Foliar spectral signatures reveal adaptive divergence in live oaks (*Quercus* section *Virentes)* across species and environmental niches

Mariana S. Hernández-Leal[1*], J. Antonio Guzmán Q.[1], Antonio González Rodríguez[2], Jeannine Cavender-Bares[1*]

[1] Department of Organismic and Evolutionary Biology, Harvard University, 22 Divinity Ave., Cambridge, MA 02138, USA.

[2] Instituto de Investigaciones en Ecosistemas y Sustentabilidad, Universidad Nacional Autónoma de México, Antigua Carretera a Pátzcuaro No.8701, Morelia, 58190, México

*Authors for correspondence:

Mariana S. Hernández-Leal:

Orcid Id 0000-0002-2042-3882

Email mshernandez@fas.harvard.edu,

Tel: 612-477-7861

Jeannine Cavender-Bares:

Email jcavender@fas.harvard.edu

Tel: 612-624-6337

Orcid Id 0000-0003-3375-9630

Total word count (excluding summary, references and legends)	7243	Acknowledgment	53
Summary	218	No. of Fig.s	7 (Figs 1- 7 in color)
Introduction	1113	No of supporting Information Files	3 (Supporting information.docx; pRDA_Results.xlxs, PST_Results.xlxs)
Materials and Methods	1974		
Results	1635		
Discussion	2276		
Conclusion	245		

Summary

1. Genomic tools have transformed our understanding of species and population genetic structure in landscapes. However, discerning the impacts of neutral and adaptive evolutionary forces remains challenging, largely due to the scarcity of tools capable of measuring a broad spectrum of phenotypic traits.
2. We used spectroscopic data from preserved leaves to test for adaptive divergence among populations of live oaks (*Quercus* section *Virentes*) across genetic and phylogenetic levels. The monophyletic lineage includes seven species that diversified under sympatric, parapatric and allopatric speciation modes. We used 427 individuals to test for isolation-by-distance (IBD) and isolation-by-environment (IBE), as well as the influences of selection and phylogenetic inertia on traits. Finally and to examine how phylogenetic signals are distributed across their foliar reflectance spectra.
3. Partial redundancy analyses (pRDA) revealed that (IBE explains more phenotypic variation than (IBD among sympatric species, particularly in certain spectral regions and traits derived from spectra. Across the phylogeny, phylogenetic generalized least squares (PGLS) models show that environmental variables—including minimum temperature of the coldest month and annual precipitation—predict traits related to stress tolerance across climatic gradients, such as lignin content and anthocyanin levels.
4. These results demonstrate that leaf reflectance spectra can be used to capture adaptive differentiation and evolutionary history across scales, offering a powerful, non-destructive tool for linking phenotype, environment, and evolutionary processes in long-lived plant lineages.

Introduction

In heterogeneous landscapes, selection processes act on populations, generating genetic differentiation and local adaptation, while gene flow—a neutral evolutionary force—limits divergence between populations (Haldane, 1948; Slatkin, 1987; Manel et al., 2003; Storfer et al., 2007). When gene flow is reduced, genetic drift plays a more significant role, especially in small and isolated populations. Thus, biodiversity arises from both adaptive and non-adaptive processes and understanding the spatial and temporal scales at which these processes operate is one of the main challenges in evolutionary biology (Bernatchez, 2016; Wellenreuther & Hansson, 2016; Luikart et al., 2018).

Advances in genomics have helped clarify the relationships between genetic diversity, spatial structure, and environmental factors (e.g., Kanaka et al. 2023, Hipp et al. 2020, Deschepper et al. 2017). However, genomic data alone are not sufficient to explain phenotypic diversity, which arises through a combination of genetic variation, natural selection, and plastic responses to the environment (Miner et al., 2015, Wood et al. 2021, Svenson et al. 2021). Phenotypes are subject to selection and influence an organism's performance in different environments driving diversification. In contrast, the genetic structure of neutral genes mainly reflects demographic processes such as genetic drift and changes in effective population size, which are the result from historical biotic and abiotic conditions throughout a species´ evolutionary history (Leoninen et al., 2013). By comparing the degree of divergence between neutral markers and quantitative traits, it is possible to test for selection to achieve a broader perspective on how populations respond to environmental change (Mackay et al., 2009; Hill & Kirkpatrick 2010). Leaf reflectance spectra provide a new opportunity to describe plant phenotypes and to test selection in comparison with neutral genetic variation. Previous studies have shown tight coupling between leaf or canopy optical properties of plants and their evolutionary relationships (Asner & Martin 2011; Cavender-Bares et al., 2016; Meireles et al., 2020; Anderegg et al., 2023). In recent decades, the use of leaf spectroscopy to quantify biological diversity as a non-invasive method has been intensified providing rapid and reliable information of leaf optical properties (Asner &

Martin, 2016; Cavender-Bares et al., 2016; Meireles et al., 2020). Leaf optical properties, such as reflectance, transmittance, and absorbance, are determined by structural and biochemical components that reflect the energy acquisition and resource allocation of plants (Ustin & Gamon, 2010; Cavender-Bares et al., 2017; Kothari & Schweiger, 2022). Spectral phenotypes have been demonstrated to evolve like quantitative traits, with some regions under strong selection and others shaped by neutral processes (Meireles et al., 2020).

This study employs leaf spectroscopy from pressed leaves and existing genetic data to investigate neutral and adaptive evolution in *Quercus* section *Virentes*, a group of seven live oak species spanning from the southeastern U.S. to Costa Rica, including Baja California Sur and Cuba populations (Q. *virginiana*, Q. *geminata*, Q. *minima*, *Q. brandegeei, Q. fusiformis*, Q. *oeloides*, and *Q. sagraeana*) (Nixon & Muller 1997; Manos et al., 1997; Cavender-Bares et al., 2011). These species inhabit low-elevation temperate zones with mild winters or seasonally dry tropical climates (Muller, 1961a; Boucher, 1983; Nixon, 1985; Cavender-Bares et al., 2015) (Fig. 1). Their broad distribution across diverse climates, hydrology, and fire regimes has driven ecological divergence through both allopatric and sympatric processes (Cavender-Bares, 2019). Phylogenetic studies (Cavender-Bares et al., 2015; Hipp et al., 2020) reveal two main clades within the lineage—one with *Q. fusiformis and Q. brandegeei*, and another with the other five species. Within the latter clade, Q. *minima*, Q. *virginiana*, and Q. *geminata* coexist sympatrically, while Q. *oeloides* and Q. *sagraeana* are allopatric. Divergence between Q. *virginiana* and Q. *geminata* is maintained by differences in flowering phenology, but Q. *geminata* and Q. *minima* have overlapping flowering periods, promoting introgression (Cavender-Bares & Pahlich, 2009). Despite indications of genetic similarity based on nuclear microsatellite and chloroplast data, differences in habitat linked to leaf function, including fire dependency and leaf traits, distinguish these species (Kurz & Godfrey, 1962; Cavender-Bares et al., 2004b, 2015). Given the large climatic gradients that live oaks collectively span, natural selection is likely to have shaped leaf functional traits within and across species.

Here, we evaluate the extent to which leaf reflectance spectra from leaves of preserved samples can detect divergence and adaptive evolution between genetic groups at different hierarchical levels. Our goals are three-fold: First, evaluate the role of isolation-by-distance and isolation-by-environment (Wright, 1943; Wang & Bradburd, 2014) in driving phenotypic divergence among sympatric species (*Q. virginiana*, *Q. geminata*, and *Q. minima*), with a focus on trait-environment relationships and gene flow. Second we investigate the association of ecological variables and phenotypic variation among closely related species that have been separated into distinct abiotic environments for millions of years (*Q. fusiformis* vs. *Q. brandegeei* and *Q. oleoides* vs. *Q. sagraeana*) to assess the relative influence of adaptive trait evolution (selection) and shared evolutionary history (phylogenetic inertia) on functional leaf traits. Finally, we seek to identify regions of the electromagnetic spectrum (i.e., 400-2500 nm) capable of capturing the phylogenetic signals within the section *Virentes*, and evaluate how the strength and distribution of this signal vary across the phylogeny. We specifically compare regions of the spectrum that are strongly linked with pigments such as the visible region (i.e., 400-700 nm), to those mostly influenced by structure in the near infrared region (i.e., 700-1200 nm, NIR), and the concentration of chemical compounds in the short-wave infrared region (i.e., 1200-2500 nm, SWIR).

We test three hypotheses about adaptive differentiation in ecological niches at different biological and geographic scales within the live oaks based on foliar spectral signatures in relation to neutral molecular markers and phylogenetic information. First, we hypothesize that closely related sympatric species (*Q. virginiana*, *Q. minima*, and *Q. geminata*) exhibit distinct leaf phenotypes that are adaptive to contrasting hydrologic microhabitats, despite gene flow (Cavender-Bares et al. 1999, Cavender-Bares and Holbrook 2001), as evidenced by greater foliar phenotypic variation than can be explained by distance alone. Across all live oak populations, we further expect that traits such as anthocyanin content, lignin content, and leaf mass per area (LMA), along with spectral features tied to leaf structure, will covary with source environmental variation in directions that support the hypothesis of habitat-driven adaptations. Second, we hypothesize that phenotypic divergence between pairs of closely related allopatric

species (*Q. fusiformis* and *Q. brandegeei* or between *Q. oleoides* and *Q. sagraeana)* are a consequence of adaptive evolution associated with ecological divergence enforced by vicariance. Finally, we hypothesize that across the *Virentes* phylogeny, spectral bands in the visible region—where selection to conserve the photosynthetic machinery is strong—will show less lability and greater phylogenetic signal than in the NIR or SWIR, where spectral bands are linked to chemical and structural aspect that we expect to be labile and to vary with divergent evolutionary pathways.

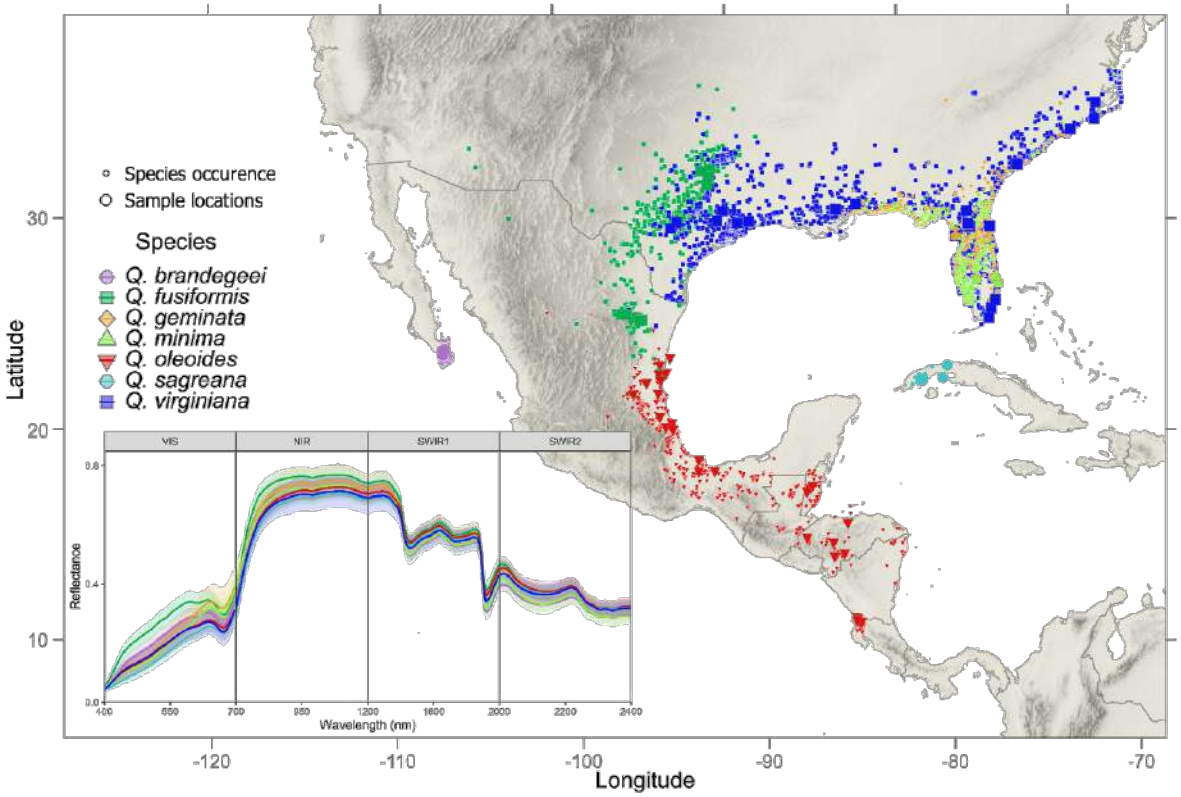


Fig. 1. Species occurrence and sample locations of seven species of *Quercus* section *Virentes*. The bottom-left panel represents the mean leaf spectral reflectance of dried, pressed samples (± standard deviation) for each species. Shown are visible (VIS), near-infrared (NIR), and first and second short-wave infrared (SWIR1 and SWIR2) spectral regions.

Material and Methods

Sample collection

Individual trees of each *Virentes* species were sampled throughout their occurrence ranges and within common gardens between 2003 and 2016 for DNA extraction and leaf phenotyping (Fig. 1). In natural populations, identification of species was based on leaf, bark and stem height characters following Muller (1961), Nixon & Muller (1997) and Kurz & Godfrey (1962). Multiple leaves from each tree were pressed and stored in a dry cabinet in the laboratory, with representative voucher specimens housed in the University of Minnesota Bell Museum of Natural History. Common garden experiments were established at the University of Minnesota Plant Growth Facilities from seeds collected in natural populations. Leaves from individual saplings from the common gardens were pressed, dried and stored in a dry cabinet (Supporting Table S1). Additional traits, like freezing and drought tolerance and differences in growth form, further reinforce reproductive isolation. (description of environment for each species can be found in Supporting Methods S1).

Neutral genetic variation and genetic structure

We used eleven nuclear simple sequence repeats (nSSR) loci previously employed for assessing neutral genetic variation in other studies (Cavender-Bares & Pahlich 2009; Cavender-Bares et al., 2011; Gugger et al., 2013; Cavender-Bares et al., 2015). We selected 427 individuals that represented the entire geographic range of the seven species and for which leaf material was well-preserved and molecular data were available. We applied a Bayesian clustering analysis to disentangle the genetic groups among the seven *Virentes* populations and among the *Q. oleoides* populations. This analysis was conducted using STRUCTURE v. 2.3.4 (Pritchard et al., 2000) with an admixture model, excluding location as prior information (Hubisz et al., 2009). Further details on genetic clustering analysis are provided in the Supporting Information (Methods S2). We calculated pairwise genetic differentiation between species and genetic groups with F_{ST} estimators using *hierfstat* Package in R (Goudet, J. 2005).

Dried leaf reflectance spectra measurements and preprocessing

Pressed samples are stored at the Harvard University Herbaria. Leaf reflectance was measured on the adaxial surface of three fully expanded leaves in each of the 427 individuals using a leaf-clip with an internal light source attached to a high-spectral resolution field spectroradiometer SVC HR-1024i (Spectra Vista Corp., Poughkeepsie, NY, USA). Leaf reflectance spectra were corrected for the splice in bands near 990 and 1900 nm using the Spectrolab package in R (Miereles et al., 2017). Subsequently, spectra were resampled to 3 nm and transformed using continuous wavelet transformation (CWT). CWT was used as a method to isolate and enhance spectral features to improve discrimination among species and phenotypes (White et al. 2025). This transformation was computed based on a second-order derivative of Gaussian using scales 2^2, 2^4, and 2^6. Wavelet scales were then summed for further analyses. Bands at the edge of the spectrometer range (< 400 nm and > 2450 nm) were excluded for further analysis due to low signal-to-noise ratios or transformation artifacts. The CWT was performed using the “wavCWT” function in the 'wmtsa' package of R (Constantine & Percival, 2016). Details on CTW spectra transformation are provided in the Supporting Information (Methods S3).

Estimation of leaf traits from spectra

We predicted leaf reflectance spectra functional traits using previous measurements of dried-leaf reflectance spectra. Specifically, we focused on six leaf structural traits: leaf mass per area (LMA; kg m^{-2}), leaf dry thickness (LDT; mm), content of cellulose (CEL; %), hemicellulose (HEM; %), soluble cell contents (SOL; %), and lignin (LIG; %). Estimations of leaf traits were performed excluding petioles from herbarium samples. The concentration of carbon fractions, including solubles (non-structural carbohydrates, cell contents like carbohydrates, lipids, pectin, starch, soluble proteins and non-protein nitrogen). hemicellulose, cellulose, and lignin (%) were obtained from sequential digestion (Fiber Analyzer 200; ANKOM Technology).

We used a common Partial least squares regression (PLSR) modeling framework to predict leaf traits from dried-leaf reflectance spectra across the full range (400-2450 nm). This framework involves three main steps: i) split the data into training and testing

datasets, ii) selection of the optimal number of components, and iii) assessment of the models. A detailed description of the employed framework is presented in Supporting (Methods S4).

Partial least-squares discriminant analysis for species classification

We employed a Partial Least Squares Discriminant Analysis (PLS-DA) model to classify dried-leaf wavelet spectra among the seven species in *Virentes*. Each species was represented by a minimum of 20 samples. The dataset was divided into 70% training (calibration) and 30% testing (validation) subsets based on the species-genetic group level. We first conducted an iterative PLS-DA analysis (50 iterations) using the bootstrap method to control sample size to determine the optimal number of components. After selecting the optimal number of components, a final PLS-DA model was generated based on 50 iterations. We assessed the classification performance using confusion matrices between species and genetic groups. PLS-DA modeling was conducted using the caret R package (Khun et al., 2020).

Bayesian clustering based on spectrally predicted traits

In order to compare species phenotypic structure, we used the unsupervised Bayesian clustering algorithm GENELAND v 4.0.9.(Guillot et al., 2009). The six traits spectrally predicted from reflectance spectra were used together with the geographic location of the populations to delimit phenotypic clusters among the seven *Virentes* species and among populations of *Q. oleoides*. To explore the possible number of phenotypic clusters, we carried out 11 independent runs with K ranging from 1 to 10 with 2×10^6 MCMC iterations and a thinning value of 1000. We subsequently fixed the K value at 6 to estimate individual probabilities of membership in each cluster for the seven species of *Virentes* and K value of 5 for *Q. oleoides*. We ran 20 additional independent runs and chose those that had the highest mean logarithmic posterior probability and post-processed them using a burn-in equivalent of 10%; each cell was about 10 km^2. Individuals with a membership probability of less than 0.6 were considered admixed. To visualize the results in GENELAND, runs of the most probable K value were summarized using CLUMPP v 1.1.2 (Jakobsson & Rosenberg, 2007).

Comparing phenotypic vs genotypic divergence

Analyzing the relationship between phenotypic and genetic variation offers valuable insights into the presence of selection, particularly through P_{ST} - F_{ST} comparisons (McKay & Latta, 2002; Merilä & Crnokrak, 2001). The nSSR-derived genetic distances (F_{ST}) were compared with quantitative trait distances (P_{ST}) for each pairwise species. Phenotypic distances were derived considering the six traits spectrally predicted from the PLSR spectral modelling, and the three main spectral regions: Visible (400-700 nm), near infrared (NIR, 700-1200 nm) and short-wave infrared (SWIR, 1200-2500 nm). From these regions we selected—as phenotypic traits—wavelengths that had high values of Variables of Importance in Projection (VIP) from the PLS-DA analysis to discriminate seven species using CWT wavelet spectra. First, we selected all wavelengths with VIP values higher than 0.8; subsequently, to have a broader selection of traits, we included bands with values higher or close to 0.4 (Wold *et al.*, 2001). We refer to these selected wavelengths as spectral traits thereafter. Detailed methods for calculating P_{ST} using spectrally predicted traits and the full spectra are provided in the Supporting Information (Method S5).

Spatial and environmental drivers of population divergence.

To determine whether trait differentiation among species follow patterns of isolation-by-distance (IBD) or isolation-by-environment (IBE), we used Redundancy Analysis (RDA) to compare genetic, spatial, and phenotypic variation to variation in environmental variables. The comparison provides a means to test whether environmental variables drive phenotypic differences beyond spatial distance effects. To select the environmental variables

For each pair of species, partial redundancy analysis was conducted to partition the explainable phenotypic variation, into those attributable to spatial (SPACE), genetic (GEN), environmental factors (ENV), and their combined effect. The full, partial and joint contributions of SPACE, ENV and GEN to the explainable phenotypic variations were estimated and tested for significance, and the most influential single explanatory variables were identified. (Supporting pRDA_results.xlsx).

For more detailed explanation of pRDA models see Supporting Information (Methods S6)

Testing for phylogenetic signal

We use phylogenetic least squares regression (PGLS) to analyze phylogenetic signals in phenotypes while simultaneously examining how these traits derived data relate to environmental variation. The analysis was conducted using the phylogenetic tree derived from RADseq data (Cavender-Bares et al. 2015) as a framework with 17 samples of different populations of the 7 different species. To obtain the variation in phenotypic characters. The mean of the individuals belonging to these populations was calculated.Spectral traits were analyzed separately for each region of the spectrum: Visible (VIS), Near-Infrared (NIR), and Short-Wave Infrared (SWIR), as well as for six key leaf traits derived from leaf spectroscopy. To gain a broader perspective, the same method was also applied to various vegetation indices that characterize features associated with biophysical and chemical properties. The indices selected were Chlorophyll Index Red Edge (CI= 750nm/710nm) as a descriptor of the chlorophyll concentration (Gitelson et al. 2003), Normalized Difference Water Index (NDWI= (835nm - 1610nm)/(835nm + 1610) as a descriptor of the leaf water content (Quemada et al. 2021), and the Anthocyanin Reflectance Index (ARI = (1 / 550nm) - (1 / 700nm)) as a descriptor of the anthocyanins concentration (Li et al. 2023). Overall, this approach allowed us to assess the phylogenetic structure and environmental associations for each spectral region, vegetation index, and predicted leaf trait independently, providing insights into how evolutionary and environmental factors shape spectral variation. Most common environmental variables used in pRDA analysis were used after removing highly correlated variables (based on Pearson's correlation $|r| \leq 0.60$) to control for multicollinearity for each pairwise species. To evaluate the correlation between environmental variables and the traits of 427 samples, and to complement the information and gain a better understanding of how environmental variables relate to the traits, a Pearson correlation was performed, as opposed to the PGLS which only used 17 individuals. The phylogenetic signal was quantified using Blomberg’s K to determine the extent to which spectral data were conserved along the phylogeny (Blomberg et

al.,2003) implemented in Phytools (Revell, 2012). Blomberg's K measures the degree to which trait variance lies within clades vs among clades as compared to a Brownian expectation. Significance was assessed using 999 tip-swap randomizations. The package phylosig in R (Revell, 2012), was chosen because it incorporates standard errors (SE) in its calculations and allows flexibility in estimating Blomberg's K with different settings. A value of K=0 indicates no phylogenetic signal, while K>1 suggests that closely related species exhibit stronger trait similarities than expected under a Brownian motion model of evolution (Blomberg et al., 2003).

To evaluate the significance of the observed K values, we compared them to null distributions generated under two models. First, a white noise (WN) model was used, where trait values were randomly permuted across the phylogeny's tips 1000 times. Second, a Brownian motion (BM) model simulated trait evolution under BM across the phylogeny 1000 times. K values falling below the 95% distribution of the simulated BM values suggest traits are less phylogenetically structured than expected under Brownian motion (Blomberg et al., 2003).

Accounting for environmental influence on phenotypes using spectral variation in leaves from a common garden

We assessed the phenotypic differences among species independently of environmental effects by using pressed leaf samples from a common garden. To do so, we measured reflectance spectra from pressed leaves from 22 individuals of each of four species (*Q. fusiformis, Q. geminata*, *Q. oleoides, Q. virginiana*). These individuals were grown in a common garden in a controlled environment at the Plant Growth Facility on the St. Paul campus of the University of Minnesota. Phenotypic distances (P_{ST}) from spectrally derived trait values were calculated in the same way as for wild populations (VIP wavelengths were not used in this case). All pairwise estimates of P_{ST}, their confidence intervals (CI) and the comparison with F_{ST} genetic distances were calculated using the package Pstat in R (Blondeau & Da Silva, 2018). F_{ST} (and CI) were estimated using the *boot.ppfst* function in *hierstat* R package.

Results

Neutral genetic variation

Pairwise genetic differentiation values using F_{ST} for nSSRs were statistically significant ($p < 0.05$) for all species pairs and ranged between 0.012 and 0.25. The highest value was found between *Q. brandegeei vs Q. geminata* (Supporting Table S2). Pairwise genetic differentiation using genetic groups identified by STRUCTURE for F_{ST} showed similar trends to those obtained at the species level (Supporting Table S2).

Functional traits from spectra

The PLSR models for estimating traits using the internal validation dataset show high to moderate accuracy and precision. The best spectrally predicted traits were LMA (R^2 = 0.88, RMSEP = 17.6), cellulose (R^2 = 0.72, RMSEP = 3.37), and thickness (R^2 = 0.65, RMSEP =0.8) Table S4, Supporting Fig. S2. Variable importance in projection (VIP) was used to identify which regions of the spectrum were important in predicting leaf traits (Supporting Fig. S3). For all spectrally predicted traits, the ranges between 660-680 and 750-780 nm (visible) (Supporting Fig. S3 a, c) showed the highest importance in predicting traits. SWIR range was important to predict lignin and soluble cell contents (carbohydrates, lipids, pectin, starch, soluble proteins and non-protein nitrogen). around 1920-2050 nm (Supporting Fig. S3 b, d). The NIR and much of the SWIR region (i.e., 1400–1850 nm) were less important for predicting leaf traits.

Bayesian clustering based on genetic data and phenotypic traits

Using the nSSRs from 427 individuals, the clustering analysis suggested seven (K = 7) most probable genetic groups with a mean lnP(K) = -3337.23 and ΔK = 96.93. This result is like those obtained in Cavender-Bares et al. (2015), indicating that the reduced sample size did not meaningfully change the genetic clustering. As before, *Q. geminata* and *Q. minima* formed a unique genetic group 1 (Fig.. 2a). Using the GENELAND algorithm, the most probable number of clusters obtained based on all spectrally predicted traits (i.e., phenotypes) was *K*= 6. Phenotypic group 2 combines individuals from five different species, showing similar phenotypic characteristics in *Q. geminata, Q. minima, Q. virginiana*, *Q. sagraeana*, and Central America populations of *Q. oleoides*

(Fig. 2b). Clustering within *Q. oleoides* using nSSR data showed four major genetic groups (K= 4) with a mean lnP(K) = -7935.1 and ΔK = 13.02. This analysis describes genetic groups divided into major geographical regions: Northeastern Mexico (N. MX), Southeastern Mexico (S. MX.), Central America (BZ and HND), and Costa Rica (CR). Phenotypic clustering using GENELAND showed five distinct phenotypic groups, suggesting differences with respect to genetic groups (Supporting Information Fig S4).

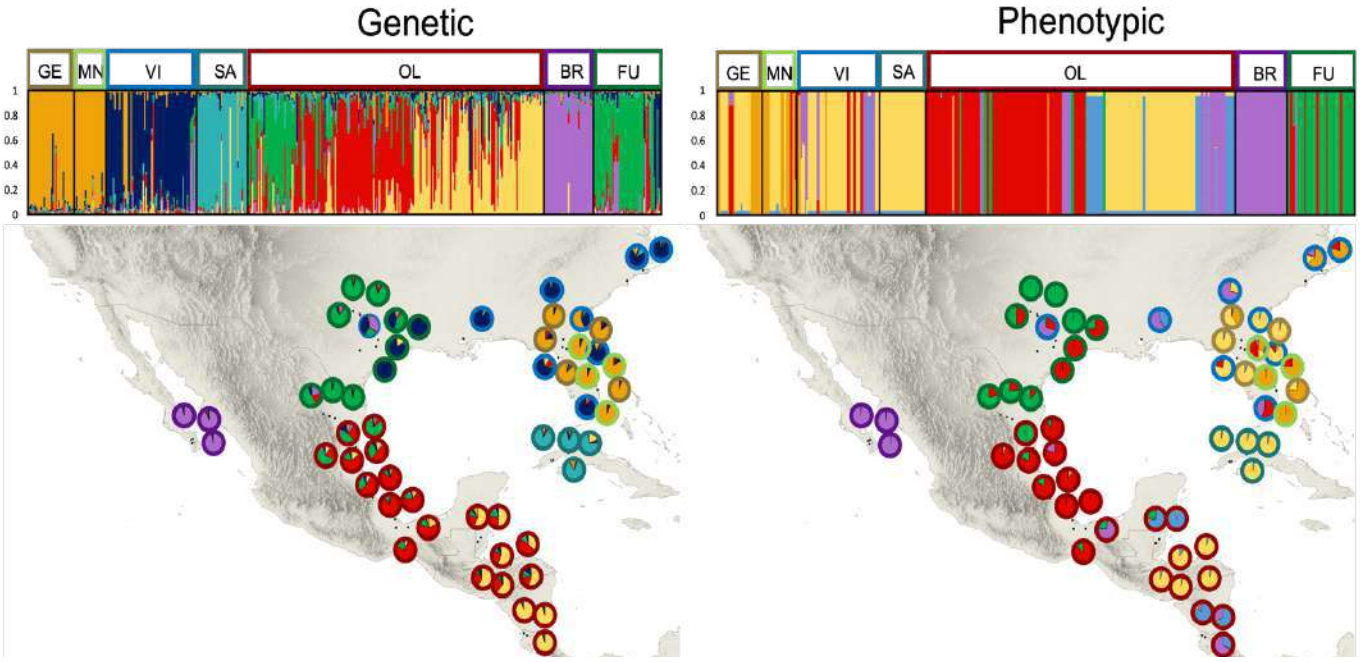


Fig. 2. Spatial distribution of genetic and phenotypic variation of seven oak (*Quercus*) species of the section *Virentes.* **a**), Seven genetic groups identified by STRUCTURE from 56 populations. **b**), Six phenotypic groups identified by GENELAND using six leaf traits derived from reflectance spectra from the same 56 populations. The percentage assignment to genetic or phenotypic groups is represented at both the individual tree level (upper bar plots) and subpopulation level (pie charts). Colors outlining the pie charts represent species of sample origin. Colors inside the pie charts represent genetic (a) or phenotypic (b) groups. GE= *Q. geminata*, MN= *Q. minima*, VI= *Q. virginiana*, SA= *Q. sagraeana*, OL= *Q. oleoides*, BR= *Q. brandegeei*, FU= *Q. fusiformis*.

Partial least-squares discriminant analysis for species classification

In contrast to the unsupervised Bayesian clustering algorithm GENELAND where *Q. geminata* and *Q. minim*a were clustered in the same genetic group, the PLS-DA classification model using spectra from pressed leaves showed high performance for discriminating individuals among the seven species of *Virentes*.This analysis correctly predicted the taxonomic identity of 295 out of 302 samples in the training dataset (Fig. 3).The model performed well in discriminating individuals among species with sympatric

distributions. For example, it achieved 85% classification accuracy for *Q. geminata*, 84% for *Q. minima*, and 85% for *Q. virginiana*.When using the SSR data (Fig. 2a), the analysis revealed high rates of admixture between individuals of *Q. geminata* and *Q. minima*.

This analysis correctly predicted the taxonomic identity of 295 of 302 samples in the training dataset (Fig. 3), showing good results discriminating individuals among species that have sympatric distributions. For example, it achieved 85% classification accuracy for *Q. geminata*, 84% for *Q. minima*, and 85% for *Q. virginiana* (85%). Accuracy, kappa, sensitivity, and specificity using 20 PLS components results are in Supporting information Table S3.

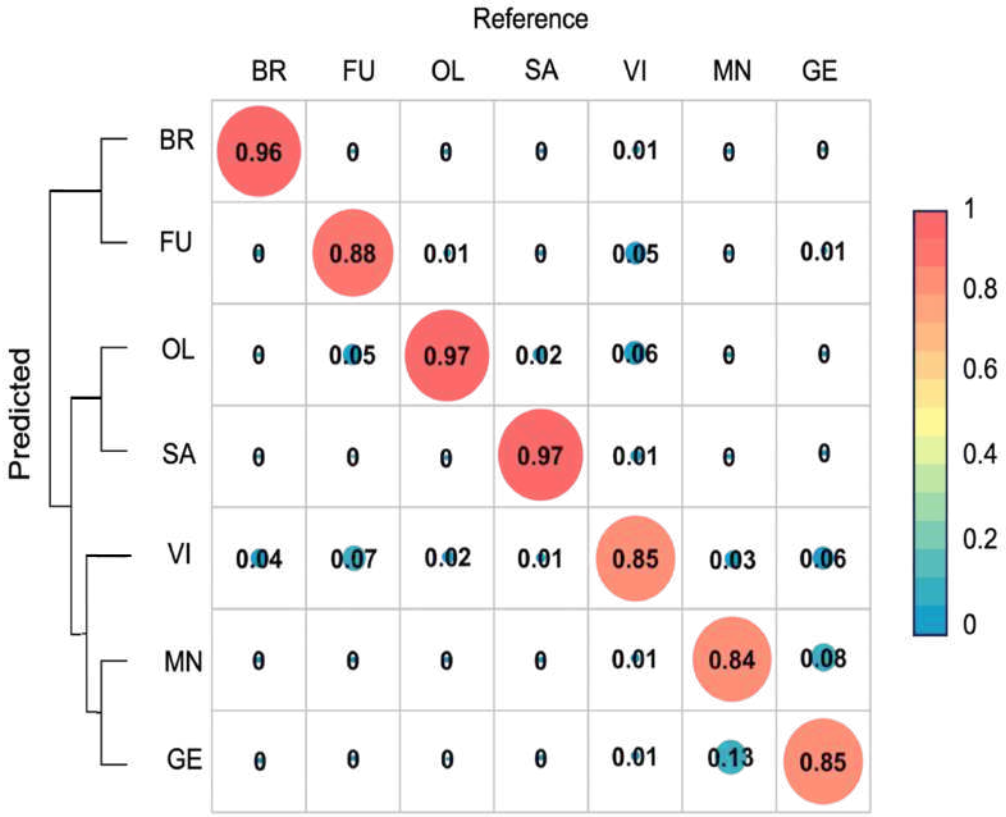


Fig. 3. Confusion matrices from PLS-DA analysis for seven *Quercus* species of section *Virentes*. Columns represent the observed identities, while rows indicate the predicted identities and phylogenetic relations among species. Values along the diagonals show the percentage of individuals accurately classified within each group, while values above or below diagonals show the percentage of misclassification.Left side phylogeny to show species relations. GE= *Q. geminata*, MN= *Q. minima*, VI= *Q. virginiana*, SA= *Q. sagraeana*, OL= *Q. oleoides*, BR= *Q. brandegeei*, FU= *Q. fusiformis*.

Comparing phenotypic vs genotypic divergence

Using leaf traits derived from spectra, sympatric species (i.e., *Q. geminata vs Q. virginiana*) showed significant differences in their P_{ST} pairwise distances ($c/h^2 > 0.25$) for almost all traits compared to F_{ST} except for cellulose. *Q. minima vs Q. virginiana* had significant differences in their P_{ST} pairwise distances for LMA, thickness, solubles and hemicellulose ($c/h^2 > 0.25$) and cellulose and lignin when ($c/h^2 > 0.5$). For the pairwise comparison between *Q. geminata* and *Q. minima,* only cellulose showed significantly higher values of P_{ST} (0.86) than F_{ST} (0.011).

Significant differences between P_{ST} -F_{ST} for sister species *Q. oleoides* and *Q. sagraeana* (F_{ST} = 0.073) showed evidence for phenotypic selection on three leaf traits: LMA (P_{ST} = 0.8), thickness (P_{ST} = 0.88), and lignin (P_{ST} = 0.93). Differentiation measures between *Q. brandegeei* and *Q. fusiformis* (F_{ST} = 0.153) were also significantly different on four traits: thickness (P_{ST} = 0.96), solubles (P_{ST} = 0.92), cellulose (P_{ST} = 0.93), and lignin (P_{ST} = 0.9). Fig. 4a shows trait difference for pairwise comparisons among species with c/h^2 values of 0.25, 0.5 and 0.75.

Using bands with high importance in discriminating species (i.e., VIP) from the PLS-DA species classification model, we found that closely related species tended to be more spectrally similar in the visible region than more distantly related species. Given that our main objective is to focus on adaptive differentiation in ecological niches at different biological and geographic scales, we only present pairwise comparisons values of phenotypic significant $P_{ST} > F_{ST}$ among sympatric species (*Q. geminata, Q. virginiana,* and *Q. minima*), allopatric sister species (*Q. fusiformis* and *Q. brandegeei* or *Q. oleoides* and *Q. sagraeana*), and widely distributed species with parapatric populations (*Q. oleoides* and *Q. fusiformis; Q. virginiana*) are shown in Fig. 4b. Detailed results for all pairwise comparisons values of phenotypic P_{ST}-F_{ST} obtained from spectrally derived traits, and their confidence intervals can be found in the Supporting Fig. S5 and Supporting information (PST_Results.xlsx).

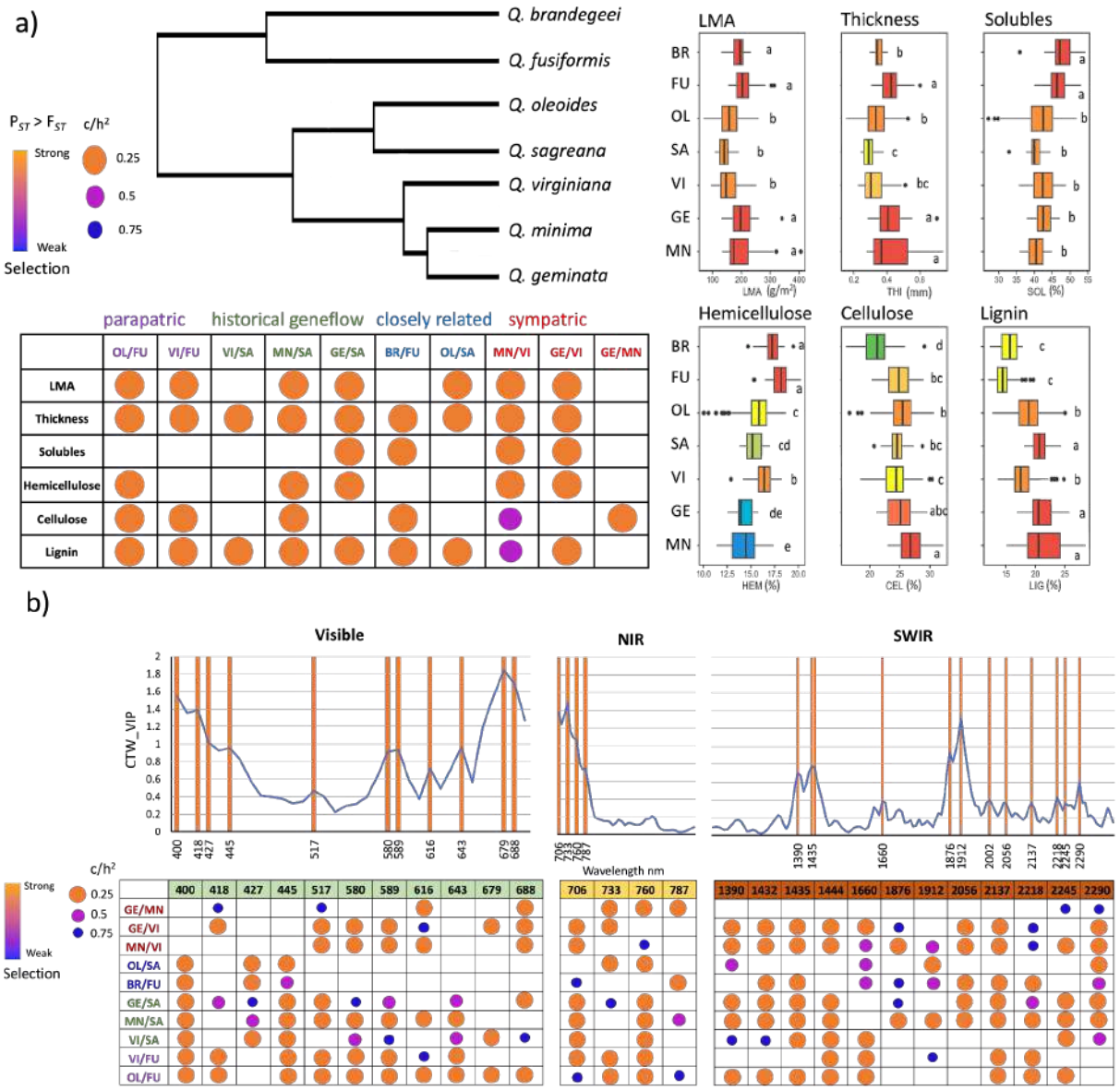
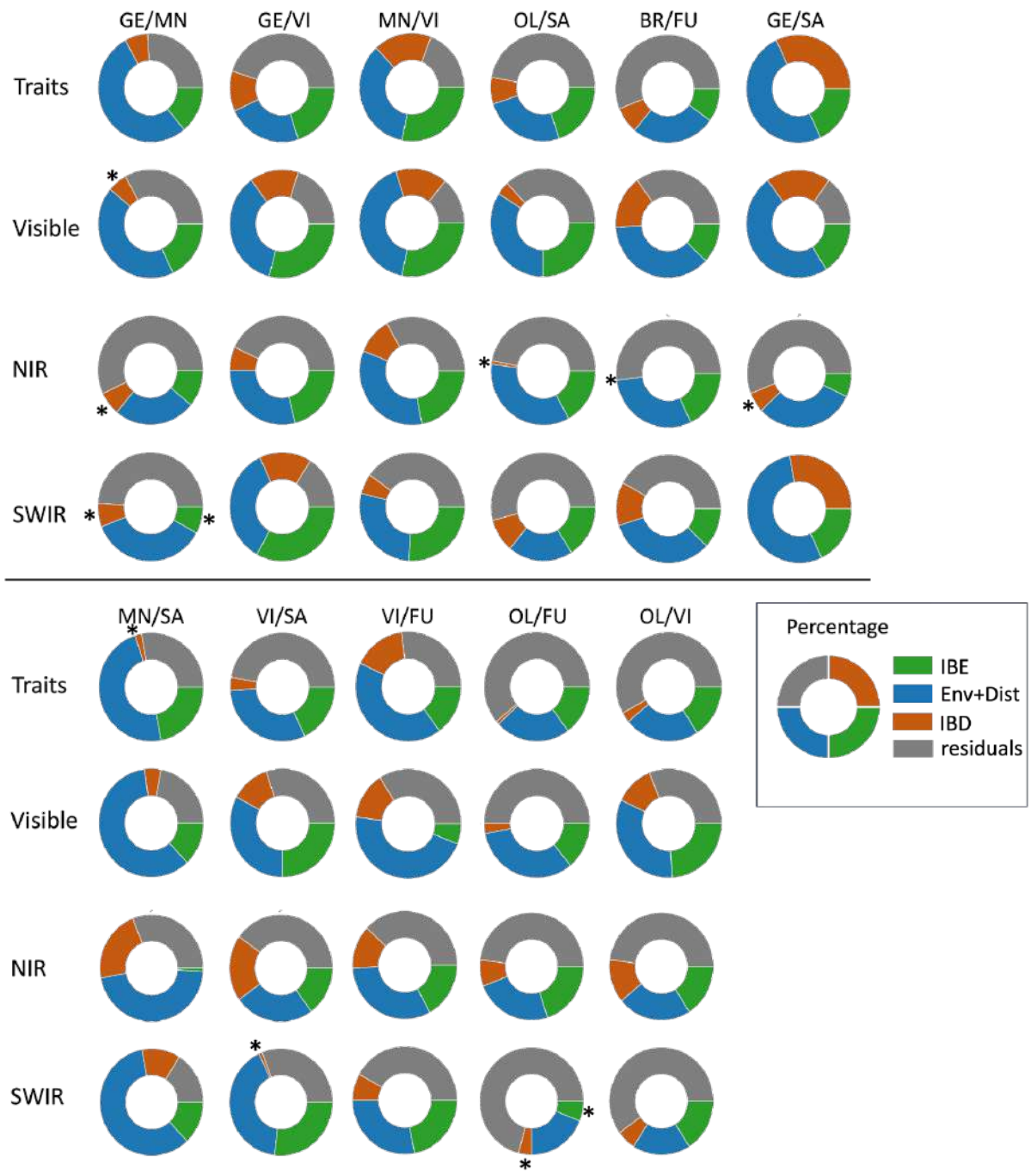


Fig. 4. Pairwise P_{ST} comparisons: Matrices represent P_{ST} plotted as a function of c/h² values selected at 0.25 (orange), 0.5 (pink) and 0.75 (blue). The optimal value of c/h² at which the lower confidence limit of P_{ST} is higher than the upper confidence limit of F_{ST}. Panel **a) spectrally predicted traits (LMA: Leaf mass area; THI: thickness; SOL: solubles; HEM: hemicellulose; CEL: cellulose; LIG: lignin)** LEFT shows a simplified phylogenetic tree inferred from RADseq data for 17 *Virentes* individuals using RAXML (Cavender-Bares et al., 2015) with the pairwise comparisons among species that were conducted using the six spectrally predicted traits. Colored pairs represent sister relationships, historical introgression between specie pairs, and/or sympatric geographic associations within the *Virentes*: Red, sympatric sister species; Blue, sister but not sympatric species; Green, historically introgressing populations; Purple, parapatric species with introgression. RIGHT Means and variance are shown with box and whisker plots for each species for the six predicted traits. Colours are associated with species means, and different letters indicate significant Student's t-test species--level differentiation (P < 0.05). Panel **b) Variable Importance of Projection (VIP) spectral bands selected within the Visible, near infrared (NIR) and short-wave infrared (SWIR) regions based on importance in discrimination among species using PLSDA,** represents plotted VIP values obtained

from the PLS-DA classification model using wavelet spectra: orange vertical lines represent wavelengths that were used as traits to calculate P_{ST} pairwise distances. Matrices are divided into Visible, NIR, and SWIR spectral regions.phylogenetic/geographic relations. GE= *Q. geminata*, MN= *Q. minima*, VI= *Q. virginiana*, SA= *Q. sagraeana*, OL= *Q. oleoides*, BR= *Q. brandegee*i, FU= *Q. fusiformis.* Only phylogenetically and geographically meaningful pairwise comparisons are shown **Leaf mass area, thickness, solubles, hemicellulose, cellulose, lignin**

Spatial and environmental drivers of population divergence

We performed multivariate redundancy analyses (RDA) to attribute explainable variation in phenotypic traits, to spatial location, environment or their joint effect. This analysis integrates two classical models of population structure, isolation-by-distance and isolation-by-environment. The results show the percentage of variation explained by environmental variables controlling space PHENO ~ ENV + Condition (GEO), spatial variables controlling environment PHENO ~ GEO + Condition (ENV), and the combination of both (PHENO ~ GEO + ENV for different species pairs (e.g., *Q. geminata* vs *Q. minima* GE/MI, *Q. geminata* vs *Q. virginiana* GE/VI). Overall, the total variation explained by environment and spatial location varies widely across species pairs, ranging from 20% (OL/SA) to 57% (GE/SA). In most cases, the variation explained by environmental variables (IBE) is greater than that explained solely by spatial factors (IBD), suggesting that environmental conditions play a more significant role than spatial variation in determining spectral traits. This is particularly evident in pairs, including GE/MI (18% vs. 6.5%) in the visible range, and GE/VI (33% vs. 16%) in the SWIR region. However, in certain cases (e.g., GE/SA), spatial factors explain a relatively high proportion of variation (31%) compared to environmental variables (20%), potentially reflecting spatial patterns linked to species' geographic distributions (Fig. 5). Complete results showing Adjusted r^2 and proportion of variance explained (PVE %) for each model, and proportion of variable explained by each environmental variable (Supporting Information, pRDA_Results.xlsx).



493

494 **Fig. 5. Redundancy analysis of divergence in pairwise *Quercus Virentes* species**. Radians
495 represent the percentage of the total variation in spectrally derived traits and spectral VIP wavelengths in
496 each region Visible, NIR and SWIR that can be explained by space:Environment: PHENO ~ ENV +
497 Condition (GEO) (green IBE); PHENO ~ GEO + Condition (ENV) (brown IBD) and their interaction
498 PHENO ~ GEO + ENV (blue). Asterisks (*) indicate that the PVE % is not significant. Environmental
499 variables peer individuals are listed in Supporting file Environmental.xlxs

500

501

Phylogenetic signal and phenotypic variation

We found significant levels of phylogenetic signal in the spectra of *Quercus* section *Virentes*, with strong signals observed in the visible and some parts of the NIR regions. Predicted traits such as hemicellulose, soluble contents, and lignin also exhibited strong to moderate phylogenetic signals (Table S6). In contrast, the SWIR region showed limited phylogenetic signals, suggesting that this region may be less influenced by evolutionary constraints.

Accounting for phylogenetic relationships among populations using PGLS, we found significant associations between traits and bioclimatic variables, including the minimum temperature of coldest month (Bio 6, Fig. 6A), the mean temperature of the wettest quarter (Bio 8), annual precipitation (Bio 12, Fig. 6B), and precipitation of the warmest quarter (Bio 18) (Table S5). Populations that occur in colder regions (e.g., within *Q. fusiformis* and *Q. virginiana*) tended to exhibit lower lignin content than those from warmer regions (within *Q. oleoides, Q. brandegeei, Q. sagraena*) (Bio 6), while populations occurring in drier conditions (Bio 12), including those of *Q. brandegeei* and *Q. fusiformis* showed lower concentrations of anthocyanins as indicated by lower values in anthocyanin index. Across all 427 individuals, Pearson correlation coefficients showed significant relationships between traits and the minimum temperature of the coldest month (Fig. 6A) for lignin (r=0.24, P=0.0001), solubles (r=0.2, P=0.0001), LMA (r=0.21, P=0. 0001), hemicellulose (r=0.24, P=0.0001), anthocyanin index(r=0.23, P= 0.0001), and for annual precipitation (Fig. 6B) for anthocyanin index (r=0.4, P=0.0001), hemicellulose (r=0.4, P=0.0001), chlorophyll index(r=0.4, P=0.0001), cellulose (r=0.21, P=0.0001), lignin (r=0.4, P=0.0001).

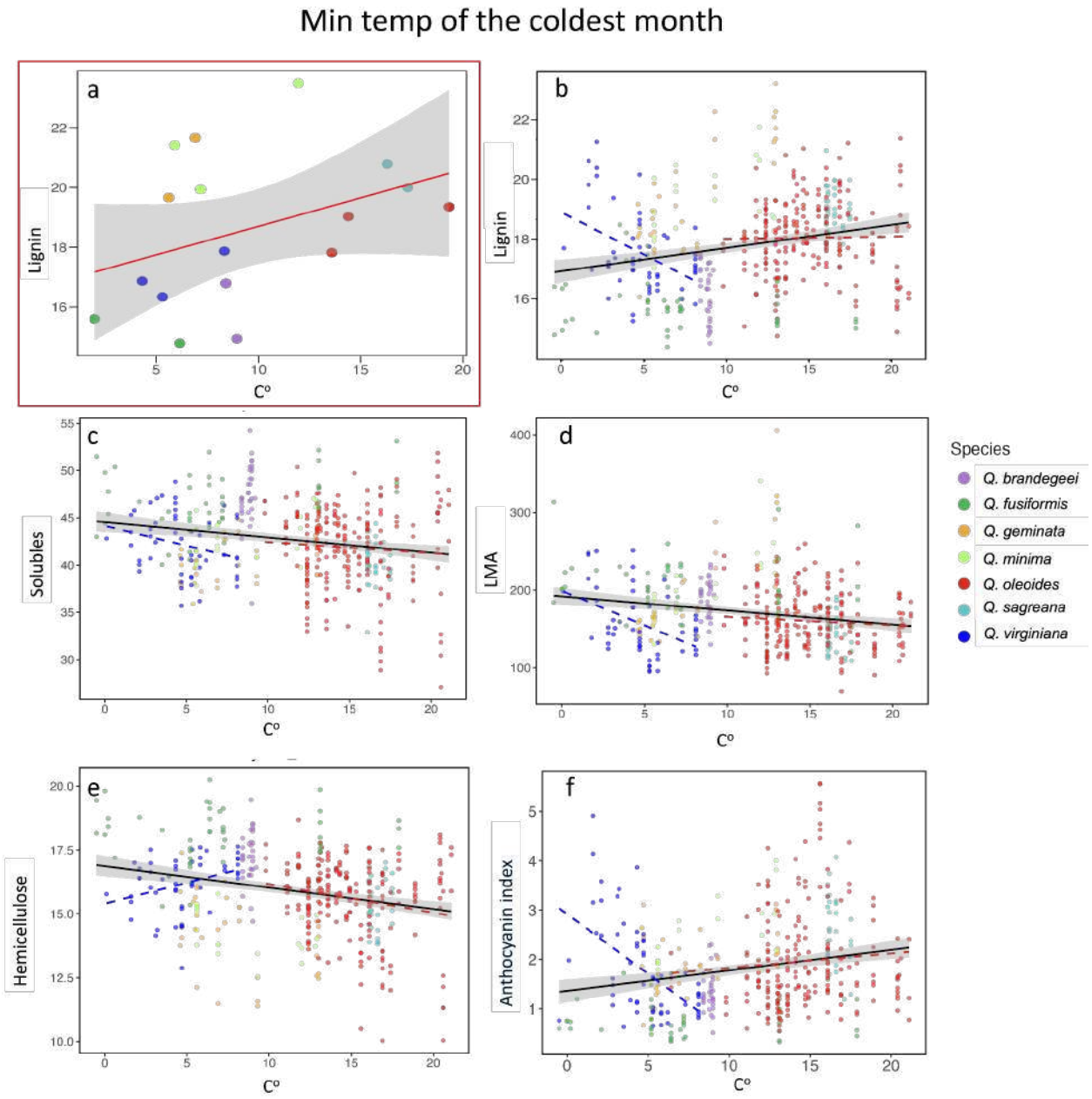


Fig. 6. Relationship between traits and minimum temperatures of the coldest month . Red square enclosed graphs represent PGLS model significant association of trait and environmental variables controlled by phylogeny using a phylogenetic tree of 17 individuals. Other graphs represent significant correlations between environmental variables and trait measures using 427 individuals. Dashed lines represent significant correlations for individuals of *Q. virginiana* (blue) and *Q. oleoides* (red)

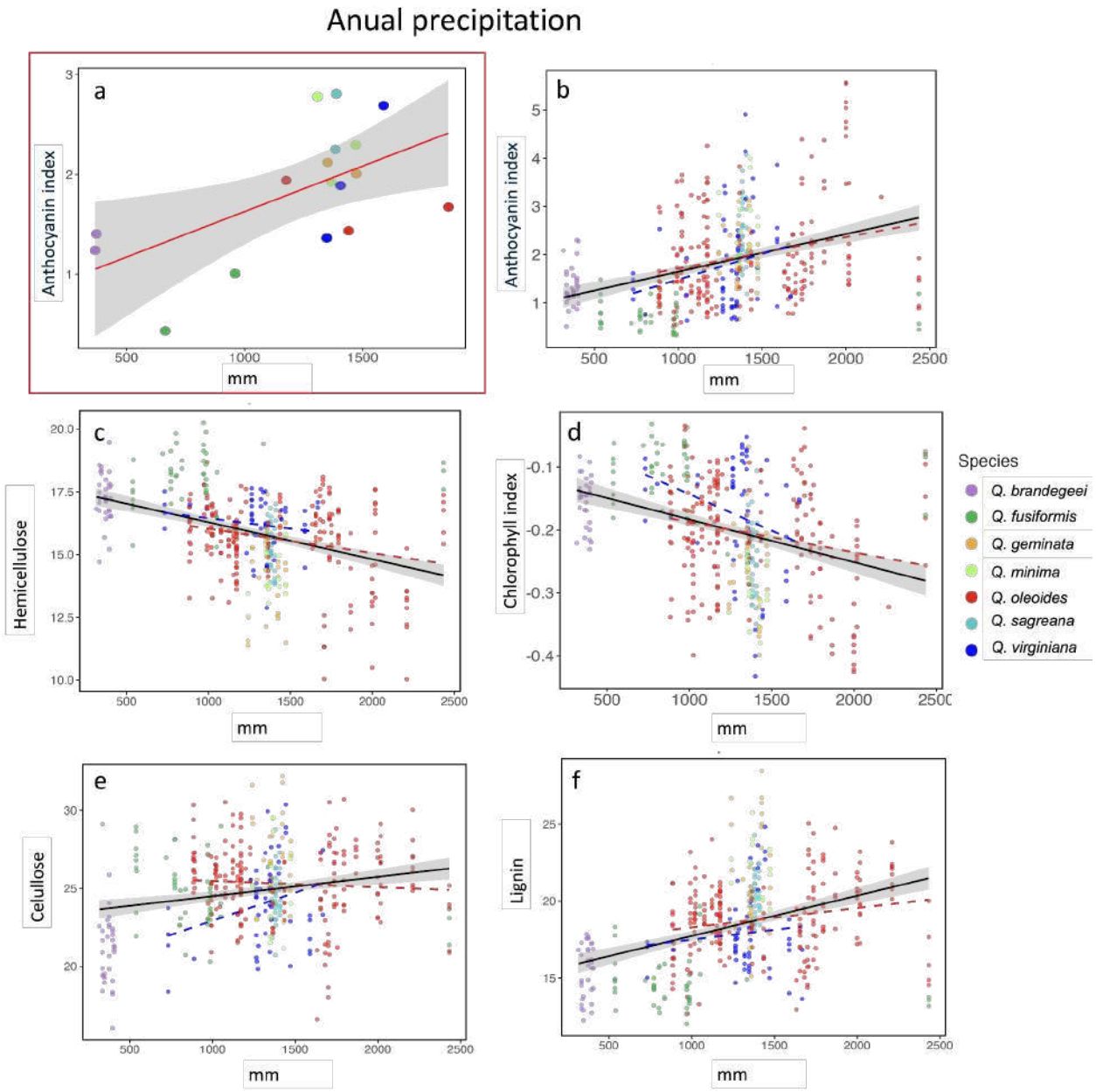


Fig. 7. Relationship between traits and annual precipitation. Red square enclosed graphs represent PGLS model significant association of trait and environmental variables controlled by phylogeny using a phylogenetic tree of 17 individuals. Other graphs represent significant correlations between environmental variables and trait measures using 427 individuals. Dashed lines represent significant correlations for individuals of *Q. virginiana* (blue) and *Q. oleoides* (red).

Accounting for environmental influence on phenotypes using spectral variation in leaves from a common garden

Common garden results for P_{ST}-F_{ST} comparisons in *Q. fusiformis, Q. geminata, Q. oleoide*s, and *Q. virginiana* generally supported the results for wild populations (shown at a c/h^2 ratio of 0.75 in Fig. S7). Notably, hemicellulose, cellulose, lignin, thickness, and LMA emerged as traits where P_{ST} exceeded F_{ST} at least in some pairwise comparisons in both common garden and wild population individuals. Not all pairwise results that were significant in wild populations remained significant in common gardens, yet many were. For example, in both common garden and wild populations, cellulose values for P_{ST} exceeded F_{ST} for Q. *virginiana vs Q. fusiformis* and for *Q. oleoides vs Q. fusiformis* but not other pairwise comparisons. For lignin, PST values exceeded F_{ST} values for *Q. geminata vs Q. virginiana* and *Q. oleoides vs Q. fusiformis* in both common garden and wild populations but not in other pairwise comparisons.

Discussion

Our study highlights the capability of spectral phenotypic data combined with neutral genetic variation to reveal evolutionary processes that have shaped diversification within *Quercus* section *Virentes*. Across the range of the live oak lineage that spans temperate and tropical environments, spectral variation provides key insights into how environmental pressures shape phenotypic traits and drive ecological divergence, complementing genetic analyses for a more comprehensive view of adaptive evolution.

Sympatric species with niche specialization show evidence for adaptive differentiation

We found that among sympatric *Quercus* species in the southeastern US, (*Quercus virginiana, Q. geminata,* and *Q. minima*), environmental selection, rather than geographic proximity, drives phenotypic divergence consistent with the hypothesis that fine-scale adaptations to contrasting local habitats enable coexistence. Our pRDA analysis shows that environmental variables explain more phenotypic variation than geographic distance, even with gene flow. Traits like LMA, lignin content, cellulose,

thickness and spectral features that differentiate species are likely under selection. These traits and phenotypic attributes may contribute to survival across hydrologically distinct microhabitats. Previous studies have shown linkages between functional trait variation and topography or microhabitat in this region. The topography of northern central Florida forms an ecological gradient where small elevation changes cause shifts in water availability (Brown, Stone & Carlisle, 1990). *Quercus* species in this area occupy habitats that span xeric sandhills to mesic river edges and ravines (Myers, 1992; Kurz & Godfrey, 1962), resulting in selective pressures that shape functional traits and adaptation to hydrological regimes (Cavender-Bares et al., 2004b;Cavender-Bares & Holbrook, 2001; Reich et al., 2003). Evolutionary divergence among Virentes, *Quercus virginiana, Q. geminata,* and *Q. minima,* has also been found despite complex introgression (Eaton et al., 2015; Cavender-Bares et al., 2015). These species maintain reproductive isolation through phenology, niche specialization, and habitat preference(Cavender-Bares et al., 2004a, b; Cavender-Bares & Pahlich, 2009). Full spectral analysis reveals strong adaptive divergence among these species, especially in the near-infrared (NIR, 700–800 nm) and visible (VIS, 616, 688 nm) regions, indicating differences in foliar structure and pigments associated with water availability (Ourcival and Rambal 1990, references). Although *Q. geminata* and *Q. minima* show little differentiation in the shortwave infrared (SWIR) and do not differ significantly in most of the structural traits (with the exception of cellulose, Fig 4a), both differ markedly from *Q. virginiana*. Lower lignin concentrations in *Q. virginiana* compared to *Q. geminata* and *Q. minima* as well as reduced LMA and lower leaf thickness, are traits differences associated with decreased sclerophylly expected in more mesic habitats (Sancho-Knapik et al., 2021; Alonso-Forn et al., 2020, 2023).

Phenotypic clustering across regions is associated with shared ancestry, historical geneflow and climatic similarity

Unsupervised clustering analyses (GENELAND) based on spectrally predicted traits place *Q. oleoides* (Central America populations), *Q. sagraeana* (Cuba), and *Q. virginiana, Q. geminata,* and *Q. minima* (Florida) into the same cluster (Fig. 2b).

Phenotypic clustering among distinct species can be a consequence of shared ancestry or historical gene flow followed by vicariance. The origin of *Q. sagraeana*, an allopatric Cuban endemic has been long investigated: one hypothesis suggests migration from Florida (Santiago-Valentín & Olmstead, 2004; Graham, 2010; Gugger et al., 2013), while another supports a Central American origin (Muller, 1955; Eaton et al., 2015). Genome-wide RADseq analyses reveal *Q. sagraeana* as a sister species to *Q. oleoides*, supporting a Central American origin with later introgression from *Q. virginiana* and *Q. geminata* (Eaton et al., 2015). Despite restricted gene flow—especially with current high sea levels (Gugger et al., 2013)—the lack of clear phenotypic structure may reflect retained ancestral polymorphisms. Phenotypic clustering may also be a consequence of similar selection pressures, given that these populations all occur in subtropical humid environments.

In contrast to clustered phenotypes associated with shared ancestry, historical geneflow and climatic similarity, populations of the allopatric sister species *Q. brandegeei* and *Q. fusiformis* show geographically structured phenotypes that align with genetic differences. *Q. fusiformis* and *Q. brandegeei* diverged about 5.2 Ma. *Q. brandegeei,* which occurs in an isolated region of southern Baja California's desert (Cavender-Bares et al. 2015) exhibits traits more suited to aridity, such as smaller and thicker leaves with higher LMA, compared to its sister species. These findings highlight ecological divergence driven by environmental selection between contrasting climates, with pRDA supporting spectral trait differentiation, especially in the NIR region Fig 5.

Collectively, the patterns reveal different evolutionary processes: contrasting environmental pressures and microhabitat-level differentiation drive phenotypic divergence, while shared history and ancestral gene flow, coupled with shared environments, promote phenotypic clustering. Plasticity is also likely to contribute to the

patterns of phenotypic variation we observe in wild populations, given the general pattern of greater P_{ST} - F_{ST} differences in wild populations compared to common garden populations (Fig. S6).

Phylogenetic history, signal, and environmental influences on phenotypic variation

In *Quercus* section *Virentes*, species distributed across temperate and subtropical climates exhibit significant phenotypic variation influenced by both phylogenetic history and environmental pressures. Using phylogenetic generalized least squares (PGLS) that included several populations per species, we found that species from colder latitudes, such as *Q. fusiformis* and *Q. virginiana*, exhibited lower lignin concentrations in their leaves. This same trend emerges when examining variation at the level of individuals. Individuals across the Virentes showed an increase in soluble sugars and hemicellulose in colder climates, consistent with findings in other plant species, including *Arabidopsis* (Panter et al., 2019; Kutsuno et al., 2022). These chemical changes found in populations from colder climates–lower lignin concentrations coupled with higher cellulose and soluble cell constituents–may be associated with biochemical modifications to the cell wall that enhance freezing tolerance by stabilizing cell structures and facilitating water movement during freeze-thaw cycles (Kutsuno et al., 2022). Lower lignin levels are linked to cell permeability, which facilitates water outflow and ice formation in extracellular spaces without damaging cells (Yamada et al. 2002, Domon et al.2013, Cass et al. 2015). The cell wall plays a crucial role in protecting the plasma membrane from extracellular freezing damage, as it serves as the primary site of ice crystal formation (Panter et al., 2020). In *Arabidopsis*, alterations in the pectin cross-link structure, lignin biosynthesis (Huang et al. 2010), and modifications in hemicellulose composition have been shown to affect basal freezing tolerance (Panter et al., 2019; Shi et al., 2014). Lignins and other phenolic compounds can also act as defense agents affecting leaf optical properties in the SWIR2 region. (Li et al. 2023, Czyż et al. 2020).

Previous studies show that minimum temperature of the coldest month predicts freezing tolerance and cold acclimation capacity in live oaks. In common garden experiments, species from temperate latitudes *Q. virginiana, Q. geminata* and *Q.*

fusiformis, demonstrated the ability to increase freezing tolerance in response to chilling growth temperatures, contrary to tropical species like *Q. oleoides* (Koehler et al., 2012; Cavender-Bares, 2007). The ability to cold acclimate and express higher freezing tolerance under temperate conditions was associated with less competitive growth rates under tropical (non-stressed) growth conditions (Koehler et al., 2012).

Across individuals significant variation in Anthocyanin Reflectance Index (ARI) is evident within *Q. virginiana* and *Q. oleoides* (Fig. 6b), reflecting their broad geographic and climatic ranges. Within *Q. virginiana*, the highest levels of anthocyanins occur in the coldest climates, but the overall trend across the Virentes is one of increasing anthocyanin content in warmer, more tropical regions. Accumulation of anthocyanins can contribute to photoprotective effects under cold conditions but can also deter herbivores (Gould 2004). Under low temperatures, anthocyanins have been shown to mitigate photodamage by intercepting light or neutralizing reactive oxygen species (Pietrini et al. 2002; Gould 2004; Hughes et al. 2012). Ramírez-Valiente et al. (2015) reported that anthocyanin levels in immature leaves of Virentes species and populations increased in response to seasonal low temperature stress but that that across all Virentes populations, those from tropical regions exhibited higher anthocyanin levels than those from temperate regions. The latter result points to the role that anthocyanins play in defense against herbivores.

Herbivore pressure is well known to increase at tropical latitudes (e.g., Coley and Barone 1996, Salazar and Marquis 2012, Tang et al. 2023), and defense chemistry has been shown to increase in more tropical regions in oaks (Pearse and Hipp 2012). In warmer climates, higher anthocyanin concentrations have been linked to reduced herbivory, including in *Q. robur* (Valdes-Correcher et al. 2025). Our findings of increasing anthocyanin levels in tropical regions indicate that anthocyanins play a greater role in protection against herbivores than in cold tolerance in the Virentes. The coupled increase in both anthocyanins and lignin concentrations in tropical regions (Fig. 6d,f) indicate that both may be important elements in defense against herbivores.

We also see a significant trend in increased anthocyanin levels with mean annual precipitation (Fig. 7). *Quercus brandegeei* and *Q. fusiformis*, which experience significant seasonal drought, have lower anthocyanins than species from more mesic environments. Ramírez-Valiente et al. (2015) reported that anthocyanin accumulation was more pronounced in mesic ecotypes under drought conditions, emphasizing that anthocyanins play a key role in mitigating photodamage through their antioxidant and light-filtering properties in environments with higher water availability because other photoprotective involving xanthophyll cycle mechanisms are absent. They concluded that mesic populations are more reliant on anthocyanin-based photoprotection, Interestingly, our results also show that *Q. brandegeei* and *Q. fusiformis* exhibited higher chlorophyll indices despite their lower anthocyanin levels, suggesting that xeric species may prioritize maintaining high photosynthetic efficiency during a shorter active growing season. The observed trade-off between anthocyanin accumulation and chlorophyll concentration in our study (Fig. 7b, d) suggests an investment in high photosynthesis in shorter active growing seasons in xeric populations (Ramírez-Valiente and Cavender-Bares 2017), on the one hand, and increased anthocyanin-based photoprotection and/or defense chemistry in more mesic conditions. We also found that individuals from drier sites showed somewhat higher LMA (Fig. 7f). Previous studies have shown that xeric ecotypes in species like *Fagus sylvatica* tend to exhibit reduced anthocyanin levels while relying on morphological adaptations such as increased LMA and higher trichome density to enhance drought tolerance (Camarero et al., 2012; García-Plazaola & Becerril, 2000a). These structural and physiological strategies, combined with higher chlorophyll levels, may support efficient energy use and photosynthetic function in regions prone to drought stress. Our findings highlight the multifunctional role of anthocyanins (Gould 2004) and the importance of environmental context in shaping leaf morphology, pigment dynamics, and carbon compounds across *Quercus* species.

Balancing evolutionary constraint and adaptive divergence in spectral traits

The strong phylogenetic signal observed in the visible spectrum (Table S6) supports the hypothesis that spectral traits associated with photosynthesis, such as pigment concentrations, are evolutionarily conserved (Meireles et al., 2020). Although closely related species such as *Q. geminata* and *Q. minima* exhibit minimal differences in the visible spectrum, these differences are still ecologically relevant, particularly in specific spectral bands associated with environmental variables. When pairwise comparisons include *Q. virginiana*, these differences become more pronounced, reflecting a dual influence of shared ancestry and local adaptation. This suggests that while the visible spectrum captures phylogenetically conserved traits, it also reveals adaptive shifts driven by local environmental pressures (Liu et al., 2015). This interplay highlights the potential for evolutionary plasticity within a conserved spectral framework. The significant environmental contribution to visible spectrum variation—particularly the influence of the topographic wetness index (PVE% supporting information pRDA_Results.xlsx)—underscores the importance of microhabitat-level adaptive divergence in closely related, sympatric species. Given the sympatry of these species, strong environmental contributions, such as those driven by differences in topographic wetness index levels, to variation in the visible spectrum may reflect niche differentiation, enabling co-occurrence by reducing competition for water resources. Variation in the visible spectrum may reflect niche differentiation, allowing co-occurrence by reducing competition. This pattern is consistent with the hypothesis that recent, sympatric divergence is often driven by strong ecological pressures acting on traits with direct functional relevance to the environment (Arnegard et al., 2015). Even for traits with high phylogenetic signals, species can fine-tune or modify their traits within the limits of their evolutionary potential to adapt to contrasting environments.

Using common gardens to decipher the genetic basis of spectral traits and the role of plasticity

The spatial autocorrelation of environmental variables can lead to geographic and neutral differences between populations correlating with environmental conditions, thereby masking adaptive evolutionary processes (Reznick & Ghalambor, 2001; Prentis

et al., 2008). Consequently, the use of common gardens is crucial for clarifying the roles of different evolutionary forces in populations, as they allow for the assessment of phenotypic variation under controlled environmental conditions. Our examination of spectral features and spectrally derived traits in common garden experiments compared to wild populations increases our understanding of how trait selection varies depending on interaction with the environment. Sympatric (*Q. geminata* and *Q. virginiana*) or parapatric (*Q. fusiformis* and *Q. oleoides*) species may undergo specialization reinforcing species boundaries and influencing the evolution of adaptive traits. While we did not explicitly test for plasticity in this study, the comparison of P_{ST} - F_{ST} in the same populations grown in a common garden or in the wild reveals that there is a significant component of phenotypic variation attributable to plasticity. However, we found that individuals of sympatric or parapatric species show similar values of P_{ST} and significant adaptive divergence, in both common gardens and wild populations, indicating the fixation of traits in these populations, perhaps as a consequence of species interactions. This phenomenon was not observed when comparing allopatric or widely distributed species, where plasticity may be more important for persistence. Plasticity within widely distributed species may confer an advantage in adapting to diverse climates, as seen in populations of *Q. oleoides*. Long-lived trees that inhabit extensive climatic ranges face the challenge of coping with highly variable climatic conditions and dynamic selective pressures on growth and stress tolerance across different space-temporal scales (Meireles et al. 2017).

Conclusions

Our research demonstrates the efficacy of using leaf-level spectra and spectrally derived traits to elucidate the roles of adaptive divergence and phenotypic plasticity in the persistence of populations of long-lived organisms. Understanding adaptive divergence among species is a long-standing challenge in evolutionary biology due, in part, to the complexity of quantifying phenotypic traits for large sample sizes. Our findings show that isolation-by-environment, rather than geographic proximity, shapes phenotypic divergence among sympatric *Quercus* species, highlighting the importance of adaptive strategies in sustaining regional coexistence among species that occupy

diverse microhabitats. In sympatric species, niche specialization plays a key role in diversification resulting in trait divergence and speciation and contributing to landscape-level coexistence among closely related species. While we found evidence for adaptive divergence in live oak species, this divergence occurred alongside substantial gene flow within local populations. This interplay between neutral and selective forces highlights the complexity of adaptation, where phenotypic differentiation arises together with genetic connectivity. Taken together, these patterns underscore the critical roles of natural selection, genetic variation and phenotypic plasticity in ensuring the long-term persistence of these species (Teixeira & Huber, 2021; Cavender-Bares, 2019). Integration of data that capture a high degree of phenotypic and genetic variation represented within and across divergent lineages is critical to deciphering the influence of genetic and environmental factors and their interactions on phenotypes. The approach we used integrates disparate fields within evolutionary and ecology biology to elucidate how interactions between landscapes and species biological traits drive biodiversity patterns.

Acknowledgements

We are grateful for the support provided by technicians Isaiah Clark and Allison Scott for their assistance in the spectral data acquisition. Funding: This study is supported by the National Science Foundation NSF DBI: 2021898 as part of the Advancing Spectral Biology in Changing Environments to understand Diversity an NSF Biology Integration Institute.

All authors contributed intellectually. J.C.B. conceived of and designed the study; collected the specimens;managed the laboratory; M.S.H.L designed the data analysis, worked on genetic, phenotypic analysis, contributed to common garden leaf scanning, designed Figures and wrote the manuscript. J.A.G.Q designed spectral trait prediction models and contributed with Figure editing, A.G.R. was involved in phenotypic analysis. M.S.H.L and J.C.B wrote the manuscript and all authors edited it.

Supporting information

Additional supporting information may be found in the online version of this article.

Table S1. Previous studies from which the genetic data were obtained and from which specimens were collected and measured for spectral data

Table S2. Pairwise values of genetic differentiation in nSSR a) among species and b) genetic groups of section *Virentes* and c) Among genetic groups of *Q.oleoides*.

Table S3. Summary statistics and SD of PLSR_DA model from pressed-leaf spectra restricted to 1400-2400 nm. *Q. oleoides*

Table S4. Summary statistics for the PLSR calibration and validation models for each leaf trait.

Table S5 Values of phenotypic P_{ST} obtained from spectrally derived traits, values for spectral bands identified by the variable importance of projection (VIP) metric, and comparisons of P_{ST} with F_{ST} and their confidence intervals.

Fig. S1 Distribution of the seven species in section *Virentes* based on species occurrence.

Fig. S2 Internal validation results for LMA, Thickness, Solubles, Hemicellulose, Cellulose and Lignin predicted from dry spectra.

Fig. S3 The variable importance in the projection (VIP) metric was calculated based on dry sample spectral data models for six traits.

Fig. S4 Spatial distribution of genetic and phenotypic variation in *Q. oleoides*.

Fig. S5. P_{ST} comparisons among species using: a) spectrally predicted traits (LMA: Leaf mass area; THI: thickness; SOL: solubles; HEM: hemicellulose; CEL: cellulose; LIG: lignin); b) spectral bands within the visible (VIS), near infrared (NIR) and short-wave infrared (SWIR) region with high importance (i.e., Variable Importance of Projection (VIP) in discriminating species using wavelet spectra

Fig S6. Spectrally predicted traits and selected wavelength relationships under different environmental conditions.

Fig. S7 P_{ST} vs F_{ST} estimates and 95% confidence intervals from quantitative predicted traits among wild and greenhouse individuals for four species in *Quercus* section *Virentes*.

Fig. S8 Phylogenetic signal detected in leaf spectra varies across wavelengths across .

Quercus Virentes species.

References

Adams, H. D., Zeppel, M. J. B., Anderegg, W. R. L., Hartmann, H., Landhäusser, S. M., Tissue, D. T., Huxman, T. E., Hudson, P. J., Franz, T. E., Allen, C. D., Anderegg, L. D. L., Barron-Gafford, G. A., Beerling, D. J., Breshears, D. D., Brodribb, T. J., Bugmann, H., Cobb, R. C., Collins, A. D., Dickman, L. T., … McDowell, N. G. (2017). A multi-species synthesis of physiological mechanisms in drought-induced tree mortality. Nature Ecology & Evolution, 1(9), 1285–1291.

Alonso-Forn, D., Sancho-Knapik, D., Fariñas, M. D., Nadal, M., Martín-Sánchez, R., Ferrio, J. P., de Dios, V. R., Peguero-Pina, J. J., Onoda, Y., Cavender-Bares, J., Arenas, T. G. Á., & Gil-Pelegrín, E. (2023). Disentangling leaf structural and material properties in relationship to their anatomical and chemical compositional traits in oaks (Quercus L.). Annals of Botany, 131(5), 789–800.

Alonso-Forn, D., Sancho-Knapik, D., Ferrio, J. P., Peguero-Pina, J. J., Bueno, A., Onoda, Y., Cavender-Bares, J., Niinemets, Ü., Jansen, S., Riederer, M., Cornelissen, J. H. C., Chai, Y., & Gil-Pelegrín, E. (2020). Revisiting the Functional Basis of Sclerophylly Within the Leaf Economics Spectrum of Oaks: Different Roads to Rome. Current Forestry Reports, 6(4), 260–281.

Anderegg, L. D. L. (2023). Why can't we predict traits from the environment? New Phytologist, 237(6), 1998–2004.

Arnegard ME, McGee MD, Matthews B, Marchinko KB, Conte GL, Kabir S, Bedford N, Bergek S, Chan YF, Jones FC, Kingsley DM, Peichel CL, Schluter D. Genetics of ecological divergence during speciation. Nature. 2014 Jul 17;511(7509):307-11. doi: 10.1038/nature13301.

Asner, G. P., & Martin, R. E. (2011). Canopy phylogenetic, chemical and spectral assembly in a lowland Amazonian Forest. New Phytologist, 189(4), 999–1012.

Asner, G. P., & Martin, R. E. (2016). Spectranomics: Emerging science and conservation opportunities at the interface of biodiversity and remote sensing. Global Ecology and Conservation, 8, 212–219.

Bernatchez, L. (2016). On the maintenance of genetic variation and adaptation to environmental change: considerations from population genomics in fishes. Journal of Fish Biology, 89(6), 2519–2556.

Boucher, D. H., Hansen, M., Risch, S., Vandemeer, J. H., & Janzen, D. H. (1983). Costa Rican natural history.

Brommer, J. E. (2011). Whither Pst? The approximation of Qst by Pst in evolutionary and conservation biology. Journal of Evolutionary Biology, 24(6), 1160–1168.

Cass, C. L., Peraldi, A., Dowd, P. F., Mottiar, Y., Santoro, N., Karlen, S. D., ... & Sedbrook, J. C. (2015). Effects of PHENYLALANINE AMMONIA LYASE

(PAL) knockdown on cell wall composition, biomass digestibility, and biotic and abiotic stress responses in Brachypodium. Journal of Experimental Botany, 66(14), 4317-4335.

Cavender-Bares J, Kitajima K, Bazzaz FA. (2004b). Multiple trait associations in relation to habitat differentiation among 17 Floridian oak species. Ecological Monographs 74: 635–662.

Cavender-Bares, J. (2007). Chilling and freezing stress in live oaks (Quercus subsection Virentes): intra- and inter-specific variation in PS II sensitivity corresponds to latitude of origin. Photosynthesis Research, 94(2–3), 437.

Cavender-Bares, J. (2019). Diversification, adaptation, and community assembly of the American oaks (Quercus), a model clade for integrating ecology and evolution. New Phytologist, 221(2), 669–692.

Cavender-Bares, J. M., & Pahlich, A. (2009). Molecular, morphological, and ecological niche differentiation of sympatric sister oak species, Quercus virginiana and Q. geminata (Fagaceae). American Journal of Botany, 96 9, 1690–1702.

Cavender-Bares, J., Ackerly, D. D., Baum, D. A., & Bazzaz, F. A. (2004a). Phylogenetic Overdispersion in Floridian Oak Communities. The American Naturalist, 163(6), 823–843.

Cavender-Bares, J., Gamon, J. A., Hobbie, S. E., Madritch, M. D., Meireles, J. E., Schweiger, A. K., & Townsend, P. A. (2017). Harnessing plant spectra to integrate the biodiversity sciences across biological and spatial scales. American Journal of Botany, 104(7), 966–969.

Cavender-Bares, J., González-Rodríguez, A., Eaton, D. A. R., Hipp, A. A. L., Beulke, A., & Manos, P. S. (2015). Phylogeny and biogeography of the American live oaks (Quercus subsection Virentes):a genomic and population genetics approach. Molecular Ecology, 24(14), 3668–3687.

Cavender-Bares, J., Gonzalez-Rodriguez, A., Pahlich, A., Koehler, K., & Deacon, N. (2011). Phylogeography and climatic niche evolution in live oaks (Quercus series Virentes) from the tropics to the temperate zone. Journal of Biogeography, 38(5), 962–981. https://doi.org/10.1111/j.1365-2699.2010.02451.x

Cavender-Bares, J., Meireles, J., Couture, J., Kaproth, M., Kingdon, C., Singh, A., Serbin, S., Center, A., Zuniga, E., Pilz, G., & Townsend, P. (2016). Associations of Leaf Spectra with Genetic and Phylogenetic Variation in Oaks: Prospects for Remote Detection of Biodiversity. Remote Sensing, 8(3), 221.

Chan, L. M., Brown, J. L., & Yoder, A. D. (2011). Integrating statistical genetic and geospatial methods brings new power to phylogeography. Molecular Phylogenetics and Evolution, 59(2), 523–537.

Chevallier, S., Bertrand, D., Kohler, A., & Courcoux, P. (2006). Application of PLS-DA in multivariate image analysis. Journal of Chemometrics, 20(5), 221–229.

Clé C, Hill LM, Niggeweg R, Martin CR, Guisez Y, Prinsen E, Jansen MA. Modulation of chlorogenic acid biosynthesis in Solanum lycopersicum;

consequences for phenolic accumulation and UV-tolerance. Phytochem. 2008;69(11):2149–56.
Constantine, W., & Percival, D. (2016). Wmtsa: Wavelet Methods for Time Series Analysis. https://cran.r-project.org/web/packages/wmtsa/index.html,.
Crispo, E., Bentzen, P., Reznick, D. N., Kinnison, M. T., & Hendry, A. P. (2006). The relative influence of natural selection and geography on gene flow in guppies. Molecular Ecology, 15(1), 49–62.
Da Silva, S. B., & Da Silva, A. (2018). Pstat: An R Package to Assess Population Differentiation in Phenotypic Traits. R J., 10(1), 447. URL https://CRAN.R-project.org/package=Pstat
Deacon, N. J., & Cavender-Bares, J. (2015). Limited Pollen Dispersal Contributes to Population Genetic Structure but Not Local Adaptation in Quercus oleoides Forests of Costa Rica. PLOS ONE, 10(9), e0138783.
Domon, J. M., Baldwin, L., Acket, S., Caudeville, E., Arnoult, S., Zub, H., ... & Rayon, C. (2013). Cell wall compositional modifications of Miscanthus ecotypes in response to cold acclimation. Phytochemistry, 85, 51-61.
Donovan, L. A., Maherali, H., Caruso, C. M., Huber, H., & de Kroon, H. (2011). The evolution of the worldwide leaf economics spectrum. Trends in Ecology & Evolution, 26(2), 88–95.
Eaton, D. A. R., Hipp, A. L., González-Rodríguez, A., & Cavender-Bares, J. (2015). Historical introgression among the American live oaks and the comparative nature of tests for introgression. Evolution, 69(10), 2587–2601.
Edwards, S. V., Shultz, A. J., & Campbell-Staton, S. C. (2015). Next-generation sequencing and the expanding domain of phylogeography. Folia Zoologica, 64(3), 187–206.
Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. International Journal of Climatology, 37(12), 4302–4315.
Forester, B. R., Jones, M. R., Joost, S., Landguth, E. L., & Lasky, J. R. (2016). Detecting spatial genetic signatures of local adaptation in heterogeneous landscapes. Molecular Ecology, 25(1), 104–120.
Garrick, R. C., Bonatelli, I. A. S., Hyseni, C., Morales, A., Pelletier, T. A., Perez, M. F., Rice, E., Satler, J. D., Symula, R. E., Thomé, M. T. C., & Carstens, B. C. (2015). The evolution of phylogeographic data sets. Molecular Ecology, 24(6), 1164–1171.
Gitelson, Anatoly A., Yuri Gritz, and Mark N. Merzlyak. "Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves." Journal of plant physiology 160.3 (2003): 271-282.
Goudet, J. (2005). <scp>hierfstat</scp> , a package for <scp>r</scp> to compute and test hierarchical F -statistics. Molecular Ecology Notes, 5(1), 184–186.
Gould KS. 2004. Nature's Swiss Army Knife: The Diverse Protective Roles of Anthocyanins in Leaves. J Biomed Biotechnol 2004(5): 314-320.

Graham, S. A. (2010). Revision of the Caribbean Genus Ginoria (Lythraceae), Including Haitia From Hispaniola 1. Annals of the Missouri Botanical Garden, 97(1), 34–90. https://doi.org/10.3417/2007028

Gugger, P. F., & Cavender-Bares, J. (2013). Molecular and morphological support for a Florida origin of the Cuban oak. Journal of Biogeography, 40(4), 632–645.

Guillot, G., & Santos, F. (2009). A computer program to simulate multilocus genotype data with spatially autocorrelated allele frequencies. Molecular Ecology Resources, 9(4), 1112–1120. https://doi.org/10.1111/j.1755-0998.2008.02496.x

Haldane, J. B. S. (1948). The theory of a cline. Journal of Genetics, 48(3), 277–284.

Hickerson, M. J., Carstens, B. C., Cavender-Bares, J., Crandall, K. A., Graham, C. H., Johnson, J. B., Rissler, L., Victoriano, P. F., & Yoder, A. D. (2010). Phylogeography's past, present, and future: 10 years after Avise, 2000. Molecular Phylogenetics and Evolution, 54(1), 291–301.

Hickerson, M. J., Dolman, G., & Moritz, C. (2005). Comparative phylogeographic summary statistics for testing simultaneous vicariance. Molecular Ecology, 15(1), 209–223.

Hill, W. G., & Kirkpatrick, M. (2010). What Animal Breeding Has Taught Us about Evolution. Annual Review of Ecology, Evolution, and Systematics, 41, 1–19.

Hipp AL, Manos PS, Hahn M, Avishai M, Bodénès C, Cavender-Bares J, Crowl AA, Deng M, Denk T, Fitz-Gibbon S, et al. 2020. Genomic landscape of the global oak phylogeny. New Phytologist 226(4): 1198-1212.

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

Hubisz, M. J., Falush, D., Stephens, M., & Pritchard, J. K. (2009). Inferring weak population structure with the assistance of sample group information. Molecular Ecology Resources, 9(5), 1322–1332.

Jakobsson, M., & Rosenberg, N. A. (2007). CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics, 23(14), 1801–1806.

Kanaka KK, Sukhija N, Goli RC, Singh S, Ganguly I, Dixit SP, Dash A, Malik AA. 2023. On the concepts and measures of diversity in the genomics era. Current Plant Biology 33: 100278.

Kaproth, M. A., Fredericksen, B. W., González-Rodríguez, A., Hipp, A. L., & Cavender-Bares, J. (2023). Drought response strategies are coupled with leaf habit in 35 evergreen and deciduous oak (Quercus) species across a climatic gradient in the Americas. New Phytologist, 239(3), 888–904.

Koehler, K., Center, A., & Cavender-Bares, J. (2012). Evidence for a freezing tolerance–growth rate trade-off in the live oaks (Quercus series Virentes) across the tropical–temperate divide. New Phytologist, 193(3), 730–744.
Kothari, S., & Schweiger, A. K. (2022). Plant spectra as integrative measures of plant phenotypes. Journal of Ecology, 110(11), 2536–2554.
Kothari, S., Beauchamp-Rioux, R., Blanchard, F., Crofts, A. L., Girard, A., Guilbeault-Mayers, X., Hacker, P. W., Pardo, J., Schweiger, A. K., Demers-Thibeault, S., Bruneau, A., Coops, N. C., Kalacska, M., Vellend, M., & Laliberté, E. (2023). Predicting leaf traits across functional groups using reflectance spectroscopy. New Phytologist, 238(2), 549–566.
Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., & Engelhardt, A. (2020). Caret: classification and regression training. R package version 6.0-86. 2020. URL http://CRAN. R-project. org/package= caret.
Kurz, H., & Godfrey, R. K. (1962). Trees of northern Florida. University of Florida, Gainesville.
Kutsuno, T., Chowhan, S., Kotake, T., & Takahashi, D. (2023). Temporal cell wall changes during cold acclimation and deacclimation and their potential involvement in freezing tolerance and growth. Physiologia Plantarum, 175(1), e13837.
Lande, R. (1976). Natural selection and random genetic drift in phenotypic evolution. Evolution, 30(2), 314–334.
Lande, R. (1992). neutral theory of quantitative genetic variance in an island model with local extinction and colonization. Evolution, 46(2), 381–389.
Lang, P. L. M., Willems, F. M., Scheepens, J. F., Burbano, H. A., & Bossdorf, O. (2019). Using herbaria to study global environmental change. New Phytologist, 221(1), 110–122.
Leinonen, T., Cano, J. M., Mäkinen, H., & Merilä, J. (2006). Contrasting patterns of body shape and neutral genetic divergence in marine and lake populations of three spine sticklebacks. Journal of Evolutionary Biology, 19(6), 1803–1812.
Leinonen, T., McCairns, R. J. S., O'Hara, R. B., & Merilä, J. (2013). QST–FST comparisons: evolutionary and ecological insights from genomic heterogeneity. Nature Reviews Genetics, 14(3), 179–190.
Li, C., Czyż, E. A., Halitschke, R., Baldwin, I. T., Schaepman, M. E., & Schuman, M. C. (2023). Evaluating potential of leaf reflectance spectra to monitor plant genetic variation. Plant Methods, 19(1), 108.
Li, X., Wei, Z., Peng, F., Liu, J., & Han, G. (2023). Non-destructive prediction and visualization of anthocyanin content in mulberry fruits using hyperspectral imaging. Frontiers in Plant Science, 14, 1137198.
Liu, B., Wang, X., Cao, Y., Arora, R., Zhou, H., & Xia, Y. (2020). Factors affecting freezing tolerance: a comparative transcriptomics study between field and artificial cold acclimations in overwintering evergreens. The Plant Journal, 103(6), 2279–2300.
Liu, H., Xu, Q., He, P. et al. (2015). Strong phylogenetic signals and phylogenetic niche conservatism in ecophysiological traits across divergent lineages of Magnoliaceae. Sci Rep 5, 12246 (

Liu, H., Ye, Q., Simpson, K. J., Cui, E., & Xia, J. (2022). Can evolutionary history predict plant plastic responses to climate change? New Phytologist, 235(3), 1260–1271.

Luikart, G., England, P. R., Tallmon, D., Jordan, S., & Taberlet, P. (2003). The power and promise of population genomics: from genotyping to genome typing. Nature Reviews Genetics, 4(12), 981–994.

Lynch, M. (1990). The Rate of Morphological Evolution in Mammals from the Standpoint of the Neutral Expectation. The American Naturalist, 136(6), 727–741.

Mackay, T. F. C., Stone, E. A., & Ayroles, J. F. (2009). The genetics of quantitative traits: challenges and prospects. Nature Reviews Genetics, 10(8), 565–577.

Manel, S., Schwartz, M. K., Luikart, G., & Taberlet, P. (2003). Landscape genetics: combining landscape ecology and population genetics. Trends in Ecology & Evolution, 18(4), 189–197.

Manos PS, Doyle JJ, Nixon KC (1999) Phylogeny, biogeography, and processes of molecular differentiation in Quercus

Meireles, J. E., Cavender-Bares, J., Townsend, P. A., Ustin, S., Gamon, J. A., Schweiger, A. K., Schaepman, M. E., Asner, G. P., Martin, R. E., Singh, A., Schrodt, F., Chlus, A., & O'Meara, B. C. (2020). Leaf reflectance spectra capture the evolutionary history of seed plants. New Phytologist, 228(2), 485–493.

Meireles, J. E., Schweiger, A. K., & Cavender-Bares, J. M. (2017). Spectrolab: Class and Methods for Hyperspectral Data. R package version 0.0. 2. (0.0. 2.).

Merilä, J., & Crnokrak, P. (2001). Comparison of genetic differentiation at marker loci and quantitative traits. Journal of Evolutionary Biology, 14(6), 892–903.

Muller, C. H. (1955) "The origin of Quercus on Cuba." Revista de la Sociedad Cubana de Botánica 7 (41-47.

Muller, C. H. (1961a). The Live Oaks of the Series Virentes. American Midland Naturalist, 65(1),

Muller, C. H. (1961b). The origin of Quercus fusiformis Small1. Journal of the Linnean Society of London, Botany, 58(370), 1–12.

Myers, R. L. (Ed.). (1992). Ecosystems of Florida. University of Central Florida Press.

Myers, R. L. 1990. Scrub and High Pine. Pages 150-193 in R. L. Myers and J. J. Ewel, editors. Ecosystems of Florida. University of Central Florida Press, Orlando.

Nardini, A. (2022). Hard and tough: the coordination between leaf mechanical resistance and drought tolerance. Flora, 288, 152023

Nixon KC, Muller CH (1997) Quercus Linnaeus sect. Quercus White oaks. In: Flora of North America North of Mexico (eds Committee FoNAE), pp. 436–506. Oxford University Press, New York City, New York.

Nixon KC. (1985). A Biosystematic Study of Quercus Series Virentes (the live oaks) with Phylogenetic Analyses of Fagales, Fagaceae and Quercus.University of Texas.

Ourcival, J. M., Joffre, R., & Rambal, S. J. N. P. (1999). Exploring the relationships between reflectance and anatomical and biochemical properties in Quercus ilex leaves. The New Phytologist, 143(2), 351-364.

Pérez-Harguindeguy, N., Díaz, S., Garnier, E., Lavorel, S., Poorter, H., Jaureguiberry, P., Bret-Harte, M. S., Cornwell, W. K., Craine, J. M., Gurvich, D. E., Urcelay, C., Veneklaas, E. J., Reich, P. B., Poorter, L., Wright, I. J., Ray, P., Enrico, L., Pausas, J. G., de Vos, A. C., … Cornelissen, J. H. C. (2016). Corrigendum to: New handbook for standardised measurement of plant functional traits worldwide. Australian Journal of Botany, 64(8), 715.

Pfennig, D. W. (2021). Phenotypic Plasticity & EvolutionCauses, Consequences, Controversies, CRC Press.

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of Population Structure Using Multilocus Genotype Data. Genetics, 155(2), 945–959.

Quemada, C., Pérez-Escudero, J. M., Gonzalo, R., Ederra, I., Santesteban, L. G., Torres, N., & Iriarte, J. C. (2021). Remote sensing for plant water content monitoring: A review. Remote Sensing, 13(11), 2088.

Raeymaekers, J. A. M., Chaturvedi, A., Hablützel, P. I., Verdonck, I., Hellemans, B., Maes, G. E., De Meester, L., & Volckaert, F. A. M. (2017). Adaptive and non-adaptive divergence in a common landscape. Nature Communications, 8(1), 267.

Ramírez-Valiente, J. A., & Cavender-Bares, J. (2017). Evolutionary trade-offs between drought resistance mechanisms across a precipitation gradient in a seasonally dry tropical oak (Quercus oleoides). Tree Physiology, 37(7), 889–901.

Ramirez-Valiente, J. A., Koehler, K., & Cavender-Bares, J. (2015). Climatic origins predict variation in photoprotective leaf pigments in response to drought and low temperatures in live oaks (Quercus series Virentes). Tree Physiology, 35(5), 521–534.

Revell, L. J. (2012). phytools: An R package for phylogenetic comparative biology (and other things). Methods in Ecology and Evolution, 3(2), 217-223.

Sancho-Knapik, D., Escudero, A., Mediavilla, S., Scoffoni, C., Zailaa, J., Cavender-Bares, J., Álvarez-Arenas, T. G., Molins, A., Alonso-Forn, D., Ferrio, J. P., Peguero-Pina, J. J., & Gil-Pelegrín, E. (2021). Deciduous and evergreen oaks show contrasting adaptive responses in leaf mass per area across environments. New Phytologist, 230(2), 521–534.

Santiago–Valentin, E., & Olmstead, R. G. (2004). Historical biogeography of Caribbean plants: introduction to current knowledge and possibilities from a phylogenetic perspective. TAXON, 53(2), 299–319.

Scheiner, S. M. (1993). Genetics and Evolution of Phenotypic Plasticity. Annual Review of Ecology and Systematics, 24, 35–68.

Sexton, J. P., Hangartner, S. B., & Hoffmann, A. A. (2014). Genetic isolation by environment or distance: which pattern of gene flow is most common? Evolution, 68(1), 1–15.
Silva, S. B. da, & Silva, A. da. (2018). Pstat: An R Package to Assess Population Differentiation in Phenotypic Traits. The R Journal, 10(1), 447.
Slatkin, M. (1987). geneflow and the Geographic Structure of Natural Populations. Science, 236(4803), 787–792.
Spitze, K. (1993). Population structure in Daphnia obtusa: quantitative genetic and allozymic variation. Genetics, 135(2), 367–374.
Storfer, A., Murphy, M. A., Evans, J. S., Goldberg, C. S., Robinson, S., Spear, S. F., Dezzani, R., Delmelle, E., Vierling, L., & Waits, L. P. (2007). Putting the 'landscape' in landscape genetics. Heredity, 98(3), 128–142.
Sultan, S. E. (1987). Evolutionary Implications of Phenotypic Plasticity in Plants. In Evolutionary Biology (pp. 127–178). Springer US.
Sultan, S. E. (2000). Phenotypic plasticity for plant development, function and life history. Trends in Plant Science, 5(12), 537–542.
Svensson, E. I., Arnold, S. J., Bürger, R., Csilléry, K., Draghi, J., Henshaw, J. M., ... & Runemark, A. (2021). Correlational selection in the age of genomics. Nature ecology & evolution, 5(5), 562-573.
Teixeira, J. C., & Huber, C. D. (2021). The inflated significance of neutral genetic diversity in conservation genetics. Proceedings of the National Academy of Sciences, 118(10).
Ustin, S. L., & Gamon, J. A. (2010). Remote sensing of plant functional types. New Phytologist, 186(4), 795–816.
Valdés-Correcher, E., Kadiri, Y., Bourdin, A., Mrazova, A., Bălăcenoiu, F., Branco, M., ... & Castagneyrol, B. (2025). Effects of climate on leaf phenolics, insect herbivory, and their relationship in pedunculate oak (Quercus robur) across its geographic range in Europe. Oecologia, 207(4), 1-13.
Vogt, G. (2020). Disentangling the environmentally induced and stochastic developmental components of phenotypic variation. In Phenotypic Switching (pp. 207–251). Elsevier.
Wellenreuther, M., & Hansson, B. (2016). Detecting Polygenic Evolution: Problems, Pitfalls, and Promises. Trends in Genetics, 32(3), 155–164.
Westoby, M., & Wright, I. J. (2006). Land-plant ecology on the basis of functional traits. Trends in Ecology & Evolution, 21(5), 261–268.
White, D. M., Cavender-Bares, J., Davis, C. C., Guzmán Q., J. A., Kothari, S., Robles, J. M., & Meireles, J. E. 2025. Seeing herbaria in a new light: Leaf reflectance spectroscopy unlocks predictive trait and classification modeling in plant biodiversity collections. EcoEvoRxiv.
Wright, I. J., Reich, P. B., Westoby, M., Ackerly, D. D., Baruch, Z., Bongers, F., Cavender-Bares, J., Chapin, T., Cornelissen, J. H. C., Diemer, M., Flexas, J., Garnier, E., Groom, P. K., Gulias, J., Hikosaka, K., Lamont, B. B., Lee, T., Lee, W., Lusk, C., … Villar, R. (2004). The worldwide leaf economics spectrum. Nature, 428(6985), 821–827.

Wright, S. (1949). The genetical structure of populations. Annals of Eugenics, 15(1), 323–354.

Yamada, T., Kuroda, K., Jitsuyama, Y., Takezawa, D., Arakawa, K., & Fujikawa, S. (2002). Roles of the plasma membrane and the cell wall in the responses of plant cells to freezing. Planta, 215, 770-778.

Zamudio, K. R., Bell, R. C., & Mason, N. A. (2016). Phenotypes in phylogeography: Species' traits, environmental variation, and vertebrate diversification. Proceedings of the National Academy of Sciences, 113(29), 8041–8048.

***New Phytologist* Supporting Information**

Article title: **Foliar spectral signatures reveal adaptive divergence in live oaks (*Quercus* section *Virentes)* across species and environmental niches**

Authors: Mariana S. Hernández-Leal[1*], J. Antonio Guzmán Q.[1], Antonio González Rodríguez[2], Cavender-Bares[1*]

The following Supporting Information is available for this article:

Fig. S1 Distribution of the seven species in section *Virentes* based on species occurrence.

Table S1. Previous studies from which the genetic data were obtained and from which specimens were collected and measured for spectral data.

Methods S1 Environmental Characteristics of the *Virentes* Lineage

Methods S2 Bayesian clustering using STRUCTURE software.

Method S3 Continuous Wavelet Transform

Methods S4 PLSR modeling framework to predict leaf traits from dried-leaf reflectance spectra.

Methods S5 *Comparing phenotypic vs genotypic divergence.*

Method S6 ***Spatial and environmental drivers of population divergence***.

Table S2. a) Results of hierarchical analyses of molecular variance (AMOVA) for 7 species of the *Virentes* section in 64 populations based on eleven nSSRs loci. b) Pairwise F_{ST} values of genetic differentiation in nSSR a) among species and b) genetic groups of section *Virentes*.

Table S3. Summary statistics and SD of PLSR_DA model from pressed-leaf spectra restricted to 1400-2400 nm.

Table S4. Summary statistics for the PLSR calibration and validation models for each leaf trait.

Fig. S2 Internal validation results for LMA, thickness, solubles, Hemicellulose, cellulose andlLignin predicted from dry spectra.

Fig. S3 The variable importance in the projection (VIP) metric was calculated based on dry sample spectral data models for six traits.

Table S5 Summary statistics for the PGLS models for each Spectrally predicted traits and selected wavelengths SOL: solubles; LIG: lignin

x

Fig. S5. P_{ST} comparisons among species using a) spectrally predicted traits (LMA: Leaf mass area; THI: thickness; SOL: solubles; HEM: hemicellulose; CEL: cellulose; LIG: lignin); b) spectral bands within the visible (VIS), near infrared (NIR) and short-wave infrared (SWIR) region with high importance (i.e., Variable Importance of Projection (VIP) in discriminating species using wavelet spectra

Fig S6. Spectrally predicted traits and selected wavelength relationships under different environmental conditions.

Fig. S7 P_{ST} vs F_{ST} estimates and 95% confidence intervals from quantitative predicted traits among wild and greenhouse individuals for four species in *Quercus* section *Virentes*.

Fig. S8 *Phylogenetic* signal detected in leaf spectra varies across wavelengths across . *Quercus Virentes* species.

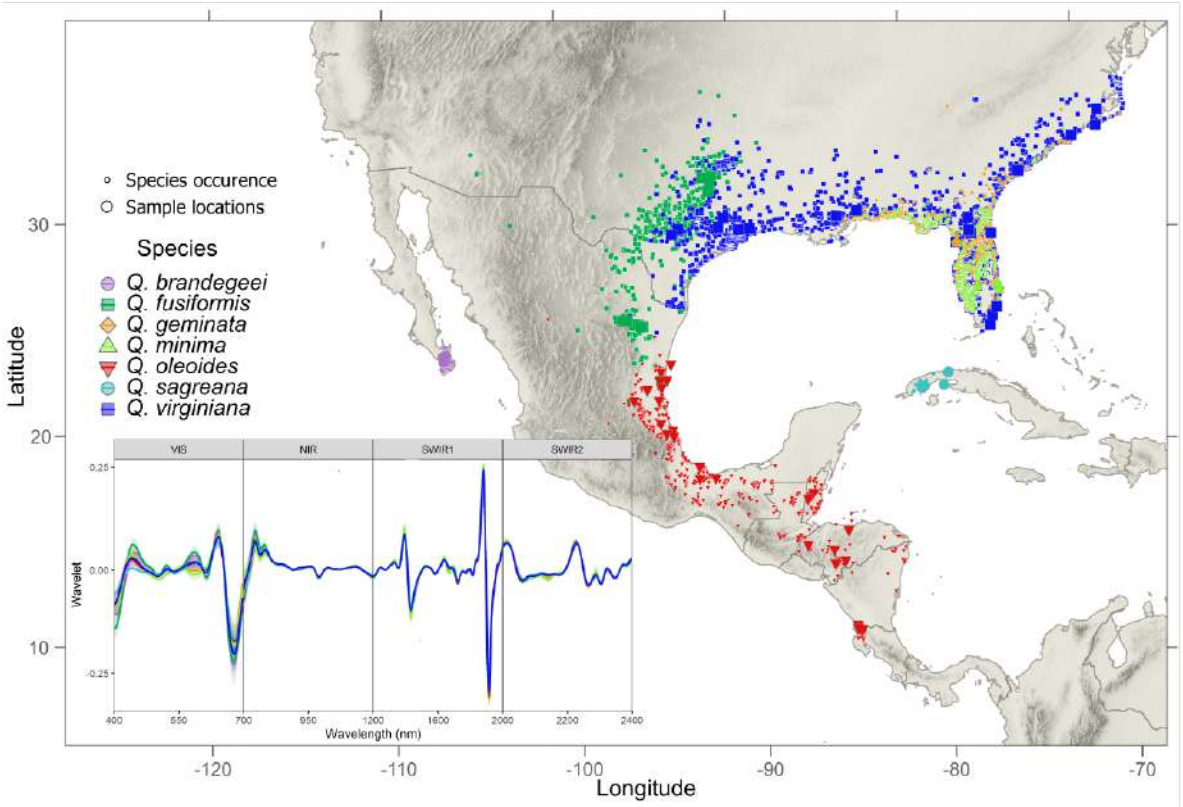


Fig. S1. Distribution of the seven species in section *Virentes* based on species occurrence modified from Cavender-Bares *et al.* 2015. The panel represents CWT mean spectra and 95% quantiles. for each species within the visible (VIS), near-infrared (NIR), and first and second spectral regions of the short-wave infrared (SWIR1 and SWIR2).

Supporting Methods

Table S1 Previous studies from which the genetic data were obtained and from which specimens were collected and measured for spectral data. Studies from which three individual leaves were taken to measure the dry spectrum. Leaves were collected from wild populations and common gardens for DNA extraction, and ecophysiology experiments. A portion of the leaves were pressure dried and stored in a dry cabinet at the University of Minnesota.

Species	Data	Comments	Study
Q. geminata, Q. virginiana	nSSRs		Cavender-Bares, & Pahlich,2009
Q. geminata, Q. virginiana, Q. oleoides, Q. fusiformis	Dried leaves for spectra measurments	Only the leaves of individuals on warm treatment were measured	Koehler *et al.* 2011
Q. oleoides, Q. virginiana	nSSRs		Cavender- Bares *et al.* 2011
Q. sagreana	nSSRs		Gugger & Cavender-Bares, 2013
Q. oleoides	Dried leaves for spectra measurments	Only leaves from well-watered individuals were measured from spectral data	Ramírez-Valiente *et al.* 2017

Methods S1: *Environmental Characteristics of the Virentes Lineage*

The *Virentes* lineage is distinguished within the genus *Quercus* by its restriction to low-altitude habitats, generally on well-drained sandy or volcanic tuff soils (Muller 1961a; Boucher 1983; Nixon 1985; Cavender-Bares et al. 2004a). The species in this lineage share key morphological synapomorphies, such as fused cotyledons and fused stellate

trichomes (Candolle 1862; Engelmann 1876-1877; Lewis 1911; Coker 1912; Camus 1936-1938). All Virentes species are wind-pollinated and inter-fertile, with exceptionally high wood density (Nixon 1985; Nixon & Muller 1997) They maintain green foliage throughout winter in the southeastern U.S. and Texas or during the dry season in Central America, with a leaf lifespan of approximately one year. Species within this lineage exhibit varying degrees of tolerance to freezing and drought, which influences their distribution and migration patterns (Cavender-Bares 2007; Cavender-Bares & Pahlich 2009; Koehler et al. 2012).

1. *Quercus virginiana*

Quercus virginiana is a large, long-lived tree that grows in a variety of soil types, from moist to well-drained soils, and it can tolerate both alkaline and salty soils. Unlike *Q. minima* and *Q. geminata, Q. virginiana* is less drought-tolerant (Cavender-Bares et al. 2004). Some populations of *Q. virginiana* have shown the ability to tolerate short-term freezing temperatures, withstanding temperatures as low as -18°C to -12°C, making it more cold-tolerant than tropical oaks like *Q. oleoides* but less so than temperate oaks.(Cavender-Bares & Pahlich 2009; Koehler et al. 2012, Fontes et al in preparation).This cold tolerance has allowed it to establish in areas with mild winters.

2. *Quercus geminata*

Quercus geminata is a smaller tree relative to *Q. virginiana*, with pubescent leaves that reduce water loss, making it well-suited to arid conditions. It has a robust root system that provides greater stability in sandy soils. The tree prefers sandy, well-drained soils typical of coastal dunes and deep sandy areas. Compared to *Q. virginiana*, it is more tolerant of salt and drought. It can also produce resprouts from its roots, enabling it to form clonal colonies. This combination of traits gives *Q. geminata* a competitive advantage in sandy coastal areas exposed to strong winds and occasional drought.

3. *Quercus minima*

Quercus minima is a low-growing shrub that rarely exceeds 2 meters in height. It reproduces both by seed and through underground rhizomes, allowing it to form dense

clonal colonies. This species is found in the coastal plains of the southeastern United States, particularly in Florida. It prefers well-drained soils and thrives in thickets and prairies that experience frequent fires. Due to its reliance on rhizome regeneration, Q. minima is more fire-resistant than *Q. geminata* and *Q. virginiana*, as its underground shoots allow for rapid post-fire recovery.

4. Quercus brandegeei

*Quercus brandegee*i is an endemic and endangered oak species restricted to the mountainous region of the Sierra La Laguna in Baja California Sur, Mexico. Unlike other oaks, this tree is limited to growth in ephemeral riverbeds, which are seasonally flooded by hurricane waters, making water availability a crucial factor for its survival. (Denvir and Westwood, 2016; Cavender-Bares et al., 2015) *Q. brandegeei* exhibits a clustered spatial distribution, concentrating near these water sources. It relies heavily on riparian environments, where fluctuations in moisture are essential for its survival. The conservation status of *Q. brandegeei* is of significant concern due to its limited range and dependence on specific water availability (Carrero et al.2020).

5. Quercus fusiformis

Quercus fusiformis, also known as Texas live oak, is native to the southern United States and parts of northern Mexico (Muller, 1961). This species is adapted to dry, mountainous terrain and can be found as far west as Arizona. It is highly tolerant of heat and drought, thriving in arid environments where other oaks struggle. It exhibits frost tolerance to temperatures as low as -12°C , with its survival in colder regions depending on local microclimates and the age or health of the tree (Ramirez-Valiente et al. 2015). Unlike most other species, its evergreen nature allows it to retain leaves year-round, which helps shield the bark and inner tissues from freezing temperatures.

6. *Quercus oleoides*

Quercus oleoides, also known as the tropical live oak, is a key species in the seasonally dry tropical forests of Central America, with its range extending from northern Mexico through Costa Rica. It prefers warm, dry climates with limited exposure to freezing

temperatures. This species thrives in nutrient-poor, sandy soils or volcanic tuff and forms part of the monodominant stands of tropical dry forests (Cavender-Bares, 2005). Unlike temperate oak species, *Q. oleoides* is not highly tolerant to cold. It can endure brief, mild cold spells, but prolonged frost or subfreezing temperatures are lethal to the species. This thermal limitation reflects its adaptation to tropical and subtropical conditions (Cavender-Bares et al. 2011).

7. *Quercus sagraeana*

Quercus sagraeana, also known as the Cuban oak, is a medium-sized evergreen tree that is endemic to western Cuba. It is the only oak species native to the Caribbean. This species thrives in the seasonally dry tropical biome, where it is adapted to warm, dry conditions with pronounced wet and dry seasons. Its habitat includes the Cuban pine forests ecoregion, and its survival is linked to its ability to endure the seasonal variability of water availability *(*Gugger & Cavender-Bares 2013).

Summary of Cold Tolerance

Highly Cold-Tolerant Species:

Quercus fusiformis: Tolerates frost down to -12°C but depends on microclimatic conditions for survival in colder areas.

Quercus virginiana: Can survive short-term freezing conditions down to -18°C to -12°C*).*

Low to Moderate Cold Tolerance:

Quercus geminata: Adapted to warm, coastal environments, tolerates mild cold but not extreme frost.

Low or No Cold Tolerance:

Quercus minima: Primarily adapted to Florida's fire-prone, subtropical habitats.

Quercus oleoides: Prefers warm, tropical conditions; does not survive prolonged subfreezing temperatures.

Quercus sagraeana: Occupies warm, seasonally dry tropical environments in Cuba, with no adaptation to freezing temperatures.

Method S2 Bayesian clustering using STRUCTURE software.

Bayesian analysis was used both to define the genetic groups within the seven species and to define the groups within the broadly distributed *Quercus oleoides* species. We allowed *K* groupings to range from 1 to 9 based on previous results (Cavender-Bares, *et al.* 2015) with independent runs and a burn-in of 2×10^5 steps followed by 2×10^6 Markov Chain Monte Carlo (MCMC) iterations using multilocus data for all individuals. The most probable value of *K* was identified with the Δ*K* statistic (Evanno *et al.,* 2005) with the online version of Structure Harvester v 0.6.94 (Earl & Von Holdt, 2012). We consider the percentage of assignment of an individual to a specific cluster to consider its belonging to that genetic group. Individuals with assignment values (Q_i) between 0.2-0.5 were considered as admixed, where *i* represents the assignment of individuals to the i^{th} cluster.

Analysis of molecular variance (AMOVA) comparing species, and genetic groups analysis were performed with ARLEQUIN v3.5.1.2 (Excoffier & Lischer, 2010) using the infinite alleles model (F_{ST}). Additionally, we calculated pairwise genetic differentiation between the species and genetic groups with F_{ST} estimators using *hierfstat* Package in R (Goudet, J. 2005).

Method S3 Continuous Wavelet Transform

The Continuous Wavelet Transform (CWT) is a method used to analyze localized variations of power within a signal at multiple scales, enabling the decomposition of a signal into its frequency components (Graps 1995). Unlike Fourier transforms, which analyze signals in a fixed global frequency domain, the CWT provides a multi-scale representation, making it particularly suited for analyzing complex, non-stationary signals such as reflectance spectra.

In our study, we applied the CWT to [briefly describe what the spectra or data represent, e.g., leaf reflectance spectra] to identify spectral patterns associated with [specific traits or properties]. The CWT transforms the original signal into a set of wavelet coefficients that correspond to different scales (or frequencies). Each scale reflects patterns or features at a specific resolution, with smaller scales capturing finer details and larger scales representing broader trends.

Summing Wavelet Scales:

To summarize information across relevant scales, we summed the wavelet coefficients within predefined ranges of scales. These ranges were chosen based on [criteria, e.g., prior studies, known spectral regions related to traits]. Summing across scales provides an aggregate metric that captures the overall contribution of specific spectral features to the traits of interest. This approach is analogous to integrating over a frequency band but tailored to the multi-scale nature of wavelet analysis.

Method S4 PLSR modeling framework to predict leaf traits from dried-leaf reflectance spectra.

For each sample, we measured full-range reflectance spectra (350–2,500 nm) of the leaves. To construct the models, we measured the following leaf structural and chemical traits on a percentage of the samples of *Quercus Virentes* species plus other *Quercus* species: LMA (kg/m2), thickness (mm), carbon fractions (soluble cell contents, hemicellulose, cellulose, and lignin; %), and concentrations of a variety of elements (Al, C, Ca, Cu, Fe, K, Mg, Mn, N, Na, P, Zn; % or mg/g). We used a PLSR modelling framework to predict each trait from pressed-leaf spectra across the full range (400–2,400 nm). PLSR is suited to handle spectral datasets, which have many collinear predictors, because it projects the spectral matrix onto a smaller number of orthogonal latent components in a way that maximizes the ability to predict the response variable. Our methods for model calibration and validation largely follow Burnett *et al.* (2021). To avoid the imbalance given by the number of samples of each species, the data were split hierarchically using the species as an index to maintain a proportion close to 60% samples for calibration and 40% for independent validation datasets (Appendix Table 1 list the number of individuals used for each trait prediction, the smallest number of

components selected). We quantified model performance using R2 and root mean squared error (RMSE) between measurements and mean predictions. We also report the RMSE as a percentage of the 2.5% trimmed range of measured values (%RMSE), which we used rather than the entire range (as in e.g. Burnett *et al.*, 2021) for robustness to outliers. For each trait, we also tested whether the magnitude of residuals (observed minus predicted) in the validation dataset varied among leaves with different discoloration scores. We performed all statistical analyses in R v. 3.6.3 (R Core Team, 2020) and used package pls v. 2.7.1 (Mevik *et al.*, 2019) for PLSR modelling.

Method S5 Comparing phenotypic vs genotypic divergence.

The Q_{ST} index introduced by Lande (1992) and Spitze (1993) is intended as an analog to F_{ST} by measuring the degree of phenotypic variance among populations over a set of quantitative traits rather than at a specific locus. F_{ST} can be used as a null hypothesis by assuming that the value of F_{ST} measured through neutral loci is the value of divergence between populations due to drift and migration. Assuming neutrality (no selection), we thus expect that $F_{ST} = Q_{ST}$, meaning that divergence between traits (or phenotypic characters) could be achieved by drift alone. If $Q_{ST} > F_{ST}$ the inference is that quantitative traits show a higher level of differentiation than expected by genetic drift, assuming directional selection by favoring different phenotypes (i.e., heterogeneous selection). If $Q_{ST} < F_{ST}$ trait divergence among populations is less than expected by drift alone, indicating the influence of natural selection, but one that is selecting for the same optimum in different populations (i.e., stabilizing selection).

In wild populations where imposing a breeding design is challenging, the Q_{ST} index is often approximated by P_{ST} (Leinonen et al., 2006). The difference between Q_{ST} and P_{ST} is that the latter is calculated from phenotypic variance components with no distinction between the relative contribution of genetic and environmental variation:

$$P_{ST} = \frac{\frac{c}{h^2}\sigma_b^2}{\frac{c}{h^2}\sigma_b^2 + 2\sigma_w^2} \qquad \text{(eqn. 1)}$$

where σ_b^2 and σ_w^2 are the phenotypic variance components between and within populations, h^2 is heritability (i.e., the proportion of phenotypic variance due to additive

genetic effects), and *c* is an estimate of the proportion of the total variance due to additive genetic effects across populations (Brommer, 2011).

Given that traits were measured in wild populations rather than in common garden experiments designed to estimate heritability (h^2) and additive genetic variation (c), and that the P_{ST} approximation of Q_{ST} is dependent on how well *c* and *h* are theoretically calculated, we estimated the c/h^2 ratio following Seeholzer & Brumfield (2018):

$$\frac{c}{h^2} = \frac{-2F_{ST(upper)}\,\sigma^2_{W(upper)}}{\sigma^2_{B(lower)} + (F_{ST(upper)} - 1)} \qquad \text{(eqn 2)}$$

Where $\sigma^2_{W(upper)}$ is the upper confidence interval (CI) value for the within-populations phenotypic variance, $\sigma^2_{B(lower)}$ is the lower CI value for the between populations phenotypic variance and $F_{ST\,(upper)}$ is the upper CI value estimated from the genetic markers (F_{ST}). Ratios of c/h^2 closer to zero are considered robust evidence that P_{ST} (phenotypic differentiation among populations) surpasses F_{ST} (genetic differentiation among populations), indicating a deviation from neutral expectations (Brommer, 2011). Following this rationale, we interpret c/h^2 ratios less than 0.25 as strong evidence for natural selection driving phenotypic differentiation, ratios of 0.26–0.50 as moderate selection, and values between 0.51–0.75 as weak selection. We interpret values approaching or exceeding one as very weak or no selection. In the latter case, differentiation results primarily from random changes (i.e., genetic drift), plasticity, and environmental effects (Brommer, 2011, Cruz-Nicolas *et al.* 2019). Several studies have detected that the leaf weight-area ratio exhibits allometric growth; (Niklas *et al.* 2007, 2009; Li *et al.* 2008; Niklas & Cobb 2008, Sun *et al.* 2017; Lin *et al.* 2018). Since both the traits predicted from the spectrum and the wave-lengths are affected by leaf area and density, all phenotypic traits were transformed using the Aitchison log-ratio transformation (Aitchison, 1986).

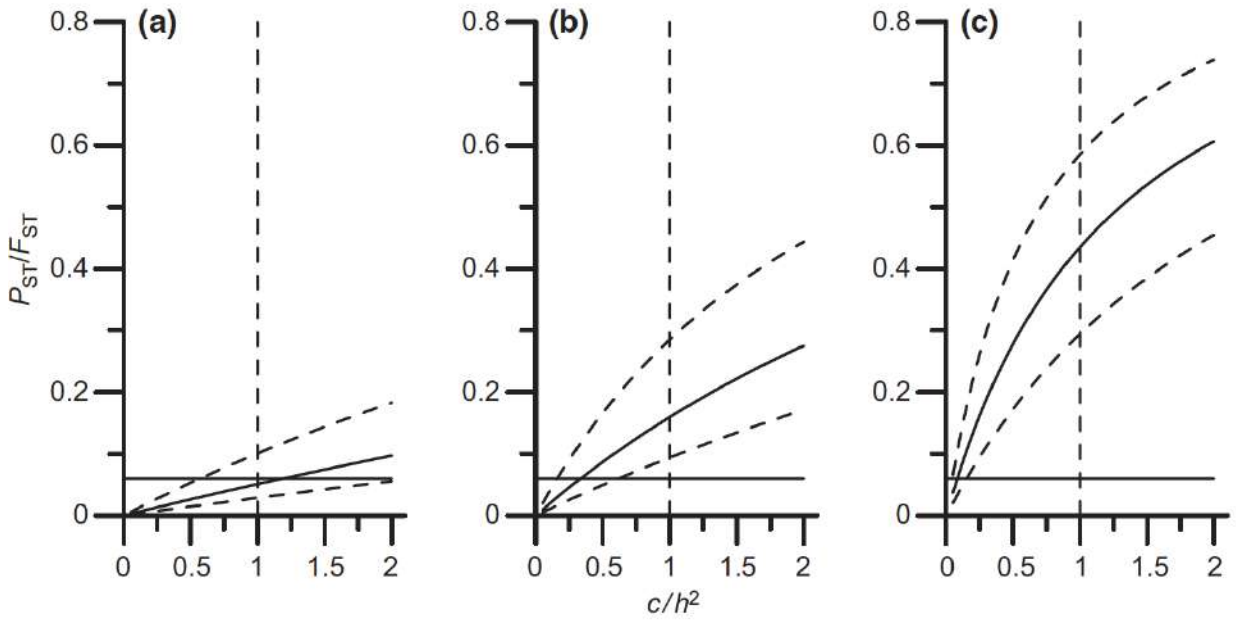


Fig. 1 Plots illustrating how a comparison between neutral differentiation and the estimate of P_{ST} [sensu eqn (2)] depends on the c/h2 ratio. In each plot, the dashed vertical line indicates the point of the 'null assumption' c = h2 for estimating P_{ST}. The horizontal line marks the upper confidence estimate of the neutral divergence estimated as F_{ST} (= 0.06 in each plot). Estimates of P_{ST} and its lower and upper 95% confidence intervals are plotted. Panel (a) indicates low P_{ST} (95% CI at c = h2: 0.029–0.107), where PST clearly does not differ from F_{ST}. Panel (b) is for a trait where $P_{ST} > F_{ST}$ for the null assumption (95% CI: 0.0938–0.285), but the significance of this difference is not very robustas the lower confidence estimate of P_{ST} overlaps with the upper confidence estimate for F_{ST} when c/h2 = 0.63. Panel (c) indicates a trait with strong phenotypic divergence (P_{ST} 95% CI at c = h2: 0.2946–0.586), and the difference in P_{ST} and F_{ST} is fairly robust as their confidence intervals only overlap when c/h2 = 0.17.(Taken from Brommer 2011).

Method S6 ***Spatial and environmental drivers of population divergence****.*

To quantify environmental variation across distribution of the seven species of Virentes we used the 19 Environmental variables extracted from the WorldClim database and the HYDRO1k (USGS, 2024) at 30s arc resolution. Hydro 1k raster data sets are the hydrologically correct DEM, derived flow directions, flow accumulations, slope, aspect, and a compound topographic (wetness) index. Before any analyses, the pool of variables was reduced from 24 to 6 after removing highly correlated variables (based on

Pearson's correlation |r| ≤ 0.60) to control for multicollinearity for each pairwise species. Since *Q. oleoides* is the most widely distributed species, pairwise comparison with other species was made only considering populations that have genetic or phenotypic concordance with other populations. For example: we only consider Northern Mexico populations of *Q. oleoides* that show admixture with *Q. fusiformis*. Only *Q. oleoides* populations of Central America were used to compare pairwise pRDA with Florida and Cuban populations of *Q. virginiana, Q. sagreana, Q. minima* and *Q. geminate*. These six variables represented variation in temperature, precipitation, and soil water capacity. Geography was characterized by the latitude and longitude of the population. Genetic distance was calculated between all 427 individuals, pairwise Nei's D (Nei, 1972) calculated using GENALEX were collapsed using PCoA, and the first five PCoA axes were treated as composite variables to represent each group of data.

We quantified the degree to which phenotypic variation was assessed using a variance partitioning technique to estimate individual and shared contributions of each variable (Borcard et al., 1992). This approach uses partial RDAs to estimate the proportion of explained variance for each predictor variable, independently and combined, out of the total explained variance. Each model included phenotype as the response variable and all combinations of genetic, environmental, and geographic variation as predictors. Phenotype was characterized by four variables: the six traits predicted from the PLSR spectral modelling, and the VIP wavelength bands that had PST > FST significant values for each pairwise species. The total phenotypic variance explained (PVE) by the predictors and adjusted r^2(i.e., individual contribution in terms of total PVE) was estimated with the varpart function in the vegan package in R. Model significance, when appropriate, was assessed using independent RDA and pRDAs, permutation-based ANOVAs (n = 999), and significance thresholds of α ≥ 0.05. All measurements were centered and standardized prior to analyses.

.

Table S2. a) Results of hierarchical analyses of molecular variance (AMOVA) for 7 species of the *Virentes* section in 64 populations based on eleven nSSRs loci. b) Pairwise F_{ST} values of genetic differentiation in nSSR a) among species and b) genetic groups of section *Virentes*. Values in bold numbers are significant $P < 0.05$. GE= *Q. geminate*, MN= *Q. minima*, VI= *Q. virginiana*, SA= *Q. sagreana*, OL= *Q. oleoides*, BR= *Q. brandegee*i, FU= *Q. fusiformis.*

a)

Level	df	SS	VC	%
Among species	6.0	500.1	0.4	9.2**
Among populations within species	57.0	627.3	0.3	6.5**
Among individuals within populations	605.0	2792.6	0.6	13.7**
Within individuals	669.0	2226.5	3.3	70.6**
Total	1337.0	6146.5	4.7	

df, degrees of freedom; SS, Sum of Squares; VC, Variance Components; *p <0.05, **p < 0.001

b)

	GE	MN	VI	OL	SA	BR	FU
GE		**0.011**	**0.130**	**0.162**	**0.233**	**0.272**	**0.175**
MN			**0.111**	**0.138**	**0.214**	**0.244**	**0.146**
VI				**0.059**	**0.100**	**0.167**	**0.079**
OL					**0.073**	**0.185**	**0.046**
SA						**0.243**	**0.146**
BR							**0.152**
FU							

Table S3. Summary statistics and SD of PLSR_DA model from pressed-leaf spectra restricted to 1400-2400 nm. Using Raw spectral data, vector Normalized and Continuous Wavelet Transform

	Raw	Vector Normalized	CWT
Accuracy	0.92 ± 0.011	0.91 ± 0.012	0.92 ± 0.011
Kappa	0.88 ± 0.018	0.86 ± 0.019	0.88 ± 0.017

Sensitivity	0.88 ± 0.10	0.87 ± 0.10	0.87 ± 0.10
Specificity	0.98 ± 0.016	0.98 ± 0.017	0.89 ± 0.014

Table S4. Summary statistics for the PLSR calibration and validation models for each leaf trait. #S: number of samples, #C: number of components. Performance statistics include R^2: the fit between the observed values and the predicted values, RMSE: root mean square error, %RMSE: percent root mean square error, MAE: mean absolute error. spectrally predicted traits (LMA: Leaf mass area; THI: thickness; SOL: solubles; HEM: hemicellulose; CEL: cellulose; LIG: lignin)

				Calibration model				**Validation model**			
Trait	**# S**	**# C**	**data range**	**R2**	**RMSE**	**%RMSE**	**MAE**	**R2**	**RMSE**	**%RMSE**	**MAE**
LMA g/m-2	146	6	33.8 – 273.6	0.87	19.21	0.15	14.51	0.88	17.6	0.14	13.25
THI mm	151	6	0.101 – 0.75	0.82	0.06	0.2	0.04	0.65	0.08	0.28	0.06
SOL %	84	8	36.5 – 61.9	0.86	2.4	0.05	1.87	0.51	4.29	0.09	3.63
HEM %	74	8	11.3 – 0.9	0.79	1.06	0.06	0.83	0.25	2.07	0.12	1.61
CEL %	97	9	8.6 – 32.6	0.84	2.28	0.11	1.78	0.72	3.37	0.17	2.64
LIG %	96	8	7.1 – 23.8	0.74	2.72	0.17	0.14	0.59	2.72	0.17	0.17

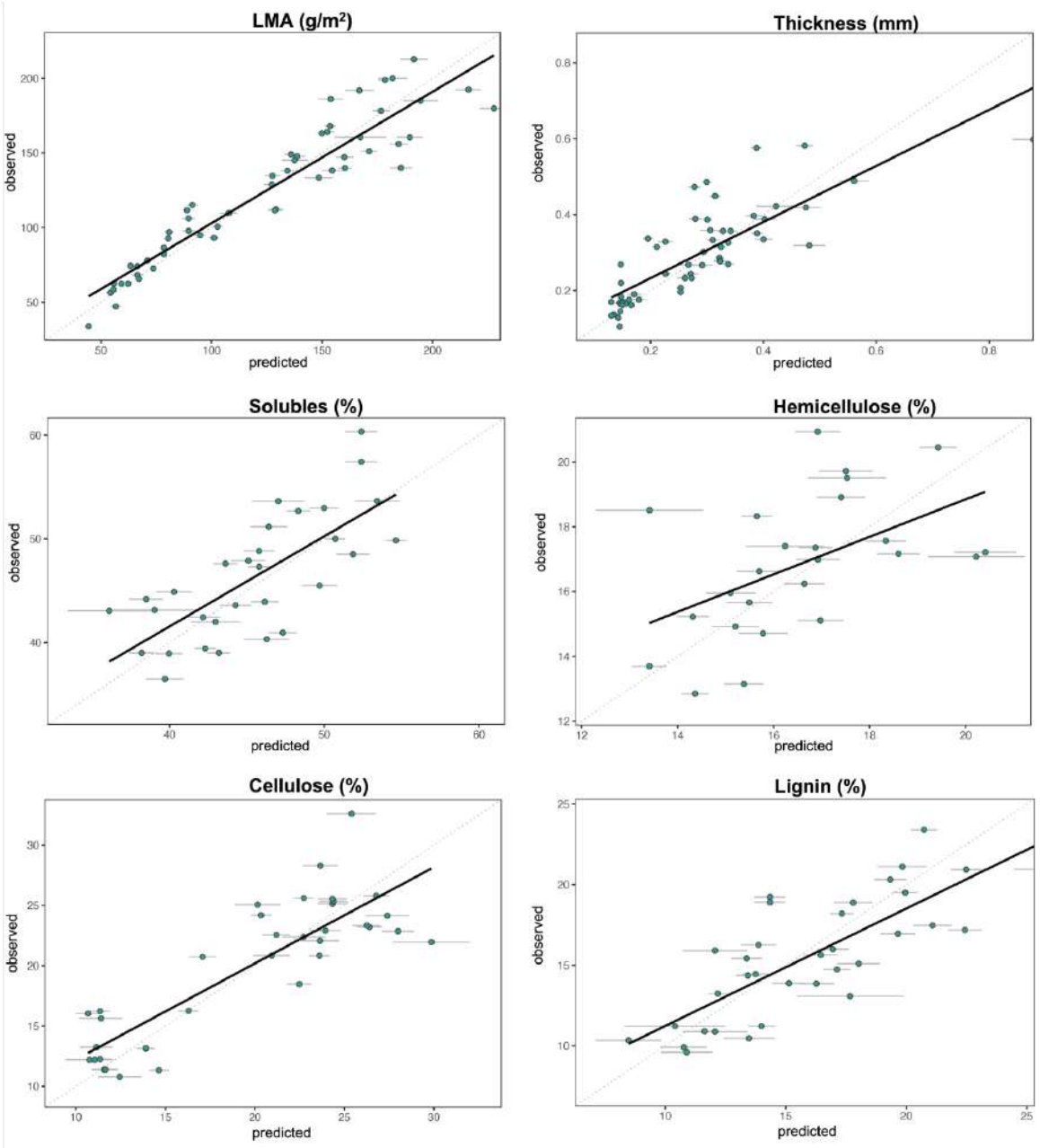


Fig. S2 Internal validation results for LMA, Thickness, Solubles, Hemicellulose, Cellulose and Lignin predicted from dry spectra. The error bars for each data point are 95% confidence intervals calculated from the distribution of predictions based on the ensemble of 1000 iterations.

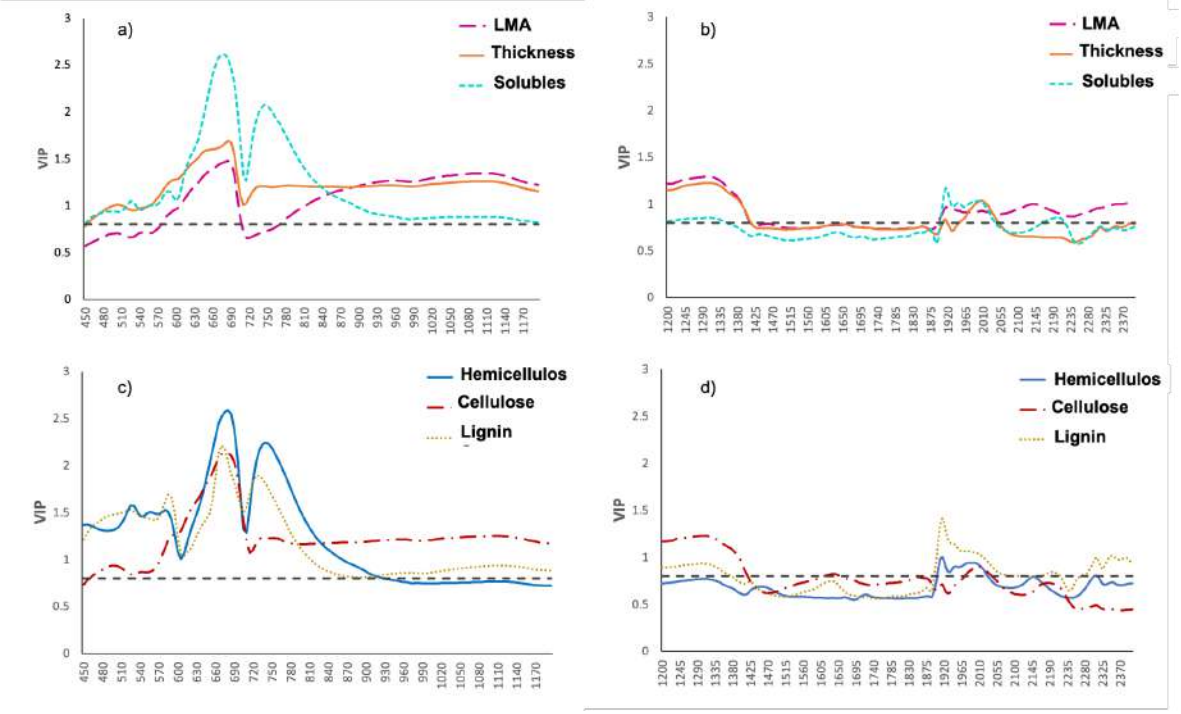


Fig. S3. The variable importance in the projection (VIP) metric was calculated based on dry sample spectral data models for six traits. a) and **c**) represent wavelengths from 400 to 1200 nm VIP, **b**) and **d**) 1200 to 2400 nm VIP. The dashed horizontal line at 0.8 represents a heuristic threshold for importance, as suggested by Burnett *et al.* (2021).

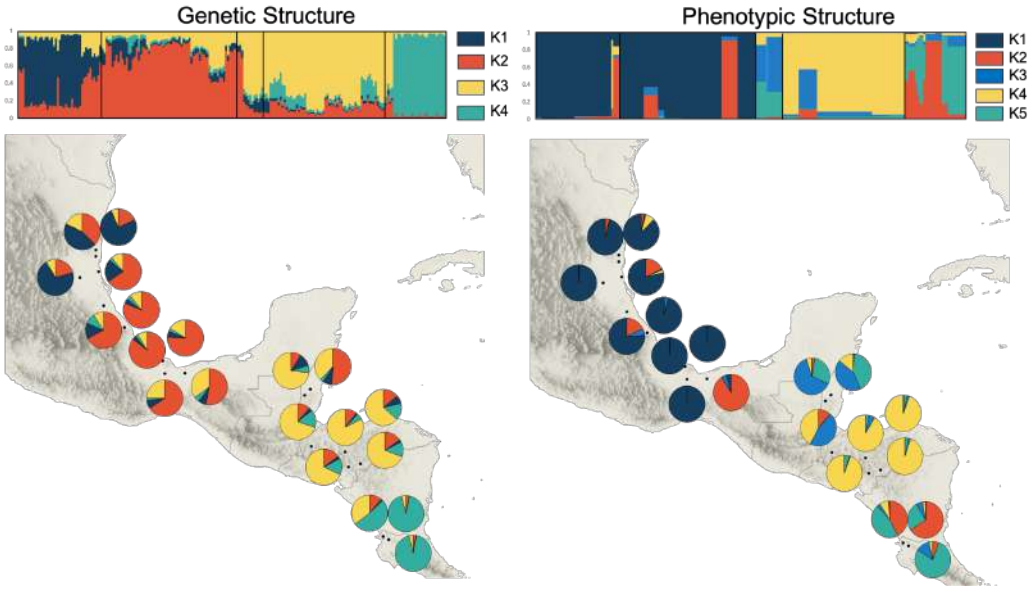


Fig. S4. **Spatial distribution of genetic and phenotypic variation in *Q. oleoides*. a)** Four genetic groups were identified using STRUCTURE (Pritchard 2000) from 123 individuals and **b)** five phenotypic groups identified by GENELAND (Guillot et al., 2009) using six spectrally derived leaf traits (leaf mass area, thickness, solubles, hemicellulose, cellulose, lignin) derived from dried leaf spectra from all individuals. The percentage assignment to genetic or phenotypic groups is represented at both the individual tree level (upper bar plots) and subpopulation level (pie charts).

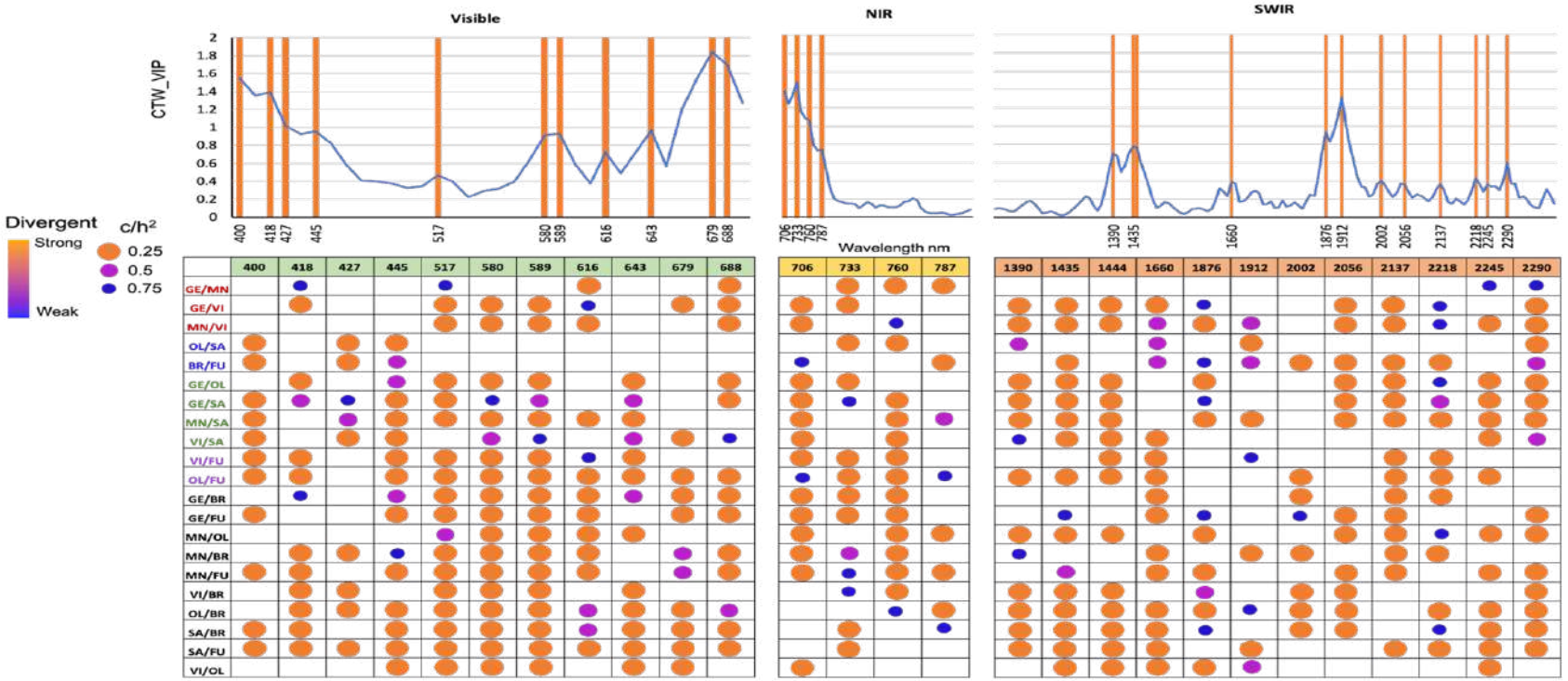
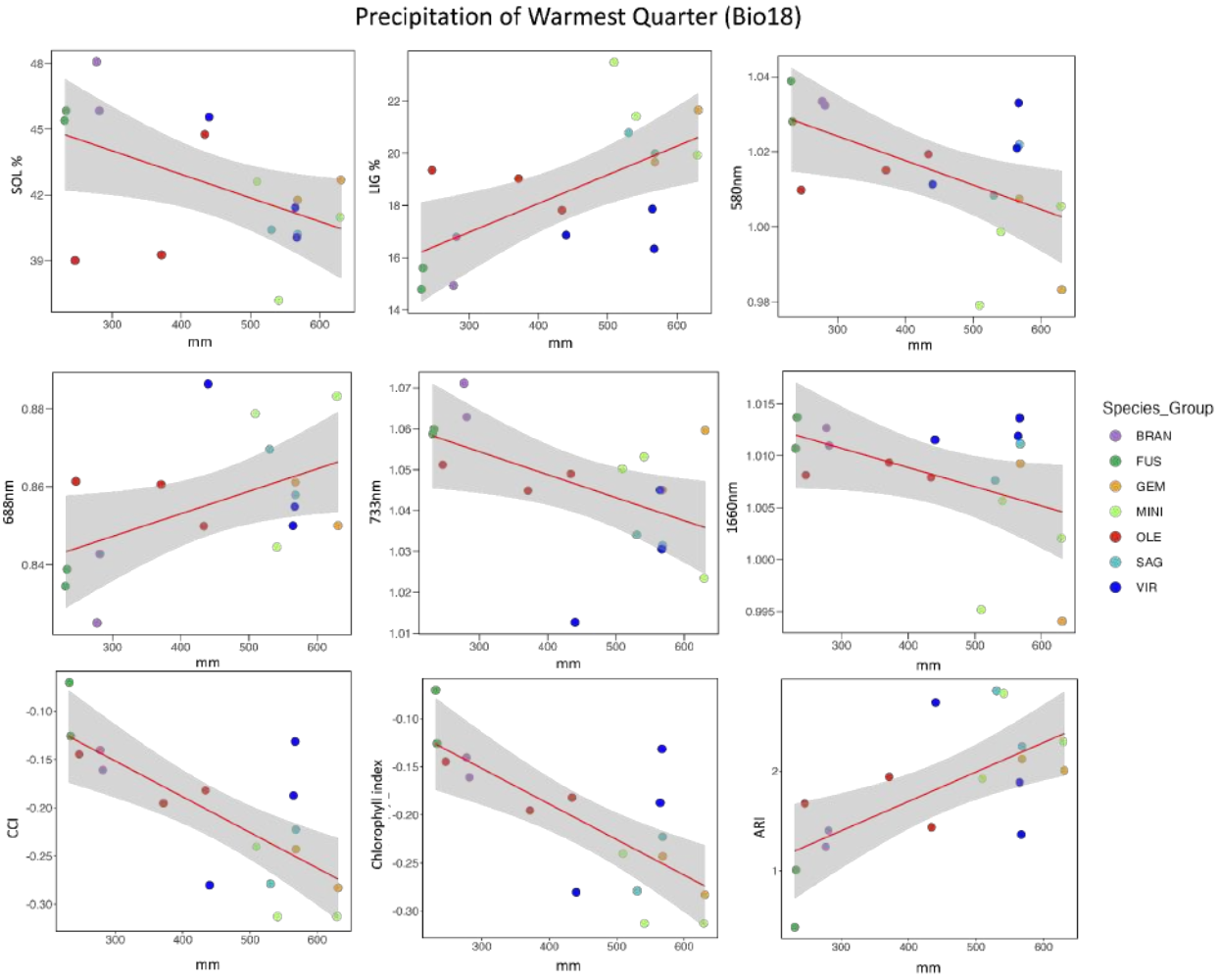


Fig. S5. P_{ST} comparisons among species using: a) spectrally predicted traits (LMA: Leaf mass area; THI: thickness; SOL: solubles; HEM: hemicellulose; CEL: cellulose; LIG: lignin); b) spectral bands within the visible (VIS), near infrared (NIR) and short-wave infrared (SWIR) region with high importance (i.e., Variable Importance of Projection (VIP) in discriminating species using wavelet spectra . Matrices represent P_{ST} plotted as a function of c/h^2 values of 0.25 (orange), 0.5 (pink) and 0.75 (blue). The optimal value of c/h^2 at which the lower confidence limit of P_{ST} is higher than the upper confidence limit of F_{ST} was chosen as the critical value of c/h^2 at which P_{ST} exceeds F_{ST}. The lower this critical value, the more robust inferences of selection are to environmental effects. Panel **a**) shows a simplified phylogenetic tree inferred from RADseq (min 20) data for 27 *Virentes* individuals using RAXML (Cavender-Bares et al., 2015) with the pairwise comparisons among species that were conducted using the six spectrally predicted traits. Colored lines represent sister relationships, historical introgression between specie pairs, and/or sympatric geographic associations within the *Virentes*: Red, sympatric sister species; Blue, sister but not sympatric species; Green, historically introgressing populations; Purple, parapatric species with introgression. Only phylogenetically and geographically meaningful pairwise comparisons are shown. Panel **b**) represents plotted VIP values obtained from the PLS-DA classification model using wavelet spectra: orange vertical lines represent wavelengths that were used as traits to calculate P_{ST} pairwise distances. Matrices are divided into VIS, NIR, and SWIR spectral regions. Contrary to panel **a**) all pairwise comparisons among species are shown, but name colors represent the same phylogenetic/geographic relations. GE= *Q. geminata*, MN= *Q. minima*, VI= *Q. virginiana*, SA= *Q. sagreana*, OL= *Q. oleoides*, BR= *Q. brandegee*i, FU= *Q. fusiformis.*

Table S5. Statistical associations between various plant traits, environmental variables, and spectral reflectance at specific wavelengths. The results include standard error (Std_Error), significance level (p_value), coefficient of determination (R^2), and adjusted R^2. Additionally, potential biochemical compounds linked to each spectral band are provided.

	Trait	Environmental Variable	Std_E	p	R^2	Adjusted R^2	Potential Associated Trait (Concentration/Reflectance)
Traits	SOL %	Bio_18	0.15	0.001	0.58	0.56	
	LIG %	Bio_6	0.11	0.04	0.03	0.29	
		Bio_18	0.15	0.001	0.64	0.62	
VIS	580nm	Bio_18	0.00	0.001	0.51	0.47	Carotenoids, including β-carotene (+/-)
	688nm	Bio_18	0.000	0.003	0.466	0.431	anthocyanins
NIR	733nm	Bio_18	0.00	0.001	0.44	0.40	Reflectance linked to leaf structure and mesophyll thickness. (-/+)
	760nm	Bio_14	0.00	0.01	0.37	0.33	potential influence of leaf structural traits. (-/+)
	787nm	Bio_8	0.00	0.04	0.25	0.20	Leaf thickness, Leaf Mass Area (LMA), (-/+)
SWIR	1660nm	Bio_18	0.00	0.02	0.33	0.28	Lignin and cellulose (-/+)
		Wetness index	0.01	0.04	0.25	0.20	
	2050nm	Bio_8	0.00	0.001	0.47	0.43	Cellulose, lignin (OH and CH bond absorptions) (-/+)
INDEX	CCI	Bio_18	-0.78	0.14	0.0005	0.68	
	Chlorophyll	Bio_18	-0.78	0.14	0.0005	0.68	
	Chlorophyll	Bio_8	-0.46	0.16	0.01	0.36	
	ARI1	Bio_18	0.80	0.15	0.00	0.64	
	ARI1	Bio_12	0.64	0.19	0.00	0.43	
	ARI1	Bio_8	0.54	0.15	0.00	0.45	


Precipitation of Warmest Quarter (Bio18)
SOL %
LIG %
580nm
688nm
733nm
1660nm
CCI
Chlorophyll index
ARI
mm
Species_Group
BRAN
FUS
GEM
MINI
OLE
SAG
VIR

438

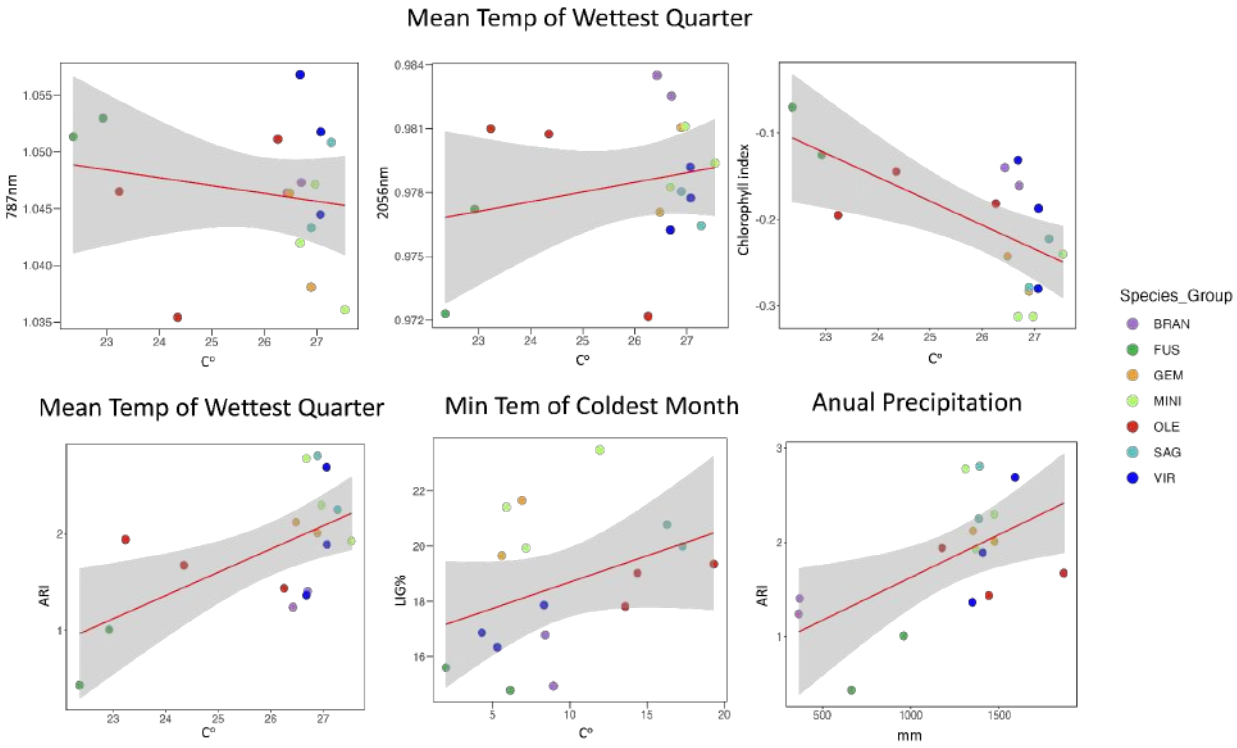


Fig S6. Spectrally predicted traits and selected wavelength relationships under different environmental conditions. Phylogenetic generalized least squares (PGLS) models account for relatedness across species. Coefficients for all models are given in Supporting Information **Table S5**

Table S6. Phylogenetic signal calculated using Blomberg's K (K), K_{WN} white noise and K_{BM} Brownian motion estimated for the seven species in *Quercus Virentes* phylogeny, where regions with significant signal *(p-value < 0.05) ns, no significative

Region	wavelength/trait	*K*	*p*_value K	$p\ (K > K_{WN})$
VIS	400	0.67	**	**
	418	1.11	**	**
	427	0.52	**	**
	445	0.64	**	**
	517	0.59	**	**
	580	0.59	**	**
	589	0.60	**	**
	616	0.90	**	**
	643	0.39	ns	**
	679	0.51	**	**
	688	0.49	**	**
NIR	706	0.83	**	**
	733	0.58	**	**
	760	0.44	**	**
	787	0.38	ns	**
	985	0.30	ns	**
SWIR	1390	0.23	ns	**
	1432	0.21	ns	**
	1435	0.23	ns	**
	1444	0.28	ns	**
	1660	0.34	ns	**
	1876	0.16	ns	**
	1912	0.18	ns	**
	2056	0.26	ns	**
	2137	0.29	ns	**
	2218	0.21	ns	**
	2245	0.16	ns	**
	2290	0.14	ns	**
Traits	LMA	**0.24**	ns	**
	thickness	**0.25**	ns	**
	solubles	**0.69**	**	**
	hemicellulose	**1.79**	**	**
	cellulose	**0.30**	ns	**
	lignin	**0.92**	**	**

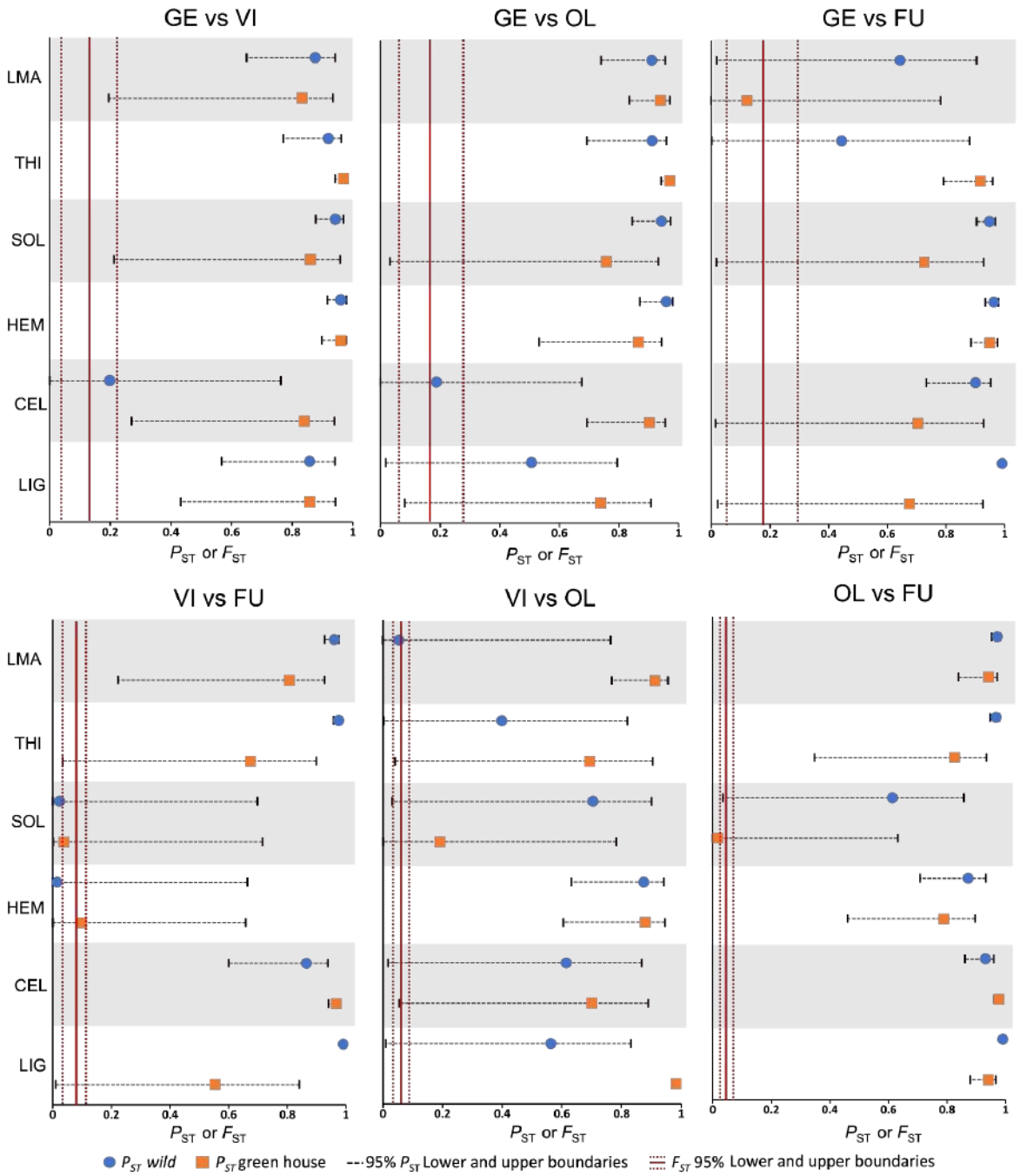


Fig. S7. P_{ST} vs F_{ST} estimates and 95% confidence intervals from quantitative predicted traits among wild and greenhouse individuals for four species in *Quercus* section *Virentes*. Graphs represented in colored blue circles (wild), and orange squares (greenhouse) mean P_{ST} values plotted as a function of $c/h^2 = 0.75$. GE= *Q. geminate*, VI= *Q. virginiana*, OL= *Q. oleoides*, FU= *Q. fusiformis.* Predicted traits (LMA) Leaf mass area, (THI) thickness, (SOL) solubles, (HEM) hemicellulose, (CEL) cellulose, and (LIG) lignin.

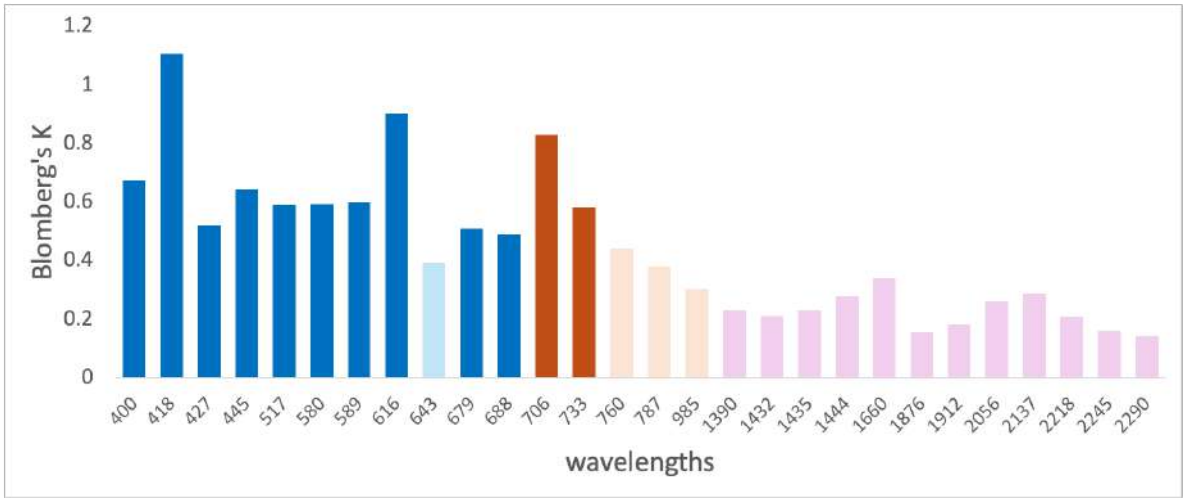


Fig. S8 Phylogenetic signal detected in leaf spectra varies across wavelengths across . *Quercus Virentes* species. Phylogenetic signal calculated using Blomberg's K (K) estimated, where regions with significant signal (P-value < 0.05) dark colors. Not significant values are colored with dim colors. Blue visible; red NIR; purple SWIR

References

Aitchison, J. (1986). The Statistical Analysis of Compositional Data. Springer Netherlands.

Brommer, J. E. (2011). Whither Pst? The approximation of Qst by Pst in evolutionary and conservation biology. Journal of Evolutionary Biology, 24(6), 1160–1168.

Burnett, A. C., Anderson, J., Davidson, K. J., Ely, K. S., Lamour, J., Li, Q., Morrison, B. D., Yang, D., Rogers, A., & Serbin, S. P. (2021). A best-practice guide to predicting plant traits from leaf-level hyperspectral data using partial least squares regression. Journal of Experimental Botany, 72(18), 6175–6189.

Cavender-Bares, J., González-Rodríguez, A., Eaton, D. A. R., Hipp, A. A. L., Beulke, A., & Manos, P. S. (2015). Phylogeny and biogeography of the American live oaks (*Quercus* subsection *Virentes*): a genomic and population genetics approach. Molecular Ecology, 24(14), 3668–3687.

Carrero C, Jerome D, Beckman E, Byrne A, Coombes A, Deng M, González Rodríguez A, Van Sam H, Khoo E, Nguyen N, Robiansyah I, Rodríguez Correra H, Sang J, Song Y-G, Strijk J, Sugau J, Sun W, Valencia-Ávalos S, Westwood M. 2020. *The Red List of Oaks 2020.*

Cruz-Nicolás, J., Giles-Pérez, G., González-Linares, E., Múgica-Gallart, J., Lira-Noriega, A., Gernandt, D. S., Eguiarte, L. E., & Jaramillo-Correa,

J. P. (2019). Contrasting evolutionary processes drive morphological and genetic differentiation in a subtropical fir (Abies, Pinaceae) species complex. Botanical Journal of the Linnean Society. https://doi.org/10.1093/botlinnean/boz077

Denvir, A. & Westwood, M. (2016.) Quercus brandegeei, Encino Arroyo. The IUCN Red List of Threatened Species 2016: e.T30726A2795363. https://dx.doi.org/10.2305/IUCN.UK.2016- 3.RLTS. T30726A2795363.en

Earl, D. A., & vonHoldt, B. M. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genetics Resources, 4(2), 359–361.

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

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

Koehler, K., Center, A., & Cavender-Bares, J. (2012). Evidence for a freezing tolerance–growth rate trade-off in the live oaks (*Quercus* series *Virentes*) across the tropical–temperate divide. New Phytologist, 193(3), 730–744.

Li, G., Yang, D., & Sun, S. (2008). Allometric relationships between lamina area, lamina mass and petiole mass of 93 temperate woody species vary with leaf habit, leaf form and altitude. Functional Ecology, 22(4), 557–564. https://doi.org/10.1111/j.1365-2435.2008.01407.x

Mevik, B., Wehrens, R. and Liland, K. H. (2020). pls:Partial Least Squares and Principal Component Regression. R package version 2.7-3. https://CRAN.Rproject.org/package=pls

Nei, M. (1972). Genetic distance between populations. *The American Naturalist, 106*(949), 283–292

Niklas, K. J., Cobb, E. D., & Spatz, H. (2009). Predicting the allometry of leaf surface area and dry mass. American Journal of Botany, 96(2), 531–536. https://doi.org/10.3732/ajb.0800250

Niklas, K. J., Cobb, E. D., & Spatz, H.-C. (2009). Predicting the allometry of leaf surface area and dry mass. *American Journal of Botany*, *96*(2), 531–536.

Ramírez-Valiente, J. A., & Cavender-Bares, J. (2017). Evolutionary trade-offs between drought resistance mechanisms across a precipitation gradient in a seasonally dry tropical oak (*Quercus* oleoides). Tree Physiology, 37(7), 889–901

Seeholzer, G. F., & Brumfield, R. T. (2018). Isolation by distance, not incipient ecological speciation, explains genetic differentiation in an Andean songbird (Aves: Furnariidae: Cranioleuca antisiensis, Line-cheeked Spinetail) despite near threefold body size change across an environmental gradient. Molecular Ecology, 27(1), 279–296. https://doi.org/10.1111/mec.14429

Shi, P., Liu, M., Ratkowsky, D. A., Gielis, J., Su, J., Yu, X., Wang, P., Zhang, L., Lin, Z., & Schrader, J. (2019). Leaf area–length allometry and its implications in leaf shape evolution. Trees, 33(4), 1073–1085. https://doi.org/10.1007/s00468-019-01843-4

Teixeira, J. C., & Huber, C. D. (2021). The inflated significance of neutral genetic diversity in conservation genetics. Proceedings of the National Academy of Sciences, 118(10). https://doi.org/10.1073/pnas.2015096118

Wold, S., Sjöström, M., & Eriksson, L. (2001). PLS-regression: a basic tool of chemometrics. Chemometrics and Intelligent Laboratory Systems, 58(2), 109–130.