

Supplement 1: Methods and results

Methods: Biogeography and Climatic Niche

Stratigraphy

We follow the latest version (v. 2022/02) of the International Chronostratigraphic Chart (<https://stratigraphy.org/>; Cohen et al. 2013, updated). The major change has been the re-recognition of lower/early, middle, and upper/late Neogene and Quaternary strata as formally accepted and named units, in some cases between series/epochs and stages/ages (Aquitania + Burdigalian = Early Miocene; Langhian + Serravallian = Middle Miocene; Tortonian + Messinian = Late Miocene; Gelasian + Calabrian = Early Pleistocene).

Fossilized Birth-Death analysis

For fossilized birth-death (FBD) analyses, the RAD-seq matrix was reduced to 29 tips, a single tip per species within both sects. *Cerris* and *Ilex*, with the exception of *Q. cerris*, for which two individuals were kept that did not group together in any analyses and might represent cryptic species. Loci were retained if they were present in at least 10 individuals. A NEXUS file was exported using the RADAMI package, including 47 additional lines of undetermined positions (coded as “?”), one per fossil included in the FBD analyses.

FBD analyses were conducted in BEAST2 (Bouckaert *et al.*, 2014). Markov chain Monte Carlo (MCMC) runs of 50 million generations each were run from ten independent random starting points on each of three random draws from the uniform distribution of the fossil age ranges. Analyses were conducted using a nucleotide substitution model that allows for rate variation and invariant sites ($\Gamma + I$), with the shape parameter (α) and proportion of invariant positions estimated, and four gamma categories. The relaxed log normal clock was used, with the clock rate estimated. Priors were specified using defaults with these exceptions: origin of the clade was set at an initial position of 58 million years (Ma), with a range of 49 to 60 Ma, based on previous results (Hipp *et al.*, 2020) that suggest the origin of subgenus *Cerris* to be close to the origin of the genus; and sampling proportion was estimated with bounds between 0.4 and 0.6, with a starting proportion of 0.45. Clade priors were specified by assigning each fossil in section *Cerris* or *Ilex* to one of thirteen pre-defined clades identified in the maximum likelihood analysis as MRCA priors, enforcing monophyly of the clade including the fossils and extant taxa (Supplementary Table 3). The position of the three *Cyclobalanopsis* fossils was designated by assigning all taxa except the *Cyclobalanopsis* fossils to a single “not *Cyclobalanopsis*” clade. Scripts for exporting data, BEAST2 XML

configuration files, and RAD-seq data matrices are all archived in the code repository for this paper (prerelease v0.92-2: <https://github.com/andrew-hipp/cerris-fbd>; <https://doi.org/10.5281/zenodo.6595965>), along with output files and instructions for executing the scripts.

Köppen Signatures

We used grid-weighted ‘Köppen signatures’ (Denk *et al.*, 2013; Bouchal *et al.*, 2018; Grímsson *et al.*, 2018), henceforth ‘Köppen profiles’, to summarize the climate niches occupied by species of *Cerris* and to investigate climate niche evolution within and among subsections of *Cerris* (Table 1; Supplementary Data 3). A Köppen profile reflects the proportional coverage of the various Köppen-Geiger climate zones (cf. Kottek *et al.*, 2006; Peel *et al.*, 2007) by a modern species based on gridded distribution data

Today, *Cerris* oaks thrive in three main climate types: Subtropical to mild-temperate “warm temperate” *C* climates with temperatures of the coldest month, $T_{\min} > -3^{\circ}\text{C}$ and $< +18^{\circ}\text{C}$, (cool-)temperate “snow climates” with pronounced winter cold, *D*, $T_{\min} < -3^{\circ}\text{C}$, and ‘equatorial’ (= tropical) *A* climates (marginally, Indochina) with $T_{\min} > +18^{\circ}\text{C}$. The second letters in the two-/three-letter climate formula used in the Köppen-Geiger classification indicate fully humid (*f*) summer-dry (*s*), or winter-dry (*w*) conditions. In addition, for equatorial climates, the second letter *m* indicates monsoon climates. For details of the three-letter code, see Kottek *et al.* (2006) and Peel *et al.* (2007). Some species further occur in arid *B* climates (Table 1; Supplementary Table 4, Supplementary Data 23). We opted for using the Köppen-Geiger system, which does not recognize a subtropical zone, instead of the Köppen-Trewartha system because there are no ready-to-use high-resolution Köppen-Trewartha (Trewartha and Horn, 1980) climate grids. Furthermore, the Scotese *et al.* (2014) palaeo-climate zones used to cross-check climate preference of fossil-taxa (see below) correspond to the Köppen-Geiger system, distinguishing five major zones: ‘tropical’ (= equatorial *A*-climates), ‘arid’ (= *B*-climates), ‘warm (temperate)’ (\approx warm temperate *C*-climates), ‘cool (temperate)’ (\approx *Dxa/Dxb*, snow climates with summer) and ‘cold’ (continental-dry *D* and *E* climates without summer).

Modern species distributions were connected to fossil distributions by using georeferenced occurrence data for each species, downloaded from the GBIF database (www.gbif.org; Supplementary Data 3). Each data set was checked for natural distribution outliers (e.g. specimens from botanical gardens outside the natural distribution range of a species). Published chorological data were used to detect these outliers (e.g. Browicz and Zieliński, 1982; Caudullo *et al.*, 2017; San-Miguel-Ayanz *et al.*, 2016; Fang *et al.*, 2009). The

cleaned georeferenced occurrence data were then plotted onto 5 arc minutes Köppen-Geiger grid (1986–2010 data; Rubel *et al.*, 2017) to establish Köppen profiles for all species of section *Cerris*; and on major terrestrial biome maps (Olson *et al.*, 2001; Supplementary Data 3) to assess species' forest biome preferences. For the Köppen-Geiger plots, the georeferenced dataset was filtered so that multiple occurrences in a single grid cell were only counted once (labelled 'unique grid cells' in the diagrams). Likewise, for the biome plots, georeferenced data were filtered so that multiple occurrences with the same coordinates were treated as single occurrences (labelled 'unique localities' in the diagrams).

The georeferenced data and the Köppen-Geiger map with resolution of 5 arc minutes were processed using the 'Sample Raster Values Toolbox' in QGIS Version 3.16.4-Hannover (<http://www.qgis.org>). The biome shape files were processed using the 'Geoprocessing Tool' in QGIS. The biomes and Köppen-Geiger climates occupied by extant members of section *Cerris* are shown as maps generated with QGIS and as frequency (proportional distribution) diagrams. The raw coordinate data are archived in EXCEL spreadsheet format (Supplementary Data 3: Tabulated Data S3-1).

In addition, historical climate data (1970–2000) were compiled for 2,779 georeferenced occurrences for members of *Quercus* sect. *Cerris* using WORLDCLIM vers. 2.1 (<https://www.worldclim.org/data/worldclim21.html>) at a resolution of 30 seconds (Supplementary Data 3: Tabulated Data S3-3; ca. 1 km²; Fick *et al.*, 2017)). To characterize the climate envelope of modern-day species, we plotted the mean temperature of the coldest month (MTCM) against precipitation during the winter quarter (PCQ; WORLDCLIM variable BIO19), as well as monthly temperature and precipitation averages (MMT, MMP) and monthly minimal temperatures (MTmin; Supplementary Data 3).

Maximum likelihood reconstructions of major climate niches and main biomes

Based on the quantitative assessment of biome and climate zone preferences of the modern-day species, we binned extant and fossil species into five basic categories, accounting either for biome or climate zone preferences (Table 1; Supplementary Data 3). Our generalisation and categorisation make use of the terminology and concepts introduced by Schroeder (1998; cf. Denk *et al.*, 2013) and allow (i) direct comparison of biome and climate zones preferences, which are commonly correlated but not synonymous, and (ii) relation of quantitative modern-day categorisation qualitatively to our fossil-taxon set. Towards that end, we first defined for each fossil-taxon of section *Cerris* (columns *Biome/Major Köppen climate type* in Supplementary Table 2) the putative covered biomes: 'Tropical & Subtropical Coniferous

Forest’, ‘Tropical & Subtropical Moist Broadleaf Forest’, ‘Mediterranean Forests, Woodlands, and Scrubs’, and ‘Temperate Conifer Forest.’ We also specified climate zones for each fossil taxon, distinguishable in the fossil record as five principal combinations: *Cf* + *Cw* (may include *Aw* for oldest fossils); chiefly *Cf*; *Cf/Cw* extending into *Df/Dw*; *Cf* + *Df*; and *Cf* + *Cs*. The scoring is based on our knowledge about the fossil floras (place, time, plant association) and general background information about the climatic history of the northern hemisphere (Scotese *et al.*, 2014). We then identified common patterns and assigned the following five categories to the modern-day and fossil-taxa (columns *Category according biome*; *Category according climate zone* in Supplementary Table 2; matrix in *MainClimates* in Supplementary Data 4: Data File S4-1; characters #10 –biome class, and #11 – climate class).

0–Moist-Subtropical: Associated exclusively with the Tropical and Subtropical Moist Broadleaf Forests biome; fossil species that can be associated with *Cf* climates during global greenhouse phases or at low latitudes, i.e. exclusively within the southern half of the ‘warm’ zones (\approx modern-day humid subtropics), possibly extending in the extinct ‘boreal tropical’ palaeo-climate zones in Scotese *et al.* (2014). In case of modern-day *Cerris* oaks, species are associated almost exclusively with summer-moist climates with hot summers (*Cfa*, *Cwa* climates). The only modern species with an accordingly characteristic climate niche is the East Asian *Q. chenii*, firmly restricted to the *Cfa* climate of central-eastern China.

1–Meridio-Nemoral: Associated with the ecotone between Tropical and Subtropical Moist Broadleaf Forests and Temperate Broadleaf and Mixed Forests biomes. Fossil-species are linked exclusively to *Cf* climates and are placed within Scotese *et al.*’s (2014) ‘warm(-temperate)’ palaeo-climate zones. A modern-day Meridio-Nemoral species has its main distribution in subtropical to temperate climates with ample precipitation in the hot summer (typically *Cfa* + *Cwa* extending into warm *Cfb/Cwb* variants).

2–Nemoral: Either restricted to Temperate Broadleaf and Mixed Forests or extending into both Tropical and Subtropical Moist Broadleaf Forests and Temperate Coniferous Forests biomes; fossil species associated with fully humid *Cf* and *Df* climates ranging from ‘warm’ into ‘cool(-temperate)’ palaeo-climate zones. Nemoral species are fully temperate species preferring mild climates with ample precipitation during the typically long (\sim 8 months with MTCM $> 10^{\circ}\text{C}$) growing season. Key taxon for this category is *Fagus* (beech), which has its climax distribution in fully temperate *Cfb* climates, extending into subtropical *Cfa* lowland and, latitudinally and altitudinally, into cool-temperate *Dfa/Dfb* climates. Fossil assemblages of this category would typically contain broadleaved oaks of subgenus *Quercus* associated

with beech and maples with palmate leaves (as found in modern-day species of *Acer* sects *Acer*, *Macrophylla*, *Palmata* p.p., and *Platanoidea*). The only modern-day *Cerris* oak that qualifies for this category is *Q. variabilis*, being the most widespread species of section *Cerris* with a strong preference for the Temperate Broadleaf and Mixed Forests biome and the full range of summer-humid warm temperate (*Cf*, *Cw*) climates.

3–Meridional: Generalists tolerating summer-drought, otherwise similar to the Nemoral category. Fossil-species of this category can be associated with Tropical and Subtropical Moist Broadleaf Forests *and* biomes with seasonal drought/ water stress (Mediterranean Forests, Woodlands, and Scrub and Temperate Coniferous Forests biomes); they thrive in fully humid *Cf* as well as summer-dry *Cs* climates. The natural distribution of modern-day Meridional species is restricted to the subtropical belt but characterized by a split preference for both fully Mediterranean lowland climates (unlike in species of the previous categories), summer-dry with hot summers (*Csa*), and their altitudinal, potentially moister successions: summer-dry (*Csb*) or perhumid (*Cfb*) climates with warm summers. In contrast to the species categorized as ‘Mediterranean’, they can endure substantial winter cold and frost phases, with some species extending into the boreal, continental, summer-dry snow climates found at mid-latitudes in Western Eurasia. They do not usually form part of coastal vegetation but prefer higher altitudes and the interior.

4–Full-Mediterranean: Summer-drought tolerant specialists restricted to summer-hot and winter-mild biomes and climates. Modern-day species of this category are restricted to Mediterranean Forests, Woodlands, and Scrub biome (or Temperate Coniferous Forests biome: *Q. afares*) and hot, summer-dry *Csa* climates, while only marginally extending into moister biomes and moister (Mediterranean-*Cfa*) or colder (*Csb*) climates. Further, Full-Mediterranean species can be found in hot or cold ‘steppe’ climates (*BSh*, *BSk*).

Ancestral states of the unordered five-state categorical climate character were reconstructed under the *MkI* model in MESQUITE v 2.75 (Maddison and Maddison, 2011). We used two different input trees: (i) the original dated tree for standard top-down reconstruction of ancestral states, i.e. using only the information scored for the modern-day species; (ii) the dated tree with nodes and tips added to account for states of fossil-taxa. Fossil-taxa that could be associated with a distinct branch (lineage) were treated as sister lineages and used to break down the according branch. We used the oldest possible age of the fossil-taxon as age of the putative MRCA, and the youngest possible age to define the MRCA-added-tip distance. The

resultant Mesquite-NEXUS file is included in the Github repository/ Zenodo submission (Supplementary Data 5).

Tests for inter-species gene flow

Evidence of introgressive hybridisation was investigated using Patterson's *D*-statistic test (Durand *et al.*, 2011) as implemented in IPYRAD (Eaton and Overcast, 2020). The following text describes the rational, experimental set-up and summarizes the main results.

Background:

D-statistics

To test for introgression and hybridisation using the *D*-statistic tests, one needs SNP data from at least four tips: an outgroup defining the root of the three tested tips (O or P4); a tip (P3) representing one potentially introgressing population or species; and two other tips (P2 and P1) of which one is hypothesized to introgress with P3, the other of which is not. SNPs are subsetting to the included tips and include only binary SNPs that form one of two patterns: ABBA, where A represents one nucleotide and B the other, and the pattern references P1 to P4 in order; or BABA. The relative frequencies of ABBA and BABA are compared, and Z-scores on bootstrap-resampled are used to calculate a P-value (which we report in the paper using Holm-correction for each *D*-statistic test performed to correct for multiple testing). A significant over-abundance of the ABBA pattern suggests that phylogenetic discordance implied by the SNPs is likely due to introgression between P3 and P2; a significant over-abundance of the BABA pattern suggests introgression between P3 and P1. A non-significant *D*-statistic result suggests that phylogenetic discordance implied by the SNPs can be accounted for as an outcome of incomplete lineage sorting.

Our tests focused on two primary hypotheses: the hybrid origins of *Q. afares* (as proposed in, e.g., Mir *et al.* 2006, Welter *et al.* 2012, Mhamdi *et al.* 2013) and of *Q. crenata*. We specifically tested for admixture in *Q. afares* and *Q. crenata* because both taxa have been traditionally viewed as (F1) hybrids: *Q. afares* as *Q. suber* × *Q. canariensis* (the latter belonging to sect. *Quercus*; e.g. Mir *et al.*, 2006); *Q. crenata* as *Q. cerris* × *Q. suber*. In addition, we performed follow-up tests for admixture between *Q. ilex* and *Q. suber* (as identified in Simeone *et al.* 2018), *Q. cerris* and *Q. afares*, and among subsections *Aegilops*, *Suber*, and *Libani*.

Quercus afares hybrid origins

The only non-morphological evidence for a hybrid origin of *Q. afares* was the study of Mir *et al.* (2006), which included data from the two postulated (based on morphology and geography) donors of *Q. ×afares*: the Cerris oak *Q. suber*, molecularly, a distant relative, and the white oak *Q. canariensis*, sym-/parapatric with *Q. afares*, a species from a different subgenus and much different evolutionary and geographic origin (see Denk and Grimm, 2010; see also the Fagaceae phylogenies and data compiled by Zhou *et al.*, 2022). Without including an additional tip as outgroup not involved in the tested hybrid scenario, e.g. *Q. ilex*, Mir *et al.* were unable to perform a topology test, so their evidence is based on isozyme allele frequencies. The test is particularly interesting because a section *Cerris* × section *Quercus* hybrid is unlikely, at least in natural populations. While the occurrence of (±ancient) inter-sectional hybrids can be assumed for sections of the same subgenus, there has been so far no evidence for mixing between the two main lineages, the exclusively Old World subgenus *Cerris* and the originally New World subgenus *Quercus* in areas (Eurasia) where they grow sympatrically today or in the past (Denk and Grimm, 2010; Simeone *et al.*, 2018). In contrast, species groups and lineages characterised by sharing of plastid haplotypes such as Mediterranean *Cerris* and *Ilex* oaks and subtropical-tropical East Asian cycle cup (sect. *Cyclobalanopsis*) and *Ilex* oaks (Simeone *et al.*, 2016, 2018; Zhou *et al.*, 2022), or for which past introgression has been inferred using nuclear data (sections *Ponticae* and *Quercus*, McVay *et al.*, 2017; sect. *Protobalanus* and *Quercus* in western North America, cf. ITS data of Manos *et al.*, 2001) always were members of the same subgenus. Inter-generic reticulation possibly can be found even within major core Fagaceae lineages such as *Notholithocarpus* and American members of subgenus *Quercus* (P. Manos, pers. comm., 2012): subgenus *Quercus* and the western North American relict genera *Notholithocarpus* and *Chrysolepis* share the same plastome lineage (Zhou *et al.*, 2022). In contrast, there is up to date no evidence at all for inter-subgeneric gene flow (past or present) in oaks, despite the fact that e.g. species of section *Quercus* frequently can be found in sympatry with (widespread) species of sections *Cerris* and *Ilex* across Eurasia.

Quercus crenata hybrid origins

Hybridisation and introgression are more likely in the case of *Q. crenata*, postulated to be a *Q. cerris* × *Q. suber* hybrid or *Q. cerris* populations introgressed by *Q. suber*, because of its intermediate morphology, specifically the presence of well-developed cork layers in *Q. crenata* (Pignatti, 1982; Schwarz, 1936–1939). Also in this case, broadly-sampled nuclear

spacer data did not produce any evidence so far for *Q. cerris*-unique (diagnostic) sequence variants in *Q. crenata*, while strongly supporting an (inclusive) common origin (holophyly, i.e. monophyly in a strict sense; cf. Ashlock, 1971) of *Q. crenata* and *Q. suber* (Denk and Grimm, 2010; Simeone *et al.*, 2018; Piredda *et al.*, 2020; the current study using phylogenomic data from carefully selected placeholder tip-set).

Methods

RAD-seq data were clustered *de novo* in IPYRAD v 0.9.84 for 34 individuals comprising all section *Cerris* individuals, all *Q. ilex* individuals, four individuals from section *Cyclobalanopsis*, *Quercus canariensis*, and *Notholithocarpus densiflorus* as an outgroup, using the same clustering thresholds used above but retaining loci with a minimum of 4 samples, to increase the number of potential loci usable for *D*-statistic tests. Tests were conducted by generating all possible combinations of the outgroups, the first potential introgressor individuals in our dataset (first species listed in each comparison) as taxon p3, the second potential introgressor individuals in our dataset as taxon p2, and all other taxa compatible with the pectinate topology used in the *D*-statistic test as taxon p1 (see Durand *et al.*, 2011, fig. 1), with ABBA as the expected dominant pattern under the introgression hypothesis being tested in each case. One test (“canariensis-afares”) was limited to 16 tests with a range of possible sisters because the unconstrained test caused errors. *Z*-scores were calculated for each test, and two-tailed *p*-values for *Z*-scores were corrected for multiple test biases using Holm-Bonferroni correction in R, with correction applied separately for each of the ten hypotheses investigated. For all hypotheses performed, effect of the P3 or P1 taxon was investigated by summarizing *D*-stats, *Z*-scores, and corrected *p*-values for the tests in which each taxon was included. Scripts for performing and summarizing analyses are in the ‘D_stats’ subfolder of <https://github.com/andrew-hipp/cerris-fbd> (<https://doi.org/10.5281/zenodo.6595965>).

A total of 13 hypotheses are summarized in Table 3 of the manuscript text, which summarizes the results over all tests performed for each hypothesis; and Supplementary Table 6, which summarizes test results for specific focal individuals in either the P3 or P1 position for the tests performed. Summaries are provided by indicating the number of tests performed, and then *D*, *Z*, and Holm-Bonferroni-corrected *P*, all with mean and 95% quantiles, as well as the percent of tests significant at $P \leq 0.01$. Test numbers in Supplementary Table 6 correspond to test numbers reported in Table 3 of the main text.

Two additional tests were performed but not included in the results or discussion of the main paper for clarity: a test of relative importance of introgression with the two *Q. crenata* individuals included, using all species potentially introgressing with *Q. crenata* as tested above with each of the *Q. crenata* individuals as sister species. This test result is reported in Supplementary Table 6. The second is a five-taxon (“partitioned”) *D*-statistic test evaluating the direction of introgression with *Q. crenata* and potential introgressants. This is reported in Supplementary Table 7.

Taxa included in D-statistic tests performed

The following summarize what taxa were included in each test performed, and how many topologies (*n*) were evaluated for each test. Details of which individuals were included are in the associated Python scripts (as indexed and detailed in https://github.com/andrew-hipp/cerris-fbd/blob/main/D_stats/finalForPaper/AAA.testsSummarized.md)

- A1 -- cerris-crenata (n = 500): *Q. cerris* with *Q. crenata*
 - p4 - *Q. chenii*, *Q. variabilis*, and *Q. acutissima* as outgroups
 - p3 - *Q. cerris* potential introgressor 1
 - p2 - *Q. crenata* potential introgressor 2
 - p1 - *Q. suber* as sister species to *Q. crenata*
- A2 -- crenata-cerris (n = 2560): *Q. crenata* with *Q. cerris*
 - p4 - *Q. chenii*, *Q. variabilis*, and *Q. acutissima* as outgroups
 - p3 - *Q. crenata* potential introgressor 1
 - p2 - *Q. cerris* potential introgressor 2
 - p1 - undefined --- anyone who fits the topology
- B1 -- canariensis-afares (n = 16): *Q. canariensis* with *Q. afares*
 - p4 - *Notholithocarpus* as outgroup
 - p3 - *Q. canariensis* potential introgressor 1
 - p2 - *Q. afares* potential introgressor 2
 - p1 - *Q. libani* [sister to *Q. afares*]
- B2 -- suber - afares (n = 700): *Q. suber* with *Q. afares*
 - p4 - *Q. chenii*, *Q. variabilis*, and *Q. acutissima* as outgroups
 - p3 - *Q. suber* potential introgressor 1
 - p2 - *Q. afares* potential introgressor 2

- p1 - undefined --- anyone who fits the topology
- C1 -- *Aegilops-crenata* (n = 450 tests): introgression of subsect. *Aegilops* with *Q. crenata*
 - p4 - *Q. chenii*, *Q. variabilis*, and *Q. acutissima* as outgroups
 - p3 - *Aegilops* oaks - *Q. macrolepis*, *Q. brantii*, *Q. ithaburensis* potential introgressor 1
 - p2 - *Q. crenata* potential introgressor 2
 - p1 - *Q. suber* as sister species to *Q. crenata*
- C2 -- *crenata* - *afares* (n = 280 tests): *Q. crenata* and *Q. afares*
 - p4 - *Q. chenii*, *Q. variabilis*, and *Q. acutissima* as outgroups
 - p3 - *Q. crenata* potential introgressor 1
 - p2 - *Q. afares* potential introgressor 2
 - p1 - undefined --- anyone who fits the topology
- C3 -- *suber-cerris* (n = 6400 tests): *Q. suber* and *Q. cerris*
 - p4 - *Q. chenii*, *Q. variabilis*, and *Q. acutissima* as outgroups
 - p3 - *Q. suber* potential introgressor 1
 - p2 - *Q. cerris* potential introgressor 2
 - p1 - undefined --- anyone who fits the topology
- D -- *Libani-crenata* (n = 350): subsect. *Libani* with *Q. crenata*
 - p4 - *Q. chenii*, *Q. variabilis*, and *Q. acutissima* as outgroups
 - p3 - *Q. libani*, *Q. trojana*, or *Q. afares* potential introgressor 1
 - p2 - *Q. crenata* potential introgressor 2
 - p1 - *Q. suber* as sister species to *Q. crenata*
- E1 -- *cerris-afares* (n = 270): *Q. cerris* with *Q. afares*
 - p4 - *Aegilops* oaks as outgroups - *Q. macrolepis*, *Q. brantii*, *Q. ithaburensis*
 - p3 - *Q. cerris* potential introgressor 1
 - p2 - *Q. afares* potential introgressor 2
 - p1 - *Q. libani* 628 - *Q. trojana* 585
- E2 -- *cerris-afares* | *libani* (n = 90): *Q. cerris* with *Q. afares*, *Q. libani* as p1
 - p4 - *Aegilops* oaks as outgroups - *Q. macrolepis*, *Q. brantii*, *Q. ithaburensis*
 - p3 - *Q. cerris* potential introgressor 1
 - p2 - *Q. afares* potential introgressor 2
 - p1 - *Q. libani* 628
- E3 -- *cerris-afares* | *trojana* (n = 90): *Q. cerris* with *Q. afares*, *Q. trojana* as p1

- p4 - Aegilops oaks as outgroups - *Q. macrolepis*, *Q. brantii*, *Q. ithaburensis*
- p3 - *Q. cerris* potential introgressor 1
- p2 - *Q. afares* potential introgressor 2
- p1 - *Q. trojana* 585
- E4 -- *cerris-trojana* (n = 90): *Q. cerris* with *Q. trojana*, *Q. libani* as p1
 - p4 - Aegilops oaks as outgroups - *Q. macrolepis*, *Q. brantii*, *Q. ithaburensis*
 - p3 - *Q. cerris* potential introgressor 1
 - p2 - *Q. trojana* 585
 - p1 - *Q. libani* 628
- F -- *suber-ilex* (n = 6,240): *Q. suber* with *Q. ilex*
 - p4 - sect. *Cyclobalanopsis* og
 - p3 - *Q. ilex* potential introgressor 1
 - p2 - *Q. suber* potential introgressor 2
 - p1 - any species sister to p2 w/ respect in context of p4 and p3 used
- Supplemental Test 1 – *crenata* vs. *crenata* (n = 165)
 - - p5 - *Q. chenii*, *Q. variabilis*, and *Q. acutissima* as outgroups
 - - p3_1 = p4 - varies (see below)
 - - p3_2 = p3 - varies (see below)
 - - p2 - *Q. crenata* potential introgressor 2
 - - p1 - *Q. suber* as sister species to *Q. crenata*
- Supplemental Test 2 – 5-taxon *D*-statistic tests (n = 3)
 - - p5 - *Q. chenii*, *Q. variabilis*, and *Q. acutissima* as outgroups
 - - p3_1 = p4 - varies (see scripts)
 - - p3_2 = p3 - varies (see scripts)
 - - p2 - *Q. crenata* potential introgressor 2
 - - p1 - *Q. suber* as sister species to *Q. crenata*

Results: summary of D-statistic tests

D-statistic tests recover strong and consistent support for admixture between *Q. cerris* and *Q. crenata* (100% of tests significant at $p \leq 0.01$; Table 3). Five-taxon *D*-statistic tests (Supplemental Test 2) suggest that the direction of past or recent introgression has been from *Q. cerris* to *Q. crenata*, as the alleles that are shared by *Q. cerris* and either section *Aegilops* or section *Libani* are also significantly shared with *Q. crenata* relative to *Q. suber*

(Supplementary Table 7). There is also strong and consistent support for admixture between *Q. cerris* and *Q. afares* when *Q. libani* is used as the sister to *Q. afares* (Test E2: 98.9% of tests significant; Table 3). When *Q. trojana* (Italy, Balkans, Turkey) is used as the sister to *Q. afares* (endemic to Tunisia, N. Algeria), only the two westernmost *Q. cerris* individuals show significant levels of admixture: OAK-MOR-591 (Latium, Italy) and OAK-MOR-736 (Samsun, Turkey; 88.9% of tests significant; Test E3, Supplementary Table 6). There is minimal evidence of admixture/ introgression between *Q. ilex* and *Q. suber* (Test F: 19.0% of tests significant; Table 3). Partitioning the section *Aegilops* (p3)-*Q. crenata* (p4) test (Test C1) by individuals in the P3 position indicates that two of the *Q. ithaburensis* individuals (OAK-MOR-599, OAK-MOR-735; Levante) may have been weakly involved in introgression with *Q. crenata* (endemic to Tyrrhenian side of Italy; 61.5–88.0% of tests significant) in the past, while the other section *Aegilops* individuals (SE. Italy, Turkey, Israel) were more strongly involved (99.3–100% of tests significant; Table 6). Partitioning the “*crenata* vs. *crenata*” test (Supplemental Test 7) by P3 individuals demonstrates that *Q. crenata* TUS13-003 (OAK-MOR-593), which falls sister to *Q. crenata* 1977-541 (OAK-MOR-986) + *Q. suber*, has been disproportionately affected by admixture from the *Cerris* core clade (Supplemental Table 6).

Tests involving *Q. afares* showed no evidence of introgression with *Q. canariensis* in any topologies (Test B1, Table 3), but high variability in test results for *Q. suber*, with P-values ranging from <0.001 to 0.859 (Test B2, Table 3). Subsequent tests between subsections *Aegilops* and *Suber* (C1, C2, C3) as well as *Libani* and *Q. crenata* (D) partitioned tests (“*cerris* – *afares* | *libani*”; “*cerris* – *afares* | *trojana*”) show that some ABBA-patterns probably represent a deeper history of introgression or other source of imbalanced shared ancestral alleles rather than recent introgression. Against the background of the FBD-dated tree (main-text Fig. 4), the fossil and niche history (following sections), the *D*-statistics point to ancestral gene flow from the precursors of subsection *Cerris* into those of subsection *Libani*, especially in the central Mediterranean region. We interpret this result as a likely outcome of introgression during the crown diversification of the west Eurasian *Cerris*, resulting in phylogenetic discordance regarding resolution of the subsections rather than recent introgression,

However, there is evidence for more recent geographically structured introgression that bears additional study with more samples. We observe generally higher and more discriminative *D*-values for admixture when *Q. libani* is used than when *Q. trojana* is used as p4 in the test. The most involved *Q. cerris* individuals in the latter test (E3) are OAK-MOR-

591 from Italy ($D \geq 0.22$; 89% of tests significant; geographically closest *Q. cerris*) and, to nearly the same level, the northern Turkish OAK-MOR-736 ($D \geq 0.18$; 89% of tests significant; Supplementary Table 6). These two individuals furthermore constitute the subsection *Cerris* subclade that includes also the Iranian *Q. castaneifolia*, a species with a distinctly primitive leaf morphology not unlike *Q. afares* (Supplementary Data 3) and the widespread fossil-species *Q. kubinyi*.

The D estimates for *Q. crenata* differ considerably from those of *Q. afares*. We find low to high levels of admixture for all tested scenarios (Table 3), continuously increasing along the phylogenetic tree and peaking – as in the case of *Q. afares* – in the Italian and northern Turkish individuals of *Q. cerris*. At least two reticulation scenarios might explain such a result:

Scenario 1—*Quercus crenata* and its precursors, representing the first diverging and westernmost section *Cerris* lineage (main-text Figs 2–3), were repeatedly affected by introgression when newly diverging eastern Mediterranean lineages radiated (illustrated in main-text Fig. 4). The D -statistic results would thus reflect in this scenario a history of gene flow with other sympatric western Eurasian subsections (main-text Fig. 4; Supplementary Fig. 4).

Scenario 2—Alternatively, recent introgression of *Q. cerris* with sympatric *Q. crenata* in its western range may account for the observed D -statistics. The inferred admixture from other lineages under this scenario would reflect patterns shared by *Q. cerris* and its sister lineage, subsection *Libani* (D -values approaching the level as seen in southeasternmost *Q. cerris* individuals from the Levante) and the phylogenetically more distant subsection *Aegilops*.

We consider Scenario 2—modern-day introgression from *Q. cerris* into *Q. crenata*—the less likely: the partitioned D -statistics identified OAK-MOR-593 as the *Q. crenata* individual most affected by admixture from outside subsection *Suber*. Yet this individual exhibits relatively low root-tip distance within its clade (Fig. 1). We would expect relatively recent introgression to introduce tree-incompatible signal into the RAD-seq data, inflating the tip branch length beyond that observed in the less-impacted second *Q. crenata* individual, OAK-MOR-986. We consequently consider either (1) ancient introgression among the precursors of modern subsections of west Eurasian *Cerris* or (2) demographic processes that result in

imbalanced maintenance of ancestral polymorphisms to be a more plausible explanation for the significant *D*-statistic results.

References

- Ashlock PD. 1971. Monophyly and associated terms. *Systematic Zoology* **20**:63–69.
- Denk T, Grimm GW. 2010. The oaks of western Eurasia: Traditional classifications and evidence from two nuclear markers. *Taxon* **59**: 351–366.
- Manos PS, Zhou ZK, Cannon CH. 2001. Systematics of Fagaceae: phylogenetic tests of reproductive trait evolution. *International Journal of Plant Sciences* **162**: 1361–1379.
- McVay JD, Hipp AL, Manos PS. 2017b. A genetic legacy of introgression confounds phylogeny and biogeography in oaks. *Proceedings of the Royal Society B* **284**: 20170300.
- Mir C, Toumi L, Jarne P, Sarda V, Di Giusto F, Lumaret R. 2006. Endemic North African *Quercus afares* Pomel originates from hybridisation between two genetically very distant oak species (*Q. suber* L. and *Q. canariensis* Willd.): evidence from nuclear and cytoplasmic markers. *Heredity* **96**: 175–184.
- Pignatti S. 2017-2019. Flora d'Italia. Milano: Edagricole, New Business Media.
- Piredda R, Grimm GW, Schulze E-D, Denk T, Simeone MC. 2020. High-throughput sequencing of 5S-IGS in oaks: Exploring intragenomic variation and algorithms to recognize target species in pure and mixed samples. *Molecular Ecology Resources* **21**:495–510.
- Schwarz O. 1936–1939. *Monographie der Eichen Europas und des Mittelmeergebietes*. Feddes Repertorium regni vegetabilis. Berlin-Dahlem: Sonderbeiheft D.
- Simeone MC, Grimm GW, Papini A, Vessella F, Cardoni S, Tordoni E, Piredda R, Franc A, Denk T. 2016. Plastome data reveal multiple geographic origins of *Quercus* Group *Ilex*. *PeerJ* **4**: e1897.
- Simeone MC, Cardoni S, Piredda R, Imperatori F, Avishai M, Grimm GW, Denk T. 2018. Comparative systematics and phylogeography of *Quercus* section *Cerris* in western Eurasia: inferences from plastid and nuclear DNA variation. *PeerJ* **6**: e5793.
- Zhou BF, Yuan S, Crowl AA, Liang YY, Shi Y, Chen XY, An Q-Q, Kang M, Manos PS, Wang B. 2022. Phylogenomic analyses highlight innovation and introgression in the continental radiations of Fagaceae across the Northern Hemisphere. *Nature Communications* **13**: 1–14.