestry. Ellipses represent the 95% confidence intervals for leaf trait variation within **(A)** each species and **(B)** each species and their hybrids.

Other leaf traits were significantly different between these species and their hybrids. A MANOVA testing the differences between seven other leaf traits (leaflet number, leaf minor length, leaf major length, leaf perimeter, petiole length, SLA, and LMA) and the two species and their hybrids was highly significant (p-value < 0.009558, df = 3). Subsequent ANOVAs revealed that leaflet number, leaf major length, and leaf perimeter were the significantly different leaf traits among species and hybrids (Table S1). Pure *J. hindsii* samples exhibited more leaflets (2.1218 difference in means, Tukey posthoc test CI = 0.4175 to 3.8261), longer leaves (4.2114 cm difference in means, Tukey posthoc test CI = 0.4833 to 7.9395), and much longer petioles (14.82 cm difference in means, Tukey posthoc test CI = 5.2006 to 24.4322), than pure *J. californica* samples. All other leaf traits were not significantly different between species and hybrids (Table S2, Figure S2). Leaf trait variation is largely consistent between these genetic clusters when plotted on a PCA, even when including hybrid individuals (Figure 4).

### *Genetic differentiation is high between species, and genomic diversity is higher in J. hindsii.*

Moving window calculations for nucleotide diversity ($\boldsymbol{\pi}$), biallelic richness, and observed heterozygosity ($H_o$) for both species reveals that *Juglans californica* has lower genetic diversity across its species range compared to *Juglans hindsii* (Figure 5*).* This difference in genomic diversity is likely a function of the number of SNPs in each genetic dataset (*Juglans californica* n= 352,264 vs. *Juglans hindsii* n= 139,739). *Juglans californica* exhibits slight variation in genetic diversity across the landscape, except for a higher nucleotide diversity, biallelic richness, and observed heterozygosity in eastern regions of the species range where hybridization occurs at greater rates (Figure 5A). There is also a region of lower genetic diversity across all metrics in the Riverside area near the region of highest genetic diversity. The population centers of *J. californica* where they are most common, in the Santa Monica Mountains and Repetto Foothills, exhibit relatively intermediate levels of genetic diversity. *J. hindsii* exhibits greater variation in genetic diversity across its species range. However, the lowest values observed in biallelic richness, nucleotide diversity, and observed heterozygosity exceed the highest values observed in *J. californica*. Interestingly, the speculated pre-colonial species range of *J. hindsii* in Contra Costa, Sacramento, and Napa Counties exhibits among the lowest genetic diversity values observed range-wide. The highest bouts of genetic diversity for *J. hindsii* occur far from these pre-colonial sites in San Luis Obispo County (Figure 5), where greater hybridization with *J. californica* is reflected in genetic structure results (Figure 2). Finally, a genome-wide weighted $F_{st}$ of 0.52755 and a Weir and Cockerham mean $F_{st}$ estimate of 0.24713 indicate substantial genetic differentiation between pure *Juglans californica* and *Juglans hindsii* individuals. This level of differentiation is often observed between populations that have been reproductively isolated for extended periods and suggest distinct evolutionary paths for each species (Igarashi et al., 2018; Sendell-Price et al., 2020).

### *Different environmental variables underlie genetic variation in each California walnut species*

Gradient Forest analyses reveal that genetic variation in *J. californica* is much more strongly associated with climate than *J. hindsii*. Genetic variation in the latter species is much more strongly shaped by geography than bioclimatic variables (Figure 6). After selecting bioclimatic variables based on $R^2$ weighted importance and removing highly correlated variables (Figure S3), only one bioclimatic variable, BIO6 or the minimum temperature of the coldest month, was shared between each species in the final set of bioclimatic variables selected to conduct GEA analyses. The final set of bioclimatic variables used for each species is listed in Table 1

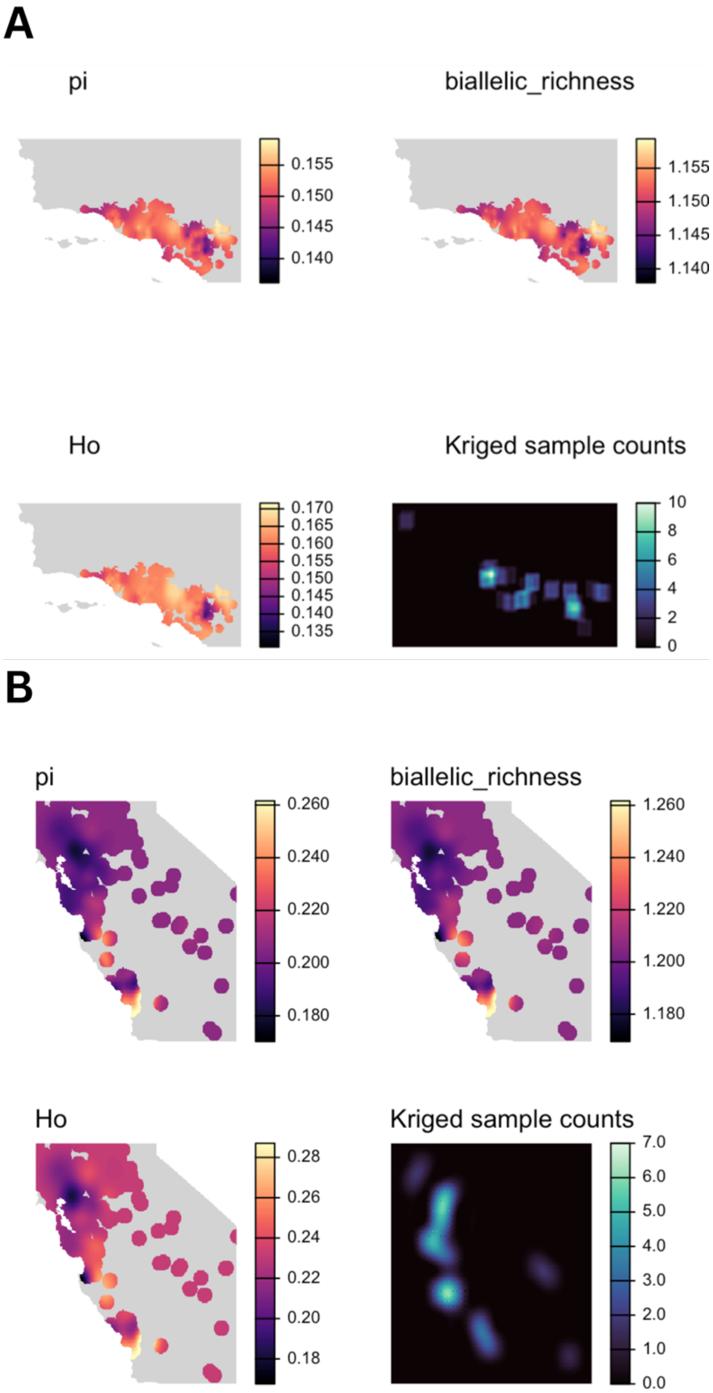


**Figure 5. Genetic diversity across the species ranges for (A) *Juglans californica* and (B) *Juglans hindsii*.** Nucleotide diversity (Pi) reflects the average number of nucleotide differences per site within a population. Biallelic richness Observed heterozygosity (Ho) describes the proportion of individuals that are heterozygous at a given loci. Sample count distribution across species range (*Juglans californica* n=45 and *Juglans hindsii* n=33). Regions with high $H_O$ and pi can covary with larger sample counts, but this is not always the case as in *Juglans californica*.

**Table 1. The final set of bioclimatic variables was used for GEA analyses for each species.** These variables were selected based on $R^2$ weighted importance resulting from a Gradient Forest model using bioclimatic and geographic variables as independent variables and genomic variation as the dependent variable.

| ***Juglans californica*** | | ***Juglans hindsii*** | |
|---|---|---|---|
| **BIO4** | Temperature Seasonality | **BIO2** | Mean Diurnal Range (Mean of monthly (max temp - min temp)) |
| **BIO6** | Min Temperature of Coldest Month | **BIO6** | Min Temperature of Coldest Month |
| **BIO10** | Mean Temperature of Warmest Quarter | **BIO9** | Mean Temperature of Driest Quarter |
| **BIO13** | Precipitation of Wettest Month | **BIO8** | Mean Temperature of Wettest Quarter |
| **BIO18** | Precipitation of Warmest Quarter | **BIO19** | Precipitation of Coldest Quarter |

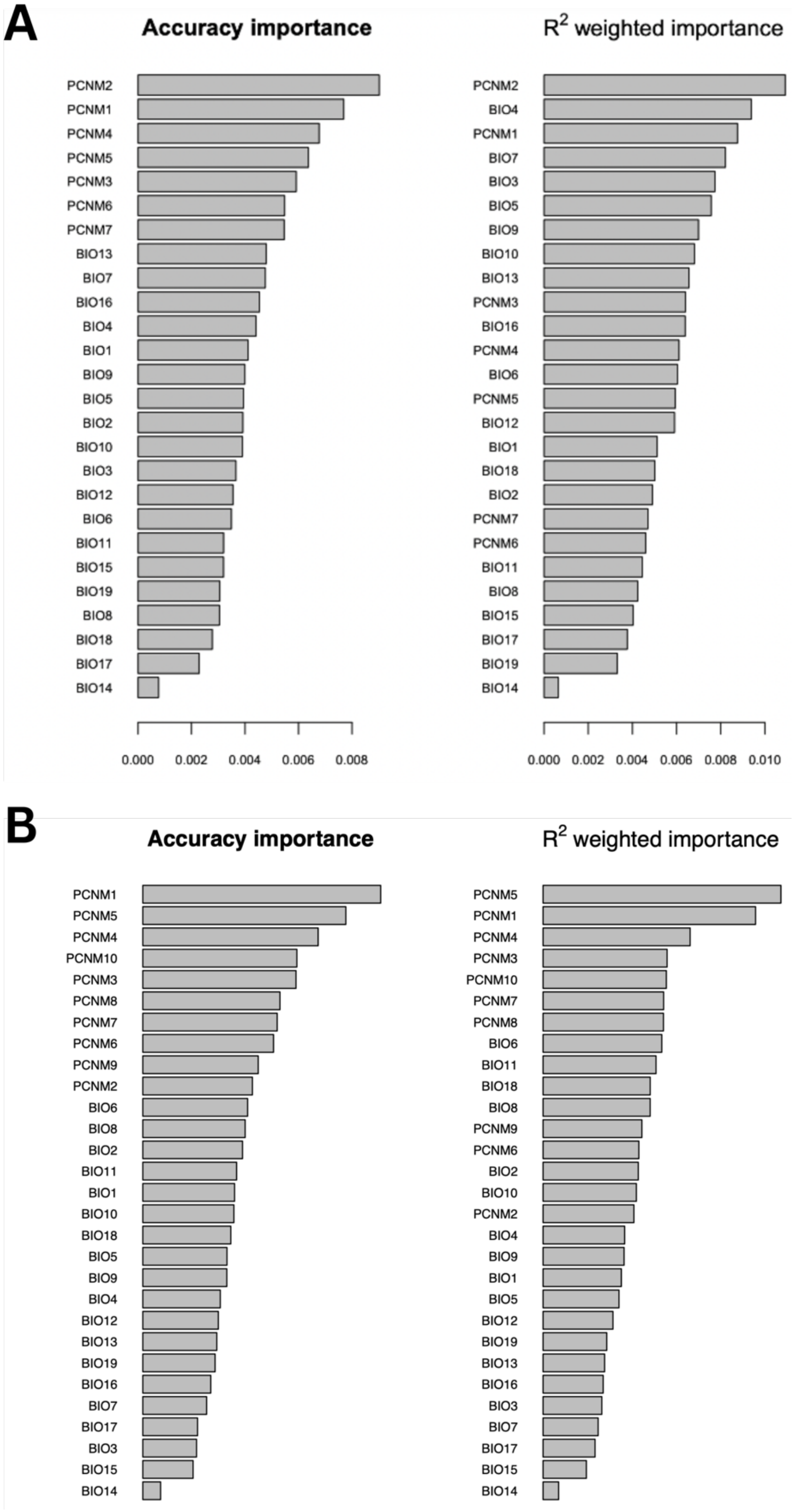


**Figure 6. Bar plots of the relative weighted importance (y-axis) of predictor variables in predicting patterns of genomic diversity using Gradient Forest.** A) *Juglans californica* and B) *Juglans hindsii*.

Individual genotypes show strong associations with bioclimatic variables. In *Juglans californica*, the southernmost samples in Orange County load strongly onto temperature variables BIO6 (Minimum Temperature of Coldest Month) and BIO10 (Mean Temperature of Warmest Quarter) in both RDA1 and RDA2 or RDA3. The northernmost samples of *J. californica* load strongly onto precipitation variables BIO13 (Precipitation of Wettest Month) and BIO18 (Precipitation of Warmest Quarter). Samples at relatively intermediate latitudes largely occupy the ordination space between these eigenvectors for temperature and precipitation. In *J. hindsii*, the southernmost samples of San Luis Obispo County generally load strongly to the temperature variable BIO2 (Mean Diurnal Range). Genetic variation in northern samples is strongly associated with BIO9 (Mean Temperature of Driest Quarter) and BIO19 (Precipitation of Coldest Quarter). Both species demonstrate consistent clustering by latitude across genomic associations with bioclimatic variables, indicating that geography might play a significant role in shaping the bioclimatic responses of these species (Figure 7). Partial redundancy analyses in both species show that climate variables are still significant explanatory variables in shaping genomic variation when geography is controlled. In both species, climate plays a significantly large role in shaping variance explained (Table 2).

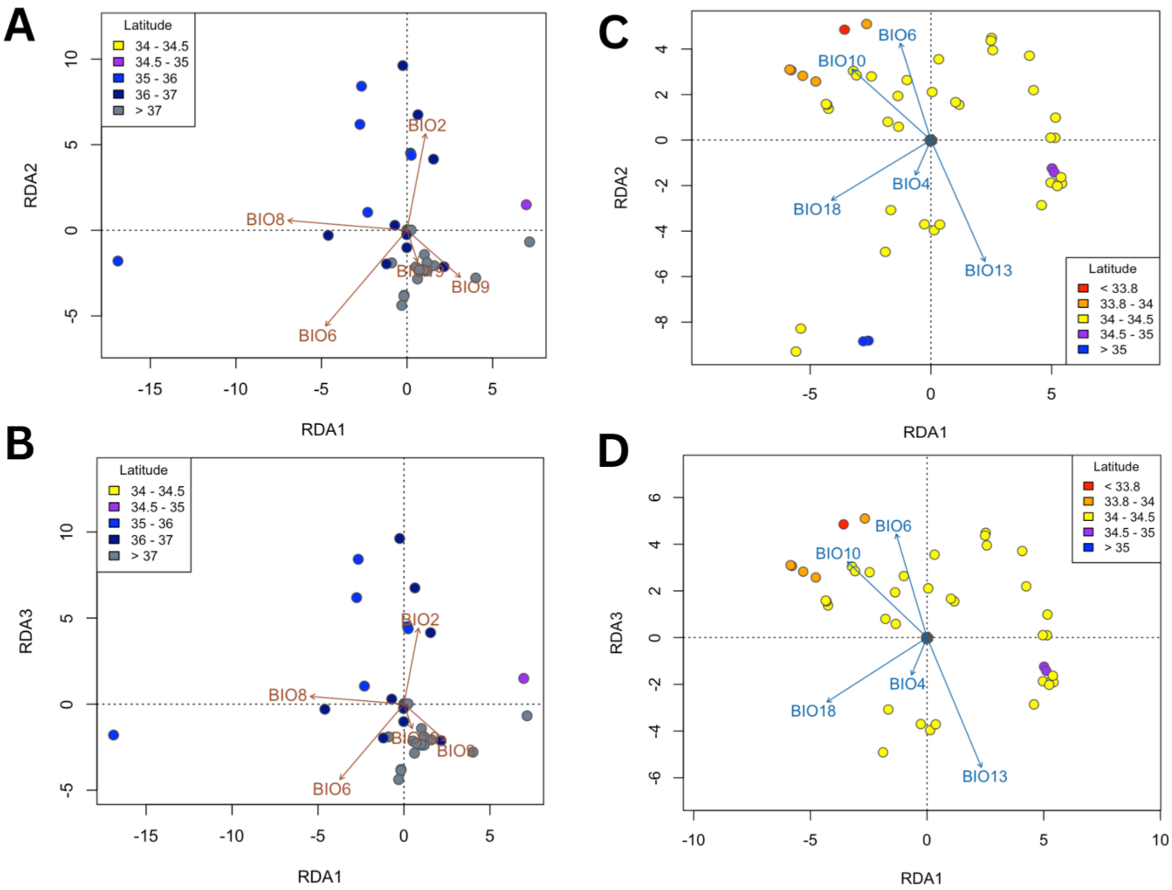


**Figure 7. RDA triplots showing relationships between genotypes and bioclimatic variables.** In both *J. hindsii* (A and B) and *J. californica* (C and D) genomic variation is characterized by strong climate gradients. BIO6, or the minimum temperature of the coldest month, appears to have a strong influence on both species. Individuals are colored by latitude, showing clustering of individuals from similar latitudes.

**Table 2. RDA and pRDA testing for significance on climate, geography, and collinearity**. For both *Juglans californica* and *Juglans hindsii*, climate is the dominant factor explaining the majority of the variance in the data, with highly significant p-values indicating strong effects. Geography also explains a substantial portion of the variance but is not statistically significant for either species. Collinearity has a negligible impact on the variance for both species. Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

| ***Juglans californica*** | | | | ***Juglans hindsii*** | | | |
|---|---|---|---|---|---|---|---|
| **Model** | **P-Value** | **Inertia** | **Variance Explained** | **Model** | **P-Value** | **Inertia** | **Variance Explained** |
| **Full model** | 0.001 *** | 59,258 | --- | **Full model** | 0.009 ** | 33,651 | --- |
| **Climate** | 0.003 ** | 41,712 | 70.39% | **Climate** | 0.02 * | 23,716 | 70.48% |
| **Geography** | 0.182 | 16,682 | 28.15% | **Geography** | 0.061 | 9,798 | 29.12% |
| **Collinearity** | --- | 864 | 1.46% | **Collinearity** | --- | 137 | 0.40% |

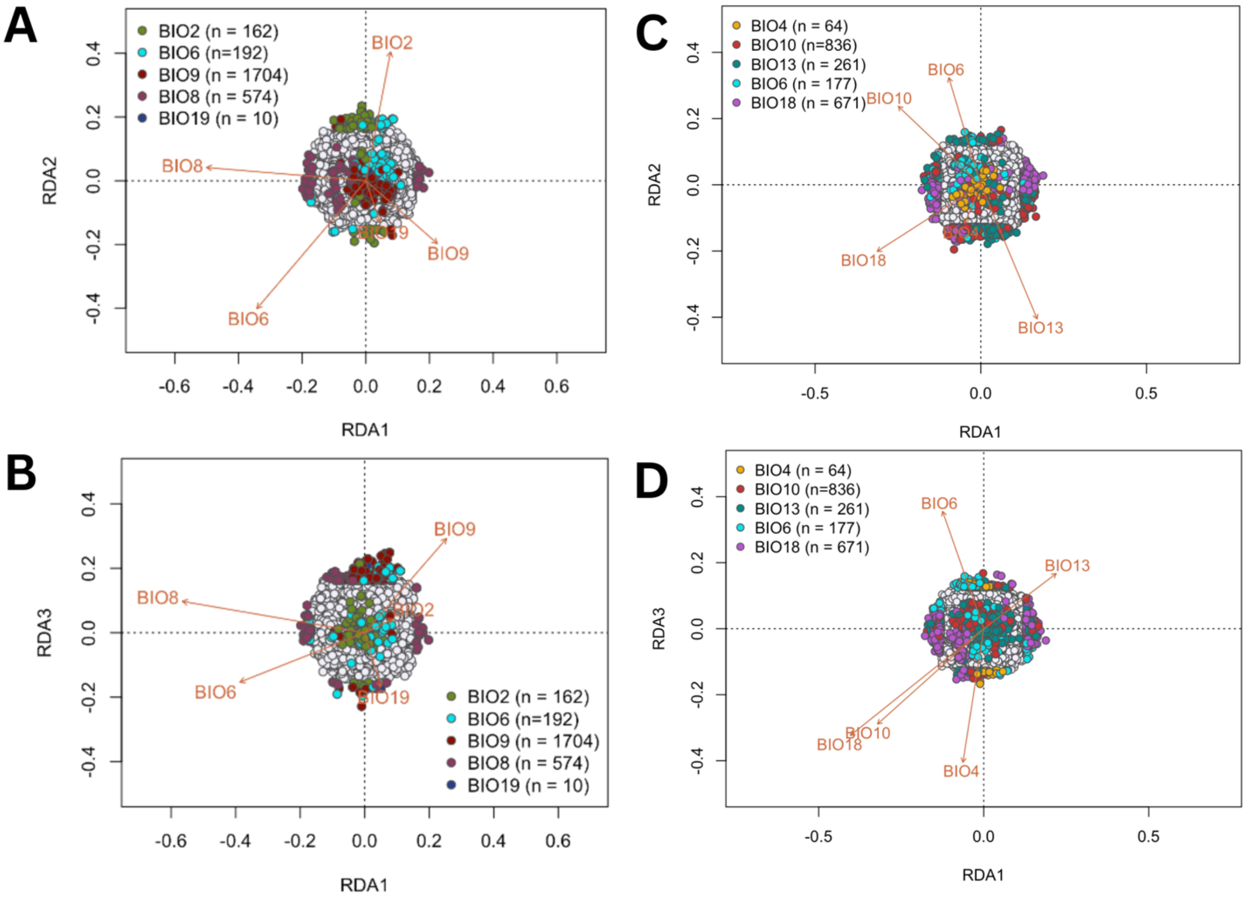


**Figure 8. Using redundancy analyses (RDA) to identify candidate SNPs involved in local adaptation. *Juglans hindsii* (A and B) and *Juglans californica* (C and D).** Each point in the ordination space represents one SNP. Arrows indicate the strength and relationship between individual SNPs and environmental variables. Outlier loci are colored to the environmental variable in which they load significantly.

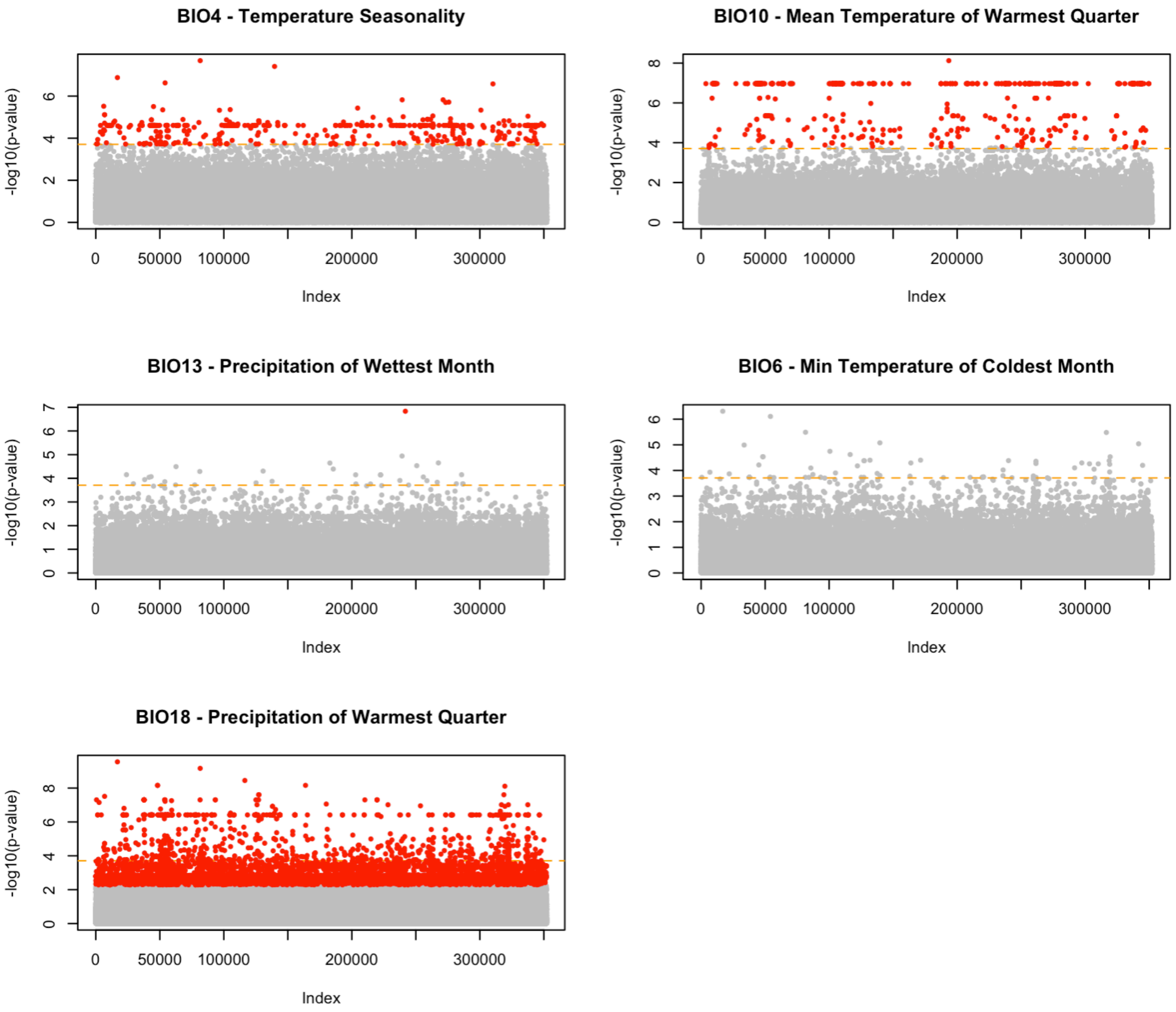
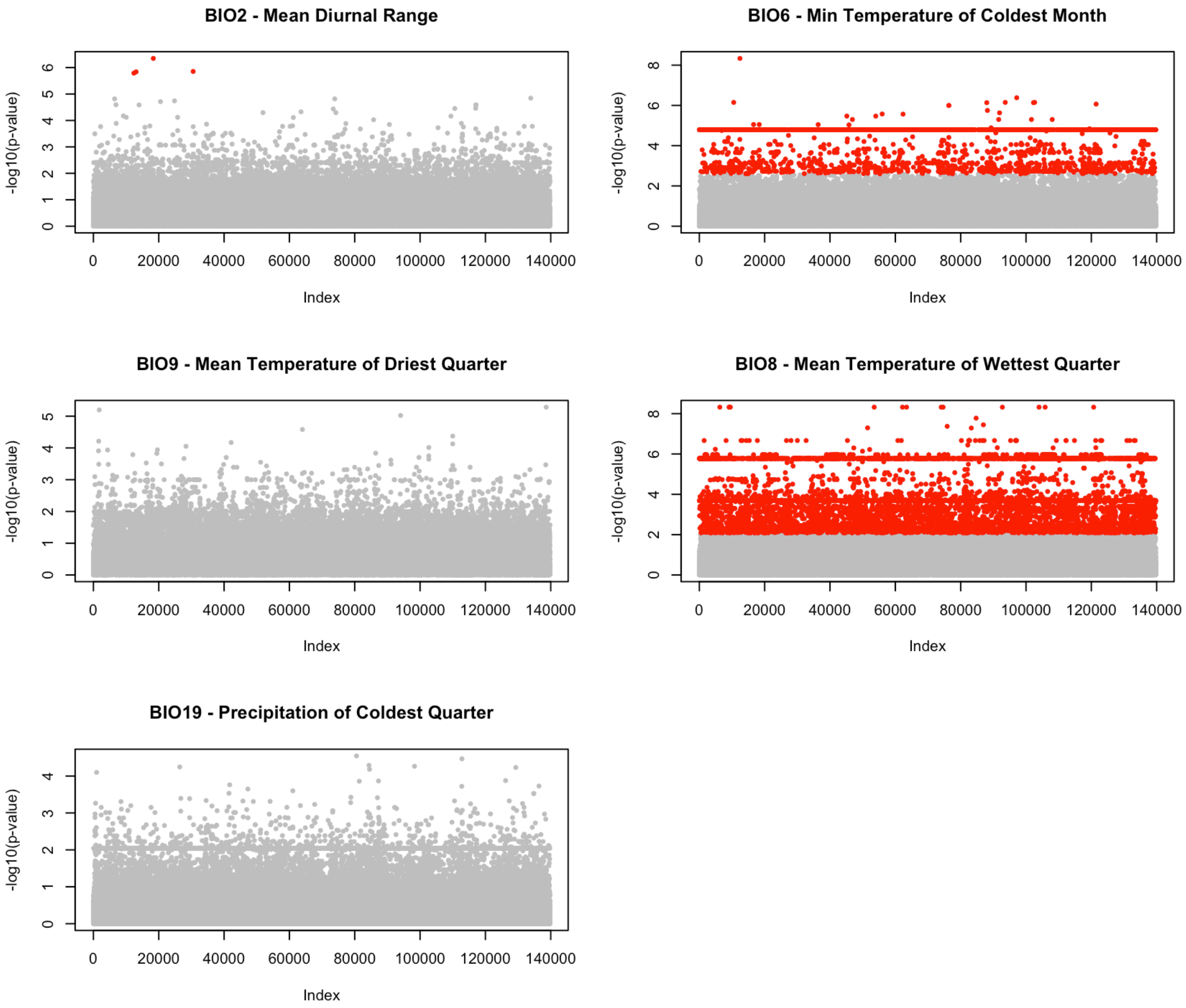


**Figure 9. Latent factor mixed models to identify climate-associated candidate SNPs in J. *californica*.** One univariate test for outlier loci was conducted per environmental variable for a total of five LFMMs. Due to multiple significance tests, p-values generated for each LFMM model were converted to q-values to reduce the false discovery rate < 10%. SNPs that remained significant are colored in red. # of outlier SNPs per bioclimate variable: BIO18 – n = 11314, BIO4 – n = 624, BIO10 – n = 119, BIO13 – n = 1. BIO6 did not identify any outlier SNPs after adjusting for multiple testing.


BIO2 - Mean Diurnal Range

BIO6 - Min Temperature of Coldest Month

BIO9 - Mean Temperature of Driest Quarter

BIO8 - Mean Temperature of Wettest Quarter

BIO19 - Precipitation of Coldest Quarter

**Figure 10. Latent factor mixed models to identify climate-associated candidate SNPs in *J. hindsii*.** One univariate test for outlier loci was conducted per environmental variable for a total of five LFMMs. Due to multiple significance tests, p-values generated for each LFMM model were converted to q-values to reduce the false discovery rate < 10%. SNPs that remained significant are colored in red. # of outlier SNPs per bioclimate variable: BIO8 – n = 5793, BIO6 – n = 2974, and BIO2 – n = 4. BIO9 and BIO19 did not identify any outlier SNPs after adjusting for multiple testing.

### *Genomic scans for climate adaptive loci*

Using RDA to detect outlier SNPs, most *J. californica* SNPs are associated with the two bioclimatic variables characterizing the hottest and driest parts of the year (BIO10 – Mean Temperature of the Warmest Quarter and BIO18 – Precipitation of the Warmest Quarter). The other three variables are associated with temperature seasonality and bioclimatic variables describing characterizing the coldest and wettest parts of the year and are related to fewer detections (Figure 8). For *J. hindsii*, most SNPs are associated with BIO9, the mean temperature of the driest month, suggesting that summer temperatures strongly shape local adaptation for both California walnut species.

Latent factor mixed models detected proportionally more climate-associated SNPs in each species compared to RDA models. For *J. californica*, BIO18 (precipitation of the warmest quarter) detected the largest number of SNPs compared to all other bioclimate variables, with 11,314 SNPs (Figure 9). BIO4 (temperature seasonality) and BIO10 (mean temperature of the warmest quarter) were the second and third largest outlier detections, respectively. Notably, BIO18 and BIO10 were also among the largest outlier loci detections under RDA analyses (Figure 8), indicating a robust and consistent association between precipitation and mean temperature of the warmest quarter and genomic variation within *J. californica*. There were no outlier loci under an LFMM model using BIO6 (min temperature of the coldest month) as an explanatory variable, unlike under the RDA model, where 177 outlier SNPs were loaded onto BIO6.

Using LFMM and for *J. hindsii*, BIO8 (mean temperature of the wettest quarter) and BIO6 (min temperature of the coldest month) identified the greatest number of outliers, with 5793 SNPs and 2974 SNPs, respectively. BIO8 was also a significant bioclimate variable for this species under RDA analyses. Under RDA analyses, BIO9 (mean temperature of the driest quarter) detected the greatest amount of candidate loci for *J. hindsii* (Figure 8), but no outliers were detected under an LFMM model including this bioclimate variable (Figure 10).

Between both GEA analyses, RDA detected 2,009 candidate SNPs, and LFMM detected 12,058 candidate SNPs for *J. californica*. RDA detected 2,642 candidate SNPs, and LFMM detected 8,771 outliers for *J. hindsii*. Between these two datasets, 877 SNPs were detected by both GEA approaches in *J. californica*, and 663 SNPs were shared by both GEA approaches in *J. hindsii*.

## Discussion

### *Genetic structure sheds light on the distribution of each taxon across California*

Our dataset of Consortium of California Herbaria field-identified *Juglans californica* samples reveals a spatially explicit genetic structure pattern that closely aligns with the vicariance observed in natural populations of *Juglans californica* and *Juglans hindsii*. Jepson himself described this stark distribution gap between the taxa of approximately 275 miles between Ventura County and Mt. Diablo, likely a primary reason for his re-negotiation of each taxon's rank to the species level from his earlier conviction that these were subspecies (Jepson, 1908 & 1923). All of our northern samples exhibit leaf morphological features that resemble or are exclusive to those of *J. hindsii* (Baldwin et al., 2012). While these misidentifications are due to the general confusion caused by the consistent re-classification of California walnuts from varieties to species and back through time (Jepson 1908 & 1923; Hickman 1993; Baldwin et al., 2012), we argue that these taxa should not be synonymous with one another. Regardless of their rank, our genetic structure results indicate that each taxa represents distinct conservation units and species that must be protected and studied independently. Our subsequent analyses revealing clear differences in morphology, genetic diversity, and patterns of adaptation provide further evidence that these taxa are ecologically and genetically distinct, and any future studies conflating the two will introduce significant confounding variables based on distinct patterns of adaptation and genetic structure and differentiation. We agree that these taxa are species and have referred to these taxa as such.

Our genetic structure results also suggest that the putative range of *J. californica* may be more restricted than previously described or, at minimum, subject to potential introgression with *J. hindsii*. Several *J. californica* x *J. hindsii* or pure *J. hindsii* individuals occur in the Eastern and Central portions of *J. californica's* range (Figure 2). More astoundingly, we could not detect any pure *J. californica* individuals south of Orange County. While the species distribution of *Juglans californica* is described as extending as far south as Baja California, we only found individuals who derive most of their ancestry from *J. hindsii* and a third unknown species within San Diego County (Figure 2). More sampling and genotyping are needed in this region to corroborate this result. However, these spatial genetic clustering patterns suggest that *J. californica*'s species range may be restricted only to the Transverse Ranges and exclude the Peninsular Ranges. Given this more limited species distribution that largely co-occurs with the Los Angeles Metropolitan area, the conservation status of *J. californica* may be more dire than current protections afford the species, as the impact of urbanization on current and future suitable habitat losses are likely underestimated (Riordan et al., 2015).

The only other recent molecular study on California walnuts also found several pure or F1 hybrids of *J. californica* and *J. hindsii* occurring in *J. californica*'s range (Potter et al., 2018). Potter also found evidence of F1 hybridization of *J. hindsii* with other non-native walnut species, the Eastern black walnut (*J. nigra*) and the English walnut (*J. regia*), across the *J. hindsii* species range (2018). Other black walnuts, Arizona walnut (*J. major)* and Texas walnut (*J. macrocarpa*), occur in neighboring states, and there is evidence that these species have also previously hybridized with *J. hindsii* (Flora of North America Editorial Committee, 2022; Potter, 2018). Our third unknown genetic cluster may be caused by signatures of one or a combination of these species, but further inquiry is warranted.

When first describing the two California species as varieties, Jepson speculated that these species were widely moved across the state by California's First Nations peoples for centuries before the entry of European colonists to the region (Jepson, 1908). While no direct evidence supports this hypothesis, *J. hindsii* nuts are larger and provide more nut meat than *J. californica* (Baldwin et al., 2012), perhaps incentivizing the pre-colonial cultivation and trade of this species into southern California. Regardless of their pre-colonial and pre-human distributions, pure *J. hindsii* and a *J. hindsii* x *J. regia* hybrid are currently and widely used as the dominant rootstock for commercial walnut production in California (McGranahan & Catlin, 1987). Given the size of California's walnut industry, the need to produce new rootstock to replace aging orchards throughout the state could provide yet another avenue for the movement of *J. hindsii* into *J. californica*'s range. Given the absence of population substructure among *J. hindsii* samples across the state despite the geographical separation between northern and southern individuals, southern individuals were likely introduced recently through human activity (Figure S1, Figure 2). Otherwise, if these southern individuals were part of natural populations of *J. hindsii*, substantial gene flow would be required between the northern and southern regions to overcome sympatry with J. californica. Additionally, reproductive barriers between species would need to exist to maintain the significant genomic divergence observed between the two species.

Interspecific gene flow between the two sister species could provide new genetic variation that could aid in local adaptation and stimulate greater evolutionary potential under climate change for either California walnut species (Hamilton & Miller, 2015). Still, hybrids could also become less fit for changing climate than pure individuals (Muhlfeld et al., 2009; Barreto et al., 2013). If these southern *J. hindsii* were only recently introduced and insufficient time has passed for hybrid offspring to backcross into the species gene pool, hybridization could also result in genetic swamping, which could eliminate the species identity of *J. californica* in its native range (Rutherford et al., 2019).

***Comparing the signals for adaptation in California walnuts based on genomic variation***

Genomic variation provides a broader range of traits that natural selection can act upon. When the environment changes, some of this genomic variation may confer advantages under the new conditions. As such, high genetic diversity can enhance local adaptation by enabling rapid selection of genotypes with heritable phenotypic variation and plasticity (Lavergne & Molofsky, 2007). Each California walnut species exhibits unique patterns in genomic diversity that can signal distinct evolutionary potential pathways under climate change. Genomic diversity across all metrics measured is higher in *J. hindsii,* but most individuals exhibit uniformly low genomic diversity to others across its species range, except for southern areas where hybridization with likely introduced *J. californica* occurs (Figure 2; Figure 5). In contrast, genomic diversity is lower in *J. californica* than *J. hindsii*. However, genomic diversity across *J. californica*'s range is consistently and relatively high between individuals, except where hybridization with *J. hindsii* has been detected. In other words, hybridization in each species seems to have different effects on genomic variation, increasing genomic variation in *J. hindsii* and decreasing genomic diversity in *J. californica*. If genomic variation enhances the likelihood of local adaptation under climate change, hybridization may have opposite effects on each California walnut species' evolutionary potential, reducing the raw material needed for rapid adaptation to environmental change in *J. californica* and increasing adaptive potential to climate change in *J. hindsii*. However, because *J. hindsii*'s whole-genome sequences were aligned to *J. californica*'s reference genome, several outcomes and potential issues due to misalignments may arise that could explain these patterns in genomic diversity. In highly conserved regions between the sister species, alignment will likely be robust and accurate. However, in variable regions (due to divergence, insertions, deletions, mutations, or significant structural variants), alignments will be less accurate and full of mismatches, gaps, or other alignment errors (Valiente-Mullor et al., 2021). This could also explain the significant differences between the final genomic datasets after quality-control filtering (Table S1); *J. hindsii*'s dataset could be limited to the independent variation that is conserved with *J. californica* (*J. californica* n = 352,264 SNPs vs. *J. hindsii* n = 139,739 SNPs).

GEA analyses of the two walnut species reveal that different environmental variables underlie genomic variation in each species, suggesting that they will likely exhibit very different responses to climate change. For *Juglans californica*, precipitation in the warmest quarter, temperature seasonality, and mean temperature in the warmest quarter produced the greatest number of outlier loci. For *J. hindsii*, the mean temperature of the driest month, the mean temperature of the wettest quarter, and the minimum temperature of the coldest month were the best explanatory variables detecting the greatest number of adaptive loci. Because California walnuts are winter-deciduous and are most photosynthetically active in the early Spring to late

Summer, most of their fitness is inextricably tied to the climate they experience in the summer. Given the warmer conditions of southern California summers, the future success of *J. californica* may depend on the capacity of the species to tolerate longer periods of drought and temperature extremes. Like many other large-seeded plants, walnuts produce recalcitrant seeds (i.e., seeds that cannot tolerate desiccation or freezing; Eira & Walters, 1993) that require long, wet, and cold winters before germination. The strong signals for local adaptation with the bioclimatic variables that underlie these germination cues and the risks of seed failure in *J. hindsii* suggest that the future of this species to adapt to climate change may be sensitive to changes in the optimal winter conditions required to ensure successful reproduction and dispersal for the species.

## Conclusion & Future Recommendations

*Juglans californica* and *Juglans hindsii* may both serve as foundation species for one of California's most threatened ecosystems, but they are otherwise morphologically, genomically, and evolutionarily distinct species with different patterns in local adaptations. The spatial patterns in our genetic structure results suggest that *Juglans californica* has a more limited species range than previously described due to the complete absence of pure individuals within the Peninsular Ranges. Signals of local adaptation in *Juglans californica* are strongly associated with summer temperature and water availability. In contrast, *Juglans hindsii* exhibits robust patterns of local adaptation associated with winter temperature and precipitation. Each California black walnut taxon will likely respond to climate change differently, depending on how these seasonal climate variables change. As such, California black walnuts must be afforded species-specific protections based on their individual vulnerabilities to habitat loss and climate change. We recommend the following for future research directions and conservation priorities:

1. *Local protections for natural populations of southern California black walnuts must be strengthened.* With a more limited species range for *Juglans californica* that largely overlaps with highly urban areas in southern California, previous estimates for habitat loss due to urbanization are likely significantly underestimated. Many of the remaining pure individuals of this species persist only in highly urban areas or urban-wildland areas in Los Angeles, Riverside, Orange, and Ventura County, where development pressures remain high. Local governments must work in concert to preserve as much of the open spaces where these species occur, and open spaces characterized as walnut woodlands must be prioritized.
2. *Potential for introgressive hybridization in southern California black walnuts*. Given the presence of pure and admixed *Juglans hindsii* within the species range of *Juglans californica*, there may be a risk for increased introgression within *Juglans californica*, especially if these *Juglans hindsii* individuals were only recently introduced. For conservation practitioners, we recommend ongoing monitoring of new generations of southern California walnuts to determine if hybridization within *Juglans californica* occurs over time. While there is leaf trait convergence between the species, our study shows that abaxial vein axil hairs are an excellent diagnostic phenotype strongly associated with individuals deriving some or all of their ancestry to *Juglans hindsii*. Examining if this phenotype increases in natural populations over time can be a cost-effective and immediate means of diagnosing hybrid individuals from pure individuals of *J. californica*. For researchers, we recommend future studies investigating how

introgression may alter the overall fitness of either species, especially in the context of climate change. The occurrence of both species within the same region, whether artificial or not, presents a unique opportunity to examine the potential emergence of a hybrid zone among California walnuts that may allow adaptive traits to be introgressed across species barriers and provide the genomic variation required to adapt to future climate scenarios.

3. *Assessing maladaptation to climate change based on signals of local climate adaptation*. Using our set of climate-associated loci, researchers can estimate the risk of maladaptation to future climate by estimating the genomic offset between current allele frequencies and those predicted to be optimal under future climate scenarios. Such models can empower conservation practitioners to execute genome-informed seed transfers or assisted migrations to new areas where the chances of survival and adaptation for either species can be maximized in the wake of climate change.

## Appendix

**Table S1. Summary of the quality control (QC) steps applied to SNP data for different VCF files.** Each row details how the number of SNPs is reduced at each step and the final percentage of missing data. A significant reduction in SNPs from the original VCF to the final set indicates the stringent QC filters applied. The Jcal dataset maintains more SNPs than the Jhin dataset after each QC step, suggesting possible differences in data quality or sample characteristics.

| QC parameter | ***J. californica*** **(n = 45)** | ***J. hindsii*** **(n = 33)** | **Full Dataset (n = 104)** |
|---|---|---|---|
| **Original VCF** | 8,777,656 | 8,777,656 | 8,777,656 |
| **Biallelic only** | 7,570,051 | 7,570,051 | 7,570,051 |
| **Mean DP 5** | 7,491,566 | 7,391,850 | 7,505,240 |
| **GENO DP 5** | 7,491,566 | 7,391,850 | 7,505,240 |
| **MAF 0.01** | 4,243,996 | 2,809,921 | 5,399,813 |
| **Missing 0.9** | 3,777,872 | 2,025,088 | 4,675,593 |
| **LD Pruned $R^2 > 0.1$** | 352,264 | 139,739 | 302,573 |
| **Final Set of SNPs** | **352,264** | **139,739** | **302,573** |
| **Missing Data** | **1.79%** | **2.65%** | **2.63%** |

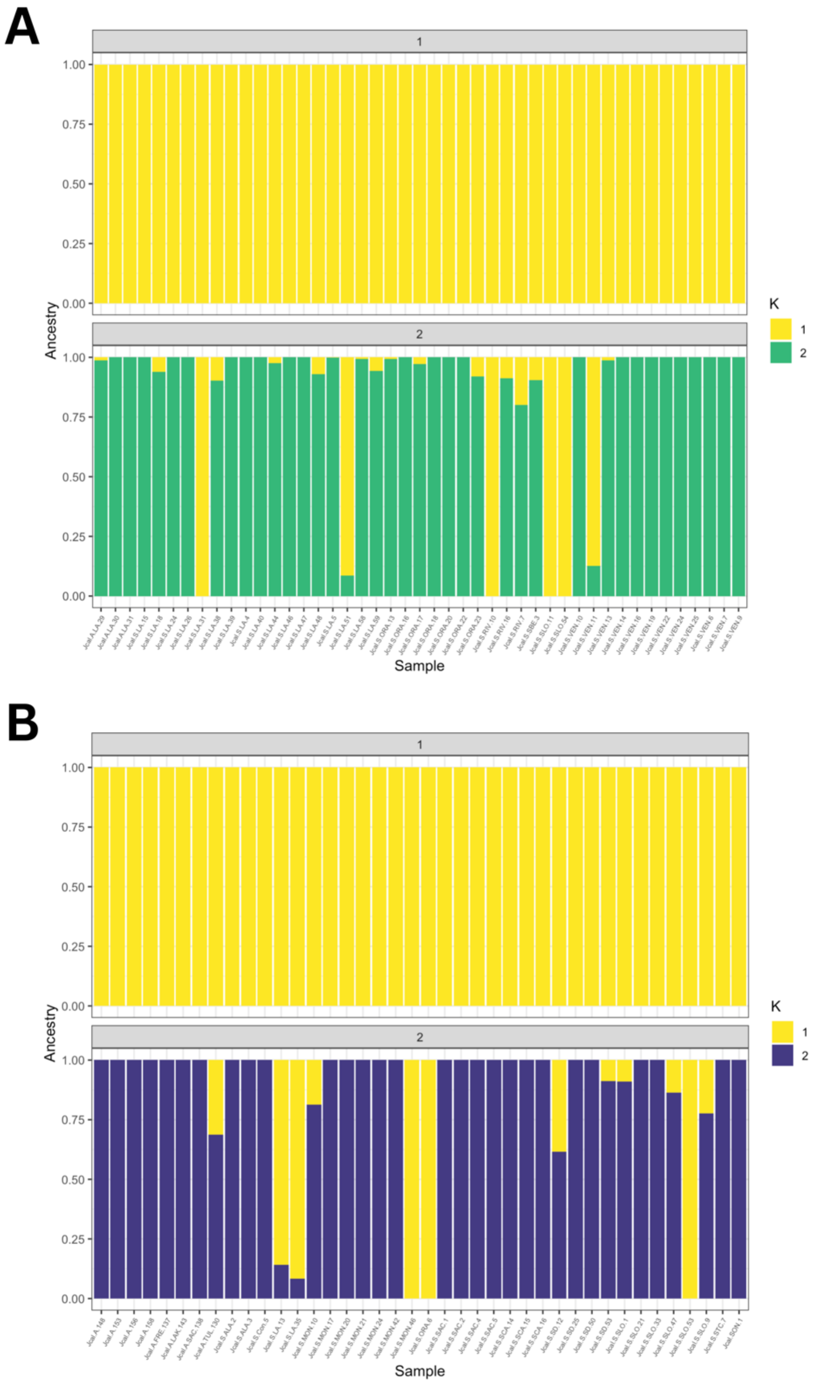


**Figure 4** Individual ancestry proportions from genotype data containing pure and hybrid individuals were estimated using ADMIXTURE 1.3 under different models of K, each denoting a new arbitrary ancestral lineage. Each bar represents the ancestry proportions of a given sample under K number of lineages. K=1 was the best fit model for **(A)** *Juglans californica* and **(B**) *Juglans hindsii* according to the ten-fold CV error value = 0.31029 and 0.25687 respectively. Hybrids were detected under models K=2.

**Table S2. ANOVA results for leaf trait variation between *J. californica* and *J. hindsii* samples.** The Shapiro-Wilk test results indicate that none of the leaf traits are normally distributed (all p-values < 0.05). Despite the lack of normality, the ANOVA test results show significant differences in the leaflet number, leaf major length, leaf perimeter, and petiole length between the two species, with the petiole length showing the most substantial difference (p < 0.001). The two species have no significant differences in leaf minor length, SLA, and LMA.

| Leaf trait | Shapiro-Wilk test statistic (W) | P-value (Normality) | ANOVA P-value |
|---|---|---|---|
| **Leaflet number** | 0.96193 | 0.01396 | 0.00514 ** |
| **Leaf minor length** | 0.93729 | 0.00049 | 0.564 |
| **Leaf major length** | 0.93462 | 0.00035 | 0.0146 * |
| **Leaf perimeter** | 0.9448 | 0.001285 | 0.00914 ** |
| **Petiole length** | 0.95503 | 0.005165 | 0.00083 *** |
| **SLA** | 0.91432 | 3.435e-05 | 0.448 |
| **LMA** | 0.95374 | 0.004309 | 0.466 |

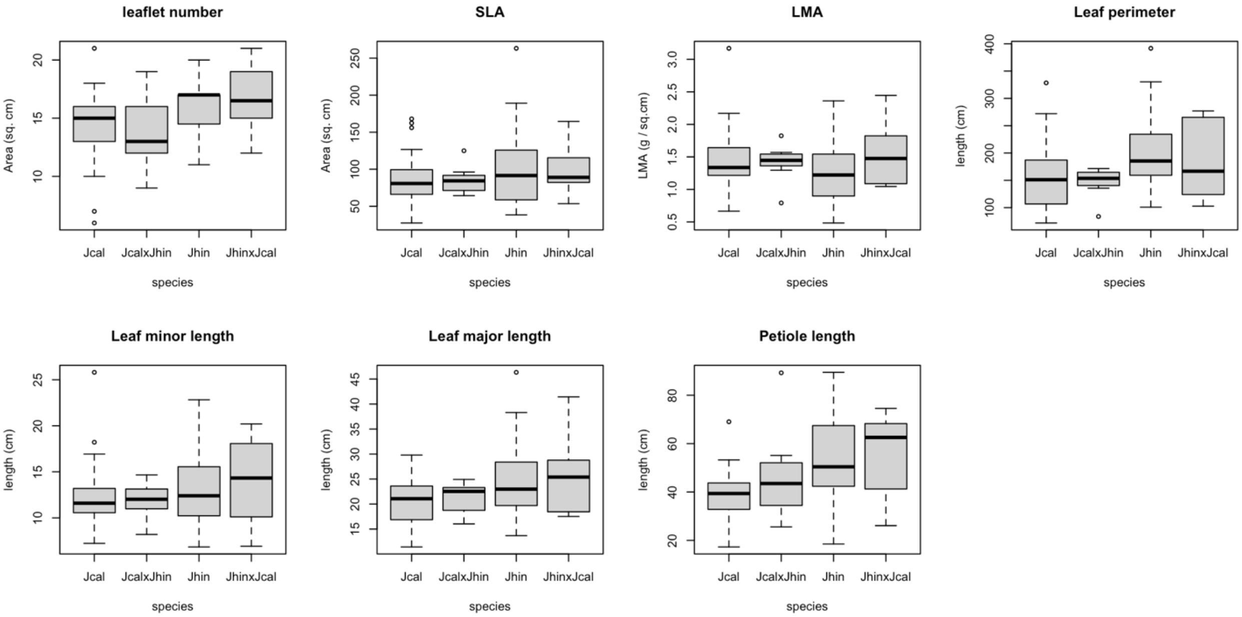


**Figure S2. Boxplots of seven leaf traits differentiating *J. californica* and *J. hindsii*.** Species assignments for leaf samples were derived from ADMIXTURE and PCA results. Jcal n=39, JcalxJhin n=7, Jhin n = 32, JhinxJcal n=6.

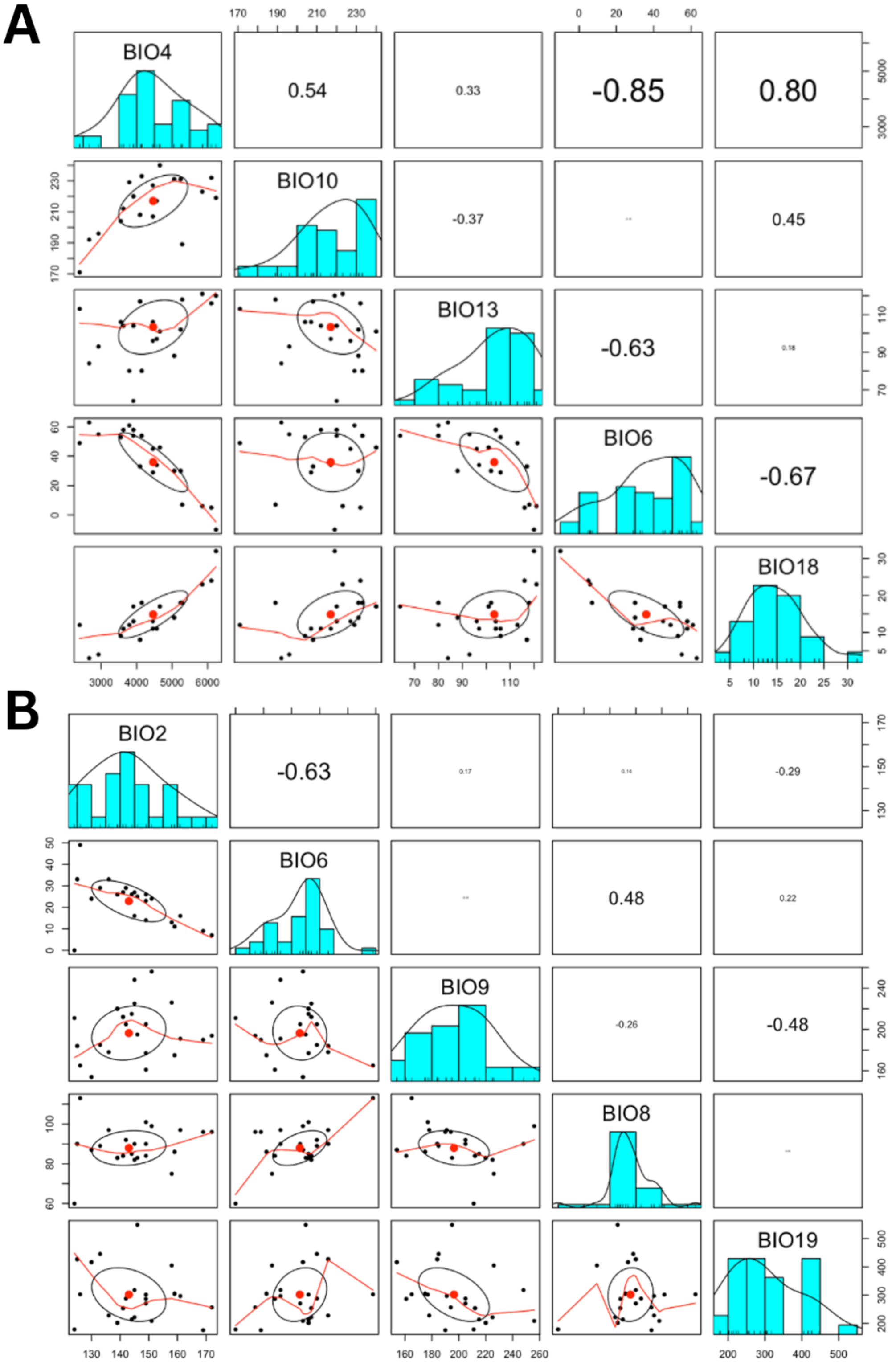


**Figure S3. Correlation matrix for environmental variables used in California walnuts GEA analyses. (A)** *Juglans californica* and **(B)** *Juglans hindsii*. To reduce variance inflation in RDA models, we attempted to remove environmental variables that were highly correlated and preferentially selected variables with greater $R^2$ values under a Gradient Forest model.

## Literature Cited

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

Abràmoff, M. D., Magalhães, P. J., & Ram, S. J. (2004). Image processing with ImageJ. *Biophotonics International*, 11(7), 36-42.

Aitken, S. N., Yeaman, S., Holliday, J. A., Wang, T., & Curtis-McLane, S. (2008). Adaptation, migration or extirpation: climate change outcomes for tree populations. *Evolutionary Applications*, 1(1), 95-111.

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

Allendorf, F.W., Hohenlohe ,P.A., and G. Luikart. (2010). Genomics and the future of conservation genetics. *Nature Reviews Genetics*, 11: 697–709.

Anderson, E.N. (2002). Some preliminary observations on the California black walnut (*Juglans californica). Fremontia - A Journal of the California Native Plant Society*, 30(1), 12-19.

Baldwin, B. G., Goldman, D. H., Keil, D. J., Patterson, R., & T.J. Rosatti (Eds.). (2012). The Jepson manual: vascular plants of California. *Univ of California Press*.

Barreto, F. S., & Burton, R. S. (2013). Elevated oxidative damage is correlated with reduced fitness in interpopulation hybrids of a marine copepod. *Proceedings of the Royal Society B: Biological Sciences*, 280(1767), 20131521.

Bisbing, S. M., Urza, A. K., Buma, B. J., Cooper, D. J., Matocq, M., & Angert, A. L. (2021). Can long-lived species keep pace with climate change? Evidence of local persistence potential in a widespread conifer. *Diversity and Distributions*, 27(2), 296-312.

Bishop, A. P., Chambers, E. A., & Wang, I. J. (2023). Generating continuous maps of genetic diversity using moving windows. *Methods in Ecology and Evolution*, 14(5), 1175-1181.

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

California Nature Plant Society (CNPS). (2017). California Native Plant Society Inventory of Rare and Endangered Plants of California. Sacramento, CA. Available at: http://www.rareplants.cnps.org/ [accessed 28 March 2023].

California Nature Plant Society (CNPS). 2021. A Manual of California Vegetation Online – Juglans California Forest and Woodland Alliance. Available at: https://vegetation.cnps.org/alliance/33. [accessed 28 March 2023]

Capblancq, T., & Forester, B. R. (2021). Redundancy analysis: A Swiss Army Knife for landscape genomics. *Methods in Ecology and Evolution*, 12(12), 2298-2309.

Eira, M. T. S., & Walters, C. (1993). Recalcitrant seed physiology and storage. *Seed Science Research,* 3(4), 289-303.

Ellis, N., Smith, S. J., & Pitcher, C. R. (2012). Gradient forests: calculating importance gradients on physical predictors. *Ecology*, 93(1), 156-168.

Ferchaud, A. L., Normandeau, E., Babin, C., Præbel, K., Hedeholm, R., Audet, C., ... & Bernatchez, L. (2022). A cold-water fish striving in a warming ocean: insights from whole-genome sequencing of the Greenland halibut in the Northwest Atlantic. *Frontiers in Marine Science*, 9, 992504.

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

Fitz-Gibbon, S., Mead, A., O'Donnell, S., Li, Z. Z., Escalona, M., Beraut, E., ... & Sork, V. L. (2023). Reference genome of California walnut, *Juglans californica*, and resemblance with other genomes in the order Fagales. *Journal of Heredity*, 114(5), 570-579.

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

Flora of North America Editorial Committee, eds. (2022). Flora of North America north of Mexico, [Online]. *Flora of North America Association*. Available: http://www.efloras.org/flora_page.aspx?flora_id=1. [36990]

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

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

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

Hamilton, J. A., & Miller, J. M. (2016). Adaptive introgression as a resource for management and genetic conservation in a changing climate. *Conservation Biology*, 30(1), 33-41.

Hickman, J. C. (Ed.). (1993). The Jepson manual: higher plants of California. *Univ of California Press*.

Hijmans, R. J., Van Etten, J., Cheng, J., Mattiuzzi, M., Sumner, M., Greenberg, J. A., ... & Hijmans, M. R. J. (2015). Package 'raster'. *R package*, 734, 473.

Igarashi, Y., Zhang, H., Tan, E., Sekino, M., Yoshitake, K., Kinoshita, S., ... & Asakawa, S. (2018). Whole-genome sequencing of 84 Japanese eels reveals evidence against panmixia and support for sympatric speciation. *Genes*, 9(10), 474.

Jepson, W. L. (1908). The distribution of *Juglans californica* Wats. Bull. *S. Calif. Acad. Sci*. 7:23-24

Jepson, W. L. (1923). *Juglans* in A manual of the flowering plants of California. *Associated Students Store, University of California*. 279.

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

Joost, S., Bonin, A., Bruford, M. W., Després, L., Conord, C., Erhardt, G., & P. Taberlet. (2007). A spatial analysis method (SAM) to detect candidate loci for selection: towards a landscape genomics approach to adaptation. *Molecular Ecology*, 16(18), 3955-3969.

Keeley, J. E. (1990). Demographic structure of California black walnut (*Juglans californica*; Juglandaceae) woodlands in southern California. *Madroño*, 237-248.

Lavergne, S., & Molofsky, J. (2007). Increased genetic variation and evolutionary potential drive the success of an invasive grass. *Proceedings of the National Academy of Sciences*, 104(10), 3883-3888.

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

Longcore, T. and N. Noujdina. (2022). Conservation of California Walnut in the Eastern Santa Monica Mountains. The Urban Wildlands Group. 1-25.

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

McGranahan, G.H. and P.B. Catlin. (1987). *Juglans* rootstocks, p. 411–450. In: R.C. Rom and R.F. Carlson (eds.). Rootstocks for fruit crops. *Wiley*, New York.

Muhlfeld, C. C., Kalinowski, S. T., McMahon, T. E., Taper, M. L., Painter, S., Leary, R. F., & Allendorf, F. W. (2009). Hybridization rapidly reduces fitness of a native trout in the wild. *Biology Letters*, 5(3), 328-331.

Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'hara, R., . . . Wagner, H. (2013). Package 'vegan'. *Community Ecology Package*, version, 2(9), 1-295.

Potter, D., Bartosh, H., Dangl, G., Yang, J., Bittman, R., & J. Preece. (2018). Clarifying the conservation status of Northern California black walnut (*Juglans hindsii*) using microsatellite markers. *Madroño*. 65: 131-140.

Quinn, R. D. (1989). The status of walnut forests and woodlands (*Juglans californica*) in Southern California. Endangered Plant Communities of Southern California. *Southern California Botanists*, 42-54.

Revelle, W., & Revelle, M. W. (2015). Package 'psych'. *The comprehensive R archive network*, *337*(338).

Riordan E. C., Gillespie T. W., Pitcher L., Pincetl S.S., Jenerette G. D., 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.

Rutherford, S., van Der Merwe, M., Wilson, P. G., Kooyman, R. M., & Rossetto, M. (2019). Managing the risk of genetic swamping of a rare and restricted tree. *Conservation Genetics*, 20, 1113-1131.

Sandercock, A. M., Westbrook, J. W., Zhang, Q., & Holliday, J. A. (2023). The road to restoration: Identifying and conserving the adaptive legacy of American chestnut. *bioRxiv*, 2023-05.

Sendell-Price, A. T., Ruegg, K. C., Anderson, E. C., Quilodrán, C. S., Van Doren, B. M., Underwood, V. L., ... & Clegg, S. M. (2020). The genomic landscape of divergence across the speciation continuum in island-colonising silvereyes (*Zosterops lateralis*). G3: *Genes, Genomes, Genetics*, 10(9), 3147-3163.

Shryock, D. F., Havrilla, C. A., DeFalco, L. A., Esque, T. C., Custer, N. A., & T. E. Wood. (2017). Landscape genetic approaches to guide native plant restoration in the Mojave Desert. *Ecological Applications*, 27(2): 429–445.

Smith, R. E. (1909). Report of the Plant Pathologist and Superintendent of Southern California Stations, July 1, 1906 to June 30, 1909. Agricultural Experiment Station Bulletin No. 203. *The University Press*, Berkeley, CA

Sork, V. L., Davis, F. W., Westfall, R., Flint, A., Ikegami, M., Wang, H., & Grivet, D. (2010). Gene movement and genetic association with regional climate gradients in California valley oak (*Quercus lobata* Née) in the face of climate change. *Molecular Ecology*, 19(17), 3806-3823.

Sork, V. L., Aitken, S. N., Dyer, R. J., Eckert, A. J., Legendre, P., 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 & Genomes*, 9.4: 901-911.

Steane, D. A., Potts, B. M., McLean, E., Prober, S. M., Stock, W. D., Vaillancourt, R. E., and M. Byrne. (2014). Genome-wide scans detect adaptation to aridity in a widespread forest tree species. *Molecular Ecology*, 23.10: 2500-2513.

Stone, D. E., Oh, S. H., Tripp, E. A., & Manos, P. S. (2009). Natural history, distribution, phylogenetic relationships, and conservation of Central American black walnuts (*Juglans* sect. *Rhysocaryon*) 1. *The Journal of the Torrey Botanical Society*, 136(1), 1-25.

Storey, J., Bass, A., Dabney, A., & Robinson, D. (2015). Package ‘qvalue’.

Stritch, L. and M. Barstow. (2019). *Juglans californica*. The IUCN Red List of Threatened Species. 2019: 1-10.

Team, R. C., Team, M. R. C., Suggests, M. A. S. S., & Matrix, S. (2018). Package stats. *The R Stats Package*.

Tibor, D. P. (Ed.). (2001). *California Native Plant Society's inventory of rare and endangered plants of California* (No. 1). California Native Plant Society.

Timbrook, J., & C. Chapman. (2007). Chumash ethnobotany: Plant Knowledge among the Chumash people of Southern California. *Santa Barbara Museum of Natural History*.

Valiente-Mullor, C., Beamud, B., Ansari, I., Francés-Cuesta, C., García-González, N., Mejía, L., ... & González-Candelas, F. (2021). One is not enough: on the effects of reference genome for the mapping and subsequent analyses of short-reads. *PLoS Computational Biology*, 17(1), e1008678.

Van der Auwera, G. A., & O'Connor, B. D. (2020). *Genomics in the cloud: using Docker, GATK, and WDL in Terra*. O'Reilly Media.

Vellend, M., Verheyen, K., Jacquemyn, H., Kolb, A., Van Calster, H., Peterken, G., & M. Hermy. (2006). Extinction debt of forest plants persists for more than a century following habitat fragmentation. *Ecology*, 87: 542-548.

Wilken-Robertson, M. (2018). Kumeyaay ethnobotany: Shared Heritage of The Californias. *Sunbelt Publications, Inc*.