Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

New insights into the adaptive chloroplast capture in European white oaks

Dimitra Sergiadou
dis@ign.ku.dk
University of Copenhagen
Daniel Wood
Royal Botanic Gardens
Luke T Dunning
University of Sheffield
Erik D Kjaer
University of Copenhagen
Jill K Olofsson
University of Copenhagen

Research Article

Keywords: Quercus, chloroplast capture, Last glacial maximum, phylogenomics

Posted Date: November 5th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11318303/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Abstract

Chloroplast genomes are frequently used to reconstruct the evolutionary history of plants, including detection of introgression and chloroplast capture events. Here, we focus on two oak species, *Quercus robur* and *Q. petraea*, which lack differentiation in several chloroplast markers. This was assumed to result from chloroplast captures during their dispersal from refugia after the last glacial maximum (LGM). We aim to reassess chloroplast variation within and between these species by *de novo* assembling 114 oak plastomes. Additionally, we combine our data with available plastome data from 91 Fagaceae species to reconstruct a dated phylogeny and estimate the divergence time between oak chloroplast lineages. Positive selection analysis on 69 chloroplast genes was also conducted to test whether chloroplast capture is adaptive. Plastome phylogenetic analyses reveal three oak haplotypes with no species delimitation. Molecular dating analysis indicates older splits among the three main haplotypes (2.68 – 0.5 Mya) and younger splits within lineages (0.39 – 0.04 Mya), suggesting that chloroplast capture could have occurred prior to the LGM in some lineages. Positive selection is observed in two chloroplast genes, with the selection on the *rbcL* gene possibly linked to the adaptive evolution of the RuBisCO enzyme. Hence, signs of a selective sweep could further explain the observed limited genetic diversity in oak plastomes. This study provides valuable insights into oak plastome evolution, setting a significant base for upcoming studies, including nuclear genome and protein-level studies to better understand cytonuclear interactions and conceivable environmental adaptations.

Introduction

Despite the increasing use of nuclear genome datasets in plant evolutionary studies (e.g. Johnson et al. 2019; Couvreur et al. 2019; Zhou et al. 2022), the plastid genome remains the cornerstone in angiosperm phylogenetics (Ruhfel et al. 2014; Shaw et al. 2014). Its haploid, non-recombinant, and (with few exceptions) uni-parental inherited nature provides clearer signatures of the age and relatedness across species (Rousseau-Gueutin et al. 2018), making it suitable for providing insights into interspecific gene flow in sympatric species (Rieseberg and Soltis 1991). Its significance also lies in its ability to detect patterns of ancient hybridization or introgression, particularly those involving chloroplast capture events (e.g. Petit et al. 2002; Premoli et al. 2012; Pham et al. 2017; Yang et al. 2021; Zhou et al. 2022). These events refer to the replacement of one species' chloroplast genome (plastome) by that of another species through hybridization and repeated backcrossing, a common phenomenon in several plant families (Rieseberg and Soltis 1991). While chloroplast capture is not restricted to closely related species, it is most likely to occur among sympatric species that are reproductively compatible (Liu et al. 2020). It is usually inferred from phylogenetic analyses and the conflict between the nuclear and plastome topology. There are several cases of distinct species delimitation based on nuclear markers, whereas the corresponding plastome topology appears non-monophyletic among and within different species and lineages (e.g. Petit et al. 2002; Acosta et al. 2010; Pham et al. 2017; Liu et al. 2020; Zhou et al. 2022). This cytonuclear discordance has been the topic of many studies, revealing the evolutionary history of cytoplasmic genomes which can provide essential insights into speciation and adaptation in both plants and animals (e.g. Sloan et al. 2016; Postel & Touzet 2020). Studying these incongruences can for example help us determine if certain cytoplasmic types (introgressive hybrids, chloroplast captures) are involved in (local) adaptive processes (e.g. Postel & Touzet 2020; Larson et al. 2026). As chloroplast and mitochondrial genes can be subject to positive selection, cytoplasmic capture could potentially offer an evolutionary advantage to some species (Tsitrone et al. 2003; Muir and Filatov 2007). Despite the high number of relevant studies on angiosperms, the relationships and plastid evolutionary histories of several species' complexes remain unclear or incomplete. Here, we use whole chloroplast genome data to reassess chloroplast variation among European white oak species (*Quercus*). Additional divergence time analysis of the Fagaceae family and positive selection analysis on selected chloroplast genes allow us to explore more thoroughly the pattern of shared chloroplast haplotypes and its possible effect on the species, while shedding more light on the evolutionary history of these species' plastomes.

The Fagaceae family is a large group of flowering plants in the Northern Hemisphere (Kremer et al. 2012) that includes over 900 species across 8 genera (POWO, 2026). It holds substantial ecological, cultural, and economic value, providing various ecosystem services from biodiversity to wood, biomass, and food production, including economically significant species such as oaks, beeches, and chestnuts (Kremer et al. 2012). During the Eocene, a period with cooling climatic conditions, Fagaceae species expanded and eventually dominated what is today Eurasia and America (Zhou et al. 2022). The well preserved and abundant fossil record of Fagaceae has been extensively utilized in phylogenetic studies, investigating the origin and migration of species, as well as estimating divergence times in modern taxa (Crepet and Nixon 1989; Manos and Stanford 2001).

The genus *Quercus* (oaks) is among the most studied genera within the Fagaceae family, comprising more than 400 species predominantly found in the Northern Hemisphere (Gil-Pelegrín et al. 2017; Backs and Ashley 2021). *Quercus* has a very complex taxonomy arising among others from its numerous species, significant morphological variation, and frequent hybridization (Denk et al. 2017). Therefore, and given its status as a key clade of woody angiosperms, *Quercus* phylogeny and evolution has always been central to various scientific studies (e.g. Aldrich and Cavender-Bares 2011; Gugger et al. 2021; Liu et al. 2021; Zhou et al. 2022). While trying to untangle complex phylogenetic/taxonomic relationships within *Quercus*, several studies have employed both nuclear and chloroplast markers throughout the years, uncovering the phylogenomic incongruence between the nuclear and plastid genome (e.g. Whittemore and Schaal 1991; Petit et al. 2002; Simeone et al. 2016; Simeone et al. 2018; Yang et al. 2021). More precisely, in a recent study by Zhou et al. (2022), nuclear genomes indicated that *Quercus* is a monophyletic genus, while plastome topologies were not monophyletic and often varied. Closely related species were combined in some subclades and more distinct taxonomic groups in others, suggesting historical hybridization and chloroplast capture events as a more probable scenario compared to incomplete lineage sorting (ILS) and ancestral polymorphisms. Consistent with these findings, other studies have frequently identified hybridization within *Quercus* sections (Kremer and Hipp 2020) and documented chloroplast capture events in sympatric lineages including the European white oaks *Q. robur* and *Q. petraea* (Petit et al. 2002; Pham et al. 2017).

The pedunculate oak (*Q. robur*) and sessile oak (*Q. petraea*) are distributed in Eurasia with *Q. robur* having a slightly more expanded distribution to the North and East (Eaton et al. 2016). The two species vary morphologically and have slightly different ecological niches where *Q. robur* favors more moist, alkaline, and nutrient-rich soils, and tolerates periodic flooding. In contrast, *Q. petraea* shows greater drought resilience and tends to have an adaptive advantage in rockier, acidic environments (Saintagne et al. 2004; Eaton et al. 2016). Despite phenotypic differences (e.g., leaf stalks, acorn shape), the two species often hybridize, producing individuals with intermediate characteristics (Eaton et al. 2016). Initial indications of extensive plastome sharing between these species, with no obvious species split based on chloroplast genes, were uncovered through analyses of a polymorphic plastome region, complemented by restriction fragment analysis (Petit et al. 1997; Petit et al. 2002). It has been hypothesized that the notable sharing of plastome haplotypes results from parallel chloroplast captures following their migration/dispersal from the refugia after the end of the LGM (Petit et al. 1997). According to this theory, rare long-distance seed dispersal by *Q. robur* led initially to the formation of *Q. robur* populations. Then, in some cases, more frequent short-distance pollen dispersal by *Q. petraea* followed, possibly resulting in asymmetric nuclear gene flow and shared chloroplast haplotypes (Petit et al. 1997). This hypothesis has not yet been explored using whole plastomes in a range-wide sampling context or through divergence time analysis to determine the timing of the split between the chloroplast lineages. Range-wide sampling could cover the complete species range and specifically target populations with individuals from mixed lineages. Furthermore, alternative hypotheses or additional factors - such as positive selection on chloroplast genes - that could also lead to plastome sharing patterns- have yet to be explored. Overall, one can argue that these species' biogeographical history and phylogenetic relationships have not yet been sufficiently resolved.

In this study, we *de novo* assemble 114 whole chloroplast genomes of European white oaks (*Q. robur*, *Q. petraea*, *Q. pubescens*, and admixed individuals) using available genome sequencing data and further evaluate the chloroplast variation within and between the species. Additionally, available plastome data representing 91 different species within the Fagaceae family were used to reconstruct and date a phylogeny, aiming to estimate whether the plastome sharing between *Q. robur* and *Q. petraea* occurred prior to or after the LGM. Positive selection on the chloroplast was examined as an ancillary factor further influencing the pattern of widespread plastome sharing. Our findings reveal new insights concerning signs of chloroplast capture events and the recolonization history of the European oaks after the LGM, followed by additional information on the evolution of some chloroplast genes.

Materials and Methods

Taxon sampling, genomic DNA extraction, and whole genome sequencing

We assembled 114 whole chloroplast genomes of *Quercus petraea*, *Q. robur, Q. pubescens*, as well as admixed individuals *of Q. petraea-Q. robur*, and *Q. petraea-Q. pubescens* originating from a total of 21 European provenances from 9 countries (Lobo et al 2026; Olofsson et al. 2026 (in revision); Table S1 and Fig. S1). Bud samples were harvested during autumn 2021 from a common garden experiment at Valby Hegn, Denmark, established in 1992. Additional chloroplast sequences from GenBank were also included (Table S1, Table S3).

Three to five buds were used to extract total genomic DNA with NucleoSpinPlantII (Macherey-Nagel, Düren, Germany) according to the manufacturer's protocol. Bud tissues were dissected, flash frozen in liquid nitrogen, and finely pulverized using a TissueLyserII (Qiagen, Hilden, Germany). The PL1 lysis buffer (400 µl), based on the CTAB method, was used and the final elusion was carried out in 60 µl of buffer PE. Total genomic DNA was quantified with Qubit and the quality of the DNA was evaluated using Nanodrop. A total of 500 ng genomic DNA per sample were shipped to IGATech (Udine, Italy) and used for paired-end Illumina sequencing (150 bp) following standard protocols for a target coverage of 6–10 Gb per sample (approximately 8–10 x of the genome). The raw sequencing data ranged from 5.8 to 14.8 Gb (mean = 9.5). AdaptorRemoval v.2.3.1 (Schubert et al. 2016) was used to quality filter the raw sequence reads, trimming adaptors and low-quality bases (Q < 20), discarding reads with over 20 ambiguous bases, and collapsing overlapping reads.

Plastome assembly, annotation, and gene extraction

Chloroplasts were assembled in NOVOPlasty v. 4.3.1, (Dierckxsens et al. 2017) using *Q. robur matK* gene (LT996900.1) as a seed. Only pair-end reads were used (ie. ignoring collapsed reads). An alignment was produced with MAFFT v.7.490 (Katoh et al., 2002) using the contigs and available reference sequences (*Q. robur*, *Q. petraea*, and *Fagus sylvatica* obtained from GenBank). All contig combinations produced for each sample were manually inspected and, if necessary, rearranged using Geneious Prime 2023.0.4 (https://www.geneious.com) to ensure the correct reading frame of the small single copy and inverted repeats. Plastome annotation and gene extraction for analysis purposes was performed in Geneious Prime 2023.0.4 using the annotated genome of *Q. petraea* (LT996899.1) as a reference. Start and stop codon positions were inspected manually. Assembly difficulties of the *ndhD* gene, possibly due to homologous sequence in the mitogenome (e.g. Gandini & Sanchez-Puerta, 2017), led to the exclusion of this region from all analyses.

Phylogenetic analysis

The above alignment of assembled plastomes was used to construct a *Quercus* phylogeny. First, the appropriate substitution model for our dataset was selected using the Smart Model Selection (SMS) as implemented in PhyML v. 1.8.1 (Lefort et al. 2017) and the GTR + G+I model was used for all following analyses. RAxML v. 8.2.11 (Stamatakis 2014) using the GTR + G+I substitution model and 100 bootstrap replicates were used to estimate a maximum-likelihood phylogeny, with *Fagus sylvatica* (MK598696.1) as the outgroup. The tree was visualized in FigTree v.1.4.2 (Rambaut 2014).

Divergence time analysis

A subset of eight samples was selected for a divergence time analysis (Table 1) due to identical chloroplast sequences between the two species in the phylogenetic analysis. The selection aimed to incorporate representatives from each of the clades identified in the phylogenetic analysis and to accurately represent the genetic diversity within both *Q. robur* and *Q. petraea* (see Results). A haplotype shared by both species is thus represented twice in the dataset (Table S3, Fig. 1a, Fig. 2b). One admixed *Q. petraea-Q. pubescens* sample from Poland (3-198-3) and a *Q. pubescens* from Turkey (3-199-1) were also included (Table 1). The 76 plastome genes referenced by Zhou et al. (2022), were examined for this analysis. However, due to possible assembly errors or genetic variability and/or indels that changed the reading frame of the protein we inferred premature stop codons in 7 genes, and thus, we only included 69 genes in our final selection (approx. 35% of the whole chloroplast genome, Table S2). Genes were extracted and concatenated in Geneious Prime 2023.0.4. The final dataset was combined with the published GenBank sequences from Zhou et al., 2022 (Table S3). *Betula pendula* (MG386401; Table S3) was used as an outgroup.

Table 1
Samples used in the dating and selection analysis.

Provenance code	Species	Country	Location	Latitude	Longitude	Samples	Origin
NA	*Q. robur*	Denmark	Bregentved	55.3	12.0	FW1, FW30	Present study
NA	*Q. robur*	Denmark	Petersgaard	55.0	12.1	FW35	Present study
185	**Admixed Q. petraea-Q. robur**	**Belgium**	**Queue de LHerse**	**50.1**	**4.4**	**2-185-3**	**Present study**
190	**Q. petraea**	**England**	**Blankeny**	**51.8**	**-2.5**	**2-190-5**	**Present study**
194	**Q. petraea**	**France**	**Bussieeres**	**47.8**	**5.5**	**2-194-4**	**Present study**
198	**Admixed Q. petraea-Q. pubescens**	**Poland**	**Sycow**	**51.2**	**17.9**	**3-198-3**	**Present study**
199	**Q. pubescens**	**Turkey**	**Bolu**	**40.9**	**31.7**	**3-199-1**	**Present study**
NA	*Q. petraea* (ref.genome)	England	NA	NA	NA	LT996899.1	GenBank Lloyd Evans,D.

Note: Regarding our dataset all samples above were included in the dating analysis; only the ones in bold were included in the selection analysis

Divergence time estimates were calculated in BEAST2 v. 2.6.3 (Bouckaert et al. 2019) within the framework of a strict molecular clock, as our data were clock-like; thus, rate variation among branches was not included in the model (Fig. S2). The configuration of the .xml file included a Yule tree prior and the designated substitution model. The calibration of the tree was based on two paleontological findings (non-monophyletic lineages on the chloroplast tree prevented the use of additional fossil calibrations that could force monophylies that only agree with the nuclear data topology); (F1) a fagaceous pollen fossil (Campanian of Wyoming, 82 – 81 Ma) which corroborates the divergence of *Fagus* lineage from the rest of the Fagaceae family (Trigonobalanoideae, Castaneoideae, Quercoideae) during the Late Cretaceous period (Grímsson et al. 2016), and (F2) Eocene-age fossil infructescences of *Castanopsis* discovered in Argentina's Laguna del Hunco, Chubut, assigning a minimum age of 52.2 Ma to the *Castanea-Castanopsis* split (Wilf et al. 2019). The fossil dates were integrated as priors in the .xml file to calibrate the minimum age of the aforementioned clades. We conducted two independent runs of 30 million generations, sampling every 1,000th generation, including a 10% pre-burnin step. Convergence of the runs was assessed using Tracer v. 1.7.1 (Rambaut et al. 2018), with an acceptance criterion of effective sample size (ESS) values > 200 for all parameters. LogCombiner v. 2.6.3 (included in the BEAST2 package) was used to combine the .log and .tree files from the two runs, and the maximum clade credibility (MCC) tree was constructed with TreeAnnotator v.2.6.3, discarding the initial 10% of trees as burn-in. The final phylogeny was visualized in FigTree v.1.4.2 (Rambaut 2014).

Inferring positive selection

We compiled a dataset of 26 samples based on the BEAST topology. The 26 individuals were selected from the branches of the tree where the newly added *Q. robur* and *Q. petraea* (eight-samples dataset) were found, including samples from the *Quercus* and *Ponticae* lineages as shown in Fig. 2b. As *Q. robur* and *Q. petraea* sequences from each of the previously identified chloroplast clades (Fig. 1) were almost identical, only the *Q. petraea* sequences from this eight-sample dataset were analyzed for selection along with the Polish and Turkish sample (Table 1). Genetic variation (SNPs and indels) was identified in 34 out of the 69 chloroplast genes which comprised the final dataset for this analysis. The rest of the genes were conserved throughout our dataset. It is well documented that *Quercus rubra* and *Fagus lucida* are phylogenetically distinct from the white oaks that were the focus in this study (Zhou et al. 2022). Therefore, these two species were used as outgroups (Table S3). Finally, the BI tree topology was edited with APE in R (Paradis and Schliep 2019) to retrieve relevant branches for the analysis.

First, the ratio of nonsynonymous (dN) to synonymous (dS) substitution rates (ω) using yn00, as implemented in the PAML v. 4.9j package (Yang 2007) was computed. Using CODEML from the same package, we then tested the site models M0 (purifying selection, $\omega < 1$), M1a (nearly neutral selection, $\omega \leq 1$), M2a (positive selection, $\omega > 1$), and branch-site models A ($0 < \omega \leq 1$) and its alternative A1 ($\omega = 1$). Model fitness was assessed via Likelihood Ratio tests (LRT), comparing M0-M1a, M1a-M2a, and A-A1, with

degrees of freedom set at 1, 2, and 0.5, respectively. Sites under positive selection were further evaluated using Bayes Empirical Bayes (BEB) analysis (Yang et al. 2005), as provided in CODEML's output.

Results

Assembly and alignment

In total, we assembled 114 chloroplast genomes from *Quercus petraea* (n = 64), *Q. robur* (n = 41), *Q. pubescens* (n = 1), and from individuals that were admixed between *Q. robur* and *Q. petraea* (n = 4), and between *Q. petraea* and *Q. pubescens* (n = 4) (Table S1). The average length of the sequences was 159,739 bp (range 159,632 – 159,831). Among the 625 variants detected in the sequences, 474 (75.84%) were single nucleotide polymorphisms (SNPs), 167 (26.72%) were insertions-deletions (indels) and 9 (1.44%) were more complex sequence variants.

Plastome phylogeny of Quercus

A phylogeny with high support (> 90 bootstrap values) on the majority of nodes was found by the Maximum Likelihood analysis (Fig. 1a). A relatively small number of SNPs defined the different clades, while the sequences were identical throughout most of the plastome.

More specifically, the genus *Quercus* is monophyletic with respect to *F. sylvatica*, however, *Q. robur* and *Q. petraea* appear to be polyphyletic and share haplotypes (Table S1). Three main plastome clades are defined, with clade A expanding from southeastern Europe (*Q. robur* and *Q. petraea* from Turkey) up to central and northern Europe (Hungary, Germany, Denmark), covering also provenances in Belgium and northern France (Fig. 1, Table S1). Two admixed *Q. petraea-Q. robur* individuals from Belgium and Germany (2-185-3 and 2-189-2 respectively) fall into the same clade (Table S1). Clade B is comprised of Danish *Q. robur* and English *Q. robur* and *Q. petraea* as well as two *Q. petraea* individuals from a provenance in Germany (Fig. 1, Table S1), covering mostly the northern and central part of Europe. The three admixed *Q. petraea-Q. pubescens* specimens from Poland are also found in clade B (Fig. 1, Table S1). Finally, clade C appears as a sister to clade B, although this relationship is not highly supported. Samples that belong to clade C are found in 16 out of the 21 provenances, and 8 countries, making it the clade with the widest distribution covering all northern and central European provenances in our dataset. It consists of *Q. petraea* from Norway, France, and Poland, *Q. robur* from Denmark, and samples from both species originating from England and Germany, along with one admixed *Q. robur-Q. petraea* sample from England (3-190-3; Fig. 1, Table S1). One *Q. petraea-Q. pubescens* specimen from Hungary (3-204-1) is also found in clade C, with the *Q. pubescens* sample from Turkey (3-199-1) located close to it, forming its own haplotype (Fig. 1, Table S1). Both *Q. robur* and *Q. petraea* species are found in all three clades, and provenances are in most cases not monophyletic (Fig. 1, Table S1).

Dating analysis

The dataset comprised 130 specimens in total, representing 91 distinct species within the Fagaceae family (Table S3). The average length of the 69 concatenated genes was 55,155 bp (range 54,165 – 55,812).

Phylogenetic reconstructions using Bayesian Inference (BI) and Maximum Likelihood (ML) produced congruent topologies, with high support on most nodes. The BEAST-derived phylogeny including time estimates is presented here (Fig. 2a). More precisely, the genera *Fagus*, *Lithocarpus*, *Castanea* and *Castanopsis* (sister clades), and *Chrysolepis* (nested among *Quercus* subgenera), emerged as monophyletic (Fig. 2a). *Trigonobalanus*, *Notholithocarpus* and *Quercus* were not inferred as monophyletic (Fig. 2a). Within *Quercus*, monophyly was recovered only for the subgenera *Lobatae* and *Cyclobalanospis* (Fig. 2a). Section *Quercus* is polyphyletic containing *Notholithocarpus*, and the sections *Ponticae*, *Virentes*, and *Protobalanus* (Fig. 2a). Similarly, the species of interest *Q. robur* and *Q. petraea* are sharing haplotypes, with samples from clades A and B forming sister subclades (Fig. 2). The Turkish *Q. pubescens* is nestled with *Q. canariensis* and *Q. pontica* (Fig. 2).

Two fossil calibrations constrained node dating (F1: *Fagus* split at 82 – 81 Ma, F2: *Castanea-Castanopsis* split at 52.2 Ma; Fig. 2a). Subsequent divergence estimates placed the *Trigonobalanus* and *Fagaceae* split (excluding *Fagus*) at ~ 54 Ma (53.97-55 Ma 95% HPD), while the *Lithocarpus*, *Notholithocarpus*, *Chrysolepis*, and *Quercus* subgenera divergence from the rest of the genera occurred

at ~ 52.4 Ma (52.24–52.57 Ma 95% HPD; Fig. 2a). The inner split among *Lithocarpus* and the younger genera was estimated to be around 19.8 Ma (18.07–22.02 Ma 95% HPD; Fig. 2a). Moreover, the crown age of section *Quercus*' was dated at ~ 12.46 Ma (11.39–13.49 Ma 95% HPD) when *Q. alba*, *Q. rugosa*, *Q. engelmmanii*, and *Q. lobata* split from the rest of the *Quercus* section (Fig. 2a). The more recent divergence of the two *Notholithocarpus* samples and the *Quercus* and *Ponticae* samples was estimated to be around 8.78 Ma (7.48–10.09 Ma 95% HPD; Fig. 2a).

Regarding the eight samples derived from the *Quercus* phylogeny, clade C's (*Q. robur* and *Q. petraea*) diverged from the other samples approximately 2.68 Ma (2.16–3.29 Ma 95% HPD), with the Turkish *Q. pubescens* sample appearing as the second oldest subclade splitting from the rest at around 2.45 Ma (1.99–2.97 Ma 95% HPD; Fig. 2b; Table 2). Samples from clade A and clade B - Poland split around 1.99 Ma (1.45–2.51 Ma 95% HPD), while Clade B and the Polish sample divergence is roughly 0.5 Ma (0.19–0.9 Ma 95% HPD; Fig. 2b; Table 2). Within clade divergence times for *Q. robur* and *Q. petraea* was calculated as follows: *Q. robur* and *Q. petraea* from clade C diverged around 0.39 Ma (0.14–0.7 Ma 95% HPD), while the respective split between clade A samples was placed at approximately 0.06 Ma (0–0.26 Ma 95% HPD; Fig. 2b). Finally, the split within clade B was dated to 0.04 Ma (0-0.17 Ma 95% HPD; Fig. 2b; Table 2).

Table 2
Time estimates between and within the eight samples used from this study

Time estimates	**Clade A**	**Clade B**	**Clade C**	***Q. pubescens***	**Poland**
	(n = 2)	**(n = 2)**	**(n = 2)**	**(n = 1)**	**(n = 1)**
Clade A	**0.06 (0-0.26)**	1.99 (1.45–2.51)	2.68 (2.16–3.29)	2.45 (1.99–2.97)	1.99 (1.45–2.51)
Clade B		**0.04 (0-0.17)**	2.68 (2.16–3.29)	2.45 (1.99–2.97)	0.5 (0.19–0.9)
Clade C			**0.39 (0.14–0.7)**	2.68 (2.16–3.29)	2.68 (2.16–3.29)
Q. pubescens				**na**	2.45 (1.99–2.97)
Poland					**na**

Note: time estimates between clades or samples are shown above the diagonal, while time estimates within *Q. robur* and *Q. petraea*

from the same clade are shown in bold on the diagonal. Numbers refer to million years and 95% HPD is shown in parentheses.

Selection analysis

In total 34 chloroplast-coding genes were evaluated for positive selection using the ω-ratio. Site model tests indicated that 26 of these genes were under purifying selection (ω < 1), with M0 as the optimal model. Five genes (*atpF*, *psaB*, *psbZ*, *rpl20*, and *rpoB*) showed neutral evolution, best represented by the M1a model. Notably, the M1a model, which allows for both purifying selection and neutral evolution, revealed varying percentages of neutrally evolving sites (ω = 1) in these genes, with the highest in *rpl20* (29%) and the lowest in *psaB* (1%). Equivalent percentages for *atpF*, *rpoB*, and *psbZ* genes were 21%, 16%, and 8% respectively.

Positive selection was detected in *ccsA*, *rbcL*, and *ycf1* genes. The *ycf1* gene, however, was omitted from subsequent analyses due to alignment gaps, potentially skewing results as PAML software interprets these as missing data. Consequently, only *ccsA* (M2a > M1a; 2ΔL = 12.76; p-value < 0.05) and *rbcL* (M2a > M1a; 2ΔL = 19.53; p-value < 0.05) were confirmed to be positively selected, with one site detected to be significantly under selection in each one of them (*ccsA* site: 94, amino acid: K, p-value = 95% and *rbcL* site: 102, amino acid: S, p-value = 95%, Table S4 and S5). More precisely, the non-synonymous amino acid Lysine (K) in the *ccsA* gene differentiates the *F. lucida* outgroup sample from the rest (Fig. S3a, Fig. 2b), and in the *rbcL* gene, Serine (S) distinguishes *Q. canariensis, Q. pontica*, *Q. robur*, 2-194-4 CladeC (*Q. petraea*), *Q. petraea*, *Q. rubra*, and *Q. pubescens* from the rest (Fig. S3b, Fig. 2b). The M2a model's site classes for both genes indicated that approximately 98.0% of the sites were under purifying selection, 2.0% had sites with ω > 1, and 0.0% with ω = 1. Branch-site models applied to *ccsA* and *rbcL* did not confirm positive selection on the foreground branches (branches with sites under selection), with the alternative A1 model (fixed ω = 1) being more suitable for both genes.

Discussion

Whole plastome phylogeny Quercus and phylogenetic relationships in Quercus

Using a large dataset of whole chloroplast genomes (114) mainly from *Quercus robur* and *Q. petraea* originating throughout Europe, our phylogeographic analysis revealed three highly supported clades with no clear species delimitation between *Q. robur* and *Q. petraea*. Samples within the same haplotype have nearly identical sequences (Fig. 1a, Table S1), with few shared variants between the haplotypes, indicating the presence of three distinct lineages (Table S1). Nevertheless, the overall genetic diversity is quite low (0.4% variable positions), as would be expected under a scenario of repeated chloroplast capture or intense selective pressure. Thus, our results are consistent with previous findings regarding the lack of chloroplast DNA variation in these species (Petit et al. 1997, 2002). Geographically, lineages are distributed in a similar but not completely identical pattern as previously reported (Petit et al. 2002) with all three of them overlapping to some extent throughout their range (Fig. 1b). Clade A follows a pan-European distribution, while clade B is more restricted to the North-East part of Europe (Denmark, England and Germany). The samples from Poland are bearing traces of introgression from *Q. pubescens* (admixed *Q. petraea-Q. pubescens* samples), forming a slightly different haplotype, expanding its distribution up to Poland (Fig. 1). Finally, clade C is found in Central and Northern Europe reaching Norway (Fig. 1b). It has been previously proposed that the distribution of the three primary haplotypes reflects the post-glacial dispersal of oaks from the different refugia (Brewer et al. 2002). Specifically, for Denmark, in line with prior studies we find all chloroplast haplotypes present, suggesting post-glacial colonization of oaks in Denmark from both eastern and western refugia (Jøhnk and Siegismund 1997). Additional samples from key areas closer to the glacial refugia and the range limits of each species could further enhance the observed geographic pattern and detect the geographic origin and borders of each haplotype. Moreover, the slight differences in the haplotypic distribution in our study compared to previous findings (Petit et al. 2002; Pakull et al. 2025) indicate the potential of a finer resolution in the geographic distribution of the haplotypes when whole plastomes are used as markers.

While hybridization between *Q. robur* and *Q. petraea* is possible in both directions (Olrik and Kjaer, 2007), female *Q. robur* with male *Q. petraea* is the most common direction in the wild (Petit et al. 2004). Therefore, the scenario of seed dispersal of *Q. robur* preceding pollen dispersal of *Q. petraea*, leading to a gradual replacement of some of the nuclear genomes and chloroplast captures is a likely scenario explaining the observed genomic patterns (Petit et al. 1997, 2002). However, our results highlight the importance of dating the divergence time within *Q. robur* and *Q. petraea* haplotypes in order to better understand the evolutionary history of these species as we show that chloroplast sharing potentially predates the post-glacial migration (see below).

Old splits of plastome lineages, but young splits between species

The recovery of the Fagaceae phylogeny based on 69 plastid genes is in congruence with previous studies (Zhou et al. 2022). Our analyses showed a highly supported Fagaceae topology, with several genera forming monophyletic groups (Fig. 2). We observed plastome sharing within *Quercus* and *Notholithocarpus* sections, echoing findings from Zhou et al. 2022 (Fig. 2a). Notably, within *Quercus* shared plastome haplotypes frequently corresponded to phylogenetically divergent clades with overlapping geographic ranges, e.g. between *Ilex-Cerris*, *Ponticae-Quercus*, *Protobalanus-Quercus*, and *Virentes-Quercus* (Fig. 2a), a pattern far from unpredicted among sympatric oak chloroplast lineages (Petit et al. 2002; Pham et al. 2017; Yang et al. 2021; Zhou et al. 2022).

The chronogram obtained from the molecular dating analysis indicates that divergence times of the main splits (*Trigonobalanus*, *Castanea-Castanopsis*, *Lithocarpus*) align closely with former estimates around 60 – 50 Ma (Zhou et al. 2022), supporting a high diversification trend/pattern after the end of the Cretaceous-Paleogene mass extinction (Wang et al. 2012; Fig. 2a). The divergence of *Lithocarpus* from the younger genera occurred around 20 Ma, with the diversification within section *Quercus* being dated at 13 Ma, coinciding with an increasing speciation period during Oligocene and Miocene (Zhou et al. 2022; Fig. 2a).

The newly incorporated *Q. robur* and *Q. petraea* samples were found to be non-monophyletic, as expected (Fig. 2b). In contrast with the previous topology (Fig. 1a), regarding the *Q. robur-Q. petraea* lineages, samples from clades A and B are now identified as sister clades with higher support on their nodes (Fig. 2b). This result is not surprising given the low support previously observed for clades' B and C relationship (Fig. 1a). It is estimated that the divergence among all main subclades formed by our newly added samples was around 2.68 Ma, further clarifying the distinction of the three plastome haplotypes (Fig. 2; Table 2). The oldest *Q. robur – Q. petraea* split within a clade is found in clade C (0.39 Ma) and corresponds to a pre-LGM period, suggesting chloroplast sharing should have occurred prior to the LGM (20.000 ya; Fig. 2b; Table 2). This finding could indicate an older split between the species' chloroplast genomes than previously thought. However, this pattern is not as clear for clades A and B, given that the time estimates

range from very recently to well before the LGM (Fig. 2; Table 2). Overall, it is worth noting that our results suggest these splits are related to the LGM period and possibly the glacial refugia in Europe. Similar patterns to those of *Q. robur* and *Q. petraea* were observed in other *Quercus* species, such as *Q. aliena*, *Q. serrata*, *Q. dentata*, and *Q. cerris*, in both our study and that of Zhou et al. (2022). Divergence time estimates of more recent plastome lineages within the same *Quercus* species can be extremely challenging using only ancient fossil calibrations, especially in cases of complete chloroplast capture events. This highlights the difficulties arising from the nuclear-cytoplasmic lineage discordance and the overall incompleteness of the fossil record, which leads to a lack of access to younger fossil calibrations in the genus. Moreover, it is important to mention that molecular dating analyses can produce different divergence estimates, depending on the dating method, calibration information and/or model settings (Saladin et al. 2017).

Hence, while the hypothesis of parallel chloroplast captures occurring post-LGM following the dispersal from the refugia cannot be dismissed, our findings suggest that the evolutionary history of plastome sharing among European oaks is more complex than previously thought. Our findings therefore call for further investigation and a more enriched analysis including samples from southern provenances, closer to the LGM refugia and more genes (or whole plastomes) alongside nuclear data for higher accuracy on divergence times and detection of introgression.

In summary, in *Q. robur* and *Q. petraea*, the plastome genetic variation appears limited, alongside the non-monophyly between the species' plastomes, both confirming the occurrence of chloroplast capture events. Additionally, our findings suggest a more recent plastome divergence compared to the divergence time of the species that has been estimated around 10 Ma (Hipp et al. 2020). This comparison along with the previously performed coalescent-based simulations regarding cytoplasmic-nuclear discordance (Zhou et al. 2022), could further support the case of hybridization and chloroplast captures rather than ILS alone. Taking into account the species' incomplete reproductive isolation, the low mutation rate of the chloroplast genome (approximately one or fewer mutations/kb/Ma; Rousseau-Gueutin et al. 2018), and the long generation times of these trees (with a minimum of 30 years to first reproduction and typically long lifespan; Ducousso et al. 1993), it is possible that their divergence estimates (both upper and lower) are relatively recent in geological terms, hence not sufficient time has passed to observe substantial levels of variation. It is highly prominent that different genome types can have different evolutionary histories, thus multiple processes could have been involved in the evolutionary history of a species or genus. Especially within genus *Quercus* and even within the Fagaceae family, such patterns are far from uncommon (Acosta et al. 2010; Liu et al. 2021; Zhou et al. 2022). This fact increases the need for finding ways to resolve such relationships in order to infer evolutionary histories more accurately and by doing so, get clearer insights into speciation and adaptation patterns. Our in-depth analyses of the *Quercus* plastomes show the value in analysing these uniparental markers in order to gain a better understanding of the evolutionary history, especially in species-complexes where the nuclear patterns can be obscured due to frequent, recurrent, and/or on-going gene flow both within and between species.

Selection of the large subunit of RuBisCo can explain low plastome diversity among European oaks

Our selection analyses revealed two genes, *ccsA* and *rbcL*, under positive selection, both with one site under selection. The *ccsA* gene is essential for the c-type cytochrome synthesis and encoding of a hydrophobic protein crucial for heme attachment (Xie & Merchant, 1996; Nakamoto et al., 2000). More specifically, the amino acid lysine (K) distinguished the outgroup sample *F. lucida*, inferring positive selection on this gene. Similarly, to our study, *ccsA* gene has been found to be under positive selection in pairwise comparisons among other *Quercus* species (Liu et al., 2021) as well as in other angiosperm genera, including orchids (Dong et al., 2018), *Speirantha gardenii* (Raman & Park, 2020), and Gentianaceae (Yang et al., 2023). These findings suggest that positively selected amino acids in the *ccsA* gene could be associated with advantageous traits, though more in-depth analysis is needed to confirm these signs of selection. The *rbcL* gene encodes the large subunit of the RuBisCo enzyme and is often used in plant phylogenetic analysis and encoding. Here, the substitution of asparagine (N) with the non-synonymous serine (S) differentiated samples from the *Q. pubescens* subclade, and the samples of *Q. robur*, *Q. petraea*, and one of our *Q. petraea* samples (clade C; Fig. 2b, Fig. S3). Hermida-Carrera et al. (2017) observed five sites under positive selection in the *rbcL* gene across different *Quercus* lineages, with these replacements likely related to various physiological and climatic factors. Similar findings of positive selection in the *rbcL* gene have been reported in other plant groups (e.g., Kapralov & Filatov, 2007; Iida et al., 2009; Yang et al., 2020; Bernabeu & Rosselló, 2021). A selective sweep could have occurred by the potentially beneficial mutations in the *rbcL* gene in *Quercus* lineages, reducing genetic diversity in the nonrecombining chloroplast (Stephan, 2019). This scenario could have led to the elimination of genetic variation across the entire haplotype upon fixation of the advantageous allele, further explaining the observed low genetic

diversity among *Q. robur* and *Q. petraea* plastomes. Additional analysis on a protein-level could further support this hypothesis. Such selective sweeps could also impact the mitogenome (Sloan et al. 2017). However, maternally inherited genomes like plastome and mitogenome, are also prone to strong genetic drift due to their lower effective population size, uniparental inheritance and their limited dispersal by seeds (Camus et al. 2022). It is therefore also possible that the reduced genetic variation observed in the *Quercus* plastomes is due to more frequent extinctions of variants during population decline.

Protein structure analysis could be used for genes like *ycf1* that showed signs of positive selection but could not be included here due to alignment issues to detect whether any of the indels can indeed produce a protein with a selection advantage. Finally, the observed genes containing stop-codons may indicate the presence of degraded genes resulting from cytonuclear coevolution and coadaptation in the chloroplast (Rousseau-Gueutin et al. 2018). However, additional analyses are required to explore these questions. While positive selection and/or genetic drift can cause the observed low genetic variation in the *Quercus* plastomes, it is also possible that cyto-nuclear interactions have a role (Rousseau-Gueutin et al. 2018; Camus et al. 2022; Larson et al. 2026). In depth analyses of plastomes can therefore shed light on evolutionary processes that form not only the plastomes but also the nuclear genome.

Conclusion

Overall, our results align with the pattern of complete chloroplast capture events, explaining the extensive sharing of chloroplast lineages between *Q. robur* and *Q. petraea*. However, our molecular dating analysis revealed evidence of older splits among the three chloroplast lineages of *Q. robur* and *Q. petraea*, and more recent splits within the lineages, suggesting that chloroplast sharing between these two species may be older than the LGM. A wider sampling to examine the distribution of haplotypes within the populations of both species would allow for further support/rejection regarding the age of these events. We also detected signs of positive selection in two plastid genes, which offers an alternative and/or complementary explanation for the low genetic diversity among European oak chloroplast genomes. Specifically, the positive selection detected on the *rbcL* gene, is potentially linked to adaptive evolution in the most dominant carbon dioxide fixing enzyme, RuBisCo. Supplementary analysis using complete chloroplast genomes along with nuclear genome analysis could provide a more comprehensive and an even more accurately dated phylogeny. Nuclear genome studies might also reveal introgression events, thus a more precise timing of chloroplast capture events. Finally, analysis of amino acids under selection from a protein function perspective could shed light on how these replacements may be associated with environmental adaptations. Taken together, these findings demonstrate that the study of cytoplasmic genome evolution reveals the diverse trajectories these genomes can follow, while offering unique research opportunities to explore fundamental interactions in nature, such as those of adaptive evolution, natural selection, and demographic history.

Declarations

Acknowledgments

The generation of the genomic data used in this study was funded by the Independent Research Fund Denmark, Case number 0217-00121B.

Funding

The generation of the genomic data used in this study was funded by the Independent Research Fund Denmark, Case number 0217-00121B.

Conflict of interest

The authors declared no conflict of interest.

Ethics approval

Not applicable.

Consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and material

All raw sequence data will be made available at NCBI's short Read Archive upon acceptance. Assembled plastomes will be deposited in a publicly available repository upon acceptance.

Code availability

Not applicable.

Author contributions

JKO and EDK conceived and designed the study. JKO and DS generated plastome data. DS performed phylogenetics, molecular dating and selection analysis with input from JKO, DW, and LTD. DS wrote the manuscript with input and feedback from all authors.

References

1. Acosta MC, Premoli AC (2010) Evidence of chloroplast capture in south American *Nothofagus* (subgenus *Nothofagus*, Nothofagaceae). https://doi.org/10.1016/j.ympev.2009.08.008. Mol Phylogenet Evol
2. Aldrich PR, Cavender-Bares J (2011) Quercus. Wild Crop Relatives: Genomic and Breeding Resources. Forest Trees. Springer, Berlin Heidelberg, Berlin, pp 89–129
3. Backs JR, Ashley MV (2021) *Quercus* Genetics: Insights into the past, present, and future of oaks. Forests. https://doi.org/10.3390/f12121628
4. Bernabeu M, Rosselló JA (2021) Molecular Evolution of *rbcL* in *Orthotrichales* (Bryophyta): Site Variation, Adaptive Evolution, and Coevolutionary Patterns of Amino Acid Replacements. J Mol Evol. https://doi.org/10.1007/s00239-021-09998-w
5. Bouckaert R, Vaughan TG, Barido-Sottani J, Duchêne S, Fourment M, Gavryushkina A et al (2019) BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. PLoS Comput Biol. https://doi.org/10.1371/journal.pcbi.1006650
6. Brewer S, Cheddadi R, De Beaulieu JL, Reille M (2002) The spread of deciduous *Quercus* throughout Europe since the last glacial period. Ecol Manag. https://doi.org/10.1016/S0378-1127(01)00646-6
7. Couvreur TL, Helmstetter AJ, Koenen EJ, Bethune K, Brandão RD, Little SA et al (2019) Phylogenomics of the major tropical plant family Annonaceae using targeted enrichment of nuclear genes. Front Plant Sci. https://doi.org/10.3389/fpls.2018.01941
8. Crepet WL, Nixon KC (1989) Earliest megafossil evidence of Fagaceae: phylogenetic and biogeographic implications. Am J Bot. https://doi.org/10.1002/j.1537-2197.1989.tb15062.x
9. Daniell H, Lin CS, Yu M, Chang WJ (2016) Chloroplast genomes: diversity, evolution, and applications in genetic engineering. Genome Biol. https://doi.org/10.1186/s13059-016-1004-2
10. Denk T, Grimm GW, Manos PS, Deng M, Hipp AL (2017) An updated infrageneric classification of the oaks: review of previous taxonomic schemes and synthesis of evolutionary patterns. In: Gil-Pelegrín E, Peguero-Pina J, Sancho-Knapik D (eds) Oaks Physiological Ecology. Exploring the Functional Diversity of Genus *Quercus* L. Tree Physiology, vol 7. Springer, Cham, pp 13–38
11. Dierckxsens N, Mardulyn P, Smits G (2017) NOVOPlasty: de novo assembly of organelle genomes from whole genome data. Nucleic Acid Res. https://doi.org/10.1093/nar/gkw955
12. Dong WL, Wang RN, Zhang NY, Fan WB, Fang MF, Li ZH (2018) Molecular evolution of chloroplast genomes of orchid species: insights into phylogenetic relationship and adaptive evolution. Int J Mol Sci. https://doi.org/10.3390/ijms19030716
13. Ducousso A, Michaud H, Lumaret R (1993) Reproduction and gene flow in the genus *Quercus L.* In: Ann For Sci, Vol. 50, No. Supplement, EDP Sciences, pp 91–106
14. Eaton EGSDJ, Caudullo G, Oliveira S, De Rigo D (2016) *Quercus robur* and *Quercus petraea* in Europe: distribution, habitat, usage and threats. In: San-Miguel-Ayanz J, de Rigo D, Caudullo G (eds) European Atlas of Forest Tree Species. Publications Office of

the European Union, Luxembourg, pp 160–163

15. Gandini CL, Sanchez-Puerta MV (2017) Foreign plastid sequences in plant mitochondria are frequently acquired via mitochondrion-to-mitochondrion horizontal transfer. Sci rep. https://doi.org/10.1038/srep43402
16. Geneious P (2023) 0.4. https://www.geneious.com
17. Gil-Pelegrín E, Peguero-Pina JJ, Sancho-Knapik D (2017) Oaks physiological ecology. Exploring the functional diversity of genus *Quercus L*. Springer, Cham
18. Grímsson F, Grimm GW, Zetter R, Denk T (2016) Cretaceous and Paleogene Fagaceae from North America and Greenland: evidence for a Late Cretaceous split between *Fagus* and the remaining Fagaceae. Acta Palaeobot. https://doi.org/10.1515/acpa-2016-0016
19. Gugger PF, Fitz-Gibbon ST, Albarrán-Lara A, Wright JW, Sork VL (2021) Landscape genomics of *Quercus lobata* reveals genes involved in local climate adaptation at multiple spatial scales. Mol Ecol. https://doi.org/10.1111/mec.15731
20. Hermida-Carrera C, Fares MA, Fernandez A, Gil-Pelegrin E, Kapralov MV, Mir A et al (2017) Positively selected amino acid replacements within the RuBisCO enzyme of oak trees are associated with ecological adaptations. PLoS ONE. https://doi.org/10.1371/journal.pone.0183970
21. Hipp AL, Manos PS, Hahn M, Avishai M, Bodénès C, Cavender-Bares J et al (2020) Genomic landscape of the global oak phylogeny. New Phytol. https://doi.org/10.1111/nph.16162
22. Iida S, Miyagi A, Aoki S, Ito M, Kadono Y, Kosuge K (2009) Molecular adaptation of *rbcL* in the heterophyllous aquatic plant *Potamogeton*. PLoS ONE. https://doi.org/10.1371/journal.pone.0004633
23. Johnson MG, Pokorny L, Dodsworth S, Botigue LR, Cowan RS, Devault A et al (2019) A universal probe set for targeted sequencing of 353 nuclear genes from any flowering plant designed using k-medoids clustering. Syst Biol. https://doi.org/10.1093/sysbio/syy086
24. Jøhnk N, Siegismund HR (1997) Population structure and post-glacial migration routes of *Quercus robur* and *Quercus petraea* in Denmark, based on chloroplast DNA analysis. Scand J Res. https://doi.org/10.1080/02827589709355394
25. Kapralov MV, Filatov DA (2007) Widespread positive selection in the photosynthetic Rubisco enzyme. BMC Evol Biol. https://doi.org/10.1186/1471-2148-7-73
26. Katoh K, Misawa K, Kuma KI, Miyata T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acid Res. https://doi.org/10.1093/nar/gkf436
27. Kremer A, Hipp AL (2020) Oaks: an evolutionary success story. New Phytol. https://doi.org/10.1111/nph.16274
28. Kremer A, Abbott AG, Carlson JE, Manos PS, Plomion C, Sisco P, Genomes et al (2012) https://doi.org/10.1007/s11295-012-0498-3
29. Larson DA, Itgen MW, Denton RD, Hahn MW (2026) Reconsidering cytonuclear discordance in the genomic age. Evol. https://doi.org/10.1093/evolut/qpaf201
30. Lefort V, Longueville JE, Gascuel O (2017) SMS: smart model selection in PhyML. Mol Ecol Evol. https://doi.org/10.1093/molbev/msx149
31. Liu BB, Campbell CS, Hong DY, Wen J (2020) Phylogenetic relationships and chloroplast capture in the Amelanchier-Malacomeles-Peraphyllum clade (Maleae, Rosaceae): Evidence from chloroplast genome and nuclear ribosomal DNA data using genome skimming. Mol Phylogenet Evol. https://doi.org/10.1016/j.ympev.2020.106784
32. Liu X, Chang E, Liu J, Jiang Z (2021) Comparative analysis of the complete chloroplast genomes of six white oaks with high ecological amplitude in China. J Res. https://doi.org/10.1007/s11676-020-01288-3
33. Lobo A, Doonan JM, Hansen JK, Kosawang C, Xu J, Olofsson JK et al (2026) Genomic prediction of superior heartwood traits in pedunculate oak (*Quercus robur* L.) using GWAS-informed Bayesian models. https://doi.org/10.1016/j.foreco.2026.123644. For Ecol Manag
34. Manos PS, Stanford AM (2001) The historical biogeography of Fagaceae: tracking the tertiary history of temperate and subtropical forests of the Northern Hemisphere. Int J Plant Sci. https://doi.org/10.1086/323280
35. Muir G, Filatov D (2007) A selective sweep in the chloroplast DNA of dioecious Silene (Section Elisanthe). https://doi.org/10.1534/genetics.107.071969. Genetics

36. Nakamoto SS, Hamel P, Merchant S (2000) Assembly of chloroplast cytochromes b and c. https://doi.org/10.1016/S0300-9084(00)00605-2. Biochimie
37. Olrik DC, Kjaer ED (2007) The reproductive success of a *Quercus petraea* x *Q. robur* F1-hybrid in back-crossing situations. Ann Sci. https://doi.org/10.1051/forest:2006086
38. Paradis E, Schliep K (2019) ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics. https://doi.org/10.1093/bioinformatics/bty633
39. Petit RJ, Pineau E, Demesure B, Bacilieri R, Ducousso A, Kremer A (1997) Chloroplast DNA footprints of postglacial recolonization by oaks. Proc Nat Acad Sci. https://doi.org/10.1073/pnas.94.18.9996
40. Petit RJ, Csaikl UM, Bordács S, Burg K, Coart E, Cottrell J et al (2002) Chloroplast DNA variation in European white oaks: phylogeography and patterns of diversity based on data from over 2600 populations. doi.org/10.1016/S0378-1127(01)00645-4. For Ecol Managhttps://
41. Petit RJ, Bodénès C, Ducousso A, Roussel G, Kremer A (2004) Hybridization as a mechanism of invasion in oaks. New Phytol. https://doi.org/10.1046/j.1469-8137.2003.00944.x
42. Pham KK, Hipp AL, Manos PS, Cronn RC (2017) A time and a place for everything: phylogenetic history and geography as joint predictors of oak plastome phylogeny. https://doi.org/10.1139/gen-2016-0191. Genome
43. Postel Z, Touzet P (2020) Cytonuclear genetic incompatibilities in plant speciation. Plants. https://doi.org/10.3390/plants9040487
44. POWO (2026) Plants of the World Online. Facilitated by the Royal Botanic Gardens, Kew. Published on the Internet; https://powo.science.kew.org/ Retrieved 29 May 2026
45. Premoli AC, Mathiasen P, Cristina Acosta M, Ramos VA (2012) Phylogeographically concordant chloroplast DNA divergence in sympatric Nothofagus ss How deep can it be? New Phytol. https://doi.org/10.1111/j.1469-8137.2011.03861.x
46. Raman G, Park S (2020) The complete chloroplast genome sequence of the *Speirantha gardenii*: Comparative and adaptive evolutionary analysis. Agronomy. https://doi.org/10.3390/agronomy10091405
47. Rambaut A (2014) FigTree v1. 4.2. University of Edinburgh, Edinburgh http://tree.bio.ed.ac.uk
48. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA (2018) Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Syst Biol. https://doi.org/10.1093/sysbio/syy032
49. Rieseberg LH, Soltis DE (1991) Phylogenetic consequences of cytoplasmic gene flow in plants. Evol Trends Plants 5:65–84
50. Rousseau-Gueutin M, Martin G, Ferreira de Carvalho J, Keller J, Aïnouche A (2018) The Intertwined Chloroplast and Nuclear Genome Coevolution in Plants. In: Ratnadewi D, Hamim (eds) Plant Growth and Regulation Alterations to Sustain Unfavorable Conditions. IntechOpen, London, p 61
51. Ruhfel BR, Gitzendanner MA, Soltis PS, Soltis DE, Burleigh JG (2014) From algae to angiosperms–inferring the phylogeny of green plants (Viridiplantae) from 360 plastid genomes. BMC Evol Biol. https://doi.org/10.1186/1471-2148-14-23
52. Saintagne C, Bodénès C, Barreneche T, Pot D, Plomion C, Kremer A (2004) Distribution of genomic regions differentiating oak species assessed by QTL detection. https://doi.org/10.1038/sj.hdy.6800358. Heredity
53. Saladin B, Leslie AB, Wüest RO, Litsios G, Conti E, Salamin N et al (2017) Fossils matter: improved estimates of divergence times in Pinus reveal older diversification. BMC Evol Biol. https://doi.org/10.1186/s12862-017-0941-z
54. Schubert M, Lindgreen S, Orlando L (2016) AdapterRemoval v2: rapid adapter trimming, identification, and read merging. BMC Res Notes. https://doi.org/10.1186/s13104-016-1900-2
55. Shaw J, Shafer HL, Leonard OR, Kovach MJ, Schorr M, Morris AB (2014) Chloroplast DNA sequence utility for the lowest phylogenetic and phylogeographic inferences in angiosperms: the tortoise and the hare IV. Am J Bot. https://doi.org/10.3732/ajb.1400398
56. Simeone MC, Grimm GW, Papini A, Vessella F, Cardoni S, Tordoni et al (2016) Plastome data reveal multiple geographic origins of *Quercus* Group Ilex. https://doi.org/10.7717/peerj.1897. PeerJ
57. 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. https://doi.org/10.7717/peerj.5793. PeerJ

58. Sloan DB, Havird JC, Sharbrough J (2017) The on-again, off-again relationship between mitochondrial genomes and species boundaries. Mol Ecol. https://doi.org/10.1111/mec.13959
59. Stamatakis A (2014) Bioinform. https://doi.org/10.1093/bioinformatics/btu033. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies
60. Stephan W (2019) Selective sweeps. Genet. https://doi.org/10.1534/genetics.118.301319
61. Tsitrone A, Kirkpatrick M, Levin DA (2003) A model for chloroplast capture. https://doi.org/10.1111/j.0014-3820.2003.tb00585.x. Evol
62. Wang W, Ortiz RDC, Jacques FM, Xiang XG, Li HL, Lin L et al (2012) Menispermaceae and the diversification of tropical rainforests near the Cretaceous–Paleogene boundary. https://doi.org/10.1111/j.1469-8137.2012.04158.x. New Phytol
63. Whittemore AT, Schaal BA (1991) Interspecific gene flow in sympatric oaks. Proc Nat Acad Sci. https://doi.org/10.1073/pnas.88.6.2540
64. Wilf P, Nixon KC, Gandolfo MA, Cúneo NR (2019) Eocene Fagaceae from Patagonia and Gondwanan legacy in Asian rainforests. https://doi.org/10.1126/science.aaw5139. Sci
65. Xie Z, Merchant S (1996) The Plastid-Encoded *ccsA* Gene Is Required for Heme Attachment to Chloroplast c-Type Cytochromes (∗). J Biol Chem. https://doi.org/10.1074/jbc.271.9.4632
66. Yang J, Kang GH, Pak JH, Kim SC (2020) Characterization and comparison of two complete plastomes of Rosaceae species (*Potentilla dickinsii* var. *glabrata* and *Spiraea insularis*) endemic to Ulleung Island, Korea. Int J Mol Sci. https://doi.org/10.3390/ijms21144933
67. Yang L, Deng S, Zhu Y, Da Q (2023) Comparative chloroplast genomics of 34 species in subtribe Swertiinae (Gentianaceae) with implications for its phylogeny. BMC Plant Biol. https://doi.org/10.1186/s12870-023-04183-1
68. Yang Y, Zhou T, Qian Z, Zhao G (2021) Phylogenetic relationships in Chinese oaks (Fagaceae, *Quercus*): Evidence from plastid genome using low-coverage whole genome sequencing. https://doi.org/10.1016/j.ygeno.2021.03.01. Genom
69. Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. https://doi.org/10.1093/molbev/msm088
70. Yang Z, Bielawski JP (2000) Statistical methods for detecting molecular adaptation. Trends Ecol Evol. https://doi.org/10.1016/S0169-5347(00)01994-7
71. Yang Z, Wong WS, Nielsen R (2005) Bayes empirical Bayes inference of amino acid sites under positive selection. Mol Biol Evol. https://doi.org/10.1093/molbev/msi097
72. Zhou BF, Yuan S, Crowl AA, Liang YY, Shi Y, Chen X et al (2022) Phylogenomic analyses highlight innovation and introgression in the continental radiations of Fagaceae across the Northern Hemisphere. Nat Commun. https://doi.org/10.1038/s41467-022-28917-1
73. Zuckerkandl E, Pauling L (1962) Molecular disease, evolution and genic heterogeneity. In: Kasha M, Pullman B (eds) Horizons in biochemistry. Academic, New York, pp 189–225

Figures

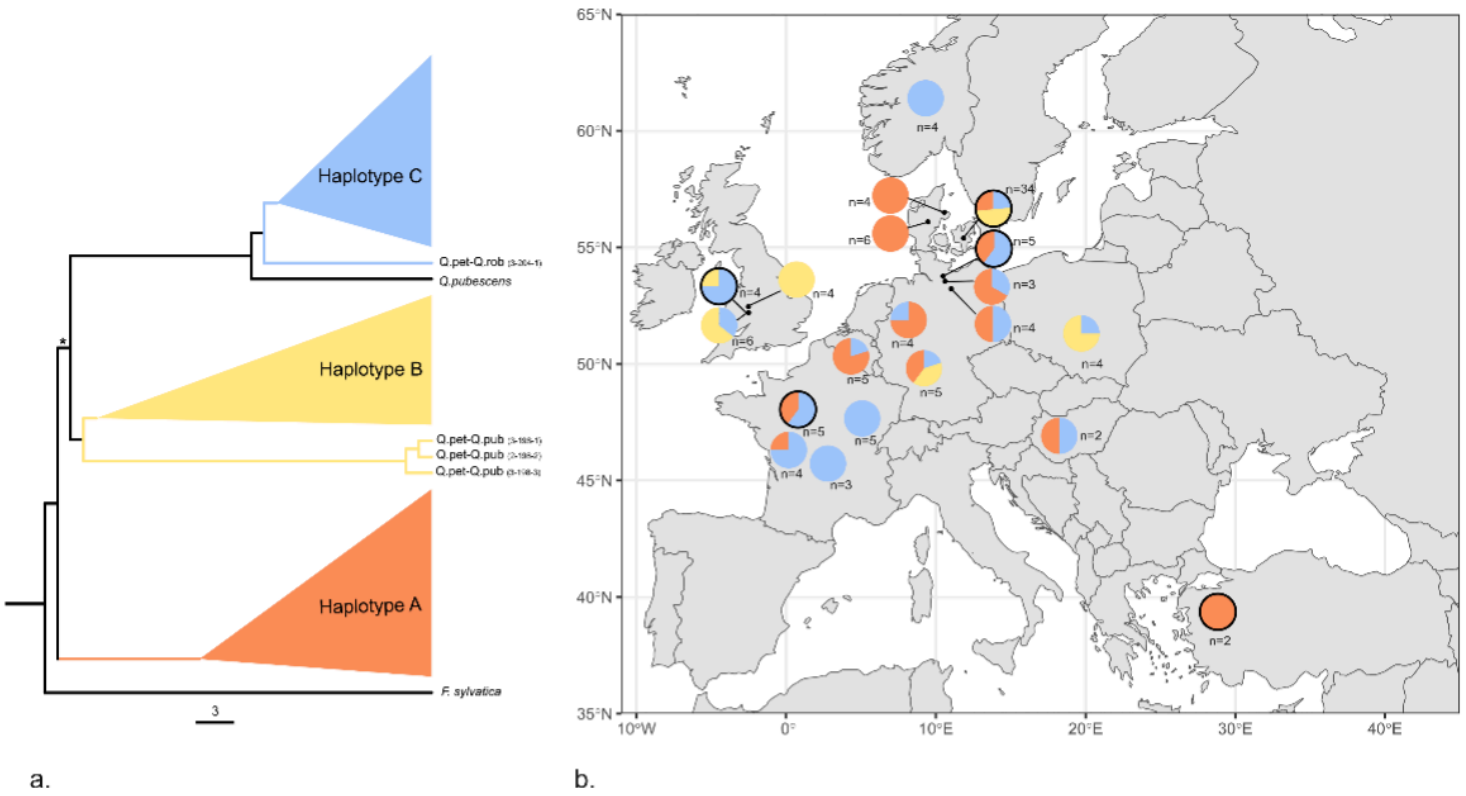


Figure 1

(a) Maximum Likelihood phylogeny based on whole chloroplast genomes. Complete tree with the outgroup *Fagus sylvatica* (beech). Monophyletic groups are collapsed in triangles, and colors represent different clades (A=orange, B=yellow, C=light blue). The haplotype names are shown in the corresponding groups on the tree. Bootstrap values >90% were produced on all main nodes, with the exception of the single labeled node. Sample code is shown in parentheses for the admixed individuals. **(b)** Map with pie charts representing the percentages of each haplotype per provenance (sampling site). Haplotypes are coloured according to the phylogenetic tree. Pie charts with a black outline correspond to provenances that include both *Q. robur* and *Q. petraea* individuals; plain pie charts include only *Q. petraea* individuals. All admixed individuals have been treated as *Q. petraea* for this graph. *Q. pubescens* (3-199-1) from Turkey is not assigned to any haplotype, thus it is not included in the map.

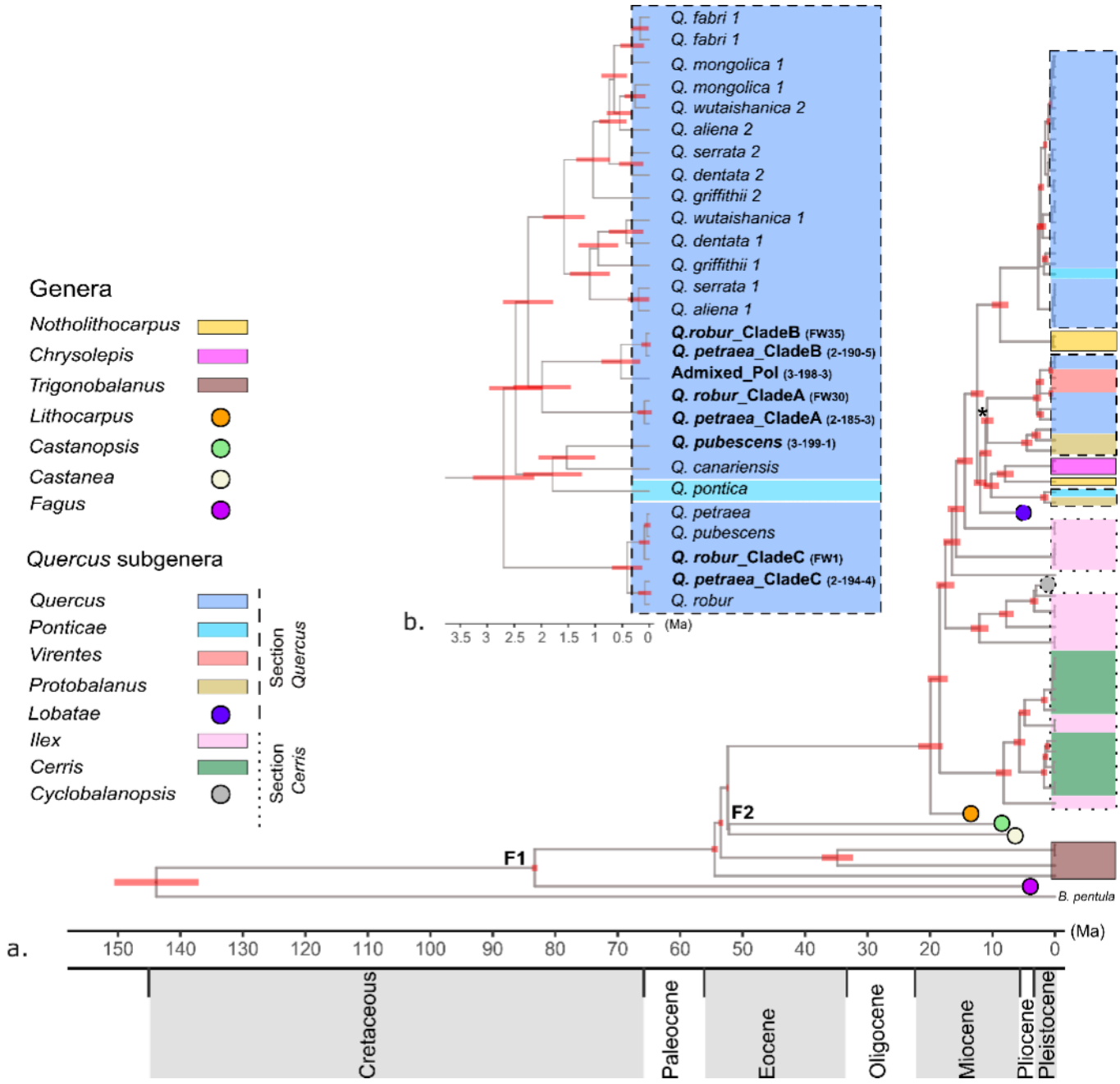


Figure 2

(a) Phylogenetic relationships and divergence time estimation of Fagaceae derived from the BEAST tree based on 69 plastid genes. Monophyletic clades are collapsed (circles), and colors represent different genera according to the legend. *Quercus*subgenera are additionally marked with different dashed borders as shown in the legend. 95% confidence intervals of divergence time estimates are shown with light red on all nodes. Posterior probabilities (>99%) were produced on all main nodes, with the exception of the single labeled node, while F1 and F2 are two fossil calibration points. *Betula pentula* (MG386401) was used as an outgroup. The time scale at the bottom is in million years (Ma). Geological epochs are also shown. **(b)** Subclade containing the branches used in the selection analyses. All samples from this study are shown in bold; the clade name corresponds to clades according to Fig.1a. Sample code is shown in parentheses. *F. lucida*, and *Q. rubra* were used as outgroups in the selection analyses but are not shown here.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- TableS1.xlsx
- Supplementarymaterial.pdf
- TableS3.xlsx