Molecular Ecology

WILEY

MOLECULAR ECOLOGY

ORIGINAL ARTICLE

Introgression, Phylogeography, and Genomic Species Cohesion in the Eastern North American White Oak Syngameon

Gabe Ribicoff[1] | Mira Garner[1,2] | Kasey Pham[1,3] | Kieran N. Althaus[1,4] | Jeannine Cavender-Bares[5,6] | Andrew A. Crowl[7] | Samantha Gray[1] | Paul Gugger[8] | Marlene Hahn[1] | Shuai Liao[1,9] | Paul S. Manos[10] | Rebekah A. Mohn[1,11] | Ian S. Pearse[12] | Nicholas R. Steichmann[1] | Ashley L. Tuffin[1,13,14] | Alan T. Whittemore[1] | Andrew L. Hipp[1,2,4]

[1]Herbarium and Center for Tree Science, The Morton Arboretum, Lisle, Illinois, USA | [2]Pritzker Lab, The Field Museum, Chicago, Illinois, USA | [3]Case Western Reserve University, Cleveland, Ohio, USA | [4]Committee on Evolutionary Biology, The University of Chicago, Chicago, Illinois, USA | [5]Harvard University Herbaria, Cambridge, Massachusetts, USA | [6]Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts, USA | [7]Florida Museum of Natural History, University of Florida, Gainesville, Florida, USA | [8]Panepistemion, Chester Township, New Jersey, USA | [9]State Key Laboratory of Plant Diversity and Specialty Crops, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou, China | [10]Department of Biology, Duke University, Durham, North Carolina, USA | [11]Department of Plant and Microbial Biology, University of Minnesota, St. Paul, Minnesota, USA | [12]U.S. Geological Survey, Fort Collins Science Center, Fort Collins, Colorado, USA | [13]Institute of Ecology and Evolution, The University of Edinburgh, Edinburgh, UK | [14]Principal's Research Group, Scotland's Rural College (SRUC), Edinburgh, UK

Correspondence: Andrew L. Hipp (ahipp@mortonarb.org)

Received: 9 February 2025 | **Revised:** 15 April 2025 | **Accepted:** 25 May 2025

Handling Editor: Daniel Ortiz-Barrientos

Funding: This work was supported by USDA Cooperative Agreement 58-8020-5-005 between A.T.W. and A.L.H.; NSF DEB Awards 2129281 to A.L.H., 2129237 to P.S.M. and 2129312 to J.C.-B.; The Morton Arboretum Center for Tree Science, which provided lab, informatic and staff support as well as research fellowship support to K.P. to develop the Hyb-Seq probes; and the University of Chicago Department of Ecology and Evolution, which supported GR's work on this project.

Keywords: admixture | genomic cluster species | hybridization | Hyb-Seq | introgression | phylogeography | *Quercus macrocarpa* (bur oak)

ABSTRACT

Hybridization and interspecific gene flow play a substantial role in the evolution of plant taxa. The eastern North American white oak syngameon, a group of approximately 15 ecologically, morphologically and genomically distinguishable species, has long been recognised as a model system for studying introgressive hybridization in temperate trees. However, the prevalence, genomic context and environmental correlates of introgression in this system remain largely unknown. To assess introgression in the eastern North American white oak syngameon and population structure within the widespread *Quercus macrocarpa*, we conducted a rangewide survey of *Q. macrocarpa* and four sympatric eastern North American white oak species. Using a Hyb-Seq approach, we assembled a dataset of 3412 thinned single-nucleotide polymorphisms (SNPs) in 445 enriched target loci including 62 genes putatively associated with various ecological functions, as well as associated intronic regions and some off-target intergenic regions (not associated with the exons). Admixture analysis and hybrid class inference demonstrated species coherence despite hybridization and introgressive gene flow (due to backcrossing of F1s to one or both parents). Additionally, we recovered a genetic structure within *Q. macrocarpa* associated with latitude. Generalised linear mixed models (GLMMs) indicate that proximity to range edge predicts interspecific admixture, but rates of genetic differentiation do not appear to vary between

Gabe Ribicoff, Mira Garner and Kasey Pham contributed equally.

putative functional gene classes. Our study suggests that gene flow between eastern North American white oak species may not be as rampant as previously assumed and that hybridization is most strongly predicted by proximity to a species' range margin.

1 | Introduction

Interspecific gene flow is common and evolutionarily significant across the tree of life, with up to 10% of animal species and 25% of plant species estimated to engage in hybridization (Mallet 2005). Botanists in particular have long recognised the evolutionary importance of hybridization and subsequent introgressive gene flow due to backcrossing (Anderson 1953; Barton 2001; Grant 1981; Stebbins et al. 1947). Hybridization has the potential to generate novel genotypic combinations more rapidly than mutation (Anderson 1949, 62, 102; De Carvalho et al. 2010; Gompert et al. 2014), particularly within syngameons, multispecies networks of partially interfertile species (Buck and Flores-Rentería 2022; Cannon and Lerdau 2022; Grant 1981; Lotsy 1917). Syngameons have been documented in reef-building corals (Mao 2020), African lake cichlids (Olave and Meyer 2020; Schliewen and Klee 2004), *Heliconius* butterflies (Mallet et al. 2007) and numerous plant clades (Buck and Flores-Rentería 2022; Ellstrand et al. 1996), including many distantly related and ecologically dominant boreal, temperate and tropical trees (e.g., Buck et al. 2023; Chhatre et al. 2018; Gardner et al. 2023; Larson et al. 2021; Linan et al. 2022; Percy et al. 2014; Sun et al. 2018; Tsuda et al. 2017). Understanding landscape, climate, habitat, and demographic contexts of interspecific gene flow is key to understanding the evolutionary importance of these complex, multispecies systems.

The oak genus (*Quercus* L.) comprises one of the best-known syngameons (Kremer and Hipp 2020; Lazic et al. 2021). This northern hemisphere clade of ca. 425 species, mostly trees, has served as a model of plant hybridization for at least 150 years (Engelmann 1878; MacDougal 1907; Palmer 1948; Wiegand 1935), syngameon dynamics for more than 75 (Dodd and Afzal-Rafii 2004; Muller 1952; Stebbins et al. 1947; Tucker 1961) and genomic patterns of adaptive introgression for the past several years (Fu et al. 2022; Leroy et al. 2020; O'Donnell et al. 2021; Zhou, Yuan, et al. 2022). The eastern North American white oaks (*Quercus* sect. *Quercus*) have been particularly important in shaping botanists' understanding of introgressive gene flow and species concepts (Burger 1975; Coyne and Orr 2004; Van Valen 1976). Nearly all eastern North American white oak species can hybridize, but they still remain ecologically, morphologically and genomically distinct (Hardin 1975; Hipp et al. 2019).

Bur oak (*Quercus macrocarpa* Michx.) has the longest latitudinal range of any eastern North American white oak, ranging from north of Winnipeg, Manitoba to south of Houston, Texas, and from Nova Scotia west to Montana (Figure 1). Bur oak is a dominant of dry, fire-prone upper Midwestern oak savannas; mesic, closed-canopy forests around the Great Lakes and Northeast; and bottomland forests along the floodplains of the south-central United States (Johnson 1990). In most of these habitats, bur oak grows in close association with at least one other eastern North American white oak species. Bur oak is sympatric with more than 12 white oak species across its range, with hybridization described between it and at least 8 naturally co-occurring species (Palmer 1948). Prior population genetic studies of bur oak have demonstrated high stand-level diversity and weak isolation-by-distance, both at smaller scales among fragmented populations (Craft and Ashley 2007; Dow and Ashley 1998) and at larger scales among rangewide samples (Hipp et al. 2019; Schnabel and Hamrick 1990; Whittemore and Schaal 1991). However, previous studies have relied upon a limited number of loci to assess nuclear patterns of population structure in the species, and none have included sufficient sampling of other species to estimate introgression across the range of the species.

To evaluate patterns and determinants of interspecific gene flow in the eastern North American white oak syngameon, we sampled bur oak and the four most prevalent white oak species sympatric with it: eastern white oak (*Q. alba* L.), a dry-mesic forest species that ranges from near the Canada-U.S. border to northern Florida and Texas; swamp white oak (*Q. bicolor* Willd.), a wet forest species mostly restricted to north of 37°; post oak (*Q. stellata* Wangenh.), common in drier locales south of approximately 40°; and chinkapin oak (*Q. muehlenbergii* Engelm.), an upland limestone specialist mostly contained within the range of *Q. alba*, but extending a bit further into eastern Kansas, Oklahoma and Texas (Figures 1 and 2). We sequenced 465 target genes in 517 individuals across 56 sites in eastern North America. We addressed the following questions in our study: (1) How prevalent is interspecific gene flow in the syngameon, and what are the spatial and ecological drivers of introgression? (2) How do levels of introgression and/or genetic differentiation vary over candidate gene classes potentially implicated in environmental adaptation? (3) How is genetic variation geographically partitioned in *Q. macrocarpa*, and does genetic variation in the species follow the ecological gradient observed from predominantly upland sites in the north to increasingly bottomland sites in the south?

2 | Methods

2.1 | Sample Collection

Sampling was designed to cover the majority of the *Q. macrocarpa* range and include co-occurring white oaks at as many sites as possible. A focused field sampling of wild populations was conducted in the summer of 2017; these wild populations were supplemented with botanical garden material from Starhill Forest Arboretum (Petersburg, IL, USA) grown from acorns of known wild provenance (thus representing wild populations our group did not visit in the field) (Figure 1). At each wild population, leaf samples were collected from up to three individuals of *Q. macrocarpa* and of each co-occurring white oak species found at the site (Figure 2). Conspecific samples were taken at least 30 m apart from each other in wild populations, when possible, to avoid genotyping close relatives. Botanical garden material was selected based on wild provenance, so spatial distance within gardens was disregarded. All trees collected from wild populations were georeferenced in the field using a hand-held GPS unit. Botanical garden samples from acorns of known wild provenance could not be

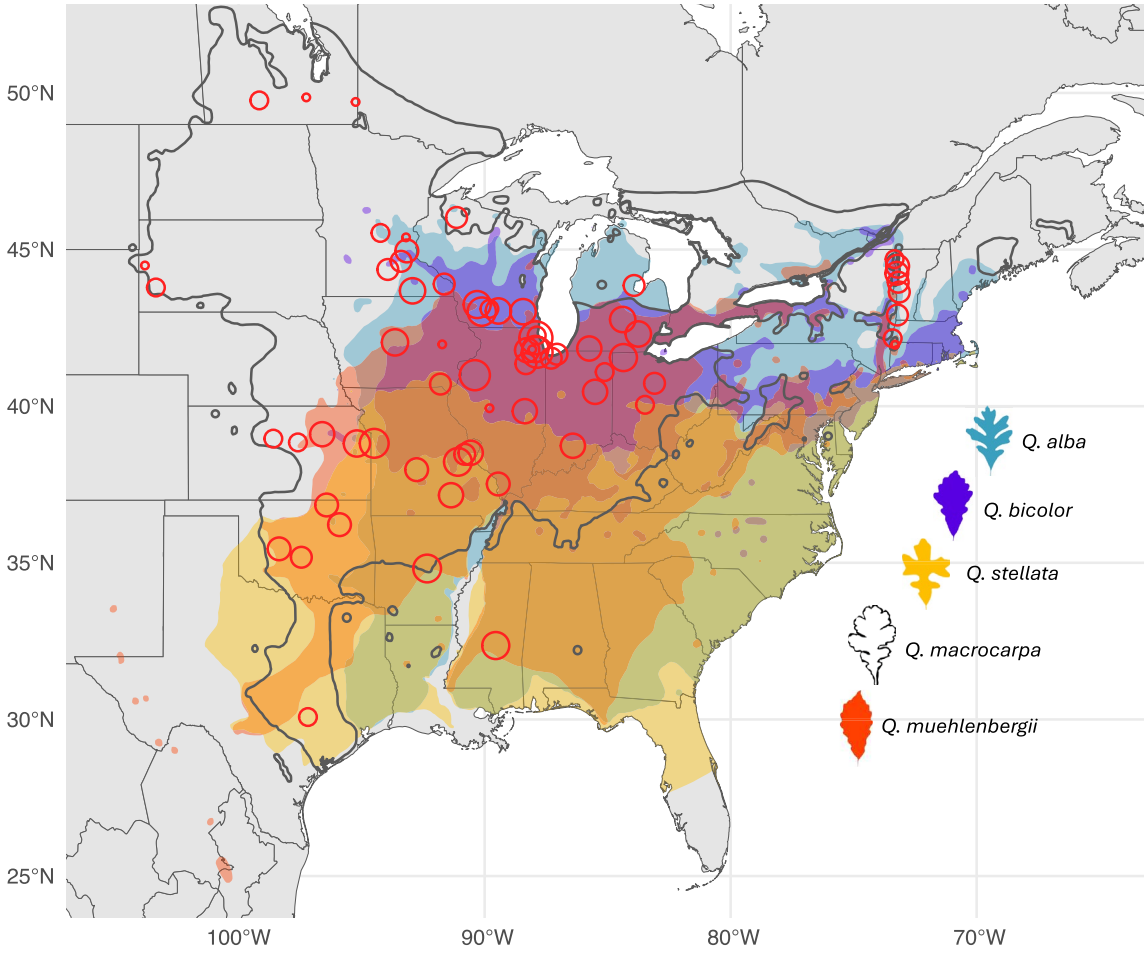

FIGURE 1 | Leaf silhouettes and geographic ranges of white oak species included in our study. Red circles indicate sampling locations, radii scaled to log-transformed sample sizes per site. Range maps are from Little's (1971) *Atlas of United States Trees*.

georeferenced precisely and were consequently excluded from spatial analyses (Table S1).

Morphological intermediates with potential hybrid ancestry were neither intentionally targeted nor avoided in the course of sampling; clearly intermediate individuals were determined to the closest morphospecies. Species determinations made in the field were checked and, if necessary, updated by examining leaf vestiture, gross leaf morphology, and twig and bud characters according to descriptions provided in the *Flora of North America* (Nixon and Muller 1997). After harvesting, leaf specimens were placed on ice in the field, then frozen at −80°C within 1 day to 1 week of collecting for later genomic analysis. Voucher samples were collected for all individuals and deposited at the Morton Arboretum Herbarium (MOR). Duplicate vouchers were deposited at the Bell Museum at the University of Minnesota (MIN) (herbarium acronyms from Thiers, Updated continuously [accessed 2025-02-02] n.d.).

2.2 | Sequence Capture Design and DNA Preparation

We selected target loci and designed sequence capture exon baits based on gene annotations of coding domain sequences in the *Q. robur* (Plomion et al. 2016, 2018) and *Q. lobata* (Sork, Fitz-Gibbon, et al. 2016) draft assemblies, as well as leaf transcriptomes of *Q. robur*, *Q. alba* and *Q. rubra* (Lesur et al. 2015) (WO454_v2, Hardwood Genomics Project, https://doi.org/10.25504/FAIRsharing.srgkaf; RO454_v2, Hardwood Genomics Project, https://doi.org/10.25504/FAIRsharing.srgkaf). We used MarkerMiner v1.0 (Chamala et al. 2015) to identify 403 long-exon, single-copy nuclear genes shared across *Quercus* references that were expected to provide high-quality phylogenetic resolution. 30 candidate genes for bud-break phenology and waterlogging response, 30 for drought tolerance and 2 for freezing tolerance were included as targets (Lesur et al. 2015; Meireles et al. 2017; Oney-Birol et al. 2018; Ueno et al. 2013). Target loci are provided as a fasta file in Appendix S1 (Table S2).

DNA was extracted from frozen leaf tissue using a DNeasy Plant Mini Kit (Qiagen) with a modified DNeasy protocol as described in Hipp and Weber (2008). DNA was quantified using a Qubit Fluorometer and visualised on an agarose gel to confirm DNA quality. DNA was fragmented to approximately 550 bp using a Covaris M220 (Covaris, Woburn, MA, USA) or NEBNExt dsDNA fragmentase (New England Biolabs, Ipswich, MA, USA). Library preparation used the Adapterama iTru system and KAPA Hyper Prep Kit (Roche Diagnostics Corporation, Wilmington, MA, USA), with KAPA Pure Beads, KAPA HyperPure Beads, or Speedbeads (Gardner et al. 2021; Glenn et al. 2016; Hale et al. 2020; Rohland and Reich 2012). Pools

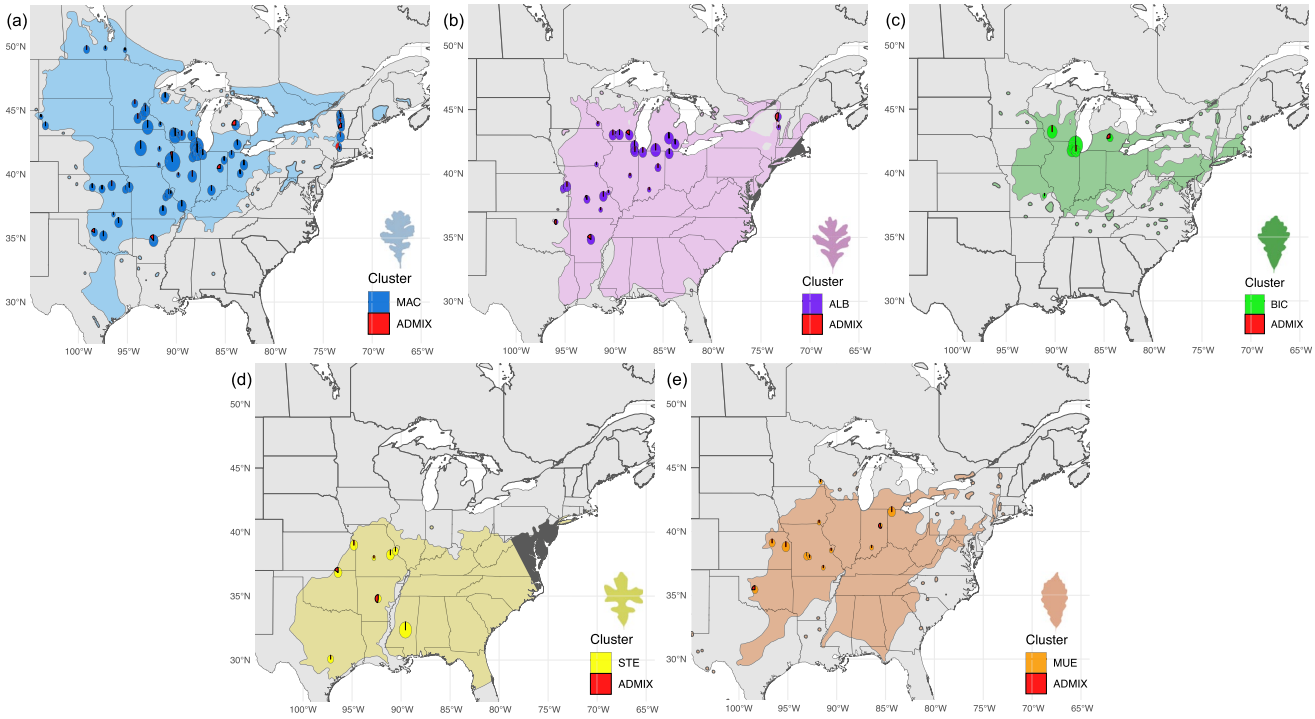

FIGURE 2 | Collection sites and genotyped samples. For the five species sampled—(a) *Q. macrocarpa*, (b) *Q. alba*, (c) *Q. bicolor*, (d) *Q. stellata* and (e) *Q. muehlenbergii*—radius of each pie chart corresponds to the log-transformed sample size per site. In each pie chart, conspecific and heterospecific *Q*-values are averaged over individuals within a site. The two *Q. macrocarpa* clusters in the $K=6$ ADMIXTURE run were summed to produce a single *Q. macrocarpa* conspecific *Q*-value (ancestry proportion) in (a). Range maps are from Little's (1971) *Atlas of United States Trees*.

of 12–16 libraries were enriched using our 465-gene custom *Quercus* MYbaits kit (Morales-Saldaña et al. 2024) and either the Mybaits v3 or v4 reagent kit (Arbor Biosciences, Ann Arbor, MI, USA). Hybridization was done at 65°C for 16 h with ½ volume of baits used per enrichment. QIAquick PCR purification kit (Qiagen) was used to purify PCR-amplified enriched libraries. Pooled libraries were quantified using a Qubit Fluorometer and run on a Bioanalyzer (Agilent, Santa Clara, CA, USA) to ascertain fragment size distributions before these were combined for sequencing. Samples were subsequently sequenced using the Illumina MiSeq (2×300 bp) or HiSeq 4000 (2×150 bp) platforms. Two-sample *t*-tests with sequencing platform as the predictor of latitude were used on all samples as well as *Q. macrocarpa* samples to test whether there was a latitudinal bias in the sequencing platform used.

2.3 | Bioinformatics Workflow

Raw FASTQ files were cleaned with Trimmomatic 0.39 (Bolger et al. 2014) to remove ILLUMINA adapters, leading and trailing bases with Phred score <15, portions of reads with an average Phred score <20 across a 4 bp sliding window, and cleaned reads shorter than 40 bp. Subsequently, reads were mapped to the *Q. lobata* genome v3.2 (Sork et al. 2022) using BWA-MEM 0.7.17 (Li and Durbin 2009). Duplicate reads were marked using Samtools 1.17 (Danecek et al. 2021). Variants were called with FreeBayes 1.3.6 (Garrison and Marth 2012) using default settings. After decomposing multinucleotide polymorphisms and retaining only biallelic SNPs, we implemented an iterative filtering pipeline adapted from O'Leary et al. (2018, pipeline FS5). Site genotyping rates and missing data per individual filters were repeatedly applied within the same pipeline according to progressively more stringent thresholds, removing low-quality variants while maximising the number of retained samples and SNPs.

Filtering was performed through a custom pipeline of vcflib 1.0.3 (Garrison et al. 2022), VCFtools 0.1.16 (Danecek et al. 2011) and BCFtools (Danecek et al. 2021); see details in methods in Appendix S1. After filtering, SNPs were thinned to a minimum distance of 500 bp using PLINK 1.9 (Purcell et al. 2007), as prior studies in oaks have demonstrated rapid decay of linkage disequilibrium, both within and across individual genes (Kremer et al. 2012; Nocchi et al. 2022; Sork, Squire, et al. 2016). Scripts to conduct variant-calling and SNP-filtering are archived on Zenodo (v 0.9–3, https://doi.org/10.5281/zenodo.15216009). Sequence reads are archived on NCBI (SRA Bioproject #1223965; http://www.ncbi.nlm.nih.gov/bioproject/1223965).

2.4 | Admixture and Spatial Genetic Analysis

Ancestry estimation was performed using the maximum-likelihood software ADMIXTURE (Alexander et al. 2009), which jointly optimises the allele frequencies of a predefined number of ancestral clusters (K) and the proportional membership ("*Q*-value") of each individual in each cluster based on SNP genotype data. Fifty replicate runs were carried out for each value of K from $K=2$ to $K=8$, and the run with the highest log-likelihood per K was selected for visualisation and

downstream analysis, as performed by Marcus et al. (2021). Evaluation of the optimal value of K was performed using the default ADMIXTURE cross-validation procedure enabled by the "--cv" flag. We visualised ADMIXTURE ancestry estimation and cross-validation results with ggplot2 (Wickham 2016) in R version 4.3.0 (R Core Team 2023). We also ran ADMIXTURE on downsampled datasets with 20 randomly sampled individuals per species and only putatively "pure" *Q. macrocarpa* individuals ($Q > 0.95$) to assess the robustness of our dataset to detect population structure within this species.

Ancestral clusters in the $K = 5$ and $K = 6$ runs corresponded almost exactly to the 5 morphospecies included in our study, with the exception of *Q. macrocarpa*, which was separated into two admixed subpopulations at $K = 6$. For subsequent analyses using ADMIXTURE Q-values, we defined the conspecific Q-value for non-*Q. macrocarpa* individuals as the maximum ancestry proportion in the $K = 6$ cluster, which almost invariably matched a sample's morphospecies determination. For samples identified as *Q. macrocarpa*, the conspecific Q-value was calculated as the sum of the Q-values for the two *Q. macrocarpa* clusters estimated in the $K = 6$ run. Heterospecific Q-values were calculated as 1—conspecific Q-value.

We tested for spatial autocorrelation of the northern *Q. macrocarpa* ancestral cluster (hereafter termed MAC_N, in contrast to the southern genetic cluster subsequently referred to as MAC_S) using global Moran's I (Moran 1950). "Pure" (conspecific Q-value > 0.95) *Q. macrocarpa* individuals were grouped by sampling site. An inverse distance matrix for spatial weighting was calculated using each site's centroid coordinate. Pairwise geographic distances between sites were calculated using the Haversine formula, assuming a spherical globe (Sinnott 1984). Moran's I was calculated with the MAC_N Q-value for each *Q. macrocarpa* individual using adegenet 2.1.10 and spdep 1.3_3 in R (Bivand 2022; Jombart 2008; Jombart and Ahmed 2011). We evaluated the significance of Moran's I via permutation test with 99,999 resamples to obtain a one-tailed p-value, corresponding to an alternative hypothesis of positive spatial autocorrelation.

To further clarify whether genetic clustering within *Q. macrocarpa* reflected either continuous genetic variation or genetic discontinuity, we examined spatial genetic variation using the Python package Fast Estimation of Effective Migration Surfaces (FEEMS) (Marcus et al. 2021). FEEMS uses deme-aggregated allele frequencies at spatial nodes to fit edge-specific effective migration estimates, inferring landscape-wide patterns of heterogeneous isolation-by-distance via a penalised likelihood approach. We subsetted the thinned SNP dataset to only precisely georeferenced *Q. macrocarpa* samples with conspecific Q-value > 0.95 and removed any resulting invariant sites ("--maf" PLINK flag) from the PLINK binary files. Demes were constructed by aggregating individuals to their nearest node in a 100 km-resolution spatial grid. We performed leave-one-out cross-validation to determine the optimal value of the model tuning hyperparameter λ, evaluating 20 values of λ from 10^{-6} (low smoothing) to 10^{2} (high smoothing). All analyses were conducted using modified FEEMS tutorial scripts published by the package authors (https://github.com/Novembrelab/feems).

2.5 | Hybrid Class Analysis

To more precisely characterise hybrids, we used NewHybrids, a Bayesian ancestry inference algorithm that uses patterns of observed genotype frequencies and inferred population allele frequencies to estimate the posterior probability of assignment to a predefined set of hybrid classes (Anderson and Thompson 2002), specifying expected genotype frequencies for multiple backcross generations (Table S3). Identifying hybrids based on genotype frequency is more tractable with loci that approach fixation between populations (Wringe et al. 2017). For this reason, and because NewHybrids can only consider pairwise species comparisons, we selected the 200 most highly differentiated loci per species pair using the getTopLoc function in the R package HybridDetective 4.3.1 (Wringe et al. 2017) based on putatively pure individuals identified in ADMIXTURE runs (conspecific Q-value > 0.95). The choice of 200 loci for analysis was based on computational limits of NewHybrids and supported by a previous finding that including more than 10 markers per linkage group (more than 120 loci in the case of oaks) provides minimally increased power to detect F1s and first-generation backcrosses (Chakraborty and Rannala 2023). All individuals with cumulative two-species Q-values > 0.95 across pairwise parental ancestral clusters were considered for each run of NewHybrids; individuals with substantial hybrid parentage from more than two species were therefore excluded, as NewHybrids does not model 3-way hybrids. For each species pair, four independent MCMC chains were run with an uninformative Jeffreys prior for a burn-in period of 300,000 iterations prior to 600,000 sampling sweeps. Traceplots and effective sample sizes, computed across all chains per pairwise analysis for the population-wide hybrid class membership proportion parameters via the R package coda 0.19–4 (Plummer et al. 2006), were inspected to evaluate MCMC convergence. As a nonparametric check on the hybrid class assignment inferred by NewHybrids, we plotted ancestry proportion against interclass heterozygosity for all individuals in the *Q. alba*—*Q. macrocarpa* and *Q. bicolor*—*Q. macrocarpa* comparisons, using triangulaR v.0.0.1 (Wiens and Colella 2024).

2.6 | Environmental Predictors

BIOCLIM rasters were downloaded at 30 arcsecond resolution from the WorldClim 2 database (Fick and Hijmans 2017). Moisture index I_m was calculated as $100 \times \frac{MAP - PET}{PET}$ (Thornthwaite 1948), where MAP is mean annual precipitation (BIO12) and PET is potential evapotranspiration extracted from the Global Aridity Index and Potential Evapotranspiration Database v3 (Zomer et al. 2022). Little's (1971) species range maps, digitised and compiled by Prasad and Iverson (2003), were downloaded as shapefiles (https://github.com/andrew-hipp/white-oak-syngameon/tree/master/data/little-maps; doi: 10.5281/zenodo.13150341). For all spatial analyses, coordinates were transformed to the North American Albers equal area conic projection to ensure accurate raster grid sizing. Projections and all subsequent spatial calculations, analyses and raster extractions were performed using the R packages sf v1.0–14 and terra v1.7–39 (Hijmans 2019–2025; Pebesma 2018).

Given range-margin uncertainty and long-distance pollen dispersal (Ashley 2021), we extended all range limits by 20 km for analyses involving range calculations. Distance from range

edge was calculated as the distance between a sample's coordinate and the nearest buffered range edge, with lake and ocean shorelines masked to avoid artificial range truncations. Number of sympatric species was calculated as the number of eastern North American white oak ranges (of the 5 species analysed in our study), aside from an individual's species determination, overlapping with the coordinates of a sample. Soil data were obtained from the USDA-NRCS Soil Survey Geographic Database (SSURGO; Soil Survey Staff, Natural Resources Conservation Service, United States Department of Agriculture, n.d.) accessed through the Web Soil Survey API using a custom R script.

2.7 | Species Distribution Modelling (SDM) and Species Co-Occurrence

Species occurrence records were obtained from the Global Biodiversity Information Facility (GBIF) database and previously curated USDA Forest Inventory and Analysis (FIA) data (Cavender-Bares et al. 2018) (https://gitlab.com/meireles/compile_fia_data/). GBIF records were filtered to remove erroneous coordinates and data points corresponding to county or state centroids, fossils and planted specimens, urban occurrences and coordinates over bodies of water using R packages rgbif 3.7.7 and CoordinateCleaner 2.0.20 (Chamberlain et al. 2024; Chamberlain and Boettiger 2017; Zizka et al. 2019). FIA data as analysed in Cavender-Bares et al. (2018) were used to summarise species co-occurrence, treating the presence of two species within any of the subplots of a single FIA plot as co-occurrence.

Preliminarily cleaned occurrence records from both sources were merged, downloaded and manually curated to remove spatial outliers in QGIS 3.28 (QGIS Development Team 2023) and thinned to a minimum distance of 20 km via spThin 0.2.0 (Aiello-Lammens et al. 2015). Study extent was defined by the 250 km-buffered convex hull of occurrence points for each respective species, with 20,000 unique, randomly sampled raster grid cells serving as a background dataset. We used all 19 BIOCLIM variables as environmental predictors in our SDMs, as Maxent's regularisation algorithm is robust to predictor multicollinearity (Merow et al. 2013). SDMs were fit using MAXENT 3.4.3 as implemented in the R package ENMeval 2.0.4 (Kass et al. 2021; Phillips et al. 2017). See methods in Appendix S1 for model-fitting details.

2.8 | Modelling Predictors of Admixture

We employed Bayesian generalised linear mixed models (GLMMs) in brms v2.20.1 (Bürkner 2017; Carpenter et al. 2017) to examine associations between environmental and ecological factors and introgression and *Q. macrocarpa* population structure. Beta regression with a logit link and default flat priors on all parameters was used for all models, and sampling site was incorporated as a random effect to control for spatial autocorrelation and reduce the possible risk of pseudoreplication (Douma and Weedon 2019). See methods in Appendix S1 for details of model-fitting and MCMC diagnostics.

To test the association of genetic differentiation in *Q. macrocarpa* with the gradient from more southern bottomlands to more northern uplands, we included as fixed effects climate and soil predictors expected to vary along moisture and temperature gradients, with expectations as follows: temperature seasonality (BIO4) and moisture index (I_m), based on our previous work demonstrating that both vary strongly among North American oak species (Hipp et al. 2018); maximum temperature of the warmest month (BIO5), as a proxy for drought severity; mean annual precipitation (BIO12), as a predictor of overall plant water requirements; and categorical soil predictors for ponding frequency, flooding frequency and drainage class, as predictors of wetland status (which we expect would differentiate *Q. alba* and *Q. stellata* as upland species most strongly from *Q. bicolor*, *Q. lyrata* and bottomland ecotypes of *Q. macrocarpa*). For modelling population structure within *Q. macrocarpa*, we analysed only individuals with conspecific Q-value > 0.95, including site as a random effect. Whole-dataset admixture was modelled by regressing heterospecific Q-value against latitude, number of sympatric species, Maxent-inferred habitat suitability, and distance from range edge, with species as a fixed categorical predictor and a random effect for site. Because we were interested in understanding the drivers of gene flow between species, rather than merely first-generation hybridization, F1s identified by NewHybrids were excluded from the admixture GLMM. See methods in Appendix S1 for modelling details.

2.9 | Admixture at Candidate Versus Background Loci

Single-nucleotide polymorphisms were annotated to indicate which candidate genes they fell within, and genes with fewer than three SNPs were excluded from jackknife analysis. To test whether genetic differentiation varied significantly between candidate gene classes, we obtained per-class F_{ST} distributions by jackknife resampling (10,000 replicates) of 15 genes per class (bud break phenology, drought tolerance, and other) and three SNPs per gene. Pairwise F_{ST} estimates (calculated per Weir and Goudet 2017) for each combination of two species were computed using the R package hierfstat v. 0.5–11 (Goudet 2005) for each jackknife replicate. We compared the means of the F_{ST} distributions for the BBP and DT classes to that of the 'Other' genes to calculate two-tailed p-values. See methods in Appendix S1 for details on establishing gene orthology, categorising genes, and analysis.

3 | Results

3.1 | Sequencing Results and SNP Recovery

Our post-filtering SNP dataset consisted of 3412 biallelic SNPs (Figure 3, black dots) with an average SNP genotyping rate of 87% across all individuals and a mean read depth of 22.0 ± 12.1 SD across all SNPs, averaged over all samples. Variants were recovered from 445 of the 465 target genes (Figure 3, red bars); nine of the targets initially identified in *Q. robur* failed to BLAST to the *Q. lobata* genome and 11 lacked any associated SNPs. Target loci and SNPs were distributed across all 12 *Quercus* chromosomes and four unsorted nuclear scaffolds

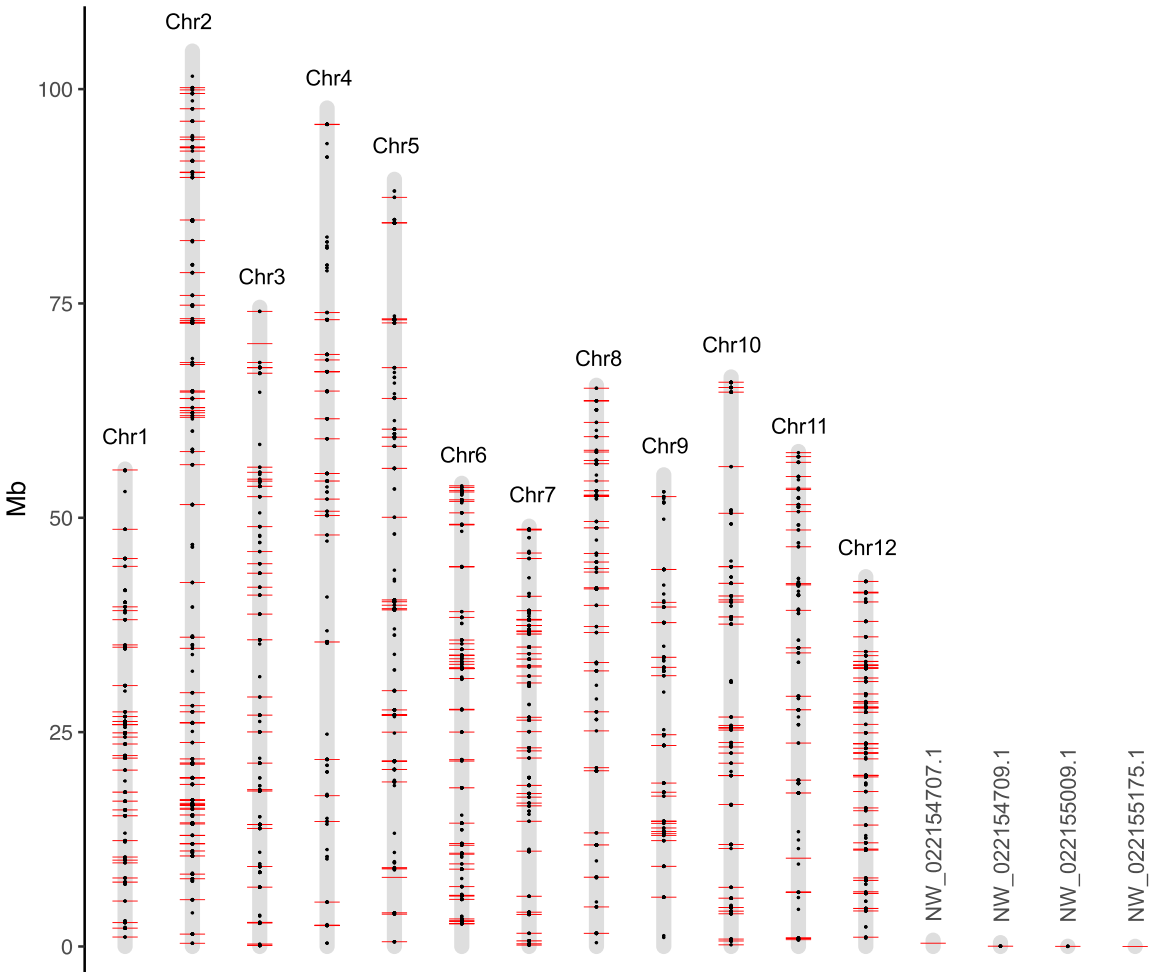

FIGURE 3 | Hyb-Seq SNP (black dots) and target gene (red bars) recovery. 3412 thinned SNPs used in final ADMIXTURE analyses and 456 genes identified from reciprocal best BLAST results (using *Q. robur* targets) in *Q. lobata*. Bars correspond to the 12 *Quercus* chromosomes and 4 unassigned genome scaffolds in the *Q. lobata* reference genome across which target genes and SNPs were identified. Chromosomes and scaffolds are scaled by size (in Mb).

(as mapped to *Q. lobata* genome v3.2). Of filtered and thinned SNPs, 22% fell outside of the target gene regions. Given that Hyb-Seq can recover high-copy, off-target nuclear regions (Weitemier et al. 2014), and because we applied stringent filtering parameters to remove erroneous and poorly resolved paralogous calls, we chose to retain off-target SNPs for subsequent analyses. ADMIXTURE results below are those from analyses conducted with the full dataset; results using only SNPs confined to target loci produced highly similar ancestry estimates (not shown). Of the 517 individuals sequenced, 364 (224 of 319 *Q. macrocarpa* [70.2%], 64 of 78 *Q. alba* [82.1%], 30 of 60 *Q. bicolor* [50.0%], 25 of 31 *Q. stellata* [80.6%] and 21 of 29 *Q. muehlenbergii* [72.4%]) passed the iterative SNP filtering pipeline (Table S1). 318 individuals with associated GPS coordinates or manual georeferences were retained for downstream spatial analyses, including GLMMs (Table 1). Sequencing platform was unbiased with respect to sample latitude across species ($t = -0.99667$, df = 184.63, $p = 0.3202$) as well as within just *Q. macrocarpa* ($t = -0.34215$, df = 230.86, $p = 0.7325$).

3.2 | Genetic Structure Across the Eastern North American White Oak Syngameon

Non-metric multidimensional scaling (NMDS) ordination of the SNP distance matrix revealed well-defined, discontinuous clusters corresponding to the 5 species sampled, with a much smaller number of admixed individuals found at varying distances between parental clusters (Figure 4a,b). ADMIXTURE's cross-validation procedure for determining the optimal number of ancestral clusters (K) returned a minimum prediction error at $K = 6$ (0.17007) and slightly higher errors for $K = 5$ (0.17232) and $K = 7$ (0.17214) (Figure 5). Cross-validation errors increased considerably for values of K smaller than 5 and larger than 7, supporting both species-level genomic delineations ($K = 5$; Figure 5a) and a population genetic divide within *Q. macrocarpa* (which appears only at $K \geq 6$; Figure 5b, Figure S1). As these genomic clusters within *Q. macrocarpa* are strongly geographically structured between northern and southern populations (see results below, *Spatial and environmental predictors of admixture and*

TABLE 1 | List of sampling sites for rangewide study.

Site	State	Latitude	Longitude	Sample size
Little Rock	AR	34.8180	−92.3360	3 Q. alba; 4 Q. macrocarpa; 2 *Q. stellata*
Litchfield	CT	41.9623	−73.3130	1 *Q. macrocarpa*
Ames	IA	42.0363	−93.6430	7 *Q. macrocarpa*
Bonaparte	IA	40.7056	−91.7947	1 Q. alba; 1 Q. macrocarpa; 1 *Q. muehlenbergii*
Cedar Rapids	IA	41.9736	−91.7239	1 *Q. macrocarpa*
Aurora	IL	41.7877	−88.2729	5 *Q. bicolor*
Chickenbristle	IL	39.8393	−88.3677	1 Q. alba; 4 *Q. macrocarpa*
Cranberry Slough	IL	41.7173	−87.8652	3 Q. alba; 4 Q. bicolor; 9 *Q. macrocarpa*
Fullersburg Woods	IL	41.8252	−87.9429	3 Q. alba; 1 Q. bicolor; 3 *Q. macrocarpa*
Galesburg	IL	40.9893	−90.3978	14 Q. macrocarpa
Goose Lake	IL	41.3571	−88.3229	3 *Q. macrocarpa*
Middlefork	IL	42.2578	−87.8816	2 *Q. macrocarpa*
Morton Arboretum	IL	41.8135	−88.0629	1 *Q. macrocarpa*
Ryerson Woods	IL	42.1818	−87.9149	3 Q. alba; 12 Q. bicolor; 10 Q. macrocarpa
Shawnee	IL	37.5163	−89.4453	4 *Q. macrocarpa*
Starhill	IL	39.9355	−89.8007	1 *Q. macrocarpa*
Dunes	IN	41.6588	−87.0575	3 *Q. alba*
Fort Wayne	IN	41.0891	−85.1199	2 *Q. macrocarpa*
Hobart	IN	41.5331	−87.2944	3 *Q. macrocarpa*
Spring Mill	IN	38.7369	−86.4133	1 Q. alba; 3 Q. macrocarpa; 1 *Q. muehlenbergii*
Taylor	IN	40.4587	−85.5082	2 Q. alba; 2 Q. macrocarpa; 1 *Q. muehlenbergii*
Baldwin	KS	38.8090	−95.1913	2 Q. alba; 2 Q. macrocarpa; 3 *Q. muehlenbergii*
Konza	KS	39.1053	−96.6030	3 Q. macrocarpa; 2 *Q. muehlenbergii*
Minooka	KS	38.9636	−98.5891	2 *Q. macrocarpa*
Salina	KS	38.8429	−97.5889	2 *Q. macrocarpa*
Shawnee Mission	KS	38.7521	−94.3685	3 Q. alba; 3 Q. macrocarpa; 2 Q. muehlenbergii; 3 *Q. stellata*
Berkshire	MA	42.1666	−73.4121	2 *Q. macrocarpa*
Assiniboine	MB	49.8578	−97.2491	1 *Q. macrocarpa*
Spruce Woods	MB	49.7610	−99.1594	2 *Q. macrocarpa*
Whiteshell	MB	49.7119	−95.2446	1 *Q. macrocarpa*
Ann Arbor	MI	42.3049	−83.7526	3 Q. alba; 3 *Q. macrocarpa*
East Lansing	MI	42.7666	−84.3877	4 Q. alba; 2 *Q. bicolor*
Pinconning	MI	43.8529	−83.9240	3 *Q. macrocarpa*
Three Rivers	MI	41.8409	−85.7457	5 *Q. alba*
Austin	MN	43.6828	−92.9292	6 *Q. macrocarpa*
Cedar Creek	MN	45.3895	−93.1949	1 *Q. macrocarpa*
Ottawa Bluffs	MN	44.3640	−93.9355	3 *Q. macrocarpa*
Quarry Park	MN	45.5304	−94.2364	2 *Q. macrocarpa*

(Continues)

TABLE 1 | (Continued)

Site	State	Latitude	Longitude	Sample size
Sakatah Lake	MN	44.6243	−93.3961	3 *Q. macrocarpa*
Twin Cities	MN	44.9557	−93.1620	4 *Q. macrocarpa*
Winona	MN	43.8992	−91.6427	1 Q. alba; 1 Q. macrocarpa; 1 *Q. muehlenbergii*
Eminence	MO	37.1573	−91.3650	1 Q. alba; 3 Q. macrocarpa; 1 *Q. muehlenbergii*
Eureka	MO	38.5122	−90.5657	1 Q. alba; 1 Q. macrocarpa; 1 Q. muehlenbergii; 2 *Q. stellata*
Ha Ha Tonka	MO	37.9732	−92.7623	2 Q. alba; 1 Q. muehlenbergii; 1 *Q. stellata*
Shaw	MO	38.4639	−90.8100	3 *Q. macrocarpa*
Sullivan	MO	38.2226	−91.0855	3 Q. alba; 1 Q. bicolor; 2 Q. macrocarpa; 3 *Q. stellata*
Bienville	MS	32.3582	−89.5472	8 *Q. stellata*
Daughmer	OH	40.7299	−83.0936	3 *Q. macrocarpa*
Goll Woods	OH	41.5522	−84.3574	3 Q. alba; 2 Q. macrocarpa; 3 *Q. muehlenbergii*
Pearl King	OH	40.0440	−83.4794	2 *Q. macrocarpa*
Hinton	OK	35.4453	−98.3543	2 Q. macrocarpa; 2 *Q. muehlenbergii*
Norman	OK	35.1771	−97.4496	3 *Q. macrocarpa*
Pawhuska	OK	36.8455	−96.4232	1 Q. macrocarpa; 3 *Q. stellata*
Tulsa	OK	36.2177	−95.8987	1 Q. alba; 3 *Q. macrocarpa*
Custer	SD	43.7887	−103.3634	2 *Q. macrocarpa*
Spearfish	SD	44.4913	−103.8090	1 *Q. macrocarpa*
Buescher	TX	30.0764	−97.1817	2 *Q. stellata*
Addison	VT	43.9580	−73.1881	3 *Q. macrocarpa*
Bennington	VT	42.9102	−73.2193	3 *Q. macrocarpa*
Burlington	VT	44.4682	−73.2109	2 Q. alba; 2 *Q. macrocarpa*
Ferrisburgh	VT	44.2398	−73.2444	1 Q. alba; 3 *Q. macrocarpa*
Grand Isle	VT	44.6925	−73.3354	2 *Q. macrocarpa*
Rutland	VT	43.6341	−73.1529	1 Q. alba; 2 *Q. macrocarpa*
Dodgeville	WI	43.0187	−90.1162	3 Q. alba; 7 *Q. macrocarpa*
Gotham	WI	43.2451	−90.3034	5 Q. bicolor; 3 *Q. macrocarpa*
Madison	WI	43.0391	−89.4394	3 Q. alba; 3 *Q. macrocarpa*
Ottawa	WI	43.0168	−88.4384	3 Q. alba; 3 *Q. macrocarpa*
Pleasant Valley	WI	43.1068	−89.8063	1 Q. alba; 1 *Q. macrocarpa*
Round Lake	WI	46.0247	−91.1393	3 *Q. macrocarpa*

Note: Sites represent the closest named locality to the population; state/province abbreviations are standard. Latitude and longitude are computed as the centroid of the coordinates for all individually georeferenced samples at a given site and rounded to 4 decimal places.

population structure), we refer to these clusters as MAC_N and MAC_S throughout the paper.

Of the 364 individuals included in the total-dataset ADMIXTURE runs, 19 (5.2%) were estimated as being F1s (17) or three-way hybrids (2) based on having their maximum *Q*-value < 0.60; all but one has at least 0.30 assignment to *Q. macrocarpa* (the exception being an inferred F1 between *Q. alba* and *Q. muehlenbergii*). The two three-way hybrids inferred include *Q. alba* as the dominant genotype ($Q = 0.52$, 0.57) and *Q. macrocarpa* and *Q. bicolor* as the only other genomic components of $Q > 0.05$ ($Q = 0.14$–0.32). Some of these might also be F2, F3 or other generation hybrids, but NewHybrid analyses demonstrate that we have minimal power to distinguish among these classes. Of the remaining samples, 6 (1.6%) have a maximum *Q*-value < 0.9, 18 (4.9%) have maximum *Q*-value < 0.95 and 56 (15.4%) have maximum *Q*-value < 0.99 (Table S4). The distribution of introgressed (not hybrid; i.e., mixed ancestry

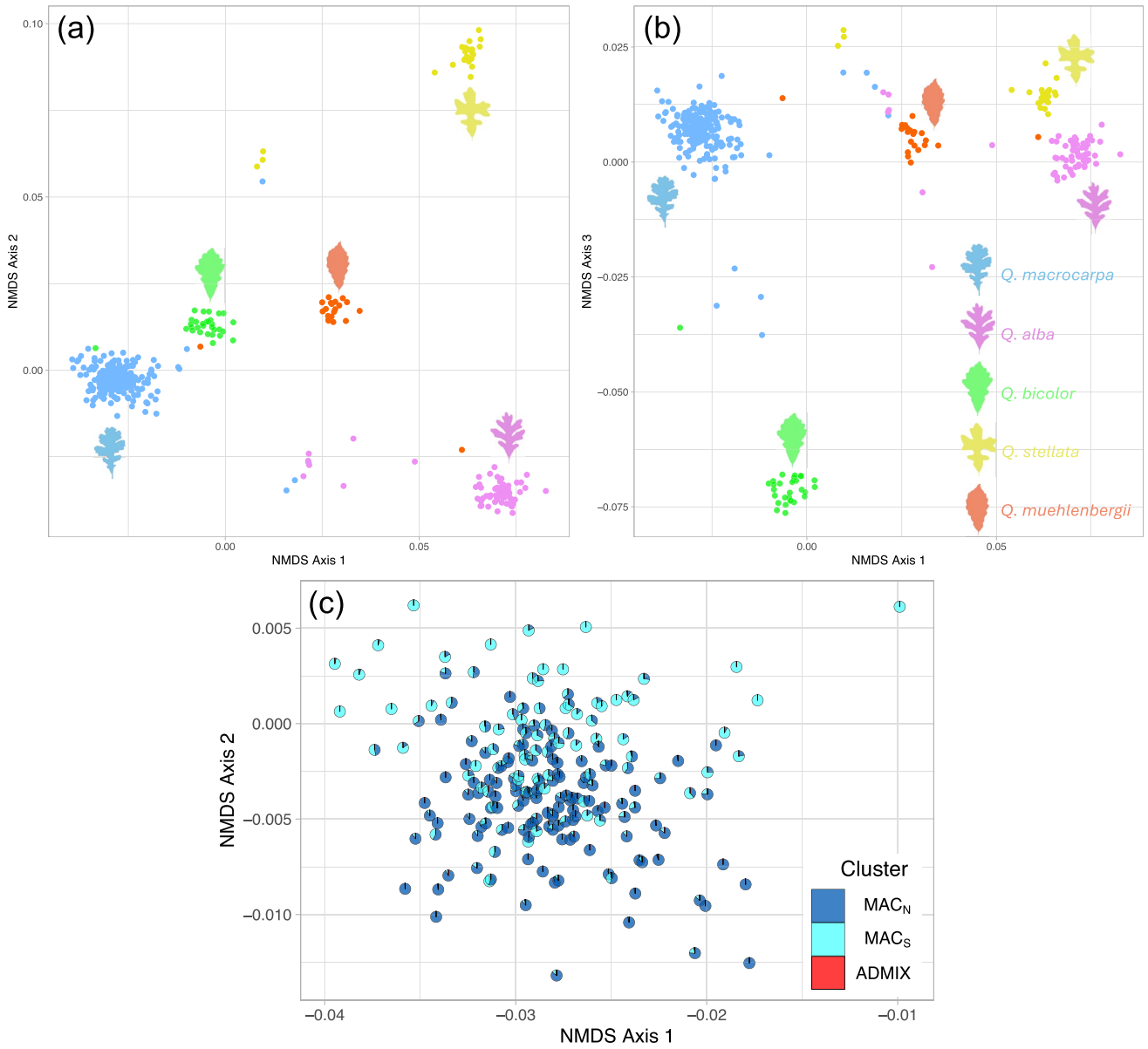

FIGURE 4 | Non-metric multidimensional scaling (NMDS) plot of SNP genetic distances. Ordination along NMDS (a) axes 1 and 2 and (b) axes 1 and 3 of all individuals that were successfully genotyped, with colours and leaf silhouettes denoting morphological species designations. (c) Ordination along axes 1 and 2 of unadmixed (conspecific Q-value > 0.95) *Q. macrocarpa* samples, with proportional membership in either of the two *Q. macrocarpa* genetic clusters inferred in the $K = 6$ ADMIXTURE run, along with heterospecific Q-values, depicted in pie chart fill.

other than inferred F1) individuals within each species at the $0.60 < Q < 0.95$ threshold was 14 out of 217 samples (6.5%) in *Q. macrocarpa*, 1 out of 58 (1.7%) in *Q. alba*, 2 out of 29 (6.9%) in *Q. bicolor*, 1 out of 22 (4.5%) in *Q. stellata*, and 0 out of 19 (0%) in *Q. muehlenbergii* or 18 out of 345 (5.2%) for the whole dataset (Table S4). Within *Q. macrocarpa*, samples were much more likely to have assignment to both MAC_N and MAC_S than to exhibit admixture with different species: of the 203 samples with $MAC_N + MAC_S > 0.95$, 119 (58.6%) had $Q < 0.95$ membership in one of the two *Q. macrocarpa* genetic clusters. However, genetic differentiation between MAC_N and MAC_S is lower than interspecific variation: pairwise F_{ST} from ADMIXTURE between inferred *Q. macrocarpa* genetic clusters was 0.103, compared to 0.292–0.468 for interspecific comparisons (Figure 5).

3.3 | Hybrid Class Assignment

362 out of 364 individuals had a summed $Q > 0.95$ across at most two species clusters and were included in pairwise NewHybrids analyses. Using a 0.9 posterior probability threshold of hybrid class assignment, 314 individuals (86.7%) were confidently classified as "pure" (unadmixed) members of their determined species, 16 (4.4%) were classified as F1s, 7 (1.9%) were classified as backcrosses (BC1, BC3, or BC of undetermined generation), and 2 (0.55%) were classified as F2s (Table S5; Data Summary S1). Only three of the seven backcrosses (0.8%) could be confidently assigned to a specific backcross generation. Additionally, discrimination between putatively pure individuals and third-generation backcrosses proved difficult for NewHybrids; 23

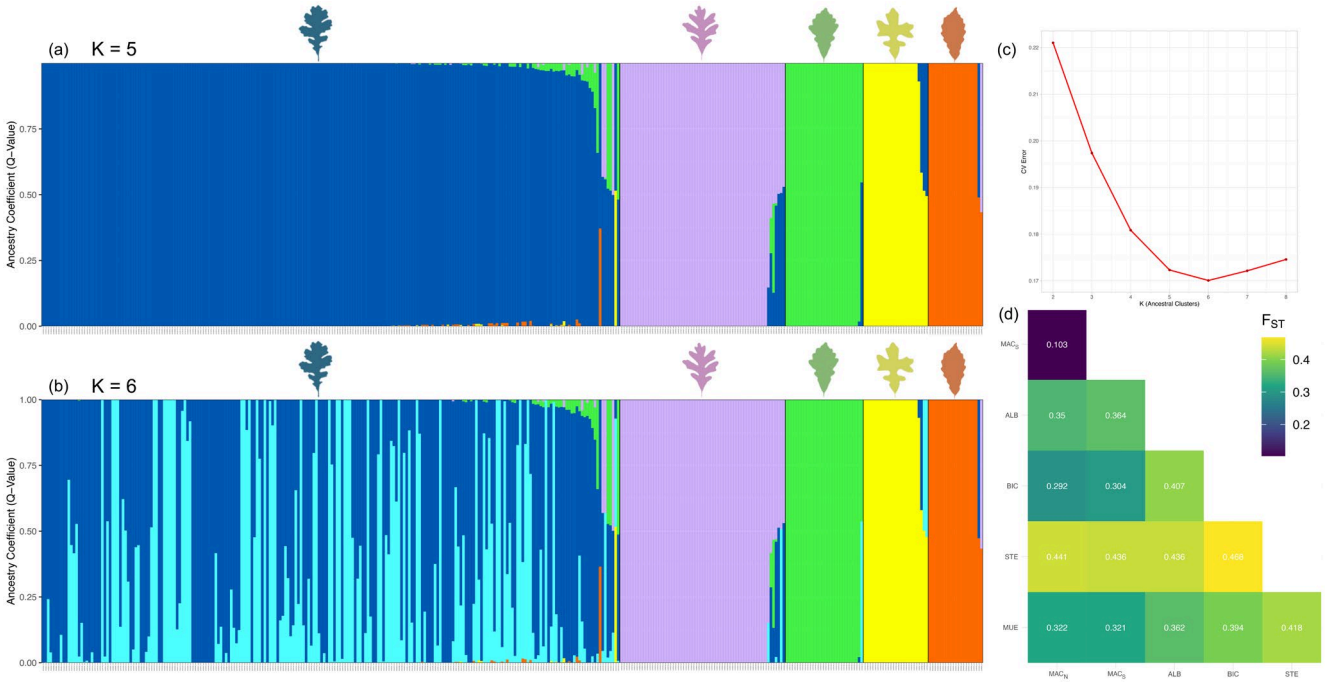

FIGURE 5 | ADMIXTURE analysis. Admixture plots for (a) $K = 5$ and (b) $K = 6$, partitioned by species (leaf silhouettes; from left to right, *Q. macrocarpa*, *Q. alba*, *Q. bicolor*, *Q. stellata*, *Q. muehlenbergii*). (c) Plot of cross-validation error as a function of K, from $K = 2$ to $K = 8$. (d) Heatmap of pairwise F_{ST} scores between ancestral clusters inferred in the $K = 6$ ADMIXTURE run.

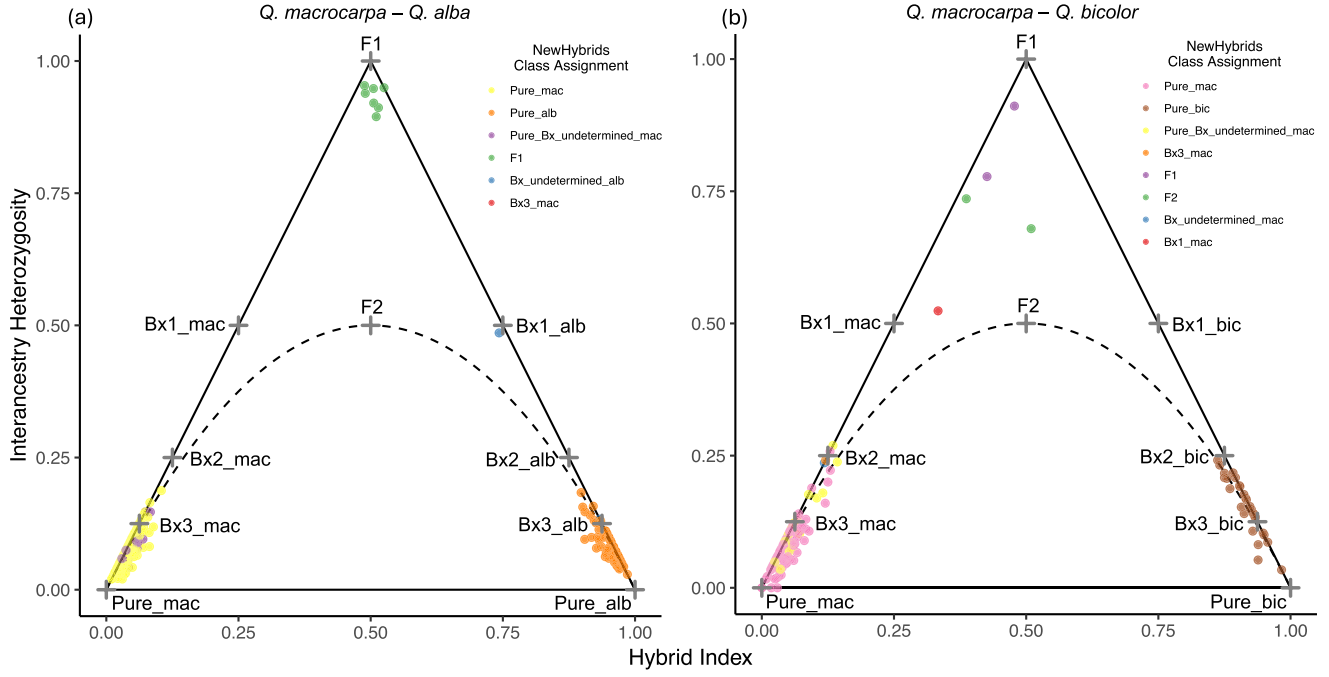

FIGURE 6 | Triangle plots. Interancestry heterozygosity plotted against hybrid index (i.e., ancestry proportion) for the (a) *Q. macrocarpa*—*Q. alba* and (b) *Q. macrocarpa*—*Q. bicolor* species pairs. Samples are coloured according to their NewHybrids consensus hybrid class assignments. Grey plus signs indicate the expected position of specified hybrid classes on the triangle plots. Expected heterozygosity as a function of ancestry fraction under Hardy–Weinberg conditions is given by the dashed parabola.

(6.3%) individuals could not be unambiguously placed into either class (Table S5). The percentage of admixed individuals in each pairwise comparison ranged from 0% (*Q. stellata*—*Q. muehlenbergii*, *Q. bicolor*—*Q. stellata*, *Q. bicolor*—*Q. muehlenbergii*, *Q. alba*—*Q. stellata*, *Q. alba*—*Q. bicolor*) to 11.6% (29 out of 248) for the *Q. macrocarpa*—*Q. bicolor* pair (Table S6; Data Summary S1). NewHybrids analyses converged across all MCMC chains, with effective sample sizes of > 100,000 for all parameters across species pairs (Table S7). Triangle plots for *Q. macrocarpa* × *bicolor* and *Q. macrocarpa* × *alba* (Figure 6) demonstrate that F1s

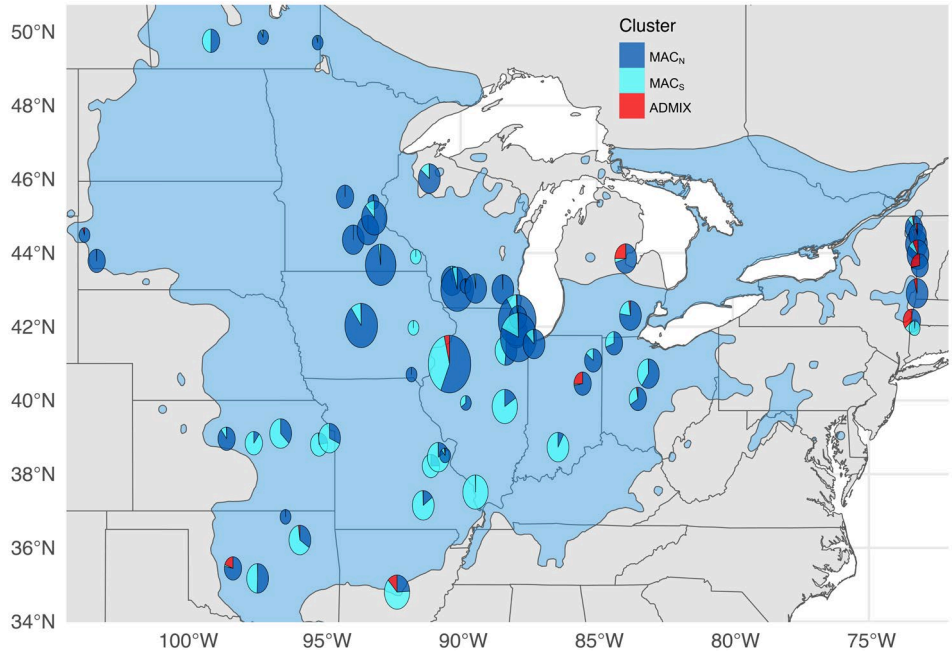

FIGURE 7 | Map of genetic cluster assignment within *Q. macrocarpa*. Radius of each pie chart corresponds to the log-transformed sample size per site. MACN and MACS correspond to the two ancestral clusters inferred within *Q. macrocarpa* in the $K=6$ ADMIXTURE run; ADMIX represents the heterospecific Q-value, i.e., $1-\left(Q_{\mathrm{MAC_S}}+Q_{\mathrm{MAC_N}}\right)$. Within each pie chart, MACN, MACS, and ADMIX Q-values are averaged over individuals within a site.

and F2s are strongly concordant with evidence from heterozygosity for both crosses. However, the inferred backcrosses beyond BC1 are not readily distinguishable from pure *Q. macrocarpa* in the cross with *Q. bicolor*, whereas the BC of undetermined generation toward *Q. alba* appears to be a plausible BC1.

3.4 | Co-Occurrence and Predictors of Admixture and Population Structure

Of the FIA plots subsetted to just those that had at least one oak, 39.5% had *Q. alba*, 14.3% had *Q. stellata*, 12.7% had *Q. macrocarpa*, 5.6% had *Q. muehlenbergii* and 2.2% had *Q. bicolor*. The most common associate of *Q. macrocarpa* was *Q. alba*, at 19.4%, followed by *Q. muehlenbergii* at 6.0%, then *Q. stellata* and *Q. bicolor*, both at 3.5%. The most common associate of *Q. alba* was *Q. stellata* (23.2%), followed by *Q. muehlenbergii* (7.5%), *Q. macrocarpa* (6.3%) and *Q. bicolor* (2.6%).

Proportional membership in ADMIXTURE-inferred *Q. macrocarpa* genetic clusters displayed significant spatial autocorrelation (Moran's $I=0.088391$, $p=0.009$), broadly corresponding to northern ($\mathrm{MAC_N}$) and southern ($\mathrm{MAC_S}$) genetic clusters (Figure 7). Cross-validation tuning of the FEEMS hyperparameter λ suggested that a modest degree of smoothing ($\lambda=2.069$) yielded the best model fit (Figure S2). Congruent with ADMIXTURE results, analysis of spatial genetic connectivity via FEEMS revealed a 300–500 km latitudinal band of reduced effective migration stretching from southern Ohio in the east to northern Kansas and southern Iowa in the west (Figure 8). Effective migration rates to the north and south were approximately 1–2 orders of magnitude larger than rates within the band of reduced effective migration.

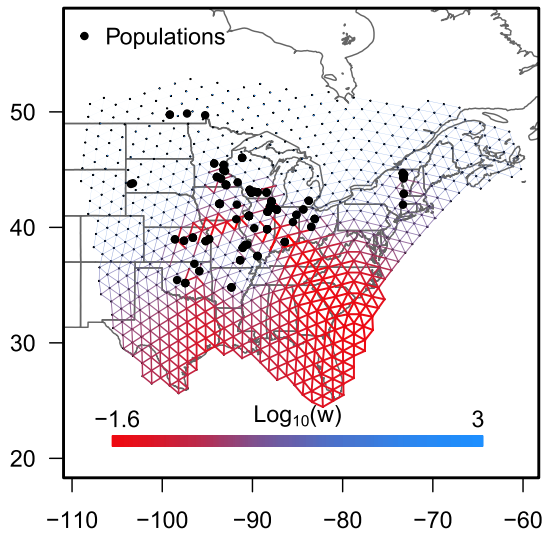

FIGURE 8 | Bur oak effective migration landscape estimated by FEEMS. Black points represent sampled populations of *Quercus macrocarpa*. Line segments forming the lattice are coloured (and thickened) according to log-transformed migration rates (w). Migration estimates are reported as the $\log_{10}$-transformed edge weights (w_{ik}) standardised against a scaling parameter w_0 estimated under a pure isolation-by-distance model, i.e., a constant value of w_{ik} across all edges.

Model selection via AICc favoured species distribution models (SDMs) with small regularisation multipliers and many feature classes (Table S8). SDMs projected across eastern North America corresponded closely to published range maps and known occurrence data for each species in our dataset (Little 1971) (Figure S3). Range edges displayed steep habitat suitability

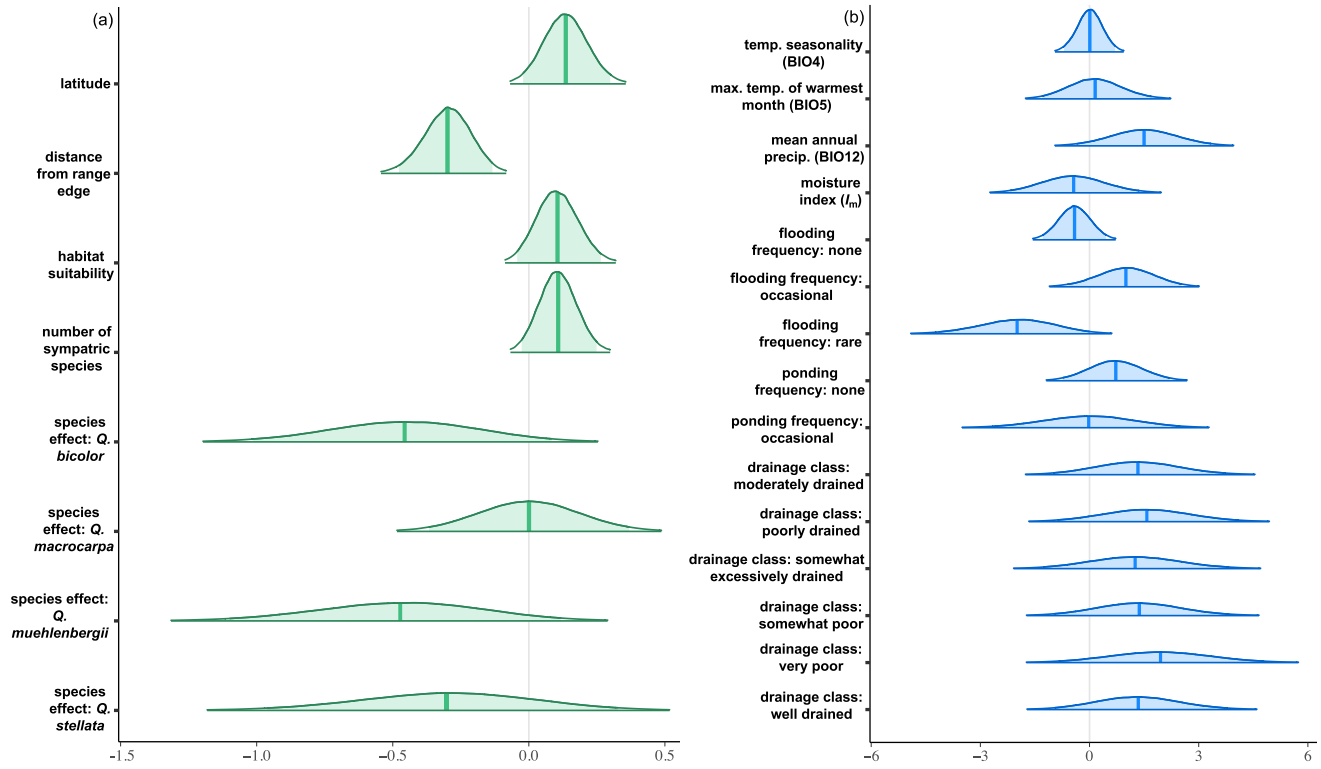

FIGURE 9 | Admixture and *Q. macrocarpa* population structure GLMM results. Posterior distributions of regression coefficients for predicting (a) admixture (heterospecific *Q*-value) across all species and (b) MAC_S *Q*-value in the unadmixed *Q. macrocarpa* dataset. Effect sizes (x-axis) are reported for centered and scaled predictor variables. Vertical lines on distributions denote posterior means. Shaded portions of distributions represent 95% credible intervals; endpoints of distributions represent 99% credible intervals.

gradients for most species, especially toward northern and western margins (Figure S3).

In the all-species admixture GLMM, distance from range edge exhibited a strong negative association with heterospecific *Q*-value ([−0.48, −0.13], 95% CI) (Figure 9a). No other environmental predictors were significantly correlated with admixture rates, nor were any of the species fixed effects, though the small number of hybrids found may limit our power to detect any such effects. Conversely, the *Q. macrocarpa* population structure GLMM revealed little predictive power for any individual environmental covariate (Figure 9b). As assessed by traceplots, posterior predictive checks and visual examination of parameter posterior distributions, our GLMMs exhibited convergence across all MCMC chains, with R-hat = 1.00 and cumulative Bulk/Tail ESS > 82,000 across all model parameters for all models (Tables S9 and S10).

3.5 | Genetic Differentiation Across Functional Classes

Rates of genetic differentiation across candidate gene classes for bud-break phenology and drought tolerance did not diverge significantly from the null distributions constructed from background (non-candidate) genes (Figure S4, Table S11).

4 | Discussion

Our study demonstrates that the five most prevalent species of the eastern North American white oak syngameon are genomic cluster species (Mallet 1995, 2020), largely predictable by their ecology and morphology. All, however, are connected by gene flow with the widespread bur oak or directly with one another. Although the species form clear, discontinuous genomic clusters, ca. 13% of the individuals we sampled are hybrids or introgressed. The eastern North American white oaks are a syngameon.

Our study also shows that genetic variation in *Q. macrocarpa* is due in part to introgression and in part to a substantial phylogeographic split separating the northern and southern populations. Our work thus supports the analysis of the eastern North American white oak syngameon made by Hardin (1975, 336) on the basis of morphology alone, but here supported by genomic data: while "a significant, although relatively minor, component of the variation is due to hybridization and localized introgression," most of the variation in the species is due to intraspecific variation that apparently has nothing to do with hybridization. Understanding the relative importance of these sources of genetic diversity will be key to predicting how bur oak, one of eastern North America's foundational oak species, will adapt to evolving environmental conditions.

4.1 | Genomic Coherence of Eastern North American White Oaks

Both the NewHybrids and ADMIXTURE analyses demonstrated that the eastern North American white oaks in our study are genomically distinct species that nonetheless hybridize and, in some cases, introgress. Of 362 samples, the NewHybrids analyses identified 16 (4.4%) as F1 hybrids and 9 (2.5%) as backcrosses or F2, for a total of 6.9% confirmed hybrids. In addition, a further 23 (6%) are of undetermined ancestry (meaning they could be either unadmixed or backcross), making 315 samples (87%) confirmed "pure" (unadmixed). The estimated frequency of hybrids in our study is much higher than estimates from historical, morphology-based studies, which generally estimated the frequency of hybrids in eastern North American oaks as less than 1% (Jensen 2002; Palmer 1948). This discrepancy may be due to both sampling bias—in at least one study, oak hybrids were found to be undersampled in herbaria relative to their prevalence in natural populations (Wu et al. 2023)—and the increased power of genomic data to detect admixture. Nonetheless, our finding recalls the claim made 50 years ago based on morphology alone that "the term 'syngameon' [in the eastern North American white oaks] does not imply unlimited and widespread gene exchange between the species. To the contrary, the gene 'flow' is but a 'trickle' in most cases" (Hardin 1975, 360).

Our admixture findings are in line with other DNA-based studies of North American oaks (e.g., Cavender-Bares et al. 2015; Craft et al. 2002; Kim et al. 2018; Ortego et al. 2017; Sullivan et al. 2016). Higher hybridization rates have been reported for oaks of Europe (e.g., Degen et al. 2023; Jensen et al. 2009; Reutimann et al. 2020; Valbuena-Carabana et al. 2005), East Asia (e.g., Liu et al. 2018; Pei et al. 2022; Qi et al. 2024; Shi et al. 2023) and Mexico (e.g., Albarrán-Lara et al. 2019; Castillo-Mendoza et al. 2019; González-Rodríguez et al. 2004; Morales-Saldaña et al. 2022; Peñaloza-Ramírez et al. 2010), but this may be influenced in part by human interplanting of species that would not grow together naturally in Europe (Stace et al. 2015) and substantially younger clade ages in Mexico (Hipp et al. 2018).

21 apparent backcrosses from *Q. bicolor* to *Q. macrocarpa* are scored as ambiguous in the NewHybrids analysis, but most make sense geographically: all *Q. macrocarpa* with ≥3.0% admixture from *Q. bicolor* are located within the range of *Q. bicolor* except for two individuals (QUE001882 from Oklahoma; QUE001635, cultivated from acorn collected in New Mexico); and four of the five *Q. macrocarpa* individuals with ≥2.0% admixture from *Q. muehlenbergii* are in the area of sympatry with that species (the exception being QUE002581, from South Dakota). However, ancestry estimation methods can fail to recover signals of ancient introgression, particularly when the magnitude of gene flow is low and occurred long enough in the past for drift, selection and recombination to fix variants and reshuffle species-specific combinations of allele frequencies (Kong and Kubatko 2021). The rates and spatial patterns of admixture observed in our dataset should be interpreted as a snapshot of one point in the history of the eastern North American white oak syngameon and possibly an under-representation of the importance of admixture.

4.2 | Spatial Patterns of Admixture

Proximity to species' range edge most strongly predicted admixture in our dataset, concordant with prior rangewide findings on introgression in *Q. alba* based on morphological data (Hardin 1975). Similar geographic patterns have been observed in poplar, pinyon pine and other temperate forest trees (Buck et al. 2023; Chhatre et al. 2018; Mimura et al. 2024) as well as clades of butterflies, toads, grasshoppers and other taxa (Barton and Hewitt 1985, 1989; Kawakami et al. 2009; Shepherd et al. 2022). These patterns align well with predictions and observations of environmental intermediacy favouring genetic admixture (Anderson 1948; Muller 1952). However, unlike these systems, which are largely characterised by parapatric hybridization, the eastern North American white oak syngameon contains many species occurring in widespread sympatry.

Range-marginal hybridization may be due to pollen swamping from larger heterospecific populations (Lepais et al. 2009), as populations reach their geographic limits, or reproductive failure due to increased stress in ecologically marginal populations, as demonstrated in *Quercus gambelii* and *Q. grisea*, where hybridization was driven in part by pollen abortion due to heat and drought stress (Williams et al. 2001). A disproportionate number of admixed individuals in our study were identified in the Hudson Valley–Lake Champlain corridor in New York and Vermont, where the eastern North American white oak species present consist of disjunct or range-marginal populations confined to lower elevation river valleys between climatically unsuitable mountain ranges. Under these circumstances, compressed distributions along topographically diverse river valleys might also facilitate higher rates of gene flow between species (Vallejo-Marín and Hiscock 2016). However, increased gene flow in the north also raises the question of whether adaptive introgression might contribute to the large geographic range and ecological variation in *Q. macrocarpa*. Adaptive introgression can facilitate species migration and persistence in novel environments (Burge et al. 2019; Cronk and Suarez-Gonzalez 2018; Dodd and Afzal-Rafii 2004; Nagamitsu et al. 2020; O'Donnell et al. 2021; Pfennig et al. 2016) and has been argued to play a key role in the northward extension of the widespread European species *Q. petraea* (Leroy et al. 2020). Climate suitability was not a significant partial predictor of admixture in our study (Figure 9a, "habitat suitability"), but edaphic factors at a finer scale than we were able to measure using the soil layers we had may be more important, as has been suggested, for example, in eastern North American *Q. ellipsoidalis* and *Q. rubra* (Khodwekar and Gailing 2017).

Even along range margins, hybrids were largely confined to a small subset of sites. Since early-generation hybridization appears rare, this heterogeneity could simply be explained by a low probability of a hybrid occurring at any given site. However, heterogeneity in local factors, such as habitat structure, disturbance (Anderson 1948; Bray 1960) or community composition, may exert strong effects on the generation and persistence of F1s and their backcrossed progeny. More detailed data on micro-scale habitat variation could help clarify whether habitat intermediacy promotes the survival of F1s and facilitates introgressive gene flow back into one or more parent populations.

4.3 | Admixture Between *Q. bicolor* and *Q. macrocarpa*

The two most closely related species in our study, bur oak (*Q. macrocarpa*) and swamp white oak (*Q. bicolor*) (Hipp et al. 2020; Larson et al. 2024), have long been thought to exchange genes relatively freely (Bray 1960; Burger 1975; Stebbins 1950, 63–64; Van Valen 1976). While our linear mixed models (GLMMs) suggest that the partial effect of species identity on admixture is non-significant (Figure 9), this may be partially due to low within-population sampling. By contrast, ADMIXTURE and NewHybrids analyses both show exceptionally high interspecific gene flow between *Q. bicolor* and *Q. macrocarpa*: well-supported F2s, 4 of the 7 demonstrated backcrosses, and 21 of the 23 samples of undetermined ancestry (unadmixed or backcross) were derived from these two species. This is particularly remarkable in light of the fact that in U.S. Forest Inventory and Analysis data, 19.4% of *Q. macrocarpa* plots also include *Q. alba*, while only 3.5% of *Q. macrocarpa* plots include *Q. bicolor*. This suggests that contemporary gene flow is much more frequent between *Q. macrocarpa* and *Q. bicolor* than expected by co-occurrence alone.

Given the close evolutionary relationship between the two species and high gene tree discordance observed among white oaks in general (Crowl et al. 2020; Larson et al. 2024; McVay, Hauser, et al. 2017; McVay, Hipp, et al. 2017), it is conceivable that our inferred admixture estimates could represent either gene flow or unsorted ancestral polymorphism. Incomplete lineage sorting between recently diverged species can masquerade as admixture when phylogeny is not accounted for (Lexer et al. 2006; Muir and Schlötterer 2005, 2006; Thomson et al. 2015). However, in our dataset, admixed individuals of *Q. macrocarpa* and *Q. bicolor* were distributed unevenly across the landscape, clustering at specific sites and extending no further than the zone of sympatry between parent species. This pattern supports interspecific gene flow over incomplete lineage sorting: under the latter, inferred admixture is expected to be widespread across the ranges of parental populations (Muir and Schlötterer 2005).

The causes of high gene flow between *Q. bicolor* and *Q. macrocarpa* as well as the strong asymmetry in introgression, almost exclusively from *Q. bicolor* to *Q. macrocarpa*, are unknown. The two recently diverged species may have lower reproductive barriers due to insufficient time for intrinsic incompatibilities to evolve. Moreover, *Q. macrocarpa* and *Q. bicolor* co-occur closely in mixed stands throughout the Midwest and portions of the Northeast, though relatively infrequently (3.5% of *Q. macrocarpa* plots). Close co-occurrence could provide ample opportunity for heterospecific pollen flow and potentially relax postzygotic selection against hybrids with intermediate ecological niches. However, most of the eastern North American white oak species in our study commonly grow in codominant mixed stands, particularly *Q. macrocarpa* and *Q. alba* (Cavender-Bares et al. 2018; Cho and Boerner 1991; Hardin 1975); and as noted above, more than 5 times as many *Q. macrocarpa* plots in the U.S. FIA data include *Q. alba* as include *Q. bicolor*. Disentangling the mechanisms maintaining genomic coherence of *Q. macrocarpa* and *Q. bicolor* would require controlled crossing experiments to tease apart the relative contributions of pollen competition, gametic incompatibility, hybrid viability and ecological selection on hybrids, among other factors.

4.4 | Homogeneity of Differentiation Across the Genome

Genetic differentiation did not vary significantly across candidate functional gene classes for any of the species pairs in our study (Figure S4). We may have lacked adequate sampling of *Quercus* individuals and target loci—our post-filtering dataset includes only 445 of the ca. 31,000–42,000 gene models inferred in a variety of annotated oak genomes to date (e.g., Ai et al. 2022; Fu et al. 2022; Kapoor et al. 2023; Larson et al. 2024; Sork et al. 2022; Zhou, Liu, et al. 2022)—to detect differential rates of divergence or gene flow across classes: adaptive differentiation or introgression may occur at a small number of loci, yielding insignificant results when averaging over an entire functional gene class consisting mostly of loci not under positive selection. Additionally, the drought-tolerance candidate genes in our marker set were selected based on orthology and literature search for California oaks (Oney-Birol et al. 2018); the bud-break phenology candidate genes were selected based on a transcriptomic study in European white oaks (Lesur et al. 2015; Ueno et al. 2013). It is unknown whether these loci are all under selection, particularly in eastern North America.

Alternatively, because we combined all samples of a given species for pairwise analyses, spatially heterogeneous patterns of gene flow or differentiation could obscure regionally significant divergence, gene flow or local adaptation. For example, adaptive introgression of drought tolerance alleles from *Q. stellata* into southwestern populations of *Q. macrocarpa* could be counterbalanced by introgression of flooding-tolerance alleles from *Q. bicolor* at northern sites. Further studies using larger marker sets and denser population sampling could provide the power to geographically partition jackknife analyses or conduct genotype-environment association tests to examine finer-scale, locus- or environment-specific patterns of (potentially) adaptive genetic differentiation across the eastern North American white oak syngameon. Finally, adaptive divergence and/or gene flow may be focused around non-coding regions, as has been demonstrated in east Asian *Quercus* (Fu et al. 2022), potentially producing null results when testing for adaptive introgression using target-capture methods for protein-coding genes.

4.5 | Phylogeography and Species Coherence of *Q. macrocarpa*

Genomic coherence of *Q. macrocarpa* is strongly supported by NMDS and ADMIXTURE analyses. Divergence between the northern and southern populations of the species is supported by ADMIXTURE and FEEMS, the latter showing a reduction in gene flow near the middle of the species' range, thus suggesting that the phylogeographic structure within *Q. macrocarpa* is not simply an outcome of isolation-by-distance. Moreover, our inclusion of a phylogenetically diverse set of widespread eastern North American white oak species (Hipp et al. 2018) suggests that population structure within *Q. macrocarpa*, beyond evidence of admixture and hybridization with the included species, represents intraspecific variation rather than gene flow from an unsampled eastern North American white oak species. Are the bur oak north and south genomic clusters distinct enough to be good species? This seems unlikely, as genetic differentiation

between inferred "pure" *Q. macrocarpa* north and south clusters (F_{ST} = 0.103) is approximately one-third as great as that between the most closely related species in our dataset, *Q. bicolor* and *Q. macrocarpa* (F_{ST} = 0.292–0.304). In combination, these analyses support our inference that, despite population structure, *Q. macrocarpa* is a single genomic cluster species, distinct from sympatric congeners.

The phylogeographic signal within *Q. macrocarpa* is not significantly predicted by climate or soil, leaving open the question of whether selection along an environmental gradient from northern upland savannas to southern bottomlands might explain genetic divergence within the species. Genetic clustering within *Q. macrocarpa* also does not correspond to known natural barriers to reproductive connectivity in eastern North America, nor to any widely documented phylogeographic patterns in the region (Soltis et al. 2006). It also does not align with a previous phylogenetic study of chloroplast haplotypes across the eastern North American white oak syngameon (Pham et al. 2017), which in fact showed weak geographic clustering, in contrast with the strong geographic structure in European oak chloroplast haplotypes (Petit et al. 2002). These findings are consistent with the previous inferences that North American oaks occupied more and smaller northern refugia and underwent less dramatic late-Pleistocene population bottlenecks than European oaks (Lumibao et al. 2017; McLachlan et al. 2005).

The genetic structure within bur oak does not correspond to any known morphological discontinuities. While latitudinal variation exists for bur oak acorn size and leaf lobing, these characters vary continuously (Desmond et al. 2021; Koenig et al. 2009). Additionally, none of the varieties previously diagnosed within *Q. macrocarpa*, which consist mostly of regional morphs throughout the Great Plains (Nixon and Muller 1997), align with the pattern of population structure we obtained. Rangewide common garden approaches hold the promise of evaluating what, if any, implications rangewide population structure holds for functional phenotypic diversity in *Q. macrocarpa*.

4.6 | Conclusions

Our study is the first genomic investigation of hybridization, introgression and population genetic structure in bur oak and sympatric species of the eastern North American white oak syngameon, a model system for examining hybridization and species coherence in temperate trees. Our study confirms that the traditional species form largely discontinuous genomic clusters, despite the presence of interspecific hybrids. Several interesting trends are suggested by our results. Levels of hybridization vary across the bur oak range, with hybrids establishing at a higher level in range-marginal populations. Other trends suggested in this dataset require further testing, since the small percentage of hybrids prevented strong statistical confirmation. Patterns of hybridization vary among species pairs. For instance, *Q. macrocarpa* showed little backcrossing with *Q. stellata* and *Q. alba*, although F1 hybrids with these species were found. On the other hand, backcrosses and F2s are much commoner for *Q. bicolor* × *Q. macrocarpa*. Backcrossing in this combination is asymmetrical, however, with *Q. bicolor* genes becoming incorporated into *Q. macrocarpa* but little evidence for the reverse. The demonstration of strong phylogeographic structure within the species suggests the possibility of either secondary contact from multiple refugia or strong divergent selection. The high amount of introgression into *Q. macrocarpa* raises the question of whether the large ecological, geographic, and morphological range of the species may be due in part to adaptive gene flow. Combined with the species distribution models presented, these data are a first step toward understanding how the interplay between genetic diversity within and among populations, genetic diversity among species and interspecific gene flow has shaped one of eastern North America's most widespread oak species.

Author Contributions

A.L.H., A.T.W. and P.S.M. designed the research. G.R., M.G. and K.P. led the field work, development of molecular tools, and data analysis. K.N.A., J.C.-B., A.A.C., S.G., P.G., M.H., S.L., R.A.M., I.S.P., N.R.S. and A.L.T. all participated materially in collecting, analysing or interpreting data. G.R. wrote the first draft of the ms with A.L.H. M.G., A.L.H., P.S.M., R.A.M. and A.T.W. collaborated on early drafts of the manuscript, and all authors participated in revising the manuscript.

Acknowledgements

This work was supported by USDA Cooperative Agreement 58-8020-5-005 between A.T.W. and A.L.H.; NSF DEB Awards 2129281 to A.L.H., 2129237 to P.S.M. and 2129312 to J.C.-B.; The Morton Arboretum Center for Tree Science, which provided lab, informatic and staff support as well as research fellowship support to K.P. to develop the Hyb-Seq probes; and the University of Chicago Department of Ecology and Evolution, which supported GR's work on this project. Sara Desmond assisted in most of the field collection. Guy and Edie Sternberg provided access to and logistical support at Starhill Forest Arboretum and have established and maintained the collection there over the course of decades; their work contributes immeasurably to our understanding of oak diversity. Permits and site access were provided by Bay County Parks and Facilities, Forest Preserve District of DuPage County, Forest Preserves of Cook County, Illinois Nature Preserves Commission, Indiana Department of Natural Resources, Ingham County Parks, Konza Prairie Biological Station, Lake County Forest Preserves, Manitoba Provincial Parks, Michigan Department of Natural Resources, Minnesota State Parks, Missouri State Parks, Missouri Department of Conservation, Ohio Department of Natural Resources, Pleasant Valley Conservancy, Prairie Moon Nursery, Shaw Nature Reserve, The Nature Conservancy, Tyson Research Center, University of Wisconsin Madison Arboretum, University of Wisconsin Milwaukee at Waukesha Field Station, US Forest Service, and Wisconsin Department of Natural Resources. Bryant Scharenbroch and Andrew Paolucci provided insights on navigating the SSURGO soil database and supplied R scripts for the acquisition and processing of data from the Web Soil Survey API. Matt Johnson and Elliot Gardner consulted in bait design and processing of Hyb-Seq data. The Herbarium and Systematics Research groups at The Morton Arboretum and Matthias Steinrücken and Trevor Price of the University of Chicago provided valuable feedback on all aspects of this work. Any use of trade, firm or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Disclosure

Benefit sharing statement: The project built a strong collaboration among numerous participants, including nine undergraduate, graduate or postdoctoral researchers who served as co-authors (including all three co-lead authors). It provided useful information on species limits

and hybridization to managers of participating natural areas where we collected tissue (see Acknowledgments) and will inform conservation decisions in at least one state. Additional benefits accrue from the sharing of our data and results on public databases as described above.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

Raw sequence data are deposited in NCBI (SRA Bioproject #1223965; http://www.ncbi.nlm.nih.gov/bioproject/1223965). All sample metadata, soils data, climatic data, and other associated data, as well as code for executing analyses, is archived in GitHub (https://github.com/gribicoff/qmac_hybseq), with releases versioned and archived in Zenodo (https://doi.org/10.5281/zenodo.15216009). The post-review version of the data and code is v0.9–3.

References

Ai, W., Y. Liu, M. Mei, et al. 2022. "A Chromosome-Scale Genome Assembly of the Mongolian Oak (*Quercus mongolica*)." *Molecular Ecology Resources* 22, no. 6: 2396–2410. https://doi.org/10.1111/1755-0998.13616.

Aiello-Lammens, M. E., R. A. Boria, A. Radosavljevic, B. Vilela, and R. P. Anderson. 2015. "spThin: An R Package for Spatial Thinning of Species Occurrence Records for Use in Ecological Niche Models." *Ecography* 38, no. 5: 541–545. https://doi.org/10.1111/ecog.01132.

Albarrán-Lara, A. L., R. J. Petit, A. Kremer, et al. 2019. "Low Genetic Differentiation Between Two Morphologically and Ecologically Distinct Giant-Leaved Mexican Oaks." *Plant Systematics and Evolution* 305, no. 1: 89–101. https://doi.org/10.1007/s00606-018-1554-8.

Alexander, D. H., J. Novembre, and K. Lange. 2009. "Fast Model-Based Estimation of Ancestry in Unrelated Individuals." *Genome Research* 19: 1655–1664. https://doi.org/10.1101/gr.094052.109.

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

Anderson, E. 1949. *Introgressive Hybridization*. John Wiley & Sons, Inc.

Anderson, E. 1953. "Introgressive Hybridization." *Biological Reviews* 28, no. 3: 280–307. https://doi.org/10.1111/j.1469-185X.1953.tb01379.x.

Anderson, E. C., and E. A. Thompson. 2002. "A Model-Based Method for Identifying Species Hybrids Using Multilocus Genetic Data." *Genetics* 160, no. 3: 1217–1229.

Ashley, M. V. 2021. "Answers Blowing in the Wind: A Quarter Century of Genetic Studies of Pollination in Oaks." *Forests* 12, no. 5: 575. https://doi.org/10.3390/f12050575.

Barton, N. 2001. "The Role of Hybridization in Evolution." *Molecular Ecology* 10: 551–568.

Barton, N. H., and G. M. Hewitt. 1985. "Analysis of Hybrid Zones." *Annual Review of Ecology and Systematics* 16: 113–148.

Barton, N. H., and G. M. Hewitt. 1989. "Adaptation, Speciation and Hybrid Zones." *Nature* 341, no. 6242: 497–503. https://doi.org/10.1038/341497a0.

Bivand, R. 2022. "R Packages for Analyzing Spatial Data: A Comparative Case Study With Areal Data." *Geographical Analysis* 54, no. 3: 488–518. https://doi.org/10.1111/gean.12319.

Bolger, A. M., M. Lohse, and B. Usadel. 2014. "Trimmomatic: A Flexible Trimmer for Illumina Sequence Data." *Bioinformatics* 30, no. 15: 2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

Bray, J. R. 1960. "A Note on Hybridization Between *Quercus macrocarpa* Michx. and *Quercus bicolor* Willd. In Wisconsin." *Canadian Journal of Botany* 38: 701–704.

Buck, R., and L. Flores-Rentería. 2022. "The Syngameon Enigma." *Plants* 11, no. 7: 895. https://doi.org/10.3390/plants11070895.

Buck, R., D. Ortega-Del Vecchyo, C. Gehring, et al. 2023. "Sequential Hybridization May Have Facilitated Ecological Transitions in the Southwestern Pinyon Pine Syngameon." *New Phytologist* 247, no. 6: 2435–2449. https://doi.org/10.1111/nph.18543.

Burge, D. O., V. T. Parker, M. Mulligan, and V. L. Sork. 2019. "Influence of a Climatic Gradient on Genetic Exchange Between Two Oak Species." *American Journal of Botany* 106, no. 6: 864–878. https://doi.org/10.1002/ajb2.1315.

Burger, W. C. 1975. "The Species Concept in *Quercus*." *Taxon* 24: 45–50.

Bürkner, P.-C. 2017. "brms: An R Package for Bayesian Multilevel Models Using Stan." *Journal of Statistical Software* 80: 1–28. https://doi.org/10.18637/jss.v080.i01.

Cannon, C. H., and M. Lerdau. 2022. "Asking Half the Question in Explaining Tropical Diversity." *Trends in Ecology & Evolution* 37, no. 5: 392–393. https://doi.org/10.1016/j.tree.2022.01.006.

Carpenter, B., A. Gelman, M. D. Hoffman, et al. 2017. "Stan: A Probabilistic Programming Language." *Journal of Statistical Software* 76: 1–32. https://doi.org/10.18637/jss.v076.i01.

Castillo-Mendoza, E., D. Salinas-Sánchez, L. Valencia-Cuevas, A. Zamilpa, and E. Tovar-Sánchez. 2019. "Natural Hybridization Among *Quercus glabrescens*, *Q. rugosa* and *Q. obtusata* (Fagaceae): Microsatellites and Secondary Metabolites Markers." *Plant Biology* 21, no. 1: 110–121. https://doi.org/10.1111/plb.12899.

Cavender-Bares, J., A. Gonzalez-Rodriguez, D. A. R. Eaton, A. L. Hipp, A. Beulke, and P. S. Manos. 2015. "Phylogeny and Biogeography of the American Live Oaks (*Quercus* Subsection *Virentes*): A Genomic and Population Genetics Approach." *Molecular Ecology* 24: 3668–3687.

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, no. 3: 565–586. https://doi.org/10.1002/ajb2.1049.

Chakraborty, S., and B. Rannala. 2023. "An Efficient Exact Algorithm for Identifying Hybrids Using Population Genomic Sequences." *Genetics* 223, no. 4: iyad011. https://doi.org/10.1093/genetics/iyad011.

Chamala, S., N. García, G. T. Godden, et al. 2015. "MarkerMiner 1.0: A New Application for Phylogenetic Marker Development Using Angiosperm Transcriptomes." *Applications in Plant Sciences* 3, no. 4: apps.1400115. https://doi.org/10.3732/apps.1400115.

Chamberlain, S., V. Barve, D. Mcglinn, et al. 2024. "rgbif: Interface to the Global Biodiversity Information Facility API. v 3.8.1 [Computer Software]." https://CRAN.R-project.org/package=rgbif.

Chamberlain, S. A., and C. Boettiger. 2017. "R Python, and Ruby Clients for GBIF Species Occurrence Data." *PeerJ Preprints* 5: e3304v1. https://doi.org/10.7287/peerj.preprints.3304v1.

Chhatre, V. E., L. M. Evans, S. P. DiFazio, and S. R. Keller. 2018. "Adaptive Introgression and Maintenance of a Trispecies Hybrid Complex in Range-Edge Populations of Populus." *Molecular Ecology* 27, no. 23: 4820–4838. https://doi.org/10.1111/mec.14820.

Cho, D. S., and R. E. J. Boerner. 1991. "Canopy Disturbance Patterns and Regeneration of *Quercus* Species in Two Ohio Old-Growth Forests." *Vegetatio* 93, no. 1: 9–18. https://doi.org/10.1007/BF00044920.

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

Craft, K. J., and M. V. Ashley. 2007. "Landscape Genetic Structure of Bur Oak (*Quercus macrocarpa*) Savannas in Illinois." *Forest Ecology and Management* 239, no. 1: 13–20. https://doi.org/10.1016/j.foreco.2006.11.005.

Craft, K. J., M. V. Ashley, and W. D. Koenig. 2002. "Limited Hybridization Between *Quercus lobata* and *Quercus douglasii* (Fagaceae) in a Mixed

Stand in Central Coastal California." *American Journal of Botany* 89, no. 11: 1792–1798.

Cronk, Q. C., and A. Suarez-Gonzalez. 2018. "The Role of Interspecific Hybridization in Adaptive Potential at Range Margins." *Molecular Ecology* 27, no. 23: 4653–4656. https://doi.org/10.1111/mec.14927.

Crowl, A. A., P. S. Manos, J. D. McVay, A. R. Lemmon, E. M. Lemmon, and A. L. Hipp. 2020. "Uncovering the Genomic Signature of Ancient Introgression Between White Oak Lineages (*Quercus*)." *New Phytologist* 226, no. 4: 1158–1170. https://doi.org/10.1111/nph.15842.

Danecek, P., A. Auton, G. Abecasis, et al. 2011. "The Variant Call Format and VCFtools." *Bioinformatics* 27, no. 15: 2156–2158. https://doi.org/10.1093/bioinformatics/btr330.

Danecek, P., J. K. Bonfield, J. Liddle, et al. 2021. "Twelve Years of SAMtools and BCFtools." *GigaScience* 10, no. 2: giab008. https://doi.org/10.1093/gigascience/giab008.

De Carvalho, D., P. K. Ingvarsson, J. Joseph, et al. 2010. "Admixture Facilitates Adaptation From Standing Variation in the European Aspen (*Populus tremula* L.), a Widespread Forest Tree." *Molecular Ecology* 19, no. 8: 1638–1650. https://doi.org/10.1111/j.1365-294X.2010.04595.x.

Degen, B., C. Blanc-Jolivet, M. Mader, V. Yanbaeva, and Y. Yanbaev. 2023. "Introgression as an Important Driver of Geographic Genetic Differentiation Within European White Oaks." *Forests* 14, no. 12: 2279. https://doi.org/10.3390/f14122279.

Desmond, S. C., M. Garner, S. Flannery, A. T. Whittemore, and A. L. Hipp. 2021. "Leaf Shape and Size Variation in Bur Oaks: An Empirical Study and Simulation of Sampling Strategies." *American Journal of Botany* 108, no. 8: 1540–1554. https://doi.org/10.1002/ajb2.1705.

Dodd, R. S., and Z. Afzal-Rafii. 2004. "Selection and Dispersal in a Multispecies Oak Hybrid Zone." *Evolution* 58, no. 2: 261–269.

Douma, J. C., and J. T. Weedon. 2019. "Analysing Continuous Proportions in Ecology and Evolution: A Practical Introduction to Beta and Dirichlet Regression." *Methods in Ecology and Evolution* 10, no. 9: 1412–1430. https://doi.org/10.1111/2041-210X.13234.

Dow, B. D., and M. V. Ashley. 1998. "High Levels of Gene Flow in Bur Oak Revealed by Paternity Analysis Using Microsatellites." *Journal of Heredity* 89: 62–70.

Ellstrand, N. C., R. Whitkus, and L. H. Rieseberg. 1996. "Distribution of Spontaneous Plant Hybrids." *Proceedings of the National Academy of Sciences of the United States of America* 93, no. 10: 5090–5093.

Engelmann, G. 1878. "About the Oaks of the United States." *Transactions of the Academy of Science of St. Louis* 3: 372–400, 539–543.

Fick, S. E., and R. J. Hijmans. 2017. "WorldClim 2: New 1-Km Spatial Resolution Climate Surfaces for Global Land Areas." *International Journal of Climatology* 37, no. 12: 4302–4315. https://doi.org/10.1002/joc.5086.

Fu, R., Y. Zhu, Y. Liu, et al. 2022. "Genome-Wide Analyses of Introgression Between Two Sympatric Asian Oak Species." *Nature Ecology & Evolution* 6: 924–935. https://doi.org/10.1038/s41559-022-01754-7.

Gardner, E. M., S. Bruun-Lund, M. Niissalo, et al. 2023. "Echoes of Ancient Introgression Punctuate Stable Genomic Lineages in the Evolution of Figs." *Proceedings of the National Academy of Sciences* 120, no. 28: e2222035120. https://doi.org/10.1073/pnas.2222035120.

Gardner, E. M., M. Garner, R. Cowan, et al. 2021. "Repeated Parallel Losses of Inflexed Stamens in Moraceae: Phylogenomics and Generic Revision of the Tribe Moreae and the Reinstatement of the Tribe Olmedieae (Moraceae)." *Taxon* 70, no. 5: 946–988. https://doi.org/10.1002/tax.12526.

Garrison, E., Z. N. Kronenberg, E. T. Dawson, B. S. Pedersen, and P. Prins. 2022. "A Spectrum of Free Software Tools for Processing the VCF Variant Call Format: Vcflib, Bio-VCF, Cyvcf2, Hts-nim and Slivar." *PLoS Computational Biology* 18, no. 5: e1009123. https://doi.org/10.1371/journal.pcbi.1009123.

Garrison, E., and G. Marth. 2012. "Haplotype-Based Variant Detection From Short-Read Sequencing." *arXiv*, 1207.3907 [q-Bio]. http://arxiv.org/abs/1207.3907.

Glenn, T. C., R. Nilsen, T. J. Kieran, et al. 2016. "Adapterama I: Universal Stubs and Primers for Thousands of Dual-Indexed Illumina Libraries (iTru & iNext)." *bioRxiv*, 049114. https://doi.org/10.1101/049114.

Gompert, Z., L. K. Lucas, C. A. Buerkle, M. L. Forister, J. A. Fordyce, and C. C. Nice. 2014. "Admixture and the Organization of Genetic Diversity in a Butterfly Species Complex Revealed Through Common and Rare Genetic Variants." *Molecular Ecology* 23, no. 18: 4555–4573. https://doi.org/10.1111/mec.12811.

González-Rodríguez, A., D. M. Arias, S. Valencia-Avalos, and K. Oyama. 2004. "Morphological and RAPD Analysis of Hybridization Between *Quercus affinis* and *Q. laurina* (Fagaceae), two Mexican Red Oaks." *American Journal of Botany* 91, no. 3: 401–409.

Goudet, J. 2005. "Hierfstat, a Package for r to Compute and Test Hierarchical F-Statistics." *Molecular Ecology Notes* 5, no. 1: 184–186. https://doi.org/10.1111/j.1471-8286.2004.00828.x.

Grant, V. E. 1981. *Plant Speciation*. 2nd ed. Columbia University Press.

Hale, H., E. M. Gardner, J. Viruel, L. Pokorny, and M. G. Johnson. 2020. "Strategies for Reducing Per-Sample Costs in Target Capture Sequencing for Phylogenomics and Population Genomics in Plants." *Applications in Plant Sciences* 8, no. 4: e11337. https://doi.org/10.1002/aps3.11337.

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

Hijmans, R. J. 2019–2025. "Spatial Data Science With R and "Terra"—R Spatial." https://rspatial.org/.

Hipp, A. L., P. S. Manos, A. González-Rodríguez, et al. 2018. "Sympatric Parallel Diversification of Major Oak Clades in the Americas and the Origins of Mexican Species Diversity." *New Phytologist* 217, no. 1: 439–452. https://doi.org/10.1111/nph.14773.

Hipp, A. L., P. S. Manos, M. Hahn, et al. 2020. "Genomic Landscape of the Global Oak Phylogeny." *New Phytologist* 226, no. 4: 1198–1212. https://doi.org/10.1111/nph.16162.

Hipp, A. L., and J. A. Weber. 2008. "Taxonomy of Hill's Oak (*Quercus ellipsoidalis*: Fagaceae): Evidence From AFLP Data." *Systematic Botany* 33: 148–158. https://doi.org/10.1600/036364408783887320.

Hipp, A. L., A. T. Whittemore, M. Garner, et al. 2019. "Genomic Identity of White Oak Species in an Eastern North American Syngameon." *Annals of the Missouri Botanical Garden* 104, no. 3: 455–477. https://doi.org/10.3417/2019434.

Jensen, J., A. Larsen, L. R. Nielsen, and J. Cottrell. 2009. "Hybridization Between Quercus Robur and *Q. petraea* in a Mixed Oak Stand in Denmark." *Annals of Forest Science* 66, no. 7: 706. https://doi.org/10.1051/forest/2009058.

Jensen, R. J. 2002. "The Oaks of Land Between the Lakes: Biodiversity and Hybridization." In *Land Between the Lakes, Kentucky and Tennessee: Four Decades of Tennessee Valley Authority Stewardship. Miscellaneous Publication Number 16 of the Center for Field Biology*, 323–334. Austin Peay State University.

Johnson, P. S. 1990. "*Quercus macrocarpa* Michx." In *Silvics of North America, Volume 2: Hardwoods: Vol. Agriculture Handbook 654*, edited by R. M. Burns and B. H. Honkala. United States Department of Agriculture Forest Service.

Jombart, T. 2008. "Adegenet: A R Package for the Multivariate Analysis of Genetic Markers." *Bioinformatics* 24, no. 11: 1403–1405. https://doi.org/10.1093/bioinformatics/btn129.

Jombart, T., and I. Ahmed. 2011. "Adegenet 1.3-1: New Tools for the Analysis of Genome-Wide SNP Data." *Bioinformatics* 27, no. 21: 3070–3071. https://doi.org/10.1093/bioinformatics/btr521.

Kapoor, B., J. Jenkins, J. Schmutz, et al. 2023. "A Haplotype-Resolved Chromosome-Scale Genome for *Quercus rubra* L. Provides Insights Into the Genetics of Adaptive Traits for Red Oak Species." *G3 (Bethesda, Md.)* 13: jkad209. https://doi.org/10.1093/g3journal/jkad209.

Kass, J. M., R. Muscarella, P. J. Galante, et al. 2021. "ENMeval 2.0: Redesigned for Customizable and Reproducible Modeling of Species' Niches and Distributions." *Methods in Ecology and Evolution* 12, no. 9: 1602–1608. https://doi.org/10.1111/2041-210X.13628.

Kawakami, T., R. K. Butlin, M. D. Adams, D. J. Paull, and S. J. B. Cooper. 2009. "Genetic Analysis of a Chromosomal Hybrid Zone in the Australian Morabine Grasshoppers (*Vandiemenella*, *Viatica* Species Group)." *Evolution* 63, no. 1: 139–152.

Khodwekar, S., and O. Gailing. 2017. "Evidence for Environment-Dependent Introgression of Adaptive Genes Between Two Red Oak Species With Different Drought Adaptations." *American Journal of Botany* 104, no. 7: 1088–1098. https://doi.org/10.3732/ajb.1700060.

Kim, B. Y., X. Wei, S. Fitz-Gibbon, et al. 2018. "RADseq Data Reveal Ancient, but Not Pervasive, Introgression Between Californian Tree and Scrub Oak Species (*Quercus* sect. *Quercus*: Fagaceae)." *Molecular Ecology* 27, no. 22: 4556–4571. https://doi.org/10.1111/mec.14869.

Koenig, W. D., J. M. H. Knops, J. L. Dickinson, and B. Zuckerberg. 2009. "Latitudinal Decrease in Acorn Size in Bur Oak (*Quercus macrocarpa*) is due to Environmental Constraints, Not Avian Dispersal." *Botany* 87, no. 4: 349–356.

Kong, S., and L. S. Kubatko. 2021. "Comparative Performance of Popular Methods for Hybrid Detection Using Genomic Data." *Systematic Biology* 70, no. 5: 891–907. https://doi.org/10.1093/sysbio/syaa092.

Kremer, A., A. Abbott, J. Carlson, et al. 2012. "Genomics of Fagaceae." *Tree Genetics & Genomes* 8: 583–610. https://doi.org/10.1007/s11295-012-0498-3.

Kremer, A., and A. L. Hipp. 2020. "Oaks: An Evolutionary Success Story." *New Phytologist* 226, no. 4: 987–1011. https://doi.org/10.1111/nph.16274.

Larson, D. A., M. E. Staton, B. Kapoor, et al. 2024. "A Haplotype-Resolved Reference Genome of *Quercus alba* Sheds Light on the Evolutionary History of Oaks [Preprint]." *bioRxiv*. https://doi.org/10.1101/2024.02.13.579671.

Larson, D. A., O. M. Vargas, A. Vicentini, and C. W. Dick. 2021. "Admixture May Be Extensive Among Hyperdominant Amazon Rainforest Tree Species." *New Phytologist* 232, no. 6: 2520–2534. https://doi.org/10.1111/nph.17675.

Lazic, D., A. L. Hipp, J. E. Carlson, and O. Gailing. 2021. "Use of Genomic Resources to Assess Adaptive Divergence and Introgression in Oaks." *Forests* 12, no. 6: 690. https://doi.org/10.3390/f12060690.

Lepais, O., R. Petit, E. Guichoux, et al. 2009. "Species Relative Abundance and Direction of Introgression in Oaks." *Molecular Ecology* 18: 2228–2242.

Leroy, T., J.-M. Louvet, C. Lalanne, et al. 2020. "Adaptive Introgression as a Driver of Local Adaptation to Climate in European White Oaks." *New Phytologist* 226, no. 4: 1171–1182. https://doi.org/10.1111/nph.16095.

Lesur, I., G. Le Provost, P. Bento, et al. 2015. "The Oak Gene Expression Atlas: Insights Into Fagaceae Genome Evolution and the Discovery of Genes Regulated During Bud Dormancy Release." *BMC Genomics* 16, no. 1: 112. https://doi.org/10.1186/s12864-015-1331-9.

Lexer, C., A. Kremer, and R. J. Petit. 2006. "Shared Alleles in Sympatric Oaks: Recurrent Gene Flow Is a More Parsimonious Explanation Than Ancestral Polymorphism." *Molecular Ecology* 15: 2007–2012.

Li, H., and R. Durbin. 2009. "Fast and Accurate Short Read Alignment With Burrows–Wheeler Transform." *Bioinformatics* 25, no. 14: 1754–1760. https://doi.org/10.1093/bioinformatics/btp324.

Linan, A. G., P. P. Lowry II, A. J. Miller, G. E. Schatz, J. C. Sevathian, and C. E. Edwards. 2022. "Interspecific Hybridization and Island Colonization History, Not Rarity, Most Strongly Affect the Genetic Diversity in Diospyros, a Clade of Mascarene-Endemic Trees." *Journal of Heredity* 113, no. 3: 336–352. https://doi.org/10.1093/jhered/esac010.

Little, E. L. 1971. *Atlas of United States Trees, Volume 1, Conifers and Important Hardwoods*. U.S. Department of Agriculture, Forest Service. https://www.biodiversitylibrary.org/item/225461.

Liu, Y., Y. Li, J. Song, et al. 2018. "Geometric Morphometric Analyses of Leaf Shapes in Two Sympatric Chinese Oaks: *Quercus dentata* Thunberg and *Quercus aliena* Blume (Fagaceae)." *Annals of Forest Science* 75, no. 4: 90. https://doi.org/10.1007/s13595-018-0770-2.

Lotsy, J. P. 1917. "La quintessence de la théorie du croisement." *Archives Néerlandaises Des Sciences Exactes et Naturelles. Série IIIB (Sciences Naturelles)* 3: 351–353.

Lumibao, C. Y., S. M. Hoban, and J. McLachlan. 2017. "Ice Ages Leave Genetic Diversity 'Hotspots' in Europe but Not in Eastern North America." *Ecology Letters* 20, no. 11: 1459–1468. https://doi.org/10.1111/ele.12853.

MacDougal, D. T. 1907. "Hybridization of Wild Plants." *Botanical Gazette* 43, no. 1: 45–58. https://doi.org/10.1086/329077.

Mallet, J. 1995. "A Species Definition for the Modern Synthesis." *Trends in Ecology & Evolution* 10, no. 7: 294–299. https://doi.org/10.1016/0169-5347(95)90031-4.

Mallet, J. 2005. "Hybridization as an Invasion of the Genome." *Trends in Ecology & Evolution* 20, no. 5: 229–237.

Mallet, J. 2020. "Alternative Views of Biological Species: Reproductively Isolated Units or Genotypic Clusters?" *National Science Review* 7, no. 8: 1401–1407. https://doi.org/10.1093/nsr/nwaa116.

Mallet, J., M. Beltrán, W. Neukirchen, and M. Linares. 2007. "Natural Hybridization in Heliconiine Butterflies: The Species Boundary as a Continuum." *BMC Evolutionary Biology* 7, no. 1: 28. https://doi.org/10.1186/1471-2148-7-28.

Mao, Y. 2020. "Genomic Insights Into Hybridization of Reef Corals." *Coral Reefs* 39, no. 1: 61–67. https://doi.org/10.1007/s00338-019-01869-7.

Marcus, J., W. Ha, R. F. Barber, and J. Novembre. 2021. "Fast and Flexible Estimation of Effective Migration Surfaces." *eLife* 10: e61927. https://doi.org/10.7554/eLife.61927.

McLachlan, J. S., J. S. Clark, and P. S. Manos. 2005. "Molecular Indicators of Tree Migration Capacity Under Rapid Climate Change." *Ecology* 86, no. 8: 2088–2098. https://doi.org/10.1890/04-1036.

McVay, J. D., D. Hauser, A. L. Hipp, and P. S. Manos. 2017. "Phylogenomics Reveals a Complex Evolutionary History of Lobed-Leaf White Oaks in Western North America." *Genome* 60, no. 9: 733–742. https://doi.org/10.1139/gen-2016-0206.

McVay, J. D., A. L. Hipp, and P. S. Manos. 2017. "A Genetic Legacy of Introgression Confounds Phylogeny and Biogeography in Oaks." *Proceedings of the Royal Society B* 284, no. 1854: 20170300. https://doi.org/10.1098/rspb.2017.0300.

Meireles, J. E., A. Beulke, D. S. Borkowski, J. Romero-Severson, and J. Cavender-Bares. 2017. "Balancing Selection Maintains Diversity in a Cold Tolerance Gene in Broadly Distributed Live Oaks." *Genome* 60, no. 9: 762–769. https://doi.org/10.1139/gen-2016-0208.

Merow, C., M. J. Smith, and J. A. Silander Jr. 2013. "A Practical Guide to MaxEnt for Modeling Species' Distributions: What It Does, and Why Inputs and Settings Matter." *Ecography* 36, no. 10: 1058–1069. https://doi.org/10.1111/j.1600-0587.2013.07872.x.

Mimura, M., Z. Tang, S. Shigenobu, K. Yamaguchi, and T. Yahara. 2024. "Genomic Introgression in the Hybrid Zones at the Margins of the Species' Range Between Ecologically Distinct Rubus Species." *Ecology and Evolution* 14, no. 11: e70476. https://doi.org/10.1002/ece3.70476.

Morales-Saldaña, S., A. L. Hipp, S. Valencia-Ávalos, et al. 2024. "Divergence and Reticulation in the Mexican White Oaks: Ecological and Phylogenomic Evidence on Species Limits and Phylogenetic Networks in the *Quercus laeta* Complex (Fagaceae)." *Annals of Botany* 133, no. 7: 1007–1024. https://doi.org/10.1093/aob/mcae030.

Morales-Saldaña, S., S. Valencia-Ávalos, K. Oyama, E. T. Sánchez, A. L. Hipp, and A. González-Rodríguez. 2022. "Even More Oak Species in Mexico? Genetic Structure and Morphological Differentiation Support the Presence of at Least Two Specific Entities Within *Quercus laeta*." *Journal of Systematics and Evolution* 60, no. 5: 1124–1139. https://doi.org/10.1111/jse.12818.

Moran, P. A. P. 1950. "Notes on Continuous Stochastic Phenomena." *Biometrika* 37: 17–23.

Muir, G., and C. Schlötterer. 2005. "Evidence for Shared Ancestral Polymorphism Rather Than Recurrent Gene Flow at Microsatellite Loci Differentiating Two Hybridizing Oaks (*Quercus* spp.)." *Molecular Ecology* 14, no. 2: 549–561.

Muir, G., and C. Schlötterer. 2006. "Moving Beyond Single-Locus Studies to Characterize Hybridization Between Oaks (*Quercus* spp.)." *Molecular Ecology* 15: 2301–2304.

Muller, C. H. 1952. "Ecological Control of Hybridization in *Quercus*: A Factor in the Mechanism of Evolution." *Evolution* 6, no. 2: 147–161.

Nagamitsu, T., K. Uchiyama, A. Izuno, H. Shimizu, and A. Nakanishi. 2020. "Environment-Dependent Introgression From *Quercus dentata* to a Coastal Ecotype of *Q. mongolica* var. *crispula* in Northern Japan." *New Phytologist* 226: 1018–1028. https://doi.org/10.1111/nph.16131.

Nixon, K. C., and C. H. Muller. 1997. "*Quercus* Linnaeus sect. *Quercus*." In *Flora of North America*, edited by Flora of North America Editorial Committee, vol. 3, 471–506. Oxford University Press.

Nocchi, G., N. Brown, T. L. R. Coker, et al. 2022. "Genomic Structure and Diversity of Oak Populations in British Parklands." *Plants, People, Planet* 4, no. 2: 167–181. https://doi.org/10.1002/ppp3.10229.

O'Donnell, S. T., S. T. Fitz-Gibbon, and V. L. Sork. 2021. "Ancient Introgression Between Distantly Related White Oaks (*Quercus* sect. *Quercus*) Shows Evidence of Climate-Associated Asymmetric Gene Exchange." *Journal of Heredity* 112, no. 7: 663–670. https://doi.org/10.1093/jhered/esab053.

Olave, M., and A. Meyer. 2020. "Implementing Large Genomic Single Nucleotide Polymorphism Data Sets in Phylogenetic Network Reconstructions: A Case Study of Particularly Rapid Radiations of Cichlid Fish." *Systematic Biology* 69, no. 5: 848–862. https://doi.org/10.1093/sysbio/syaa005.

O'Leary, S. J., J. B. Puritz, S. C. Willis, C. M. Hollenbeck, and D. S. Portnoy. 2018. "These Aren't the Loci You're Looking for: Principles of Effective SNP Filtering for Molecular Ecologists." *Molecular Ecology* 27, no. 16: 3193–3206. https://doi.org/10.1111/mec.14792.

Oney-Birol, S., S. Fitz-Gibbon, J.-M. Chen, P. F. Gugger, and V. L. Sork. 2018. "Assessment of Shared Alleles in Drought-Associated Candidate Genes Among Southern California White Oak Species (*Quercus* sect. *Quercus*)." *BMC Genetics* 19, no. 1: 88. https://doi.org/10.1186/s12863-018-0677-9.

Ortego, J., P. F. Gugger, and V. L. Sork. 2017. "Impacts of Human-Induced Environmental Disturbances on Hybridization Between Two Ecologically Differentiated Californian Oak Species." *New Phytologist* 213, no. 2: 942–955. https://doi.org/10.1111/nph.14182.

Palmer, E. J. 1948. "Hybrid Oaks of North America." *Journal of the Arnold Arboretum* 29, no. 1: 1–48.

Pebesma, E. 2018. "Simple Features for R: Standardized Support for Spatial Vector Data." *R Journal* 10, no. 1: 439–446.

Pei, Z., Q. Yang, X. Chen, et al. 2022. "Population Study Reveals Genetic Variation and Introgression of Four Deciduous Oaks at the Junction Between Taihang Mountain and Yanshan Mountain." *Forests* 13, no. 10: 1647. https://doi.org/10.3390/f13101647.

Peñaloza-Ramírez, J. M., A. González-Rodríguez, L. Mendoza-Cuenca, H. Caron, A. Kremer, and K. Oyama. 2010. "Interspecific Gene Flow in a Multispecies Oak Hybrid Zone in the Sierra Tarahumara of Mexico." *Annals of Botany* 105, no. 3: 389–399. https://doi.org/10.1093/aob/mcp301.

Percy, D. M., G. W. Argus, Q. C. Cronk, et al. 2014. "Understanding the Spectacular Failure of DNA Barcoding in Willows (Salix): Does This Result From a Trans-Specific Selective Sweep?" *Molecular Ecology* 23, no. 19: 4737–4756. https://doi.org/10.1111/mec.12837.

Petit, R. J., U. Csaikl, S. Bordacs, et al. 2002. "Chloroplast DNA Variation in European White Oaks. Phylogeography and Patterns of Diversity Based on Data From Over 2600 Populations." *Forest Ecology and Management* 156: 5–26.

Pfennig, K. S., A. L. Kelly, and A. A. Pierce. 2016. "Hybridization as a Facilitator of Species Range Expansion." *Proceedings of the Royal Society B: Biological Sciences* 283, no. 1839: 20161329. https://doi.org/10.1098/rspb.2016.1329.

Pham, K. K., A. L. Hipp, P. S. Manos, and R. C. Cronn. 2017. "A Time and a Place for Everything: Phylogenetic History and Geography as Joint Predictors of Oak Plastome Phylogeny." *Genome* 60: 720–732. https://doi.org/10.1139/gen-2016-0191.

Phillips, S. J., R. P. Anderson, M. Dudík, R. E. Schapire, and M. E. Blair. 2017. "Opening the Black Box: An Open-Source Release of Maxent." *Ecography* 40, no. 7: 887–893. https://doi.org/10.1111/ecog.03049.

Plomion, C., J.-M. Aury, J. Amselem, et al. 2016. "Decoding the Oak Genome: Public Release of Sequence Data, Assembly, Annotation and Publication Strategies." *Molecular Ecology Resources* 16, no. 1: 254–265. https://doi.org/10.1111/1755-0998.12425.

Plomion, C., J.-M. Aury, J. Amselem, et al. 2018. "Oak Genome Reveals Facets of Long Lifespan." *Nature Plants* 4, no. 7: 440–452. https://doi.org/10.1038/s41477-018-0172-3.

Plummer, M., N. Best, and K. Cowles. 2006. "CODA: Convergence Diagnosis and Output Analysis for MCMC." *R News* 6, no. 1: 7–11.

Prasad, A. M., and L. R. Iverson. 2003. *Little's Range and FIA Importance Value Database for 135 Eastern US Tree Species*. Northeastern Research Station, USDA Forest Service. http://www.fs.fed.us/ne/delaware/4153/global/littlefia/index.html.

Purcell, S., B. Neale, K. Todd-Brown, et al. 2007. "PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses." *American Journal of Human Genetics* 81, no. 3: 559–575. https://doi.org/10.1086/519795.

QGIS Development Team. 2023. *QGIS Geographic Information System*. QGIS Association. https://www.qgis.org.

Qi, M., F. K. Du, F. Guo, K. Yin, and J. Tang. 2024. "Species Identification Through Deep Learning and Geometrical Morphology in Oaks (*Quercus* spp.): Pros and Cons." *Ecology and Evolution* 14, no. 2: e11032. https://doi.org/10.1002/ece3.11032.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing, Version 4.3.0 (Version 4.3.0) [Computer Software]*. R Foundation for Statistical Computing. http://www.R-project.org/.

Reutimann, O., F. Gugerli, and C. Rellstab. 2020. "A Species-Discriminatory Single-Nucleotide Polymorphism Set Reveals Maintenance of Species Integrity in Hybridizing European White Oaks (*Quercus* spp.) Despite High Levels of Admixture." *Annals of Botany* 125, no. 4: 663–676. https://doi.org/10.1093/aob/mcaa001.

Rohland, N., and D. Reich. 2012. "Cost-Effective, High-Throughput DNA Sequencing Libraries for Multiplexed Target Capture." *Genome Research* 22, no. 5: 939–946. https://doi.org/10.1101/gr.128124.111.

Schliewen, U. K., and B. Klee. 2004. "Reticulate Sympatric Speciation in Cameroonian Crater Lake Cichlids." *Frontiers in Zoology* 1, no. 1: 5. https://doi.org/10.1186/1742-9994-1-5.

Schnabel, A., and J. L. Hamrick. 1990. "Comparative Analysis of Population Genetic Structure in *Quercus macrocarpa* and *Q. gambelii* (Fagaceae)." *Systematic Botany* 15, no. 2: 240–251. https://doi.org/10.2307/2419179.

Shepherd, L., C. Simon, S. Langton-Myers, and M. Morgan-Richards. 2022. "Insights Into Aotearoa New Zealand's Biogeographic History Provided by the Study of Natural Hybrid Zones." *Journal of the Royal Society of New Zealand* 54, no. 1: 55–74. https://doi.org/10.1080/03036758.2022.2061020.

Shi, Y., B.-F. Zhou, Y.-Y. Liang, and B. Wang. 2023. "Linked Selection and Recombination Rate Generate Both Shared and Lineage-Specific Genomic Islands of Divergence in Two Independent *Quercus* Species Pairs." *Journal of Systematics and Evolution* 62: 505–519. https://doi.org/10.1111/jse.13008.

Sinnott, R. W. 1984. "Virtues of the Haversine." *Sky and Telescope* 68, no. 2: 159.

Soil Survey Staff, Natural Resources Conservation Service, United States Department of Agriculture. n.d. "Soil Survey Geographic Database (SSURGO) | Natural Resources Conservation Service." https://www.nrcs.usda.gov/resources/data-and-reports/soil-survey-geographic-database-ssurgo.

Soltis, D. E., A. B. Morris, J. S. McLACHLAN, P. S. Manos, and P. S. Soltis. 2006. "Comparative Phylogeography of Unglaciated Eastern North America." *Molecular Ecology* 15, no. 14: 4261–4293. https://doi.org/10.1111/j.1365-294X.2006.03061.x.

Sork, V. L., S. J. Cokus, S. T. Fitz-Gibbon, et al. 2022. "High-Quality Genome and Methylomes Illustrate Features Underlying Evolutionary Success of Oaks." *Nature Communications* 13, no. 1: 2047. https://doi.org/10.1038/s41467-022-29584-y.

Sork, V. L., S. T. Fitz-Gibbon, D. Puiu, et al. 2016. "First Draft Assembly and Annotation of the Genome of a California Endemic Oak *Quercus lobata* Née (Fagaceae)." *G3: Genes, Genomes, Genetics* 6, no. 11: 3485–3495. https://doi.org/10.1534/g3.116.030411.

Sork, V. L., K. Squire, P. F. Gugger, S. E. Steele, E. D. Levy, and A. J. Eckert. 2016. "Landscape Genomic Analysis of Candidate Genes for Climate Adaptation in a California Endemic Oak, *Quercus lobata*." *American Journal of Botany* 103, no. 1: 33–46. https://doi.org/10.3732/ajb.1500162.

Stace, C. A., C. D. Preston, and D. A. Pearman. 2015. *Hybrid Flora of the British Isles*. Botanical Society of Britain and Ireland.

Stebbins, G. L. 1950. *Variation and Evolution in Plants*. Columbia University Press.

Stebbins, G. L., E. G. Matzke, and C. Epling. 1947. "Hybridization in a Population of *Quercus marilandica* and *Quercus ilicifolia*." *Evolution* 1: 79–88. https://doi.org/10.2307/2405406.

Sullivan, A. R., S. A. Owusu, J. A. Weber, A. L. Hipp, and O. Gailing. 2016. "Hybridization and Divergence in Multi-Species Oak (*Quercus*) Communities." *Botanical Journal of the Linnean Society* 181, no. 1: 99–114. https://doi.org/10.1111/boj.12393.

Sun, Y., R. J. Abbott, Z. Lu, et al. 2018. "Reticulate Evolution Within a Spruce (Picea) Species Complex Revealed by Population Genomic Analysis." *Evolution* 72, no. 12: 2669–2681. https://doi.org/10.1111/evo.13624.

Thiers, B. M. n.d. "Index Herbariorum." Updated continuously. https://sweetgum.nybg.org/science/ih/.

Thomson, A. M., C. W. Dick, A. L. Pascoini, and S. Dayanandan. 2015. "Despite Introgressive Hybridization, North American Birches (*Betula* spp.) Maintain Strong Differentiation at Nuclear Microsatellite Loci." *Tree Genetics & Genomes* 11, no. 5: 101. https://doi.org/10.1007/s11295-015-0922-6.

Thornthwaite, C. W. 1948. "An Approach Toward a Rational Classification of Climate." *Geographical Review* 38: 55–94.

Tsuda, Y., V. Semerikov, F. Sebastiani, G. G. Vendramin, and M. Lascoux. 2017. "Multispecies Genetic Structure and Hybridization in the Betula Genus Across Eurasia." *Molecular Ecology* 26, no. 2: 589–605. https://doi.org/10.1111/mec.13885.

Tucker, J. M. 1961. "Studies in the *Quercus undulata* Complex. I. A Preliminary Statement." *American Journal of Botany* 48, no. 3: 202–208. https://doi.org/10.1002/j.1537-2197.1961.tb11626.x.

Ueno, S., C. Klopp, J. C. Leplé, et al. 2013. "Transcriptional Profiling of Bud Dormancy Induction and Release in Oak by Next-Generation Sequencing." *BMC Genomics* 14, no. 1: 236. https://doi.org/10.1186/1471-2164-14-236.

Valbuena-Carabana, M., S. C. Gonzalez-Martinez, V. L. Sork, et al. 2005. "Gene Flow and Hybridization in a Mixed Oak Forest (*Quercus pyrenaica* Willd. and *Quercus petraea* (Matts.) Liebl.) in Central Spain." *Heredity* 95, no. 6: 457–465.

Vallejo-Marín, M., and S. J. Hiscock. 2016. "Hybridization and Hybrid Speciation Under Global Change." *New Phytologist* 211, no. 4: 1170–1187. https://doi.org/10.1111/nph.14004.

Van Valen, L. 1976. "Ecological Species, Multispecies, and Oaks." *Taxon* 25: 233–239.

Weir, B. S., and J. Goudet. 2017. "A Unified Characterization of Population Structure and Relatedness." *Genetics* 206, no. 4: 2085–2103. https://doi.org/10.1534/genetics.116.198424.

Weitemier, K., S. C. K. Straub, R. C. Cronn, et al. 2014. "Hyb-Seq: Combining Target Enrichment and Genome Skimming for Plant Phylogenomics." *Applications in Plant Sciences* 2, no. 9: 1400042. https://doi.org/10.3732/apps.1400042.

Whittemore, A. T., and B. A. Schaal. 1991. "Interspecific Gene Flow in Sympatric Oaks." *Proceedings of the National Academy of Sciences of the United States of America* 88: 2540–2544.

Wickham, H. 2016. *Ggplot2*. Springer International Publishing. https://doi.org/10.1007/978-3-319-24277-4.

Wiegand, K. M. 1935. "A Taxonomist's Experience With Hybrids in the Wild." *Science* 81: 161–166.

Wiens, B. J., and J. P. Colella. 2024. "triangulaR: An R Package for Identifying AIMs and Building Triangle Plots Using SNP Data From Hybrid Zones." *bioRxiv*, 2024.03.28.587167. https://doi.org/10.1101/2024.03.28.587167.

Williams, J. H., W. J. Boecklen, and D. J. Howard. 2001. "Reproductive Processes in Two Oak (*Quercus*) Contact Zones With Different Levels of Hybridization." *Heredity* 87, no. 6: 680–690. https://doi.org/10.1046/j.1365-2540.2001.00968.x.

Wringe, B. F., R. R. E. Stanley, N. W. Jeffery, E. C. Anderson, and I. R. Bradbury. 2017. "Hybriddetective: A Workflow and Package to Facilitate the Detection of Hybridization Using Genomic Data in r." *Molecular Ecology Resources* 17, no. 6: e275–e284. https://doi.org/10.1111/1755-0998.12704.

Wu, Y., A. L. Hipp, G. Fargo, N. Stith, and R. E. Ricklefs. 2023. "Improving Species Delimitation for Effective Conservation: A Case Study in the Endemic Maple-Leaf Oak (*Quercus acerifolia*)." *New Phytologist* 238, no. 3: 1278–1293. https://doi.org/10.1111/nph.18777.

Zhou, B.-F., S. Yuan, A. A. Crowl, et al. 2022. "Phylogenomic Analyses Highlight Innovation and Introgression in the Continental Radiations

of Fagaceae Across the Northern Hemisphere." *Nature Communications* 13, no. 1: 1320. https://doi.org/10.1038/s41467-022-28917-1.

Zhou, X., N. Liu, X. Jiang, et al. 2022. "A Chromosome-Scale Genome Assembly of *Quercus gilva*: Insights Into the Evolution of *Quercus* Section *Cyclobalanopsis* (Fagaceae)." *Frontiers in Plant Science* 13: 1012277. https://doi.org/10.3389/fpls.2022.1012277.

Zizka, A., D. Silvestro, T. Andermann, et al. 2019. "CoordinateCleaner: Standardized Cleaning of Occurrence Records From Biological Collection Databases." *Methods in Ecology and Evolution* 10, no. 5: 744–751. https://doi.org/10.1111/2041-210X.13152.

Zomer, R. J., J. Xu, and A. Trabucco. 2022. "Version 3 of the Global Aridity Index and Potential Evapotranspiration Database." *Scientific Data* 9, no. 1: 409. https://doi.org/10.1038/s41597-022-01493-1.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.