Botanical Sciences 103 (Special Issue): S64-S81. 2025
DOI: 10.17129/botsci.3707

Received: April 5, 2025, Accepted: June 29, 2025
On line first: August 26, 2025

Genetics / Genética

Landscape genomics of the contact zone between the Mexican red oaks *Quercus affinis* and *Quercus laurina*: genetic diversity, hybridization and local adaptation

Flor Rodríguez-Gómez[1,2], Diushi Keri Corona-Santiago[2], Ken Oyama[2*], Vanessa Izaguirre-Toriz[3], Antonio González-Rodríguez[3*]

[1] Departamento de Ingeniería Traslacional, Centro Universitario de Ciencias Exactas e Ingenierías, Universidad de Guadalajara, Guadalajara, Jalisco, Mexico.
[2] Escuela Nacional de Estudios Superiores Morelia, Universidad Nacional Autónoma deMéxico, Morelia, Michoacán, Mexico.
[3] Instituto de Investigaciones en Ecosistemas y Sustentabilidad, Universidad Nacional Autónoma de México, Morelia, Michoacán, Mexico.
*Correspondence authors: agrodrig@cieco.unam.mx; kenoyama@enesmorelia.unam.mx

Abstract
Background: Hybridization is a natural phenomenon that involves successful mating between individuals from populations that are ecologically, morphologically or physiologically different. The genus *Quercus* is characterized by a high frequency of hybridization, even among species that are notably divergent.
Questions: What are the patterns of population genomic diversity and structure, introgression and local adaptation in two hybridizing Mexican red oak species?
Studied species: *Quercus affinis* and *Quercus laurina*.
Study site and dates: Sierra Madre Oriental and eastern Trans-Mexican Volcanic Belt, Mexico, 2017-2018
Methods: We conducted an analysis of hybridization from a landscape genomics perspective using 8,106 single nucleotide polymorphisms (SNPs) to assess the extent of genomic admixture and the geographic distribution of pure and introgressed genotypes across the *Q. affinis*-*Q. laurina* contact zone.
Results: Interspecific differentiation of the two parental species was moderate (F_{ST}= 0.137). Populations in the contact zone showed the highest levels of admixture. Genotype-environment association analyses and F_{ST}-outlier methods detected between 31 and 96 putatively adaptive SNPs, mainly located in genes involved in stress response. Models of sigmoid clines fitted to 232 species-diagnostic alleles indicated an average cline center located in northern Puebla, and an average width of 387 km. Clines for 45 alleles were markedly narrower or displaced from the average center, several of which were also identified as adaptive SNPs by the outlier detection tests.
Conclusions: Our results contribute to understanding how hybridization may enhance genetic diversity and evolutionary potential in oak populations.
Keywords: Hybridization, introgression, *Quercus*, single nucleotide polymorphisms, local adaptation, Mexico.

Resumen
Antecedentes: La hibridación es un fenómeno natural que involucra el apareamiento exitoso entre individuos de poblaciones ecológica, morfológica o fisiológicamente diferentes. El género *Quercus* se caracteriza por una alta frecuencia de hibridación, incluso entre especies que son notablemente divergentes.
Preguntas: ¿Cuáles son los patrones de diversidad y estructura genómica, introgresión y adaptación local en dos especies de encinos mexicanos que se hibridan?
Especies de estudio: *Quercus affinis* y *Quercus laurina*.
Sitio y año de estudio: Sierra Madre Oriental y Faja Volcánica Transmexicana, 2017-2018.
Métodos: Realizamos un análisis de hibridación desde una perspectiva de genómica del paisaje usando 8,106 polimorfismos de un solo nucleótido (SNPs), para evaluar el grado de mezcla genómica y la distribución geográfica de los genotipos puros e introgresados en la zona de contacto entre las especies.
Resultados: La diferenciación entre las especies fue moderada (F_{ST}= 0.137). Las poblaciones en la zona de contacto mostraron los mayores niveles de mezcla genética. Los análisis de asociación genotipo-ambiente y de identificación de loci atípicos detectaron entre 31 y 96 SNPs potencialmente adaptativos, principalmente localizados en genes de respuesta al estrés. Los modelos de clinas ajustados a la frecuencia de 232 alelos especie-específicos mostraron un centro promedio localizado en el norte de Puebla, y una anchura promedio de 387 km. Las clinas de 45 alelos fueron marcadamente más angostas o estuvieron desplazadas con respecto al centro promedio.
Conclusiones: Nuestros resultados contribuyen a entender como la hibridación puede incrementar la diversidad genética y el potencial evolutivo en las poblaciones de encinos.
Palabras clave: Hibridación, introgresión, *Quercus*, polimorfismos de un solo nucléotido, adaptación local, México.

Natural hybridization is an important source of new variation and a frequent component of the evolutionary history of many species (Futuyma *et al*. 1995) and, according to recent evidence, its role in adaptive evolution is more significant in both plants and animals than previously believed (Abbott *et al.* 2016, Martin & Jiggins 2017). In fact, when genetic variation is limited within species, hybridization and introgression may be the source of new recombinant genotypes allowing rapid evolution in response to changing selective pressures. Given the current climate change rate, the positive consequences of hybridization may outweigh the negative consequences in some cases, since adaptation within an appropriate time frame may require new genetic and phenotypic combinations (Kremer *et al*. 2012).

Adaptive introgression occurs when alleles from one population are spread into a recipient population and maintained by natural selection (Suarez-Gonzalez *et al*. 2018). Introgression can offer an increased capacity for adaptation, range expansion in a changing climate and, in extreme cases, the persistence of genes at risk of extinction due to the loss of parental species (Seehausen 2013, Rius & Darling 2014). This has important implications when species are genetically impoverished, exhibit adaptational delays or have not been able to migrate in response to changing environments (Hamilton & Miller 2016). Genomic studies of introgression between related species are offering novel insights into its evolutionary consequences, underlining that it should be seen as a microevolutionary process reinforcing adaptation (Martin & Jiggins 2017, Kremer & Hipp 2019). On the contrary, interspecific introgression could break up co-adapted gene complexes, disrupt local adaptation and lead to genomic extinction (Hamilton & Miller 2016).

The goal of landscape genomics is to explain spatial patterns of genetic variation in natural populations, including the identification of locally adaptive genetic variation (Joost *et al.* 2013, Sork *et al.* 2016b). Studies in tree populations have found individual genes that show variation correlated with climate variables, such as genes involved in response to abiotic variables (i.e. heat stress, drought resistance, and others) (De Kort *et al*. 2015, Gugger *et al*. 2016, Sork *et al*. 2016b). Although most genetic differences between related species are likely to be neutral, evidence from provenance studies and landscape genomics analyses indicate that local adaptation is also important in shaping interspecific geographical patterns of genetic variation. However, demonstrating the extent to which genetic differentiation in natural populations is due to selection or demographic factors requires additional research.

Hybridization is common among oaks (*Quercus*: Fagaceae) (Sullivan *et al*. 2016). However, recurrent hybridization among oak species apparently does not lead to a loss of adaptive distinctiveness (Curtu *et al*. 2007, Peñaloza-Ramírez *et al*. 2010). Oaks thus present an opportunity to study how species coherence can be maintained with ongoing gene flow (Arnold & Bennett 1993, Dodd & Afzal-Rafii 2004). Recent population genetics studies have focused on Mexican oaks examining multiple species complexes known or suspected of hybridization (González-Rodríguez *et al*. 2004, Peñaloza-Ramírez *et al*. 2010, Valencia-Cuevas *et al*. 2015, Ramos-Ortiz *et al*. 2016).

The Mexican oaks *Quercus affinis* Scheidw. and *Q. laurina* Humb. et Bonpl. belonging to the *Lobatae* (red oak) section of the subgenus *Quercus* are not closely related species as indicated by recent phylogenetic analyses (Hipp *et al*. 2019). However, a large amount of evidence has shown that these two species show a level of hybridization and introgression (Valencia Ávalos 1994, González-Rodríguez *et al.* 2004, 2005, Ramos-Ortiz *et al.* 2016). There is clear morphological and genetic differentiation between *Q. affinis* and *Q. laurina* in most of the range of the species, as well as evidence of a secondary contact zone located in the east of the Trans-Mexican Volcanic Belt and the north of Oaxaca (Valencia 1994, González-Rodríguez *et al*. 2004, González-Rodríguez & Oyama 2005, Ramos-Ortiz *et al*. 2016). In this contact zone, there is partial inconsistency between morphological and genetic variation (Valencia 1994, González-Rodríguez *et al*. 2004, Ramos-Ortiz *et al*. 2016), which could be explained considering that when introgression occurs, some portions of the genome easily cross the specific barrier, while others do it to a minimal degree (Martinsen *et al*. 2001). This means that loci under divergent selection between the two species may be undergoing restricted introgression despite interspecific gene flow and the exchange of neutral loci.

In this study, we used genotyping by sequencing (GBS) and environmental data to assess the extent of genomic admixture and to map the geographic distribution of pure and admixed genotypes across the *Q. affinis*-*Q. laurina* contact zone. For this, we generated 8,106 single nucleotide polymorphisms (SNPs) which were then used to address

three objectives: (1) to determine population genetic diversity and structure, frequency of introgression and location of admixed genotypes across the distribution area of *Q. affinis*, *Q. laurina* and their contact zone; (2) to identify loci that could potentially be under the action of natural selection, using F_{ST}-outlier and genotype-environment association methods; (3) to visualize the patterns of population genomic turnover and its association to climatic gradients across the landscape and detect potential instances of local adaptation.

Materials and methods

Sample Collection. Leaf samples were collected from 63 individuals of *Q. affinis* and *Q. laurina* in eleven localities along a latitudinal gradient from the south of Oaxaca state to the north of Tamaulipas state in Mexico (Figure 1A). Minimum distances of 50 m were maintained between consecutive trees to avoid sampling related individuals. From each individual, 5-10 young, undamaged leaves were collected, transported on ice, and then stored in the laboratory at -80 ºC for DNA extraction.

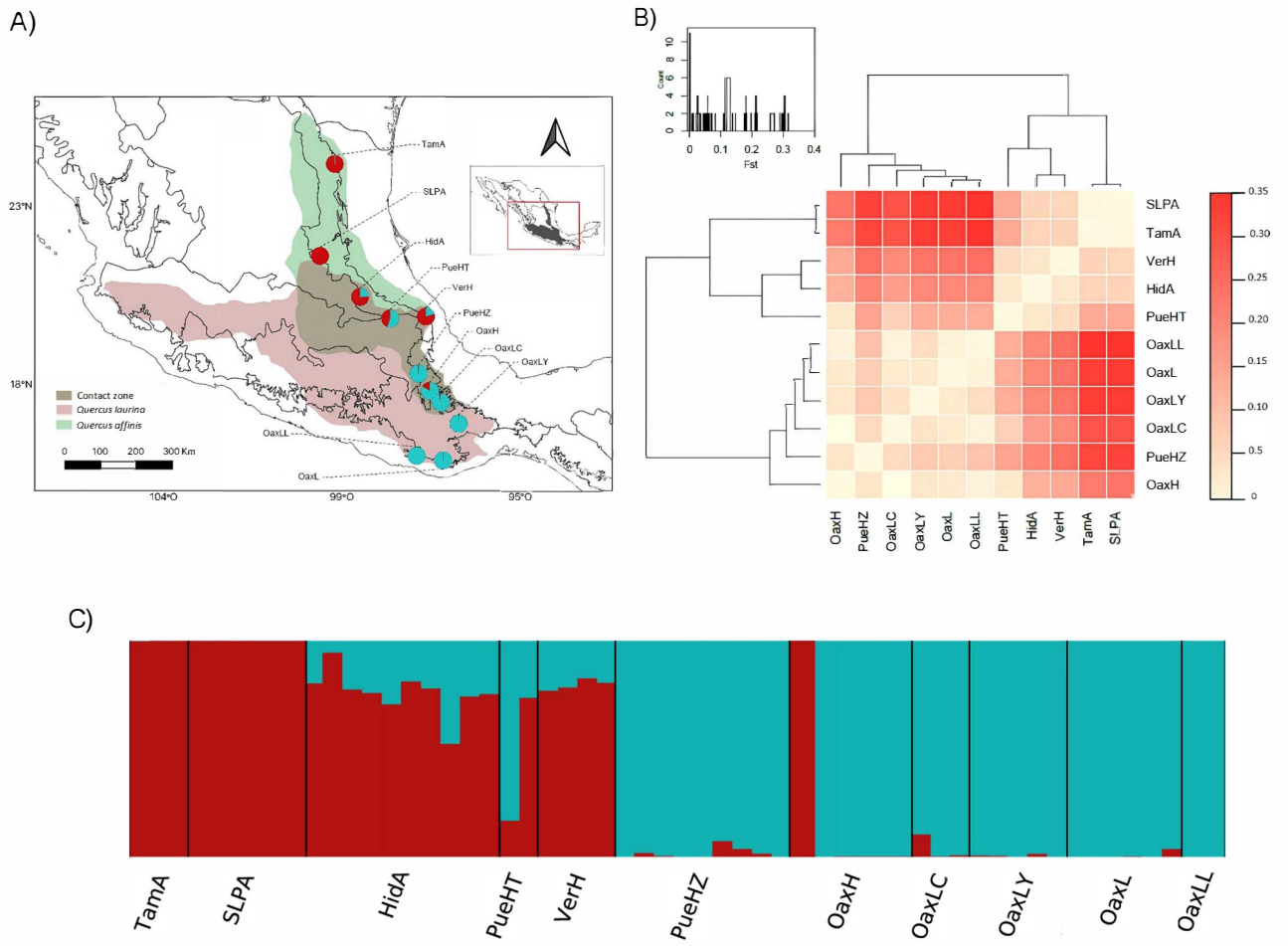

Figure 1. Distribution of *Quercus affinis* and *Q. laurina* (shaded area) and assignment proportion to the genetic groups identified in the Structure analysis ($K = 2$) in each population (A). Inner lines in the map demarcate physiographic regions. Heatmap of paired F_{ST} between populations of *Quercus affinis* and *Q. laurina* calculated with the Nei87 method (B). Genetic structure based on Bayesian assignment with the software Structure for $K = 2$ (C). Populations TamA and SLPA correspond to *Q. affinis*, OaxLY, OaxL and OaxLL to *Q. laurina*, and HidA, PueHT, VerH, PueHZ, OaxH, and OaxLC, correspond to the contact zone between the two species.

DNA extraction and GBS library preparation. DNA was extracted from 63 samples using 2X CTAB (Rogers & Bendich 1985,) and treated with RNase. Each sample was quantified using a Qubit fluorometer, quality assessed by agarose gel electrophoresis, and diluted to a concentration of 100 ng/µL. The GBS library was prepared as originally described in Elshire *et al*. (2011) with modifications used for other oak species (Martins *et al*. 2018). Briefly, each sample (20 ng/µL) was digested with the restriction enzyme *Pst1* and adapters were ligated to the DNA fragments. Adapter-ligated fragments were sequenced in an Illumina NextSeq 550 producing 150 base paired-end reads at the Laboratorio Nacional de Análisis y Síntesis Ecológica (LANASE, Escuela Nacional de Estudios Superiores Morelia, UNAM).

Bioinformatics and sequence filtering. We used Trimmomatic v. 0.39 (Bolger *et al.* 2014) to trim all sequences to 110 pb with the following filters: threads 4 -phred33 crop:110 headcrop:10 trailing:20. Using the *process_radtags* module in Stacks v. 2.0 (Catchen *et al*. 2011, 2013) we filtered trimmed Illumina reads with a minimum confidence threshold (Phred-scaled) of 20 and eliminated adapter sequences using the following parameters: --adapter_1 TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG --adapter_2 GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG --adapter_mm 2 --disable_rad_chec -r -c -q -s 20 --index_index. Using Bowtie2 (Langmead *et al*. 2009), the filtered reads were aligned to the *Quercus lobata* reference genome v. 3.0 (NCBI Accession LRBV00000000.1, and available in Valley oak project website, valleyoak.ucla.edu; Sork *et al.* 2016a). The aligned reads with the Binary Alignment Map (BAM) format were then processed for catalog locus construction and SNP identification by the *ref_map.pl* pipeline implemented in Stacks with two components (*gstacks* and *populations*). Those components identified SNPs in each aligned sample. We used VCFtools v. 0.1.1.6 (Danecek *et al.* 2011) to filter the SNPs to include only biallelic sites, to be present in at least 80 % of individuals, to have a minimum mean coverage depth of 5 and minor allele frequency (MAF) > 0.1 and to be in Hardy-Weinberg equilibrium (HWE; with a significance level of $P < 0.0001$). To remove SNPs that were in linkage disequilibrium (LD) we used Plink v. 1.90 (Purcell *et al*. 2007) with the following parameters: --indep-pairwise 50 5 0.5. Additionally, we used the *write_single_snp* command in the *populations* module of Stacks to ensure the inclusion of only one SNP per read in the database. Due to excessive missing data, seven samples were discarded, and the rest of the analyses were performed with the remaining 56 individuals (Table 1).

Table 1. Sampling localities and number of individuals analyzed from each site.

Species	Locality/State	Code	N	Longitude	Latitude	Elevation (m)
Q. affinis	El Puerto de Arrazolo, Tamaulipas	TamA	3	-99.193	23.553	1,630
	Pinal de Amoles, Quéretaro	SLPA	6	-99.563	21.24	2,320
	El Zembo, Hidalgo	HidA	9	-98.557	20.207	2,103
	Tetela de Ocampo, Puebla	PueHT	2	-97.8	19.667	2,680
Contact zone	Zoquitlán, Puebla	PueHZ	10	-97.083	18.283	2,300
	Tonayan, Veracruz	VerH	4	-96.9	19.717	1,897
	Pápalo, Oaxaca	OaxH	6	-96.8	17.85	2,900
	Comaltepec, Oaxaca	OaxLC	3	-96.525	17.558	2,600
Q. laurina	San Pedro y San Pablo Ayutla, Oaxaca	OaxLY	5	-96.075	17.025	2,100
	Suchixtepec, Oaxaca	OaxL	6	-96.467	16.1	2,600
	Lachao, Oaxaca	OaxLL	2	-97.133	16.217	2,179

Genetic diversity and population structure. To describe genetic diversity we used the packages adegenet v. 2.1.2 (Jombart 2008), pegas v. 0.13 (Paradis 2010) and hierfstat v. 0.04-22 (Goudet & Jombart 2015) in R software (R Core Team 2019). For each population, we estimated the observed heterozygosity (H_O), gene diversity (H_E) and the inbreeding coefficient (F_{IS}). Confidence intervals for F_{IS} were calculated implementing 1,000 bootstrap repetitions, and Bartlett's test of homogeneity of variances was used to identify significant differences between H_O and H_E. Principal Components Analysis (PCA) was performed using the adegenet package in R software using the cumulative variance protocol to determine the maximum number of genetic clusters. Paired F_{ST} values between populations were estimated with the hierfstat package using the *Nei87* method for calculation (Nei 1987). Hierarchical *F*-statistics were obtained to determine the overall variance components at the individuals and populations levels, and a test of significance of the effect of populations on genetic differentiation was implemented using the *test.g* module with 1,000 permutations.

Genetic structure was inferred using a Bayesian method with the software Structure v. 2.3.4 (Pritchard *et al*. 2000). For this, we evaluated *K* values from 1 to 11 performing ten independent runs for each *K*. We implemented the admixture model with a length of burn-in period of 100,000 and 500,000 Markov Chain Monte Carlo (MCMC) repetitions. We used the Δ*K* method (Evanno *et al*. 2005) implemented in Structure Harvester (Earl & von Holdt 2012) to determine the number of *K* that best describes the distribution of genetic variation in our data. Then, we used Distruct v. 1.1 (Rosenberg 2004) to display population structure plots.

Genomic cline analyses. To compare the patterns of introgression at individual loci in the hybrid zone, we used the hzar package v. 0.2.5 (Derryberry *et al*. 2014) in R using the Metropolis-Hastings Markov Chain Monte Carlo (MCMC) algorithm. For 232 species-diagnostic alleles (those with a frequency of 0 in one species and 1 in the other species; see Supplemetary material 1, Table S1) we fitted models of sigmoid clines and obtained the position of cline center and cline width for each locus. OaxL (*Q. laurina*) and TamA (*Q. affinis*) were used as reference populations. For all loci, we chose the allele with a frequency of 1 in OaxL and 0 in TamA. For each locus, we tested three models with differences in trait intervals and tail shape (observed values and none fitted; estimated values and none fitted; and estimated values and both tails estimated separately). The model with the lowest Akaike Information Criterion (AIC) score was selected as the best model. A null model was also included in the analysis. We ran each model for 100,000 MCMC using a burn-in period of 10,000. All clines were defined between 0 and 900 km. We repeated the analyses three times to ensure convergence of the parameters and used confidence intervals to compare cline center and width across all clines.

Genome scans for outlier detection. To detect candidate loci potentially under natural selection we used two outlier detection methods (Bayescan and Pcadapt). First, we applied Bayescan v. 2.1, which implements a Bayesian method based on differences in allele frequencies between populations (Foll & Gaggiotti 2008). We performed the analysis with 20 pilot runs with a lenght of 5,000 iterations, a burn-in length of 50,000 iterations and a thinning interval size of 10, prior odds set to 100 and a final run of 100,000 iterations. Then we ran the function *plot_bayescan* in R to plot and identify outlier loci based on a false discovery rate (FDR) < 0.0001. Functional annotation or genomic contexts of outlier SNPs were determined manually using the chromosome position of each SNP in the predicted gene models in the *Quercus lobata* genome (Sork *et al.* 2016a) and using UniProt/Swiss-Prot database (UniProt Consortium 2019) to identify the possible protein and function in the case of coding genes.

The second outlier method was implemented using the pcadapt v. 4.3.3 package (Privé *et al*. 2020) in R. We used $K = 2$, according to the scree plots of the proportion of explained variance (which is also in agreement to Δ*K* method, see results) to account for population structure. To correct for FDR we used the package qvalue v. 2.18.0 (Storey *et al*. 2019) in R with a cutoff of $q < 0.01$.

Genotype-environment association analysis. To detect signatures of environmental selection on SNPs, we implemented Latent Factor Mixed Models (LFMM) analysis to test for association between specific loci and climatic variables. For this, we used LFMM v. 2 (Caye *et al*. 2019) included in the Lea package (Frichot *et al*. 2013) in R.

Climatic data for sampled populations were drawn from the WorldClim database (Hijmans *et al*. 2005). Multicollinearity in the nineteen variables was detected calculating the variance inflation factor (VIF) with the package usdm v. 1.1-18 (Naimi 2023) in R and excluding variables with VIF > 5. Selected variables were bio10 (Mean Temperature of Warmest Quarter), bio11 (Mean Temperature of Coldest Quarter), bio15 (Precipitation Seasonality), bio18 (Precipitation of Warmest Quarter) and bio19 (Precipitation of Coldest Quarter).

Ten LFMM independent runs were implemented with 6,000 iterations and a burn-in of 3,000. To compute calibrated *p*-values we extracted the median |z|-scores and the χ^2 distribution for each set of associations to obtain the genomic inflation factor λ (Devlin & Roeder 1999). Visual observation of histograms of calibrated *p*-values was used to confirm latent factors effect for $K = 2$ (according to the Structure and ΔK method results, see below). To control type I error due to multiple comparisons, the FDR method (Benjamini & Hochberg 1995) was implemented. The correction of *P*-values was performed calculating the *q*-values using the package qvalue v. 2.18.0 (Storey *et al*. 2019) in R. We considered a SNP to be significantly associated with a climatic variable when $q < 0.01$ and $|z| > 2$. Then, we performed the functional annotation of the significantly associated SNPs following the same procedure explained above for those SNPs identified in the Bayescan analysis.

For visualizing the associations of spatial and climate variables with allele frequencies of populations we implemented the Gradient Forest (GF) analysis (Fitzpatrick & Keller 2015) using gradient forest package v. 1.6.1 (Ellis *et al*. 2012) in R. The model was performed for the complete set of SNPs (8,106) and for the 31 significantly climate-associated SNPs from LFMM (see results). For both GF models we used the five uncorrelated bioclimatic variables described above. The effects of spatial distribution on genetic variation was included obtaining distance-based Moran's eigenvectors (dbMEMs; Borcard *et al*. 2011, Legendre & Legendre 2012) from a geographic distance matrix using adespatial v. 0.3-8 (Dray *et al*. 2012) and adegraphics v. 1.0.15 (Siberchicot *et al*. 2017) in R. The GF analysis was performed as described in Fitzpatrick & Keller (2015): 2000 regression trees per SNP, maximum number of splits to evaluate maxLevel = $\log_2(0.368n)/2$ and a correlation threshold of 0.5 (Strobl et al. 2008, Ellis et al. 2012). Relative importance of predictors for the two SNP sets was assessed through R^2-weighted values. Changes in allele frequencies due to environmental and spatial gradients were estimated using the GF turnover function.

Additionally, we reduced multiple climatic variables into multivariate synthetic variables using non-scaled principal components analysis (PCA). Then, we performed Procrustes superimposition (Gower 1971, Jackson 1995) to compare results between the set of all SNPs and the set of climate-associated SNPs using vegan v. 2.5-6 (Oksanen *et al*. 2019) in R. This was performed to identify geographic areas where adaptive genomic variation deviates from neutral variation. Finally, a second scaled PCA for climate-associated SNPs was performed using the package pegas v. 0.13 to represent associations between predictor variables and differences in genetic composition of populations.

Results

Genomic data and samples. After discarding seven individuals with low quality data, the number of individuals genotyped per population ranged from 2 to 10 with a median of 5 per population and a total of 56 (Table 1), of which 34 were from six populations within the region that we call the contact zone (Ramos-Ortiz *et al*. 2016). In these 56 samples, the number of sequence reads obtained ranged between 1,936,179 and 6,670,057, with a mean of 2,964,218.2. We identified 8,106 SNPs after filtering for quality, depth, HWE, and LD. On average, 63 % of the samples included had less than 5.5 % of missing data, 23 % of the samples had 16.2 % of missing data, 7 % of the samples had 24.4 % of missing data and 3.6 % (two samples) had 35 % and 43 % of missing data, respectively. The mean depth of coverage per locus per sample was 26.5x.

Genetic diversity and population structure. The average gene diversity was $H_S = 0.283$. Three populations in the contact zone, HidA, PueHT and OaxH, showed the highest gene diversity ($H_S = 0.313$ to 0.323), while *Q. laurina* populations OaxLY, OaxL and OaxLL showed the lowest ($H_E = 0.236$ to 0.252) (Table 2). Overall differentiation across all populations was moderate ($F_{ST} = 0.137$; $P < 0.0001$); and pairwise F_{ST} ranged from very low to high (0.002

- 0.31), where the least differentiation was observed between the two populations of *Q. affinis* in the north of its distribution (TamA and SLPA), and the higher differentiation was observed between the southernmost population of *Q. laurina* (OaxLL) and the northern populations of *Q. affinis* (TamA and SLPA) (Figure 1B). Furthermore, there was a strong pattern of geographic structure according to Structure and PCA analyses. According to ΔK, the distribution of genetic variation is best explained with two ancestral genetic groups ($K = 2$), splitting populations of *Q. affinis* and *Q. laurina* into two distinct genetic groups (Figure 1C). Individuals of the populations HidA, PueHT and VerH showed admixture between *Q. affinis* and *Q. laurina* while a low degree of introgression was observed in the rest of the populations in the contact zone. Results obtained in the PCA analysis showed a similar pattern of geographic correspondence of individuals (Supplemetary material 2, Figure S1). PC1 explained 20 % of the variation and separated populations of *Q. affinis* (TamA and SLPA) and populations of *Q. laurina* (OaxLL, OaxL and OaxLY) at the two extremes of the axis. Admixed individuals from HidA, PueHT and VerH presented intermediate values along PC1, while the PC2 explained little variation (3.3 %) and mainly separated admixed populations from *Q. laurina* populations.

Table 2. Observed (H_O) and expected (H_S) heterozygosity and inbreeding coefficient (F) in sampled populations of *Q. affinis*, *Q. laurina* and the contact zone.

Species	Population	H_O	H_S	F confidence interval
Q. affinis	TamA	0.276	0.281	-0.002 0.036
	SLPA	0.266	0.279	0.032 0.060
Contact zone	HidA	0.286	0.313	0.076 0.098
	PueHT	0.248	0.310	-0.227 -0.131
	PueHZ	0.302	0.291	-0.052 -0.029
	VerH	0.265	0.291	0.071 0.105
	OaxH	0.278	0.323	0.129 0.155
	OaxLC	0.238	0.282	0.137 0.176
Q. laurina	OaxLY	0.233	0.247	0.042 0.077
	OaxL	0.229	0.252	0.072 0.105
	OaxLL	0.226	0.236	-0.252 -0.168
Overall		0.259	0.282	0.099

Genomic cline analyses. Sigmoid clines fitted to allele frequency changeover at 232 loci with alternative alleles fixed in reference populations of *Q. affinis* and *Q. laurina* are shown in Figure 2. The average cline center was situated at the km 430.6 (average confidence interval 362.2-517.2), counted from population OaxL, and coincided approximately with the location of population PueHT. The average cline width was 387.2 km (273.2-644.9). Six loci showed markedly narrower cline widths whose confidence intervals did not overlap with mean confidence intervals (Figure 2). Also, the cline center of 25 loci was displaced to the right of the average center and for 18 loci the center was displaced to the left of the average center. Four of these displaced loci also were among the ones with narrow clines, and of the total 45 loci with discordant clines, six were F_{ST} outliers too (see below; Supplemetary material 1, Tables S2, S3 and S4; Supplemetary material 2, Figure S2). Several loci corresponded to coding regions with functional roles such as leaf development, herbivory or pathogen response, symbiont interaction, etc.

Genome scans for outlier detection. Bayescan analysis detected 92 candidate loci under divergent selection (Figure 3) of which 73 are located in 51 genes (3 uncharacterized genes) and 19 in noncoding or unknown regions (Supplemetary material 1, Table S5). Of these 51 SNPs, twenty-nine are in exons and 22 in introns. The overall population

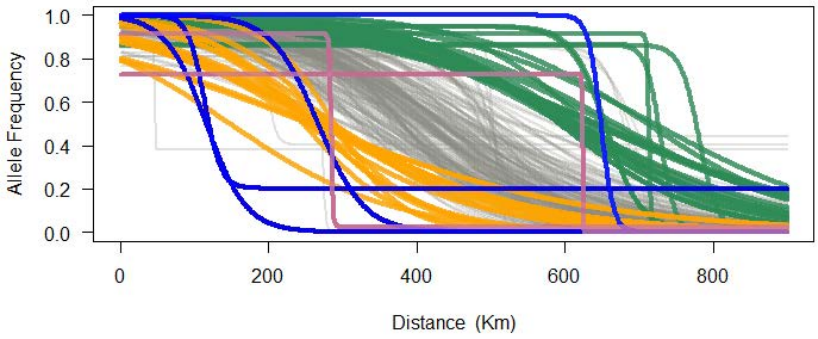

Figure 2. Models of sigmoid clines fitted to frequency changeover for 232 species-diagnostic alleles across the hybrid zone between *Q. affinis* and *Q. laurina*. Pink color indicates clines that are narrower than the average; blue, clines that are both narrower and displaced from the average cline center; and green and orange, clines that are displaced to the right and the left from the average center, respectively.

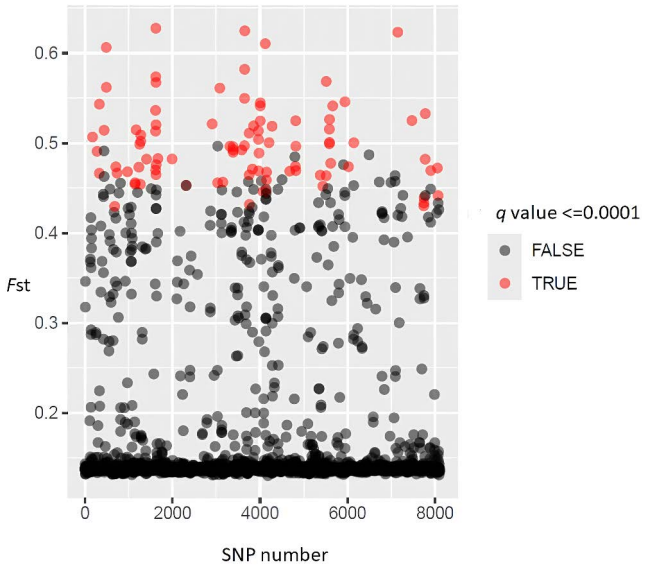

Figure 3. Bayescan plot for 8,106 SNPs of *Quercus affinis* and *Q. laurina.* Red points represent the 92 significant outlier loci detected by the analysis.

differentiation for the SNPs detected by Bayescan was considerably stronger than for the full SNP set (F_{ST} = 0.643, P < 0.0001). Pcadapt identified 96 outlier SNPs (Table S6), also considerably differentiated (F_{ST} = 0.373, P < 0.0001).

Genotype-Environment Association analysis. Using K = 2 from the Structure analysis as the number of latent factors in LFMM analysis, we found uniform distribution of the frequencies of P-values for the five bioclimatic variables included (Supplemetary material 2, Figure S3), ensuring an efficient control of the false discovery rate (François *et al*. 2016). We found 31 SNPs significantly associated (q < 0.01 and |z|-scores > 2) with climate variables (Figure 4), of which 26 are located in 19 genes (five uncharacterized genes) and five located in noncoding or unknown regions (Supplemetary material 1, Table S7). Of these 26 SNPs, ten are located in exons, while 16 are mutations located in introns. Twenty-five SNPs were associated with precipitation of coldest quarter (bio19), five SNPs to mean temperature of the warmest quarter (bio10), and two SNPs to mean temperature of the coldest quarter (bio11) (SNP 6388 was associated with both bio10 and bio11). We did not find SNPs associated with precipitation variables bio15 and bio18.

In overall, several SNPs were identified as atypical by more than one method. There were 25 SNPs in common between Bayescan and Pcadapt, and two SNPs in common between Pcapat and LFMM (Supplemetary material 2, Figure S2).

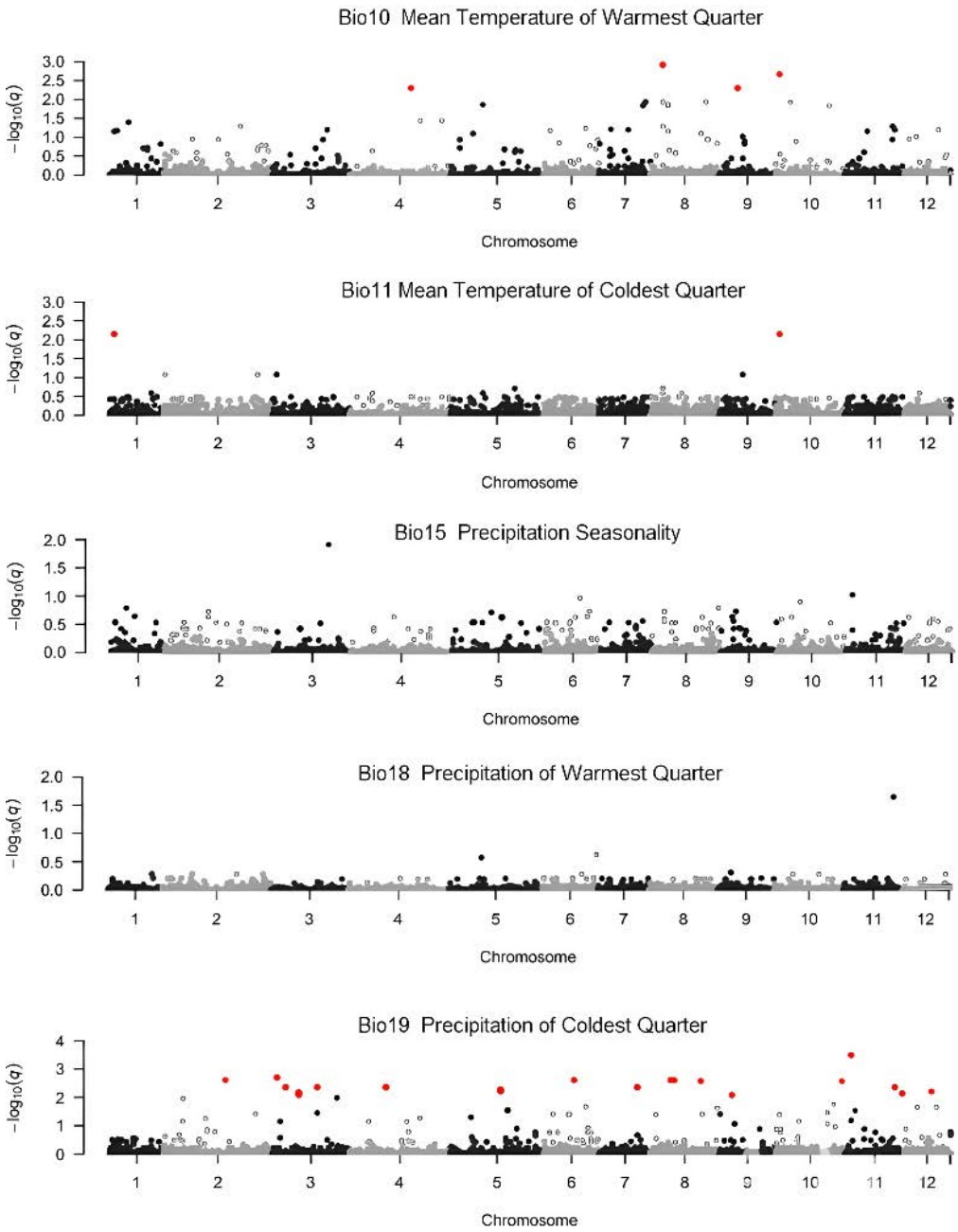

Figure 4. Results of the Latent Factors Mixed Model (LFMM) analysis. Red dots represent SNPs significantly associated with climate variables controlling for two genetic groups.

Identification of local adaptation. The GF model using all SNPs showed that dbMEM spatial variables and precipitation seasonality (bio15) were the most important predictors of the turnover in allele frequency across the landscape, while for the set of climate-associated SNPs precipitation seasonality (bio15) and precipitation of coldest quarter (bio19) were the two principal predictors, with similar R^2-weighted importance for both datasets (Supplemetary material 2, Figure S4). The model using climate-associated SNPs explained more variation along the gradients, according to R^2 values. A rapid turnover in allele frequencies was observed in SNPs associated with precipitation seasonality (bio15) and precipitation of coldest quarter (bio19) (Figure 5). PC1 (Figure 6D) separated the northern populations of *Quercus affinis* TamA and SLPA from the southern populations OaxL and

OaxLL explained mainly by the dbMEM spatial variable and precipitation seasonality (bio15), while admixed populations showed intermediate values along PC1. PC2 separated the VerH population from all the other populations, explaining allele frequency composition and divergence by the strong effect of precipitation of the coldest quarter (bio19) (Figure 6D). Admixed populations HidA and PueHT are more similar in genetic composition to *Q. affinis*, whose turnover is explained by the dbMEM spatial variable, while admixed populations PueHZ and OaxH share similar allele frequency composition with *Q. laurina* populations, explained mainly by bio15 (Figure 6D). Also, the Procrustes analysis revealed that the highest deviation between the full set (Figure 6A) and the climate-associated SNPs (Figure 6B) was in the central and eastern populations, with the highest value in VerH. These could indicate that local adaptation could be more intense in these areas (Figure 6C). Temperature variables bio10 and bio11, and precipitation variable bio18 did not have a strong effect on the genetic composition of the populations.

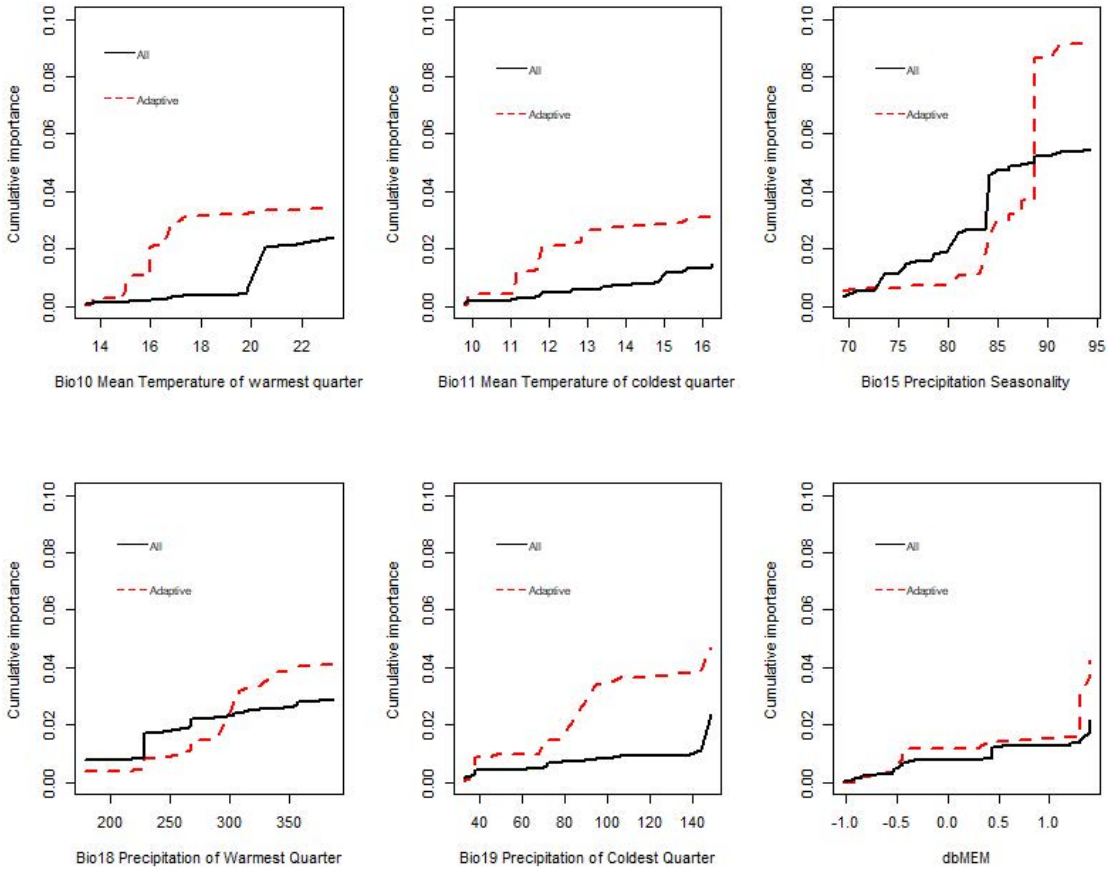

Figure 5. Comparison between allele frequency turnover in relation to values of bioclimatic variables using all SNPs and only climate-associated SNPs in *Quercus affinis* and *Quercus laurina* populations.

Discussion

The present research supports previous studies reporting hybridization between *Q. affinis* and *Q. laurina* (González-Rodríguez *et al.* 2004, González-Rodríguez & Oyama 2005, Ramos-Ortiz *et al.* 2016). However, here, the use of more than 8,000 SNPs has allowed a detailed description of genetic diversity, proportion of admixed individuals, and allele frequency composition associated with environmental gradients. Our results also agree with those obtained in other oaks (Reutimann *et al*. 2023), and are useful to understand the influence of the environment on the genetic

composition of admixed populations (Dodd & Afzal-Rafii 2004, Hamilton & Aitken 2013, Sullivan *et al*. 2016, Ortego *et al*. 2018).

Genetic structure and diversity. Quercus laurina and *Q. affinis* are not closely related taxa according to recent phylogenetic analyses (Hipp *et al*. 2019). Allopatric populations of the two species showed marked genetic differentiation, like that found for other clearly distinct oak species (Grivet *et al.* 2008, Ortego *et al*. 2018). On the other hand, we characterized admixture levels in the contact zone between *Q. affinis* and *Q. laurina* and found that three of the sampled populations combine the two parental genetic groups, HidA, PueHT and VerH. Whereas populations HidA and VerH are more related to *Q. affinis* populations, PueHT is intermediate between *Q. affinis* and *Q. laurina*. This indicates extensive levels of hybridization between the two species in the contact zone. Additionally, genetic diversity values (*H*s) were higher in these three admixed populations compared to either *Q. affinis* or *Q. laurina* populations outside the contact zone, supporting the notion that hybridization may enhance genetic diversity with potentially important ecological and evolutionary consequences (Stankowski *et al*. 2021).

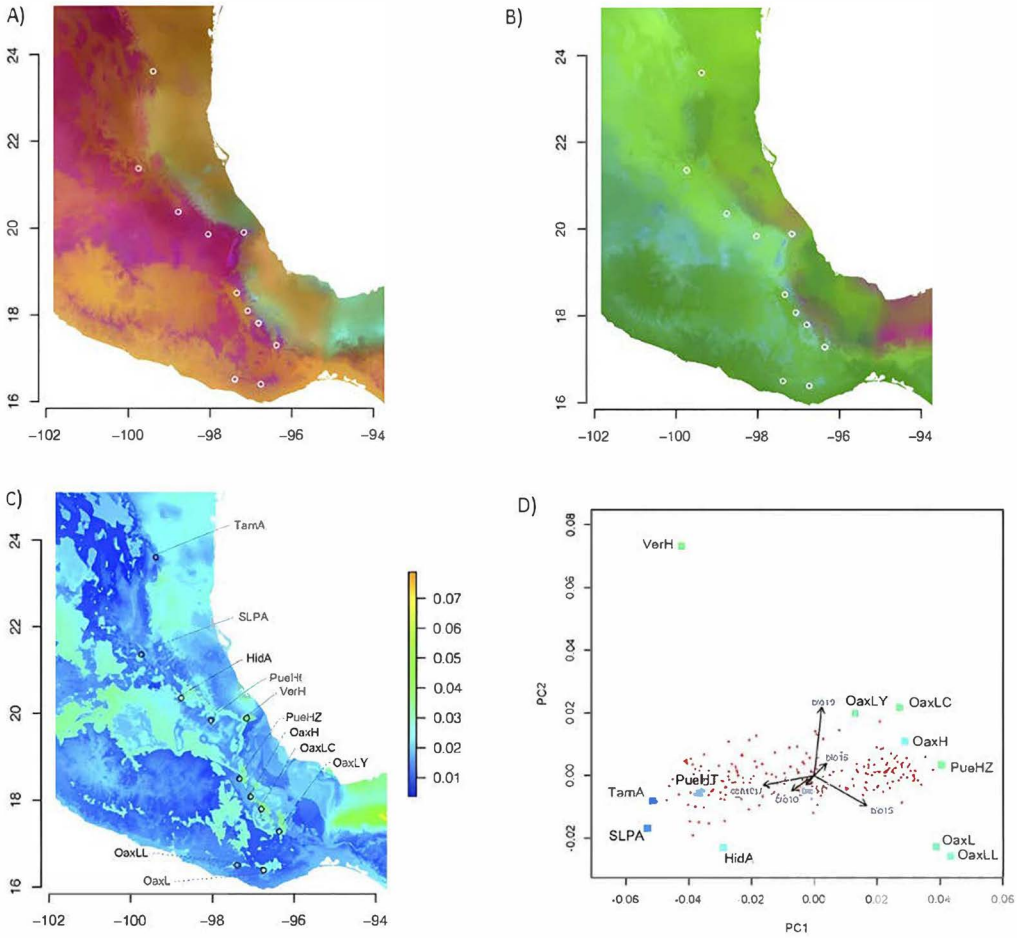

Figure 6. Gradient forest analysis showing spatial variation in genomic composition of *Quercus affinis* and *Q. laurina* populations in association to environmental gradients for the full set of SNPs (A) and climate-associated SNPs (B), and Procrustes superimposition analysis for the comparison between adaptive SNPs and the full set of SNPs (C). Colors in (A) and (B) represent the genetic turnover based on the modelled relationships of allele frequencies with environmental and spatial variables. Similar colors represent localities with similar expected genetic composition. Note that colors are not comparable between the two panels. In (C), warmer colors represent the largest difference between the two SNP sets. Principal components analysis based on climate-associated SNPs showing the effect of bioclimatic variables as vectors (D).

Local adaptation and landscape distribution of the genomic variation. The genus *Quercus* has been proposed as a model group for a species concept that relies on ecological criteria, rather than reproductive isolation, to delimit boundaries among taxa (Valen 1976). Long-distance gene flow characteristic of oaks is suggested, then, to break the ecological isolation needed for new species formation, resulting in complex hybrid systems (Whittemore & Schaal 1991, Muir & Schlötterer 2005, Petit *et al*. 2004, Valbuena-Carabaña *et al*. 2005). In contrast, barriers to gene flow in hybridizing species may be maintained through differential environmental pressures (Dodd & Afzal-Rafii 2004, Hamilton & Aitken 2013), and high population differentiation between populations may be indicative of local adaptation. In agreement with this hypothesis, oak species are frequently stratified along water and nutrient availability gradients and differ in their adaptations to drought, light availability and temperature (Cavender-Bares *et al*. 2004, Cavender-Bares 2019). In this sense, revealing the role of environmental variables in the formation of the population genetic composition is a major concern of landscape genomics (Barley *et al*. 2015, Leamy *et al*. 2016). Detecting candidate outlier loci by genome scan analysis and candidate loci with significant relationships between environmental gradients and allele frequencies is essential to identify genes potentially involved in local adaptation (Coop *et al*. 2010, Frichot *et al*. 2013). Our results revealed 92 outlier SNPs with Bayescan, 96 with Pcadapt, and 31 with LFMM which were associated with precipitation and temperature variables. Several SNPs were identified by more than one method, increasing confidence in their detection as potentially adaptive candidate loci. It was possible to identify genes in which some of those SNPs are found, most of which are well characterized coding genes in *Arabidopsis thaliana* (L.) Heynh., *Oriza sativa* L., *Nicotiana tabacum* L., and *Coffea arabica* L. (Supplemetary material 1, Tables S5 and S7) (Nozue *et al.* 1996, Li *et al*. 2001, Forouhar *et al*. 2005, Li *et al*. 2012, Li *et al*. 2016, Chang et al. 2021). These genes are related to plant biotic and abiotic stress responses and involved in processes such as DNA repair, antioxidant defense, embryo development, desiccation tolerance, plant innate and basal immunity, cell integrity of rapidly growing tissue, organ development, regulation of flowering and expression control on circadian oscillation, seed size, germination, growth and production, plant architecture determination, leaf senescence, cold shock tolerance, fungal susceptibility, carbon assimilation, respiration mechanisms, phosphorous assimilation, promotion of cell death, disease tolerance (antiviral resistance) and inflorescence architecture (more detail in Supplemetary material 1, Tables S5 and S7). These candidate genes identified in this work represent new relevant information for future studies about phenotypic response under stress conditions in oaks.

On the other hand, the high overall F_{ST} value obtained for the SNPs detected by Bayescan (0.643) and Pcadapt (0.372) indicates that other factors, different from the climatic gradients, are causing a strong differentiation at some loci. Hybrid zone theory indicates that selection against recombination between parental genomes can be either intrinsic or extrinsic (Stankowski *et al*. 2021). Intrinsic selection occurs when introgressed alleles are selected against in the heterospecific genomic context, because of the disruption of gene coadaptation and regulation networks usually reflected in a lower fitness of hybrids compared to parental individuals (Ramos-Ortiz *et al.* 2016, Stankowski *et al*. 2021). In turn, extrinsic selection occurs when the relative fitness of parental and hybrid individuals is variable across an ecological gradient due to genotype-environment interactions, with hybrids possibly showing higher fitness in some environments (Stankowski *et al*. 2021). In this context, the analysis of genomic clines was particularly revealing, because it allowed the identification of several loci with markedly narrow clines, suggesting that selection may be particularly strong against introgression of species-specific alleles in these genomic regions (Stankowski *et al.* 2021). In turn, non-coincident (displaced) clines, may indicate the action of distinct extrinsic selection pressures on different parts of the genome. However, understanding whether these patterns are due to selection by ecological factors that we did not account for (i.e. soil properties, biotic interactions, etc.), will require further studies.

Our results about genotype-environment association suggest that precipitation seasonality and precipitation of the coldest quarter are variables that exert a marked influence on the multilocus genomic composition of *Q. affinis* and *Q. laurina* populations, which is supported by the rapid allele frequency turnover at these loci, unlike background genome composition. This result is congruent with the observed response to climate variables in other Mexican oak species such as *Q. rugosa* with distribution across the Trans-Mexican Volcanic Belt (Martins *et al.* 2018). One interesting case in our results was the marked differentiation at climate-associated SNPs observed for the admixed

population VerH and mainly explained by precipitation of the coldest quarter (bio19). This population is located in the oriental face of the Sierra Madre Oriental and receives a considerable higher annual precipitation (up to 4,000 mm) compared to the other populations included in the study (600-1,200 mm) and with less seasonality, since winter precipitations represent more than 18 % of the total annual precipitation (Hernández-Cerda & Carrasco-Anaya 2004). The distinctness of this population is also supported by the Procrustes analysis, which indicated that gradient forest models differed considerably between the full set of SNPs and the climate-associated SNPs at this site.

In conclusion, in this study we have obtained a more detailed view of the genetic structure of the hybrid zone between *Q. affinis* and *Q. laurina* than was possible in earlier studies based on a few marker loci. As commonly reported for plant hybrid populations, we found evidence of increased genetic variation in three genetically admixed populations located in the contact zone. We identified a number of loci with elevated differentiation among populations, suggesting either intrinsic or extrinsic divergent selection pressures at some regions of the genome of these oak species. We also identified loci that follow climatic gradients mostly driven by precipitation of the coldest quarter. Finally, the gradient forest analysis allowed visualizing the spatial turnover of genomic composition of populations across the landscape and revealed the strong genetic distinctness of one population (VerH), with potential implications for conservation and management of these valuable tree populations.

Supplementary material

Supplemental data for this article can be accessed here: https://doi.org/10.17129/botsci.3707

Acknowledgements

We thank Jesús Llanderal-Mendoza and Violeta Patiño for laboratory support. We also thank the Laboratorio Nacional de Análisis y Síntesis Ecológica (LANASE, ENES Morelia, UNAM) and Diego Islas López for technical support in computational resources for bioinformatic analyses. We also thanks comments by two anonymous reviewers and the Associate Editor, which greatly contributed to improve the earlier version of this article.

Literature cited

Abbott RJ, Barton NH, Good JM. 2016. Genomics of hybridization and its evolutionary consequences. *Molecular Ecology* **25**: 2325-2332. DOI: https://doi.org/10.1111/mec.13685

Arnold ML, Bennett BD. 1993. Natural hybridization in Louisiana Irises: genetic variation and ecological determinants. Hybrid zones and the evolutionary process 115: 139 DOI: https://doi.org/10.1093/oso/9780195069174.003.0005

Barley AJ, Monnahan PJ, Thomson RC, Grismer LL, Brown RM. 2015. Sun skink landscape genomics: Assessing the roles of micro-evolutionary processes in shaping genetic and phenotypic diversity across a heterogeneous and fragmented landscape. *Molecular Ecology* **24**: 1696-1712. DOI: https://doi.org/10.1111/mec.13151

Benjamini Y, Hochberg Y. 1995. Controlling the false discoveryrate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)* **57**: 289-300. DOI: https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**: 2114-2120. DOI: https://doi.org/10.1093/bioinformatics/btu170

Borcard D, Gillet F, Legendre P. 2011. *Numerical Ecology with R, Second Edition*. Cham, Switzerland: Springer. ISBN: 978-3-319-71403-5

Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH. 2011. Stacks: Building and genotyping loci de novo from short-read sequences. *G3: Genes, Genomes, Genetics* **1**: 171-182. DOI: https://doi.org/10.1534/g3.111.000240

Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA. 2013. Stacks: An analysis tool set for population genomics. *Molecular Ecology* **22**: 3124-3140. DOI: https://doi.org/10.1111/mec.12354

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**: 669-692. DOI: https://doi.org/10.1111/nph.15450

Cavender-Bares J, Kitajima K, Bazzaz FA. 2004. Multiple trait associations in relation to habitat differentiation among 17 Floridian oak species. *Ecological Monographs* **74**: 635-662. DOI: https://doi.org/10.1890/03-4007

Caye K, Jumentier B, Lepeule J, François O. 2019. LFMM 2: Fast and accurate inference of gene-environment associations in genome-wide studies. *Molecular Biology and Evolution* **36**: 852-860. DOI: https://doi.org/10.1093/molbev/msz008

Chang A, Jeske L, Ulbrich S, Hofmann J, Koblitz J, Schomburg I, Neumann-Schaal M, Jahn D, Schomburg D. 2021. BRENDA, the ELIXIR core data resource in 2021: new developments and updates. *Nucleic Acids Research* **49**: D498-D508. DOI: https://doi.org/10.1093/nar/gkaa1025

Coop G, Witonsky D, Di Rienzo A, Pritchard JK. 2010. Using environmental correlations to identify loci underlying local adaptation. *Genetics* **185**: 1411-1423. DOI: https://doi.org/10.1534/genetics.110.114819

Curtu AL, Gailing O, Finkeldey R. 2007. Evidence for hybridization and introgression within a species-rich oak (*Quercus* spp.) community. *BMC Evolutionary Biology* **7**: 1-15. DOI: https://doi.org/10.1186/1471-2148-7-218

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R. 2011. The variant call format and VCFtools. *Bioinformatics* **27**: 2156-2158. DOI: https://doi.org/10.1093/bioinformatics/btr330

De Kort H, Vandepitte K, Mergeay J, Mijnsbrugge KV, Honnay O. 2015. The population genomic signature of environmental selection in the widespread insect-pollinated tree species *Frangula alnus* at different geographical scales. *Heredity* **115**: 415-425. DOI: https://doi.org/10.1038/hdy.2015.41

Derryberry EP, Derryberry GE, Maley JM, Brumfield RT. 2014. HZAR: hybrid zone analysis using an R software package. *Molecular Ecology Resources* **14:** 652-663. DOI: https://doi.org/10.1111/1755-0998.12209

Devlin B, Roeder K. 1999. Genomic control for association studies. *Biometrics* **55**: 997-1004. DOI: https://doi.org/10.1111/j.0006-341X.1999.00997.x

Dodd RS, Afzal-Rafii Z. 2004. Selection and dispersal in a multispecies oak hybrid zone. *Evolution* **58**: 261-269 DOI: https://doi.org/10.1111/j.0014-3820.2004.tb01643.x

Dray S, Pélissier R, Couteron P, Fortin MJ, Legendre P, Peres-Neto PR, Bellier E, Bivand R, Blanchet FG, De Cáceres M, Dufour AB, Heegaard E, Jombart T, Munoz F, Oksanen J, Thioulouse J, Wagner HH. 2012. Community ecology in the age of multivariate multiscale spatial analysis. *Ecological Monographs* **82**: 257-275. DOI: https://doi.org/10.1890/11-1183.1

Earl DA, von Holdt BM. 2012. STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* **4**: 359-361. DOI: https://doi.org/10.1007/s12686-011-9548-7

Ellis N, Smith SJ, Roland Pitcher, C. 2012. Gradient forests: Calculating importance gradients on physical predictors. *Ecology* **93**: 156-168. DOI: https://doi.org/10.1890/11-0252.1

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One* **6**: e19379. DOI: https://doi.org/10.1371/journal.pone.0019379

Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology* **14**: 2611-2620. DOI: https://doi.org/10.1111/j.1365-294X.2005.02553.x

Fitzpatrick MC, Keller SR. 2015. Ecological genomics meets community-level modelling of biodiversity: Mapping the genomic landscape of current and future environmental adaptation. *Ecology Letters* **18**: 1-16. DOI: https://doi.org/10.1111/ele.12376

Foll M, Gaggiotti O. 2008. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics* **180**: 977-993. DOI: https://doi.org/10.1534/genetics.108.092221

François O, Martins H, Caye K, Schoville SD. 2016. Controlling false discoveries in genome scans for selection. *Molecular Ecology* **25**: 454-469. DOI: https://doi.org/10.1111/mec.13513

Frichot E, Schoville SD, Bouchard G, François O. 2013. Testing for associations between loci and environmental gradients using latent factor mixed models. *Molecular Biology and Evolution* **30**: 1687-1699. DOI: https://doi.org/10.1093/molbev/mst063

Futuyma DJ, Shapiro LH, Harrison RG. 1995. Hybrid zones and the evolutionary process. *Evolution* **49**: 222-226. DOI: https://doi.org/10.2307/2410309

González-Rodríguez A, Arias DM, Oyama K. 2005. Genetic variation and differentiation of populations within the *Quercus affinis* - *Quercus laurina* (Fagaceae) complex analyzed with RAPD markers. *Canadian Journal of Botany* **83**: 155-162. DOI: https://doi.org/10.1139/B04-162

González-Rodríguez A, Arias DM, Valencia S, Oyama K. 2004. Morphological and RAPD analysis of hybridization between *Quercus affinis* and *Q. laurina* (Fagaceae), two Mexican red oaks. *American Journal of Botany* **91**: 401-409. DOI: https://doi.org/10.3732/ajb.91.3.401

González-Rodríguez A, Oyama K. 2005. Leaf morphometric variation in *Quercus affinis* and *Q. laurina* (Fagaceae), two hybridizing Mexican red oaks. *Botanical Journal of the Linnean Society* **147**: 427-435. DOI: https://doi.org/10.1111/j.1095-8339.2004.00394.x

Goudet J, Jombart T. 2015. hierfstat: Estimation and tests of hierarchical F-Statistics. R package version 0.04-22. https://CRAN.R-project.org/package=hierfstat. R Core Team.

Gower JC. 1971. Statistical methods of comparing different multivariate analyses of the same data. Mathematics in the Archaeological and Historical Science. *In*: Hodson FR, Kendall DG, Tautu P, eds. Proceedings of the Anglo-Roumanian Conference Mamia 1970. Edinburgh: Edinburgh University Press, pp. 138-149.

Grivet D, Sork VL, Westfall RD, Davis FW. 2008. Conserving the evolutionary potential of California valley oak (*Quercus lobata* Née): a multivariate genetic approach to conservation planning. *Molecular Ecology* **17**: 139-156. DOI: https://doi.org/10.1111/j.1365-294X.2007.03498.x

Gugger PF, Cokus SJ, Sork VL. 2016. Association of transcriptome-wide sequence variation with climate gradients in valley oak (*Quercus lobata*). *Tree Genetics and Genomes* **12**: 1-14. DOI: https://doi.org/10.1007/s11295-016-0975-1

Hamilton JA, Aitken SN. 2013. Genetic and morphological structure of a spruce hybrid (*Picea sitchensis* × *P. glauca*) zone along a climatic gradient. *American Journal of Botany* **100**: 1651-1662. DOI: https://doi.org/10.3732/ajb.1200654

Hamilton JA, Miller JM. 2016. Adaptive introgression as a resource for management and genetic conservation in a changing climate. *Conservation Biology* **30**: 33-41. DOI: https://doi.org/10.1111/cobi.12574

Hernández-Cerda M, Carrasco-Anaya G. 2004. Climatología. *In*: Luna-Vega I, Morrone J, Espinosa D, eds. *Biodiversidad de la Sierra Madre Oriental*. DF, México: Universidad Nacional Autónoma de México/ Comisión Nacional para el Conocimiento y Uso de la Biodiversidad, pp. 63-108. ISBN 970-32-1526-2

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. 2005. Very high-resolution interpolated climate surfaces for global land areas. *International Journal of Climatology: A Journal of the Royal Meteorological Society* **25**: 1965-1978.DOI: https://doi.org/10.1002/joc.1276

Hipp AL, Manos PS, Hahn M, Avishai M, Bodénès C, Cavender-Bares J, Valencia-Avalos S, Crowl AA, Deng M, Denk T, Fitz-Gibbon S, Gailing O, González-Elizondo MS, González-Rodríguez A, Grimm GW, Jiang XL, Kremer A, Lesur I, McVay JD, Plomion C, Rodríguez-Correa H, Schulze ED, Simeone MC, Victoria L. Sork. 2019. Genomic landscape of the global oak phylogeny. *New Phytologist* **226**: 1198-1212. DOI: https://doi.org/10.1111/nph.16162

Jackson DA. 1995. PROTEST: A PROcrustean Randomization TEST of community environment concordance. *Écoscience* **2**: 3. DOI: https://doi.org/10.1080/11956860.1995.11682297

Jombart T. 2008. Adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**: 1403-1405. DOI: https://doi.org/10.1093/bioinformatics/btn129

Joost S, Vuilleumier S, Jensen JD, Schoville S, Leempoel K, Stucki S, Widmer I, Melodelima C, Rolland J, Manel, S. 2013. Uncovering the genetic basis of adaptive change: On the intersection of landscape genomics and theoretical population genetics. *Molecular Ecology* **22**: 3659-3665. DOI: https://doi.org/10.1111/mec.12352

Kremer A, Hipp AL. 2019. Oaks: an evolutionary success story. *New Phytologist* **226**: 987-1011. DOI: https://doi.org/10.1111/nph.16274

Kremer A, Ronce O, Robledo-Arnuncio JJ, Guillaume F, Bohrer G, Nathan R, Bridle JR, Gomulkiewicz R, Klein EK, Ritland K, Kuparinen A, Gerber S, Schueler S. 2012. Long-distance gene flow and adaptation of forest trees to rapid climate change. *Ecology Letters* **15**: 378-392. DOI: https://doi.org/10.1111/j.1461-0248.2012.01746.x

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology* **10**: 1-10. DOI: https://doi.org/10.1186/gb-2009-10-3-r25

Leamy LJ, Lee CR, Song Q, Mujacic I, Luo Y, Chen CY, Li C, Kjemtrup S, Song BH. 2016. Environmental versus geographical effects on genomic variation in wild soybean (*Glycine soja*) across its native range in northeast Asia. *Ecology and Evolution* **6**: 6332-6344. DOI: https://doi.org/10.1002/ece3.2351

Legendre P, Legendre L. 2012. Multiscale analysis: Spatial eigenfunctions. *In*: Legendre P, Legendre L, eds. *Developments in Environmental Modelling*. Oxford: Elsevier, pp. 859-906. DOI: https://doi.org/10.1016/B978-0-444-53868-0.50014-9

Li J, Jia D, Chen X. 2001. HUA1, a regulator of stamen and carpel identities in *Arabidopsis*, codes for a nuclear RNA binding protein. *The Plant Cell* **13**: 2269-2281. DOI: https://doi.org/10.1105/tpc.010201

Li ZY, Xu ZS, He GY, Yang GX, Chen M, Li LC, Ma YZ. 2012. A mutation in *Arabidopsis* BSK5 encoding a brassinosteroid-signaling kinase protein affects responses to salinity and abscisic acid. *Biochemical and Biophysical Research Communications* **426**: 522-527. DOI: https://doi.org/10.1016/j.bbrc.2012.08.118

Li S, Yang Z, Du X, Liu R, Wilkinson AW, Gozani O, Du J. 2016. Structural basis for the unique multivalent readout of unmodified H3 tail by *Arabidopsis* ORC1b BAH-PHD cassette. *Structure* **24**: 486-494. DOI: https://doi.org/10.1016/j.str.2016.01.004

Liu Y, Huang X, Li M, He P, Zhang Y. 2016. Loss-of-function of *Arabidopsis* receptor-like kinase BIR1 activates cell death and defense responses mediated by BAK1 and SOBIR1. *New Phytologist* **212**: 637-645. DOI: https://doi.org/10.1111/nph.14072

Martin SH, Jiggins CD. 2017. Interpreting the genomic landscape of introgression. *Current Opinion in Genetics and Development* **47**: 69-74. DOI: https://doi.org/10.1016/j.gde.2017.08.007

Martins K, Gugger PF, Llanderal-Mendoza J, González-Rodríguez A, Fitz-Gibbon ST, Zhao JL, Rodríguez-Correa H, Oyama K, Sork VL. 2018. Landscape genomics provides evidence of climate-associated genetic variation in Mexican populations of *Quercus rugosa*. *Evolutionary Applications* **11**: 1842-1858. DOI: https://doi.org/10.1111/eva.12684

Martinsen GD, Whitham TG, Turek RJ, Keim P. 2001. Hybrid populations selectively filter gene introgression between species. *Evolution* **55**: 1325-1335. DOI: https://doi.org/10.1111/j.0014-3820.2001.tb00655.x

Muir G, Schlötterer C. 2005. Evidence for shared ancestral polymorphism rather than recurrent gene flow at microsatellite loci differentiating two hybridizing oaks (*Quercus* spp.). *Molecular Ecology* **14**: 549-561. DOI: https://doi.org/10.1111/j.1365-294X.2004.02418.x

Naimi B. 2023. Package 'usdm'. Uncertainty analysis for species distribution models. . https://cran.r-project.org/web/packages/usdm/index.html (accessed March 2, 2025).

Nei M. 1987. *Molecular Evolutionary Genetics*. New York: Columbia University Press. DOI: https://doi.org/10.7312/nei-92038

Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, Mcglinn D, Minchin P, O'Hara R, Simpson G, Solymos P, Maintainer HW. 2019. vegan: Community Ecology Package. R package version 2.5-5. https://CRAN.R-project.org/package=vegan Community Ecology Package (accessed March 2, 2025).

Ortego J, Gugger PF, Sork VL. 2018. Genomic data reveal cryptic lineage diversification and introgression in Cali-

fornian golden cup oaks (section *Protobalanus*). *New Phytologist* **218**: 804-818 DOI: https://doi.org/10.1111/nph.14951

Paradis E. 2010. Pegas: An R package for population genetics with an integrated-modular approach. *Bioinformatics* **26**: 419-420. DOI: https://doi.org/10.1093/bioinformatics/btp696

Peñaloza-Ramírez JM, González-Rodríguez A, Mendoza-Cuenca L, Caron H, Kremer A, Oyama K. 2010. Interspecific gene flow in a multispecies oak hybrid zone in the Sierra Tarahumara of Mexico. *Annals of Botany* **105**: 389-399. DOI: https://doi.org/10.1093/aob/mcp301

Petit RJ, Bodénes C, Ducousso A, Roussel G, Kremer A. 2004. Hybridization as a mechanism of invasion in oaks. *New Phytologist* **161**: 151-164. DOI: https://doi.org/10.1046/j.1469-8137.2003.00944.x

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* **155**: 945-959. DOI: https://doi.org/10.1093/genetics/155.2.945

Privé F, Luu K, Vilhjálmsson BJ, Blum MG. 2020. Performing highly efficient genome scans for local adaptation with R package pcadapt version 4. *Molecular Biology and Evolution* **37**: 2153-2154. DOI: https://doi.org/10.1093/molbev/msaa053

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklarb P, de Bakkerb PIW, Dalyb MJ, Sham PC. 2007. PLINK: A tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics* **81**: 559-575. DOI: https://doi.org/10.1086/519795

R Core Team. 2021. R: A language and environment for statistical computing, v.4.1.2. R foundation for Statistical Computing. http://www.r-project.org

Ramos-Ortiz S, Oyama K, Rodríguez-Correa H, González-Rodríguez A. 2016. Geographic structure of genetic and phenotypic variation in the hybrid zone between *Quercus affinis* and *Q. laurina* in Mexico. *Plant Species Biology* **31**: 219-232. DOI: https://doi.org/10.1111/1442-1984.12109

Reutimann O, Dauphin B, Baltensweiler A, Gugerli F, Kremer A, Rellstab C. 2023. Abiotic factors predict taxonomic composition and genetic admixture in populations of hybridizing white oak species (*Quercus* sect. *Quercus*) on regional scale. *Tree Genetics & Genomes* **19**: 22. https://doi.org/10.1007/s11295-023-01598-7

Rius M, Darling JA. 2014. How important is intraspecific genetic admixture to the success of colonising populations? *Trends in Ecology and Evolution* **29**: 233-242. DOI: https://doi.org/10.1016/j.tree.2014.02.003

Rogers SO, Bendich AJ. 1985. Extraction of DNA from milligram amounts of fresh, herbarium and mummified plant tissues. *Plant Molecular Biology* **5**: 69-76. DOI: https://doi.org/10.1007/BF00020088

Rosenberg NA. 2004. DISTRUCT: A program for the graphical display of population structure. *Molecular Ecology Notes* **4**: 137-138. DOI:

Seehausen O. 2013. Conditions when hybridization might predispose populations for adaptive radiation. *Journal of Evolutionary Biology* **26**: 279-281. DOI: https://doi.org/10.1111/jeb.12026

Siberchicot A, Julien-Laferrière A, Dufour A, Thioulouse J, Dray S. 2017. adegraphics: An S4 lattice-based package for the representation of multivariate data. *The R Journal* **9**: 198–212. https://journal.r-project.org/archive/2017/RJ-2017-042/index.html

Sork VL, Fitz-Gibbon ST, Puiu D, Crepeau M, Gugger PF, Sherman R, Stevens K, Langley CH, Pellegrini M, Salzberg SL. 2016a. First draft assembly and annotation of the genome of a California endemic oak *Quercus lobata* Née (Fagaceae). *G3: Genes, Genomes, Genetics* **6**: 3485-3495. DOI: https://doi.org/10.1534/g3.116.030411

Sork VL, Squire K, Gugger PF, Steele SE, Levy ED, Eckert AJ. 2016b. Landscape genomic analysis of candidate genes for climate adaptation in a California endemic oak, *Quercus lobata*. *American Journal of Botany* **103**: 33-46. DOI: https://doi.org/10.3732/ajb.1500162

Stankowski S, Shipilina D, Westram AM. 2021. Hybrid zones, *eLS* **2**: 1–16. https://doi.org/10.1002/9780470015902.a0029355

Storey JD, Bass AJ, Dabney A, Robinson D. 2019. qvalue: Q-value estimation for false discovery rate control. R package version 2.18.0, http://github.com/jdstorey/qvalue (accessed March 5, 2025).

Strobl C, Boulesteix AL, Kneib T, Augustin T, Zeileis, A. 2008. Conditional variable importance for random forests. *BMC Bioinformatics* **9**: 1-11. DOI: https://doi.org/10.1186/1471-2105-9-307

Suarez-Gonzalez A, Lexer C, Cronk QCB. 2018. Adaptive introgression: A plant perspective. *Biology Letters* **14**: 20170688. DOI: https://doi.org/10.1098/rsbl.2017.0688

Sullivan AR, Owusu SA, Weber JA, Hipp AL, Gailing O. 2016. Hybridization and divergence in multi-species oak (*Quercus*) communities. *Botanical Journal of the Linnean Society* **181**: 99-114. DOI: https://doi.org/10.1111/boj.12393

UniProt Consortium. 2019. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Research* **47**: D506-D515. DOI: https://doi.org/10.1093/nar/gky1049

Valbuena-Carabaña M, González-Martínez SC, Sork VL, Collada C, Soto A, Goicoechea PG, Gil L. 2005. Gene flow and hybridization in a mixed oak forest (*Quercus pyrenaica* Willd. and *Quercus petraea* (Matts.) Liebl.) in central Spain. *Heredity* **95**: 457-465. DOI: https://doi.org/10.1038/sj.hdy.6800752

Valen LV. 1976. Ecological species, multispecies, and oaks. *Taxon* **25**: 233-239. DOI: https://doi.org/10.2307/1219444

Contribución a la delimitación taxonómica de tres especies del género Quercus subgénero Erythrobalanus. MSc. Thesis, Universidad Nacional Autónoma de México.

Valencia-Cuevas L, Mussali-Galante P, Piñero D, Castillo-Mendoza E, Rangel-Altamirano G, Tovar-Sánchez E. 2015. Hybridization of *Quercus castanea* (Fagaceae) across a red oak species gradient in Mexico. *Plant Systematics and Evolution* **301**: 1085-1097. DOI: https://doi.org/10.1007/s00606-014-1151-4

Whittemore AT, Schaal BA. 1991. Interspecific gene flow in sympatric oaks. *Proceedings of the National Academy of Sciences of the United States of America* **88**: 2540-2544. DOI: https://doi.org/10.1073/pnas.88.6.2540

Associated Editor: Jonás Aguirre Ligouri

Author contributions: FRG, KO and AGR conceived and designed the study; FRG. and AGR collected the samples; FRG, DKCS and VIT analyzed the data; FRG and DKCS wrote the manuscript; and all authors interpreted and discussed the results and edited the manuscript.

Supporting agencies: This study was supported by Secretaría de Ciencia, Humanidades, Tecnología e Innovación (SECIHTI), Grant 240136; Dirección General de Asuntos del Personal Académico (DGAPA-UNAM)-Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica (PAPIIT), grant numbers: IN210020, IN207417 and IV201015. F.R-G and D.K.C.-S were supported by postdoctoral scholarships from Dirección General de Asuntos del Personal Académico (DGAPA) of the Universidad Nacional Autónoma de México (UNAM).

Conflict of interests: The authors declare that there is no conflict of interest, financial or personal, in the information, presentation of data and results of this article.