Research Article

Multiple nuclear genes stabilize the phylogenetic backbone of the genus *Quercus*

FRANÇOIS HUBERT1,2, GUIDO W. GRIMM3, EMMANUELLE JOUSSELIN4, VINCENT BERRY5, ALAIN FRANC1,2 & ANTOINE KREMER1,2

1INRA, UMR BIOGECO-1202, 69 route d’Arcachon, F-33612 Cestas, France
2Université Bordeaux, UMR BIOGECO-1202, Avenue des Facultés, F-33405 Talence, France
3Swedish Museum of Natural History, Department of Palaeobiology, Box 50007, 10405 Stockholm, Sweden
4INRA, UMR 1062 CBGP Centre de Biologie pour la Gestion des Populations, Campus International de Baillarguet, F-34 988 Montferrier-sur-Lez, France
5UMR LIRMM, Equipe Méthodes et Algorithmes pour la Bioinformatique, Université Montpellier II, Place Eugène Bataillon, F-34095 Montpellier, France

(Received 26 November 2013; accepted 13 June 2014)

Phylogenetic relationships among 108 oak species (genus *Quercus* L.) were inferred using DNA sequences of six nuclear genes selected from the existing genomic resources of the genus. Previous phylogenetic reconstructions based on traditional molecular markers are inconclusive at the deeper nodes. Overall, weak phylogenetic signals were obtained for each individual gene analysis, but stronger signals were obtained when gene sequences were concatenated. Our data support the recognition of six major intrageneric groups Cyclobalanopsis, Cerris, Ilex, Quercus, Lobatae and Protobalanus. Our analyses provide resolution at deeper nodes but with moderate support and a more robust infragenic classification within the two major clades, the ‘Old World Oaks’ (Cyclobalanopsis, Cerris, Ilex) and ‘New World Oaks’ (Quercus, Lobatae, Protobalanus). However, depending on outgroup choice, our analysis yielded two alternative placements of the Cyclobalanopsis clade within the genus *Quercus*. When *Castanea* Mill. was chosen as outgroup, our data suggested that the genus *Quercus* comprised two clades corresponding to two subgenera as traditionally recognized by Camus: subgenus *Euquercus* Hickel and Camus and subgenus *Cyclobalanopsis* Ørsted (Schneider). However, when *Notholithocarpus* Manos, Cannon and S. Oh was chosen as an outgroup subgenus *Cyclobalanopsis* clustered with Cerris and Ilex groups to form the Old World clade. To assess the placement of the root, we complemented our dataset with published data of ITS and CRC sequences. Based on the concatenated eight gene sequences, the most likely root position is at the split between the ‘Old World Oaks’ and the ‘New World Oaks’, which is one of the alternative positions suggested by our six gene analysis. Using a dating approach, we inferred an Eocene age for the primary divergences in *Quercus* and a root age of about 50–55 Ma, which agrees with palaeobotanical evidence. Finally, irrespective of the outgroup choice, our data boost the topology within the New World clade, where (Protobalanus + Quercus) is a sister clade of Lobatae. Inferred divergence ages within this clade and the Cerris–Ilex clade are generally younger than could be expected from the fossil record, indicating that morphological differentiation pre-dates genetic isolation in this clade.

**Key words**: Fagaceae, gene tree, outgroup, phylogenetics, *Quercus*, species tree.

Introduction

Oaks (*Quercus* spp.) belong to one of the most widely distributed forest-tree genera. They are dominant species in many forest types ranging from temperate deciduous forests in North America, Europe and Asia, to Mediterranean and desert scrub forest in America and Europe, and finally to tropical montane forest in South America and South East Asia. Besides their economic and ecological importance, oaks are also considered in many countries as cultural and patrimonial resources. Depending on the authors, there are between 300 and 600 reported oak species, making it one of the largest tree genera of the northern hemisphere. Since the beginning of Linnaean taxonomy, the classification within the genus *Quercus* L. has been controversial, and numerous classifications have been proposed (Ørsted, 1867; Schwarz, 1936; Camus, 1936–1954; Menitsky, 1984, 2005; Nixon, 1993).
Discrepancies stem either from different nomenclatures used for the classification levels (subgenera, series, species) by the different authors, or from the lack of diagnostic morphological characters. Traditional characters in taxonomy as leaf shape and lobing, pubescence, acorn morphology, floral characters exhibit large intraspecific variation and frequent hybridization in oaks further increases their range of variation. Two nuclear-encoded molecular markers, the widely used ITS region of the 35S rDNA cistron and a 2415 bp long part of the CRABS plastid data seem to lack clear discriminating signals at the infrageneric level in oaks and other Fagaceae (Manos et al., 1999) and Nothofagaceae (Acosta & Premoli, 2010; Premoli et al., 2012). Hence, earlier phylogenetic studies (Manos et al., 1999, 2001; Manos & Stanford, 2001; Xu, 2004; Oh & Manos, 2008; Denk & Grimm, 2009, 2010) have so far failed to unambiguously resolve relationships among and within the six major groups of Quercus. Discrepancies among different studies and resolution problems could be partly attributed to incomplete (unrepresentative) taxon sampling (Manos & Stanford, 2001; Xu, 2004; Xu et al., 2005; Oh & Manos, 2008) and limited gene sampling (Manos et al., 2001; Denk & Grimm, 2010), but a further source of topological ambiguity seems to be out-group selection (Manos & Stanford, 2001; Xu, 2004; Xu et al., 2005) and a generally weak phylogenetic signal in all sequence regions analysed so far. Equally problematic is a cladistic-systematic interpretation of the molecular-based phylograms since Groups Protobalanus, Quercus and Ilex are either forming grades or clades depending on the phylogenies of these groups is pending, since morphological traits are highly homoplastic in oaks. Nevertheless, five of the six groups (Groups Protobalanus, Quercus-Lobatae, Cerris, Ilex and Cyclobalanopsis) can be unambiguously identified based on pollen ornamentation; furthermore, a number of pollen types are today exclusively found in North American red or white oaks (Solomon, 1983a, 1983b; Denk & Grimm, 2009; Denk et al., 2010, 2012). Phylogenetic inferences and species delineation in oaks using molecular data have been hindered by a high intraspecific genetic variation and low interspecific divergence in the genus observed in nuclear and plastid datasets (Kremer & Petit, 1993; Denk & Grimm, 2010; Kremer et al., 2012; Simeone et al., 2013) and frequent interspecific hybridization events (Rushton, 1993). In particular, plastid data seem to lack clear discriminating signals at the intrageneric level in oaks and other Fagaceae (Manos et al., 1999) and Nothofagaceae (Acosta & Premoli, 2010; Premoli et al., 2012). Here, we explore recently developed genomic resources for the genus (Ueno et al., 2010) to find new nuclear gene sequences for clarifying phylogenetic relationships among the six groups. We also use a comprehensive taxon

---

**Table 1.** Phylogenetic relationships between Quercus infragenic groups.

| Marker | Root | Ingroup topology | Remarks | Authors |
|--------|------|------------------|---------|---------|
| ITS    | Colombobalanus | (Cyclobalanopsis, ((Cerris, Ilex), (Lobatae, Quercus))) | Group Ilex is not recognized by the authors | Manos et al., 1999 |
|       | Trigonobalanus  | (Protobalanus, Quercus)) |         |         |
| ITS    | All Fagaceae genera | ((Cerris, Ilex), (Cyclobalanopsis, (Protobalanus, (Quercus, Lobatae)))) | Group Ilex is not recognized by the authors | Manos et al., 2001 |
| ITS    | Chrysolepis     | ((Cyclobalanopsis, (Cerris, Ilex)), (Protobalanus, (Lobatae, Quercus))) | Group Ilex is not recognized by the authors | Manos & Stanford, 2001 |
|        | Castanopsis     | (Protobalanus, (Lobatae, Quercus))) |         |         |
|        | Castanea        | (Protobalanus, (Lobatae, Quercus))) |         |         |
|        | Notholithocarpus | (Protobalanus, (Lobatae, Quercus))) |         |         |
| CRC +ITS | All Fagaceae genera | ((Cerris, (Cyclobalanopsis, Ilex)), (Protobalanus, (Lobatae, Quercus))) | Only one species of Group Ilex (Q. baronii) and Group Protobalanus is included | Oh & Manos, 2008 |
| ITS    | All Fagaceae genera except Fagus | ((Cyclobalanopsis, (Cerris, Ilex)), (Protobalanus, (Lobatae, Quercus))) | Group Lobatae nested within Group Quercus | Denk & Grimm, 2010 |
sampling covering all recognized intrageneric groups and the biogeographic distribution of the genus. In addition, we sampled two outgroup taxa to test for the effect of the outgroup selection on the ingroup topology (Castanea mollissima Blume and Notholithocarpus densiflorus (Hook. & Arn; Manos, Cannon & S. Oh). These taxa, along with Castanopsis, are most closely related to Quercus (Oh & Manos, 2008).

Our objectives were: (1) to test the monophyly of all of the infrageneric groups; (2) to improve the resolution at the deeper nodes of the phylogeny; (3) to investigate the effect of outgroups on the inferred genus root; and (4) to infer divergence ages of the major infrageneric splits.

Materials and methods
Species sampling
Leaf or bud material of 108 species of the genus Quercus, and two additional species belonging to the putatively closely related genera Notholithocarpus Manos, Cannon & S. Oh and Castanopsis and Castanea Mill. (Table 2) was collected. All samples were collected in situ from natural populations or collections (provenance tests or common garden experiments).

In situ collections of leaf or bud material in natural populations were made on one to seven individuals for each species located within a neighbourhood of less than 500 metres. The overall collection of specimens (210 individuals) covers a representative set of species from the six infrageneric groups (Groups Cyclobalanopsis, Quercus, Lobatae, Protobalanus, Cerris and Ilex) across their geographic distribution range (Table 2). A complete list of samples, their taxonomic classification and their geographic origin can be found in Table S1 (see online supplemental material, which is available from the article’s Taylor & Francis Online page at http://dx.doi.org/10.1080/14772000.2014.941037).

DNA extraction, marker selection, amplification and sequencing
Collected tissues were immediately preserved in dried silica gel or were frozen at −80°C. Genomic DNA was extracted using the Invisorb® DNA plant HTS 96 kit from Invitex Company using the manufacturer’s protocol. DNA quality was visualized on a 1% (w/v) agarose gel stained with GelRed® (Biotium, USA). DNA concentration was quantified on an eight channel Nanodrop® spectrophotometer and concentration of each sample was adjusted manually to approximately 10 ng/μL.

We used the resequencing data of a set of about 800 gene fragments in three European species: Q. petraea (Matt.) Liebl., Q. robur L. (Group Quercus) and Q. ilex L. (Group Ilex) selected within the first oak EST library constructed for the NCBI database (Ueno et al., 2010). This resequencing project of 800 fragments was conducted on 11 individuals (each) of Q. petraea and Q. robur and one individual of Q. ilex. These 800 fragments corresponded to 621 candidate genes (http://www.evoltree.eu/index.php/candidate-genes) potentially involved in adaptive

| Genus         | Infrageneric Group | Geographic origin  | Number. sampled species | Estimated number of species |
|---------------|---------------------|--------------------|-------------------------|----------------------------|
| Castanea      |                     |                    |                         |                            |
| Notholithocarpus |                   |                     |                         |                            |
| Quercus       | Cyclobalanopsis     | Asia               | 1                       | 8g                        |
|               |                     | North America      | 1                       | 1b                        |
|               |                     | Asia               | 9                       | 87c, 76c                  |
| Quercus       | Asia                | Europe and North Africa | 6                  | 25c, 15c                  |
|               |                     | North and South America | 8                 | 16d                       |
| Lobatae       | North and South America |               | 35                      | 104c                      |
| Protobalanus  | North America       |                    | 31                      | 82a                       |
| Cerris        | Asia                |                    | 31                      | 5b                        |
| Ilex          | Asia                | Europe and North Africa | 4                  | 8c                        |
|               |                     |                    | 4                       | 4d                        |
|               |                     | Europe and North Africa | 2                  | 4d                        |
| Total number of species |           |                    | 110                     |                            |

a Nixon (1993).
b Flora published on the internet (http://www.efloras.org).
c Menitsky (2005).
d Flora Europea published on the internet (http://rbg-web2.rbge.org.uk/FE/fe.html).
e Camus (1936–1954).
f Assignment of the study species to infrageneric groups (Group Cyclobalanopsis, Quercus, Lobatae and Protobalanus) was done according to Camus (1936–1954) and Valencia (2004). Assignments to Group Cerris and Group Ilex were done according to Menitsky (2005).
g Govaerts & Frodin (1998).
h Manos et al. (2008).
traits (such as growth, bud phenology, biotic interactions, etc.) and a subset of 87 control genes randomly selected in the EST database of *Quercus petraea* and *Q. robur* (Ueno et al., 2010). The sequences of the 87 control genes were analysed to select a subset of fragments, exhibiting interspecific variation between *Q. robur*, *Q. petraea* and *Q. ilex*, and no intraspecific variation. We set several criteria in order to select genes that might be useful for phylogenetic investigations. To ensure that they could be sequenced in one run, we set length ranges from 500 to 900 bp. Finally, amplification success and sequence quality were used as final decision criteria. All these tests aiming at the selection of genes were conducted on a panel of 14 individuals, comprising two individuals from each of the six infragenic groups of the genus *Quercus* and two trees of *Castanea mollissima*. Using these criteria, we selected six markers annotated throughout the Blastx searches on the Swissprot database (release July 2013) with an E-value cutoff of 10$^{-5}$ (CL1155: No hit, CL8461: No hit, CL8561: NADP-dependent malic enzyme, CL5191: 1-aminocyclopropane-1-carboxylate oxidase homolog 1, CL9715: Calnexin homologue, CL8745: No hit). The primer sequences are listed in Table 3.

**Table 3.** Gene annotation and primer sequences used for PCR amplifications.

| Gene   | Gene annotation                              | Primer$^a$ | Sequence                  |
|--------|----------------------------------------------|------------|---------------------------|
| CL1155 | No hit                                       | F          | GCTTGAGCCTTCCGAAGTG       |
|        |                                              | R          | CGCATATTGCACTCTCAACTCATC  |
| CL8461 | No hit                                       | F          | CGCTGCTTATGGTCATTTCGG     |
|        |                                              | R          | GAGTCATCATAGACAATGCGGTGG  |
| CL8561 | NADP-dependent malic enzyme                   | F          | CACCAAAATAAGGGGC          |
|        |                                              | R          | CAAATTCCTCAAAAACCTC       |
| CL5191 | 1-aminocyclopropane-1-carboxylate oxidase homologue 1 | F          | AAAAGTGCTGAGATTTTG        |
|        |                                              | R          | AAAACCCGCACTTAGTGTTCC     |
| CL9715 | Calnexin homologue                            | F          | TTGCTGTTYGGATGAAGAG       |
|        |                                              | R          | CAAAACCACCCACACAAAC       |
| CL8745 | No hit                                       | F          | CAGGATGCGATATTGGTG        |
|        |                                              | R          | TGATTGTCGGATGCTC          |

$^a$F: Forward, R: Reverse.

Sequence alignments

The obtained sequences were imported into Codon Code Aligner version 3.7 (CodonCode Corporation, Dedham, MA, USA). We only used sequences for which forward and reverse reads were of good quality. All other sequences were discarded. All sequences were edited using CodonCode Aligner version 3.7 and aligned with the Muscle algorithm (Edgar, 2004) of the Seaview software version 4 (Gouy et al., 2010). Coding parts of the markers were checked for protein coding frame shifts, to eliminate pseudogenes (Zhang & Hewitt, 1996) with Mega v4.1 (Tamura et al., 2007). Insertion-deletion events were treated as missing data.

When sequences from several individuals were available for a given species, we compiled a consensus sequence by using Bioedit v7.1.3 (Hall, 1999). This strategy allowed minimizing missing data (i.e. number of taxa for which gene sequences were missing). When for a site a nucleotide base was less frequent than 95%, the IUPAC (International Union of Pure and Applied Chemistry) ambiguity code was used. Accession numbers of consensus sequences are provided in Table S2 (see supplemental material online). In order to conduct a combined analysis we concatenated all sequences into a single alignment using Seaview. The final concatenated matrix comprised 112 taxa and six genes with a total of 2896 bp.
Phylogenetic reconstructions

Phylogenetic trees were rooted using *Notholithocarpus densiflorus* and/or *Castanea mollissima* as outgroups. Recent phylogenetic analysis of Fagaceae showed that these two genera are the closest relatives of oaks (Manos et al., 2008; Oh & Manos, 2008; Denk & Grimm, 2010).

We first performed Maximum Likelihood (ML) analyses on each individual gene and on the concatenated sequences and associated standard bootstrapping using the MPI-parallelized RAxML 7.2.8-ALPHA (Stamatakis, 2006; Stamatakis et al., 2008). Given that α, which scales the shape of the gamma distribution, and the proportion of invariable sites cannot be optimized independently from each other and following Stamatakis’ personal recommendations (RAxML manual), we used GTR + I with 25 discrete rate categories for all partitions (-m GTRGAMMA). Support was established based on 1000 standard bootstrap replicates.

For Bayesian Inference (BI) analysis (using MrBayes version 3.1.2), the best fitting model was set for each partition. Parameters of the model were treated as unknown variables with uniform prior probabilities and were estimated during the analysis; they were allowed to vary across partitions. Each Bayesian analysis included two MCMC runs, each composed of four chains, three heated and one cold. Each Markov chain was started from a random tree and run for up to $20 \times 10^6$ generations, sampling the chains every 1000th cycle. At the end of each run we considered the sampling of the posterior distribution to be adequate if the average standard deviation of split frequencies was $< 0.05$. The log-likelihood scores of sample points were plotted against generation time to determine when the chain became stationary. After discarding the ‘burn-in’ samples (25% of the trees) MCMC runs were summarized and further investigated for convergence of all parameters, using *sump* and *sumt* commands in MrBayes version 3.1.2 and the computer program Tracer version 1.4 (Rambaut & Drummond, 2007). Data remaining after discarding burn-in samples were used to generate a majority-rule consensus tree where the percentage of samples recovering any particular branch of the consensus tree represented the clade’s posterior probability (Huelsenbeck & Ronquist, 2001). Posterior probabilities (PP) of 0.95 or higher were considered as significant support. The mean, variance and 95% credibility intervals were calculated from the set of substitution parameters. RAxML and MrBayes analyses were conducted on a 150-core Linux Cluster at CBGP (Centre for Biology and Management of Populations, Montferrier-sur-Lez, France) and a 16-core Linux Cluster at INRA Bordeaux, France.

Finally, we extended the phylogenetic reconstruction by combining our sequences with sequences available in gene banks, namely ITS and CRC sequences (downloaded from NCBI GenBank, 6.10.2012). GenBank data included few to many accessions from one to several individuals per species (particularly ITS). Our data matrix was constructed using strict consensus sequences. Strict consensus sequences were computed for each species in order to concatenate the new data with data stored in GenBank.

Furthermore to detect deep phylogenetic splits that may otherwise be obscured by terminal noise, strict consensus sequences were computed for the following species clusters:

- **Group Quercus**: Four consensus sequences were computed corresponding to the major disjunct distribution areas Eastern Asia, Western Eurasia, Eastern North America (i.e. Eastern USA) and Western North America + Southern North America (i.e. Western USA + Mexico).
- **Group Lobatae**: Two consensus sequences were computed for Eastern North America and Western North America + Southern North America (i.e. Western USA + Mexico).
- **Group Ilex**: Two consensus sequences were computed for Eastern Asia (*Q. franchetii*, *Q. dolicholepis*) and Euro-Mediterranean basin (*Q. cocifera*, *Q. ilex*).
- **Group Cerris**: Two consensus sequences were computed for Eastern Asia and Western Eurasia.
- **Group Protobalanus**: One consensus sequence was computed for this Western North American endemic group (California and Oregon).
- **Group Cyclobalanopsis**: One consensus sequence was computed for all the (East Asian) species investigated.

This approach was preferred over random selection of a placeholder sequence to avoid topological bias inflicted by unrepresentative sequences or missing data. Incompatible signal in the concatenated matrix was visualized using consensus networks based on the ML bootstrap replicate trees (‘bipartition networks’; Grimm et al., 2006). Bipartition networks, in which the length of each edge is proportional to the frequency of the respective bipartition in the bootstrap tree sample, were computed with the consensus network module implemented in SplitsTree using the option ‘COUNT’ (Huson & Bryant, 2006). Additionally, we applied a gene jackknifing approach to test for the robustness of the result to the gene sampling. Bootstrapping analyses were run based on the complete data, single-gene data and 7-gene datasets excluding one partition in the original matrix. The consensus sequences dataset was finally used to optimize the position of the two putative outgroups within the optimal ingroup topology using the
Molecular dating

Divergence ages were calculated in BEAST v.1.8 (Drummond et al., 2012) on the strict consensus sequence matrix. Substitution models were initially defined for each partition as selected by Modeltest (see above). Ambiguities were treated as polymorphisms by changing the defaults setting in the .xml file generated with Beast. (Drummond et al., 2012) (‘useAmbiguities’ = ‘true’) to compensate for the use of consensus sequences. We then selected a Yule tree prior with an uncorrelated lognormal molecular clock (UCLD).

To infer absolute ages for the diversification of the main species cluster, we selected five different ingroup fossils, fossils unambiguously assignable to one of the intrageneric lineages, for age calibration (Table 4). Calibrations used a single fossil as age constraint, a lognormal age prior with the uppermost limit of the time interval as a minimum hard bound (offset) and large standard deviations (1.5 for calibrations 1, 2 and 5; 1.2 for 3a, 1.3 for 3b, 4). For most of the fossil constraints relatively exact stratigraphic ages were available, except for constraint no. 3 (first Cerris-type pollen). To compensate for the age uncertainty in this case, two calibrations were done, using the minimum and maximum possible stratigraphic age of the fossil flora. To assess the reliability of our root, we ran analyses where the ingroup was constrained to be monophyletic and another batch without constraining the ingroup. Two runs of 150 million generations with sampling every 10,000 generations were performed for each calibration/rooting scheme. The two separate runs were then combined using LogCombiner v.1.8. We ensured convergence of all parameters using Tracer 1.5 (Drummond et al., 2012). We used TreeAnnotator v.1.8 to generate a maximum clade credibility (MCC) tree and FigTree v.1.3.1 to visualize the results.

Results

PCR and sequencing results

Amplifications were successful for most but not all six nuclear genes selected (Table 5 and Table S2). Eleven species were represented by only two gene sequences, 14 by all six gene sequences. The whole dataset resulted in a supermatrix of 453 sequences out of the 672 potential sequences (=112 samples×6 genes), thus representing
67% of the theoretically possible data. Overall, the six selected gene sequences exhibited less nucleotide diversity and parsimonious informative characters than previous markers used for phylogenetic reconstruction (Table 5). The number of distinct alignment patterns ranged from 57 to 161 depending on the selected genes and outgroup (Table 5). Finally, the proportion of ambiguous nucleotide sites, which comprise intra-species polymorphism and possible sequencing errors, amounted to 0.0024, which is negligible in comparison to the between-species variation (Table 5).

**Single gene analysis**

The results of the single-gene ML analyses illustrate the differences in the signal provided by each gene region and the inclusion of either *Castanea* or *Notholithocarpus* (Figs S1–S4, see supplemental material online). The phylogenies were rooted on *Notholithocarpus* and *Castanea* except for CL8745 which was unrooted, because this gene could not be amplified in both outgroup genera. The phylogenetic analysis of CL1155 retrieved four ingroup clades with high BP supports (BP generally > 90), while members of Groups Protobalanus and Quercus clustered together in another strongly supported clade (Figs S1 and S3). Phylogenetic analyses of CL8561 recovered three of the six intrageneric groups as clades (Groups Lobatae, Quercus and Cyclobalanopsis), with BP of 98, 78 and 93, respectively (Fig. S4) when *Notholithocarpus* was used as outgroup, but only two (Groups Lobatae and Quercus) when rooted on *Castanea* (Fig. S2). The third marker, CL5191, for which amplification was less successful, was also less informative (Figs S1 and S3). It divided the species into two principal groups, one clade mixing members of Group Cerris and Cyclobalanopsis and an unsorted grade with the remainder. In addition there were some taxa that were scattered in the phylogeny (one red oak was found in the Cerris-Cyclobalanopsis clade, one member of Group Cyclobalanopsis was found as a sister taxon of the clade comprising Groups Quercus and Lobatae). The phylogenetic positions of members of Group Ilex were unresolved. Analyses of CL8461 retrieved each of the following species groups: Protobalanus, Lobatae and Cerris as monophyletic, except that Group Cerris included one species belonging to the Group Ilex (Figs S1 and S3). Phylogenetic analyses based on CL8745 only supported Group Lobatae as a clade with a BP of 96 (Figs S2 and S4). Phylogenetic analyses obtained with CL9715 retrieved Group Lobatae and Group Cerris clades (Figs S2 and S4). Overall, the single gene analyses retrieved clades comprising members of one or several infrageneric groups depending on the gene, but the relationships between these clades remained unresolved. These results prevented us from using any super tree approach (Scornavacca _et al._, 2008).

---

**Table 5.** Nuclear gene regions features for the six genes studied and CRC and ITS regions.

| Gene    | Length (bp) | All samples | Quercus only | All samples | Quercus only | All samples | Quercus only | All samples | Quercus only | All samples | Quercus only |
|---------|-------------|-------------|--------------|-------------|--------------|-------------|--------------|-------------|--------------|-------------|--------------|
| CL1155  | 550         | 83          | 79           | 0.2         | 0.19         | 0.15        | 0.14         | 124         | 126          | 57          | 53           |
| CL8461  | 210         | 65          | 61           | 0.15        | 0.13         | 0.09        | 0.09         | 53          | 57           | 155         | 111          |
| CL8561  | 653         | 79          | 76           | 0.21        | 0.17         | 0.16        | 0.14         | 161         | 114          | 133         | 113          |
| CL5191  | 529         | 48          | 46           | 0.17        | 0.14         | 0.12        | 0.12         | 93          | 96           | 98          | 94           |
| CL9715  | 389         | 93          | 89           | 0.19        | 0.17         | 0.13        | 0.12         | 98          | 94           | 95          | 95           |
| CL8745  | 565         | 86          | 86           | 0.16        | 0.16         | 0.12        | 0.12         | 94          | 94           | 95          | 95           |
| Total   | 2896        | 112         | 110          | 0.19        | 0.16         | 0.14        | 0.11         | 124         | 126          | 57          | 53           |

| Gene    | Length (bp) | All samples | Quercus only | All samples | Quercus only | All samples | Quercus only | All samples | Quercus only | All samples | Quercus only |
|---------|-------------|-------------|--------------|-------------|--------------|-------------|--------------|-------------|--------------|-------------|--------------|
| CRC     | 2957        | 112         | 110          | 0.32        | 0.32         | 0.34        | 0.34         | 687         | 687          | 687         | 687          |
| ITS     |             |             |              |             |              |             |              |             |              |             |              |
**Combined analyses**

We first inferred an ingroup tree comprising all species and using the concatenated sequences of all six genes (Fig. 1). This tree clearly separates the six major infragenic groups. Topologies of ML trees including either *Castanea* or *Notholithocarpus* were highly concordant with those obtained by BI as shown by the congruence between bootstrap values and posterior probabilities (Figs 2 and 3) and were much better resolved than the trees obtained with each individual gene. However, there was a slight topological shift depending on the outgroup species used.

Using *Castanea* as outgroup (Fig. 2) the resulting tree resolved two strongly supported clades within *Quercus* corresponding to the two subgenera recognized by Camus (1936–1954) (*Cyclobalanopsis* and *Euquercus*). The nine species of Group *Cyclobalanopsis* (= subgenus *Cyclobalanopsis* according to Camus) formed a sister clade to the remaining taxa. The clade corresponding to subgenus *Euquercus* (according to Camus) was further subdivided into two moderately supported clades: one clade comprised the ‘New World Oaks’ (Manos et al., 1999) of Groups *Quercus*, *Lobatae* and *Protobalanus* (of which Group *Quercus* includes also Old World species); the second clade was made up of species belonging to Group *Cerris* and *Ilex*. The three infragenic groups within the ‘New World Oaks’ were retrieved by ML and BI analyses as clades with strong support values. The *Cerris-Ilex* clade comprised one subclade of Euromediterranean species of Group *Ilex* (*Q. ilex*, *Q. cococifera*) and a second subclade with all species of Group *Cerris* and the two analysed East Asian species of Group *Ilex* (*Q. dolicholepis* A. Camus and *Q. franchetii* Skan), however, with poor support.

When *Notholithocarpus* was included as outgroup and used to root the tree (Fig. 3), the positions of the ingroup root and of the Group *Cyclobalanopsis* changed, resulting in two poorly supported major clades: one corresponding to the ‘New World Oaks’ as above, the other to the ‘Old World Oaks’ (Manos et al., 1999) comprising Groups: *Cyclobalanopsis*, *Cerris* and *Ilex*. The main difference between both analyses is the loss of the deeper split between the two subgenera, and the shift of Group

---

**Fig. 1.** Unrooted tree of species belonging to *Quercus*, *Castanea* and *Notholithocarpus* based on the six concatenated gene sequences (2896 pb). Numbers refer to maximum likelihood bootstrap support values (based on 550 bootstrap replicates) and Bayesian-inferred posterior probabilities (based on 10 000 sampled topologies). Arrows indicate alternative ingroup roots (Figs 2 and 3).
Fig. 2. Maximum likelihood phylogram of the genus *Quercus* based on concatenated sequences of six gene regions and rooted on the genus *Castanea* (2896 bp). Numbers close to the nodes refer to maximum likelihood bootstrap support values and posterior probabilities (based on 1000 bootstrap samples and 10 000 sampled topologies). Circles at the branches indicate the tested alternative ingroup roots. Filled grey circles correspond to placements for which the likelihood was not significantly different from the initial placement. Empty circles exhibit significantly lower likelihoods.
Fig. 3. Maximum likelihood phylogram of the genus *Quercus* based on concatenated sequences of six gene regions and rooted on the genus *Notholithocarpus* (2896 bp). Numbers close to the nodes refer to maximum likelihood bootstrap support values and posterior probabilities (1000 bootstrap samples). Circles at nodes indicate the tested alternative ingroup roots. Filled grey circles correspond to placements for which the likelihood was not significantly different from the initial placement. Empty circles correspond to significantly lower likelihoods.
Cyclobalanopsis from a ‘sister-to-the-remaining-taxa’ position to the ‘Old World Oaks’.

We tested alternative ingroup roots as shown in Figure 2 (Castanea is the outgroup) and in Figure 3 (Notholithocarpus is the outgroup) using the non-parametric Shimodaira–Hasegawa (SH) test implemented in RAxML. When Castanea was included in the matrix, the likelihood of all other topologies was significantly lower than that of the ML-preferred topology (Fig. 2), except when Castanea was connected to the basal branches within the Cerris-Ilex-Cyclobalanopsis subtree. When Notholithocarpus was selected as outgroup, the likelihood of alternative root positions was not significantly different from the ML-preferred position (Fig. 3), but was lower when Notholithocarpus was connected to the branch leading to Groups Cerris and Cyclobalanopsis. In conclusion, for both outgroups, relatively more likely topologies are obtained when the World and Old World clade.

Stability of found ingroup and outgroup relationships against signals from CRABS CLAW (CRC) gene and ITS region of the 35S rDNA cistron

To investigate the effect of gene sampling on ingroup relationships and the placement of the two outgroup taxa, we complemented our data with ITS and CRC sequences harvested from gene bank data (see Materials and methods). The 8-gene bipartition (bootstrap) network obtained with the strict group consensus sequences (Fig. 4) corroborated the backbone topology of the 6-gene ML or Bayesian inferences (Fig. 1). The two main clades, the New World and the Old World clade were strongly supported. Furthermore, there was moderate support for the ingroup-outgroup split. The subsequent gene jackknifing allowed further identification of phylogenetic footprints from each marker (Table 6). The sister relationship (and mutual monophyly) between Group Protobalanus and Group Quercus is mostly due to signal from CL1155. When this gene is removed from the analysis, the alternative Lobatae–Quercus sister relationship is favoured, although the support to a Protobalanus-Quercus relationship does not collapse to zero. Furthermore, it is the CRC sequences in combination with other partitions that generate the clear split between the ingroup Quercus and the outgroup Notholithocarpus and Castanea. The position of the two outgroup taxa shown in Fig. 4 could be confirmed using the Evolutionary Placement Algorithm (EPA) implemented in RAxML. Although the EPA analyses based on individual genes indicated different possible roots, partly with high probability (CL1155, CL8561, CL9715), the concatenated data provided unambiguous support for New World-Old World split (Table 7) and the monophyly of the genus (Fig. 4).

Molecular dating

Divergence ages were calculated on the 8-gene consensus sequence matrix based on five different ingroup fossils (Table 4). Inferred divergence ages vary strongly depending on the used constraint (Fig. 5; Table S3, see supplemental material online). The overall best fit, an average difference of 10 million years (Ma) for inferred medians to expected ages (7 Ma regarding inferred min and max values), was obtained using constraint no. 1 (Table 4), the oldest unambiguous fossil record of Group Cyclobalanop- sis. This ‘best fit’ suggests a root age of 54 (68–48) Ma for genus Quercus (Fig. 5). Divergence between the New World oaks (Groups Protobalanus, Quercus, Lobatae) and the Old World oaks (Groups Cyclobalanopsis, Ilex, Cerris) occurred shortly after, during the early Eocene. Radiations within each of these two clades lasted until the mid Miocene. Cyclobalanopsis, 49 Ma (48–52 Ma, constrained), and Lobatae, 34 (53–9) Ma, diverged already in the Eocene. Inferred divergence ages between Protobalanus and Quercus in the New World oaks, 15 (8–38) Ma, and between Cerris and (East Asian) Ilex, 10 (1–30) Ma in the Old World oaks fall within the mid to late Miocene. Using a single minimum age constraint on the relatively complex sequence data resulted in large 95% highest posterior density (HPD) intervals. Figure 5 shows that, for most nodes, the different inferences converge to an interval encompassing the inferred median age using constraint no. 1. Exceptions are the deepest ingroup nodes referring to the MRCA of all oaks, the New World oaks, and the Old World oaks, respectively. A general trend is that inferred ages tend to be younger than indicated by the fossil record, irrespective of the constraint used.

Discussion

Confirmation of the infrageneric subdivision of genus Quercus

In this study we explored phylogenetic reconstructions using a larger sample of nuclear genes than in previous investigations and an extensive taxon sampling. Up to now phylogenetic analyses were mainly based on ITS sequences (Manos et al., 1999, 2001; Manos & Stanford, 2001; Denk & Grimm, 2010) with one study also using sequences from the CRABS CLAW (CRC) gene, a single copy nuclear gene that regulates carpel development in angiosperms (Oh & Manos, 2008). Despite a lower
nucleotide diversity and less parsimony informative characters (Table 5) than ITS, our data largely confirmed results from previous studies (Manos et al., 1999, 2001; Manos & Stanford, 2001; Denk & Grimm, 2010) that recognized six major infrageneric lineages (Protobalanus, Lobatae, Quercus, Cyclobalanopsis, Cerris and Ilex) but also provided new insights into the phylogenetic relationship between these lineages. The major split between the ‘New World’ and ‘Old World’ clade (Oh & Manos, 2008; Denk & Grimm, 2010) is further supported, and placed in the early Eocene (Fig. 5). Within the ‘New World’ clade, the red oaks, Group Lobatae, are resolved as sister to a

![Bipartition network based on ML bootstrapping of consensus sequences concatenated on the eight genes. Branch (edge) lengths in this consensus network based on the bootstrap replicate tree sample are proportional to the frequency of the according split in the bootstrap replicate tree sample. Bootstrap support values are annotated for selected edges/edge bundles; colouration is used to highlight edges referring to competing phylogenetic splits. Abbreviations in taxon labels: EAS, Eastern Asia; WEA, Western Eurasia, EUM, Euro-Mediterranean basin; ENA, Eastern North America; WNA, Western North America; W/SNA, Western and Southern North America.](image-url)
Table 6. Bootstrap support to different phylogenetic partitions following jackknifing.

| Gene selection | 8 gene tree | Excluding CL1155 | Excluding CL5191 | Excluding CL8461 | Excluding CL8561 | Excluding CL8745 | Excluding CL9715 | Excluding CRC | Excluding ITS |
|----------------|-------------|------------------|------------------|------------------|------------------|------------------|------------------|--------------|--------------|
| Number of bootstrap replicates | 1000 | 600 | 150 | 1000 | 1000 | 100 | 1000 | 350 | 350 |
| Split/Clade | | | | | | | | | |
| Outgroup | ingroup | | | | | | | | |
| ‘New World’ | ‘Old World’ | | | | | | | | |
| Castanea + Cyclobalanopsis | all other | | | | | | | | |
| Subgenus Cyclobalanopsis emended | | | | | | | | | |
| Cyclobalanopsis sister to Euro-Mediterranean Ilex | | | | | | | | | |
| Cerris-Ilex clade | | | | | | | | | |
| Ilex clade | | | | | | | | | |
| Eastern Asian Ilex sister to Cerris clade | | | | | | | | | |
| Cerris clade | | | | | | | | | |
| Subgenus Quercus | | | | | | | | | |
| Protobalanus sister to Quercus | | | | | | | | | |
| Lobatae sister to Quercus | | | | | | | | | |
| Lobatae clade | | | | | | | | | |
| Quercus clade | | | | | | | | | |
| North American – East Asian subclade | | | | | | | | | |
| West USA/Mexican sister to East Asian | | | | | | | | | |
| West USA/Mexican sister to East USA | | | | | | | | | |
| Eastern USA sister to remainder | | | | | | | | | |
| Eurasian subclade | | | | | | | | | |

*a Subgenus Cyclobalanopsis including the ‘Old World’ groups Cerris, Cyclobalanopsis and Ilex. Subgenus Quercus including the ‘New World’ groups Lobatae, Protobalanus and Quercus.*
clade comprising the Golden-cup oaks and white oaks (Groups Protobalanus and Quercus) in contrast to previous studies. Within the ‘Old World’ clade, members of Group Cyclobalanopsis are most distinct, and Group Ilex is suggested to be paraphyletic to Group Cerris. Interestingly, the six infragenic groups are generally reproductively isolated. Successful natural or controlled hybridization between species from different groups have only been shown between Groups Ilex and Cerris (Boavida et al., 2001; Burgarella et al., 2009; Lumaret & Jabbour-Zahab, 2009). Very anecdotal hybridization events between species of Groups Quercus and Cerris (Schnitzler et al., 2004) have also been reported. However, interspecific hybridization is considered to occur very frequently within members of Group Quercus (Rush- ton, 1993) or Group Lobatae (Moran et al., 2012), and has been documented for Group Cerris (an arboretum hybrid between Q. suber L. and Q. trojana Webb, Denk & Grimm, 2010). The possibilities of natural hybridization have been less investigated in the other groups but intragenic and intergroup artificial hybridization experiments support the general conclusion of frequent cross-breeding within groups and very rare hybridization success between groups (Cottam et al., 1982). Similar observations were made in the Nothofagaceae by Acosta and Premoli (2010), who inferred past hybridization events by cpDNA sharing among species.

### Tree topologies and outgroup choice

Results from traditional tree inference suggest two different ingroup roots and two alternative placements of the Cyclobalanopsis clade within the genus Quercus. With Castanea as outgroup, our data suggested two clades within Quercus corresponding to the subgeneric classification of Camus (1936–1954), also partially accepted by Nixon (1993) and by Menitsky to some degree (Menitsky, 1984, 2005): subgenus Quercus (called EuxQuercus by Camus) and subgenus Cyclobalanopsis (called Cyclobalanoides by Menitsky, 2005). This rooted topology differed markedly from the one observed when Notholithocarpus was chosen as outgroup: subgenus Cyclobalanopsis clustered with Groups Cerris and Ilex to form an Old World clade (Figs 2 and 3). Instability of the genus root was also observed in previous studies, when different outgroups were used. When outgroups were chosen that are only distantly related with Quercus, for example members of the Trigonobalanus genus complex (Manos et al., 1999) or Fagus (Manos et al., 2001), Group Cyclobalanopsis was usually placed as sister clade to the remaining oaks. This is similar to the topology we obtained when Castanea was used as outgroup and based on our 6-gene data. However, when more closely related genera such as Lithocarpus Blume (Manos & Stanford, 2001) or Notholithocarpus (Oh & Manos, 2008) were used as outgroups, Group Cyclobalanopsis was reconstructed as a member of the ‘Old World Oaks’. Phylogenetic relationships between subgenus Quercus and subgenus Cyclobalanopsis were therefore ambiguously resolved and were very sensitive to the outgroup choice. In our dataset, Castanea, being more distant to Quercus than Notholithocarpus, may attract ingroup taxa that are well differentiated from the rest, such as species belonging to Group Cyclobalanopsis. Thus the split between two subgeneric clades (Fig. 2) may actually be an artefact due to long branch attraction (Bergsten, 2005). However, we cannot entirely rule out that the true phylogeny actually has its root at the longest branch. The results of the SH-test (Table 6), the phylogenetic placement algorithm (Table 7) and the general support patterns in the single genes demonstrate that the signal from Castanea is more explicit regarding its placement within the Quercus tree

---

**Table 7. Probabilities of the position of the two putative outgroups using the evolutionary placement algorithm.** The placement with the highest probability (weight) is shown; in case several alternatives showed essentially the same probability (within a range of 0.05), the number of alternatives is given and the range of the probabilities. EPA could not be applied for CL8745 because no sequences were available for Castanea and Notholithocarpus (Table S2, see supplemental material online).

| Castanea inserted at split between | Probability | Notholithocarpus inserted at split between | Probability |
|-----------------------------------|-------------|------------------------------------------|-------------|
| 8-genes                           |             | New World | Old World                              | 1.00        |
| CL1155                            |             | Cyclobalanopsis | rest                              | 1.00        |
| CL5191a                           |             | Euro-Mediterranean Ilex | rest | 0.09 | New World | Old World                              | 0.82 |
| CL8461                            |             | Various (3) | ~0.28 | Various (3) | New World | Old World                              | 0.07–0.12 |
| CL8561                            |             | East Asian Ilex | rest | 1.00 | Protobalanus | rest                              | 1.00 |
| CL9715a                           |             | Cyclobalanopsis | rest | 0.89 | Various (10) | rest                              | 0.47 |
| CRCb                             |             | Protobalanus | rest | 0.53 | Lobatae | rest                              | 0.47 |

* a No data for East Asian Ilex.
* b No data for East Asian Cerris and Euro-Mediterranean Ilex.
than that of *Notholithocarpus*. Whereas *Notholithocarpus* can be shifted within the oak phylogenetic backbone, *Castanea* is strongly attracted by Group Cyclobalanopsis, but also East Asian members of Group Ilex. This attraction is strongest in the CL1155 gene; whereas the signal in other partitions is inconclusive regarding the best placement of *Castanea* in the all-oak tree. The same gene places *Notholithocarpus* afar from *Castanea*, which directly implies that the genus *Quercus* is not monophyletic (see e.g. Manos *et al.*, 2008). This incongruence is resolved when the data are concatenated and filtered against terminal noise (Fig. 4, Tables 6 and 7).

**Phylogenetic relationships and diversification within the Old World clade**

The genetic distinctness of Group Cyclobalanopsis seen in our data is well established in the literature. Within the Cerris-Ilex clade, our analyses yielded two clades, or one grade and one clade, which corresponded to the two botanical entities: Group Ilex and Group Cerris (Figs 2 and 3). Resolution is slightly better when *Notholithocarpus* is chosen as outgroup. One of the (East) Asian Group Ilex species (*Q. franchetii*) is misplaced in both analyses, however, *Q. dolicholepis* (the second Asian Ilex species) is only placed as direct sister to Group Cerris when *Castanea* is the outgroup and with low support. Reconstructions based on cpDNA sequences (Manos *et al.*, 1999) also separated Cerris from Ilex, with one Asian Ilex species (*Q. phillyreoides* A. Gray) placed as well within the Cerris clade. Evergreen and sclerophyllous Euro-Mediterranean oaks (*Q. ilex*, *Q. coccifera* and *Q. coccifera* subsp. *calliprinos*) assigned to the subgenus *Sclerophylloides* by Schwarz (Schwarz, 1936–39, 1993) form a clade in our analyses (Fig. 2) similarly to previous work based on ITS sequence data restricted to Eurasian oak species (Samuel *et al.*, 2008).
Phylogenetic relationships and diversification within the New World clade

Like earlier investigations our analyses yielded a moderately supported New World clade. However our study is the first to show the subdivision of the New World clade into three well-supported subclades corresponding to Groups Protobalanus, Lobatae and Quercus. Phylogenetic relationships between these three groups varied depending on the gene sampling. ITS sequences suggested that Group Protobalanus is a sister clade to Groups Quercus and Lobatae, but with poor support, evidence of lineage mixing, and Group Lobatae usually nested in a Group Quercus grade (Manos & Stanford, 2001; Manos et al., 2001; Oh & Manos, 2008; Denk & Grimm, 2010). The close relationship of Groups: Lobatae and Quercus is reconstructed as c. 20 Ma (5–25 Ma; Fig. 5). This appears to be too young, since both groups can be traced back in the fossil records at least to the Eocene-Oligocene boundary (Table 4). However, even modern members of Groups Protobalanus and Quercus are genetically not perfectly isolated: Manos et al. (2001) documented an individual of Group Protobalanus heterozygous for Quercus- and Protobalanus-type ITS. Our 6-gene data highlights the uniqueness of Q. sadleriana (Figs 1–3; Figs S1–S4), a highly endemic, western North American relict (Flora of North America Committee 1997) of Group Quercus (Manos et al., 2001; Denk & Grimm, 2010). Based on the new data, Q. sadleriana may represent the evolutionary link between both groups or a survivor of an earlier radiation phase (Florissant fossil flora, eastern Rocky Mts, Bouchal, 2013; J. Bouchal, T. Denk, pers. comm.). Because of its highly incompatible signal, we refrained from including Q. sadleriana in the dating. Thus, we may have failed to infer the age of the first radiation in the Protobalanus-Quercus clade, and the inferred age (Fig. 5) mirrors the final (genetic) isolation of the modern members of both groups. Or we may also conclude that morphological differentiation pre-dates genetic isolation in this clade.

Ancient rapid divergence and late radiation obscures Quercus molecular phylogenies

As earlier phylogenetic reconstructions (Manos & Stanford, 2001; Manos et al., 2001; Oh & Manos, 2008; Denk & Grimm, 2009) our reconstructions are partly characterized by low support for deeper branches. To some degree, the overall lower support is due to equally supported, competing (alternative) relationships among the different taxonomic groups; in other cases it is due to a generally weak phylogenetic signal. None of the explored gene regions showed a phylogenetic signal that was in strong conflict with the consensus topology. Such patterns are generally considered as signatures of ancient rapid radiation followed by long time spans during which earlier splits were blurred (Rokas et al., 2005). Alternatively, similar patterns can also be obtained when the genome has been poorly explored for phylogenetic signals, or when target genes exhibit inappropriate substitution rates to detect ancient events (Whitfield & Kjer, 2008). Before exploring the rapid radiation hypothesis, we may therefore question whether our selected genes were appropriate to reveal old speciation events. Selecting genes for detecting
targeted phylogenetic signals is a challenging task. On one hand it is difficult to explore the phylogenetic signal before conducting the analysis on the whole dataset, on the other hand technical criteria such as PCR amplification success and sequence quality may prevail over phylogenetic criteria. Our selection of six protein coding genes resulted from a compromise between technical criteria and optimizing phylogenetic information. The independent analyses of these genes resulted in phylogenetic trees with generally short basal branches (Figs S1 and S2). Their individual signals are comparable to those in available datasets (CRC, ITS). Short backbone branches were also found in previous reconstructions based on ITS sequences. Despite the generally weak individual signals, the concatenated data converge towards a consensus topology, which in addition shows a distinct ingroup (genus Quercus) outgroup (Notholithocarpus and Castanea) split (Fig. 4). Congruency of patterns across genes is likely to result from evolutionary processes that impact the whole genome, and therefore would support the rapid radiation hypothesis. According to the best-fit chronogram, the first divergence within Quercus took place directly after the putative establishment of the genus close to the Paleocene–Eocene boundary; the first radiations within the ‘New World Clade’ and ‘Old World Clade’ may have been finished already by the mid-Eocene (Fig. 5, Table 4). The Oligocene (34–23 Ma) has been suggested as the time when proliferation of oaks (and other Fagaceae (Denk & Grimm, 2009; Denk et al., 2012) and Nothofagaceae (Premoli et al., 2010)) occurred with the modern diversity already established by the Miocene (Axelrod, 1983) (Table 4; Fig. 5). Based on sequence variation in non-coding plastid DNA, Manos & Stanford (2001) concluded that Old World and New World members of Group Quercus split at around 17 Ma (Burdigalian, late early Miocene); which appears too old regarding the dating based on the 8-gene nuclear gene matrix (Fig. 5). This discrepancy can be easily explained by phases of unhindered gene flow via the North-Atlantic Land Bridge in Group Quercus until at least 8 Ma (Denk et al., 2010) and Beringia until the latest Pliocene or Pleistocene interglacials. In contrast to plastids, nuclear genomes are biparentally inherited, hence, genetic signatures are propagated via pollen (partly long-distance dispersed in the wind-pollinated oaks) and seeds (short-distance dispersed by jay birds and squirrels). There are no apparent geographic barriers to gene flow in modern Lobatae (Moran et al., 2012), and also in the case of Group Cerris the range may have been fragmented quite recently (analogous to the situation in Fagus, Denk & Grimm, 2009; T. Denk, pers. comm. 2014). The late and rapid radiation of modern lineages of Fagaceae in general, and possibly of some infrageneric lineages of Quercus (Groups Cerris, Lobatae, Quercus), may be linked to the long persistence of extinct and partly widespread fagaceous lineages (Smiley & Huggins, Pseudofagus; 1981; Crepet & Nixon, various organ taxa; 1989; Denk et al., 2012, Eotrigonobalanus, Trigonobalanopsis).

Acknowledgements

Thomas Denk is acknowledged for insights into the fossil record of oaks and most valuable discussions. We are grateful to all the colleagues that contributed to the collections of samples throughout the worldwide distribution of oaks: Guylaine Blanc, Catherine Bodênes, Frédéric Brison, Henri Caron, Béatrice Chassé, Bruno Fady, Silvia Fineschi, Reiner Finkeldey, Antonio Gonzalez, Riina Jалonen, Zeky Kaya, Jarkko Koskela, Thierry Lamant, Hervé Le Bouler, Richard Michalet, Huang Min-Ren, Carlos Navarro, Greuk Pakkad, Dominique Piou, Hui-Min Liu, Ken Oyama, Hui-Xin Pan, Juan Manuel Penaloz, Gabriel Schiller, Nasrin Seyedi, Jean Timbal, Jozef Turok, Susana Valencia, Yann Vitasse, Li-An Xu, Wu-Yuan Yin, Roghayed Zolfaghari and Hui-Yu Zou. We are also grateful to Pauline Garnier-Gére for providing the gene sequences, which were obtained within a resequencing project conducted within the EU-supported EVOLTREE network of excellence (FP7, project #016322).

Funding

This research was supported by the ANR (Agence Nationale de la Recherche) project PHYLOSPACE (‘Integrating cophylogenies, area shifts and palaeomaps for inferring history of associations under climate change’). GWG was financed by the Swedish Research Council (VR-grant 2008–3726).

Supplemental data

Supplemental data for this article can be accessed here.

References

Acosta, M.C. & Premoli, A.C. 2010. Evidence of chloroplast capture in South American Nothofagus (subgenus Nothofagus, Nothofagaceae). Molecular Phylogenetics and Evolution 54, 235–242.

Axelrod, D.I. 1983. Biogeography of oaks in the Arcto-Tertiary province. Annals of the Missouri Botanical Garden 70, 629–657.

Bellarosa, R., Simeone, M.C., Papini, A. & Schirone, B. 2005. Utility of ITS sequence data for phylogenetic reconstruction of Italian Quercus spp. Molecular Phylogenetics and Evolution 34, 355–370.

Berger, S.A., Krompass, D. & Stamatakis, A. 2011. Accuracy, and web server for evolutionary placement of short sequence reads under maximum likelihood. Systematic Biology 60, 291–302.

Bergsten, J. 2005. A review of long-branch attraction. Cladistics 21, 163–193.
DENK, T. & GRIMM, G.W. 2010. The oaks of western Eurasia: a bridge in the latest Miocene. Masters thesis, University of Vienna, Vienna, Austria.

BOUCHAL, J. 2013. The microflora of the latest Eocene (Priabonian) Florissant Formation, a combined method approach. Masters thesis, University of Vienna, Vienna, Austria.

BURGARELLA, C., LORENZO, Z., JABBOUR-ZAHAB, R., LUMARET, R., GUICHoux, E., PETIT, R.I., SOTO, A. & GIULI, L. 2009. Detection of hybrids in nature: application to oaks (Quercus suber and Q. ilex). Heredity 102, 442–452.

CAMUS, A. 1936–1938. Les Cônes. Monographie du genre Quercus. Tome I. Genre Quercus, sous-genre Cyclobalanopsis, sous-genre Eucquerus (sections Cerris et Mesobalant). Paul Lechevalier Edition, Paris, France.

CAMUS, A. 1936–1934. Les Cônes. Monographie du genre Quercus et Monographie du genre Lithocarpus. Paul Lechevalier Edition, Paris, France.

COHEN, K., FINNEY, S. & GIBBARD, P. 2012. International chronostratigraphic chart, IUGS.

COTTAM, W.P., TUCKER, J.M. & SANTAMOUR, F.S. 1982. Oak hybridization at the University of Utah, Salt Lake City, UT, USA.

CREPET, W.L. & NIXON, K.C. 1989. Earliest megafossil evidence of Fagaceae: phylogenetic and biogeographic implications. American Journal of Botany 76, 842–855.

DAGHILLIAN, C.P. & CREPET, W.L. 1983. Oak catkins, leaves and fruits from the Oligocene Catehoulia Formation and their evolutionary significance. American Journal of Botany 70, 639–649.

DENK, T. & GRIMM, G.W. 2009. Significance of pollen characteristics for infrageneric classification and phylogeny in Quercus (Fagaceae). International Journal of Plant Sciences 170, 926–940.

DENK, T. & GRIMM, G.W. 2010. The oaks of western Eurasia: traditional classifications and evidence from two nuclear markers. Taxon 59, 351–366.

DENK, T., GRIMSMSON, F. & ZETTER, R. 2010. Episodic migration of oaks to Iceland: evidence for a North Atlantic ‘land bridge’ in the latest Miocene. American Journal of Botany 97, 276–287.

DENK, T., GRIMSMSON, F. & ZETTER, R. 2012. Fagaceae from the early Oligocene of Central Europe: persisting new world and emerging old world biogeographic links. Review of Palaeobotany and Palynology 169, 7–20.

DENK, T. & TAELEVA, M.V. In press. Pollen morphology and ultrastructure of Quercus with focus on Group Ilex (=Quercus Subgenus Heterobalanus (Oerst.) Meninsky): implications for oak systematics and evolution. Grana (in press).

DRUMMOND, A.J., SUCHARD, M.A., XIE, D. & RAMBAUT, A. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Molecular Biology and Evolution 29, 1969–1973.

EDGAR, R.C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32, 1792–1797.

GOVIS, M., GUINDON, S. & GASCUEL, O. 2010. SeaView Version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Molecular Biology and Evolution 27, 221–224.

GOVAERTS, R. & FRODIN, D.G. 1998. World Checklist and Bibliography of Fagales (Betulaceae, Corylaceae, Fagaceae and Tiddefrandeaceae). Royal Botanic Gardens, Kew, UK.

GRIMM, G.W., RENNER, S.S., STANAMAKIS, A. & HEMLEBEN, V. 2006. A nuclear ribosomal DNA phylogeny of acer inferred with maximum likelihood, splits graphs, and motif analysis of 606 sequences. Evolutionary Bioinformatics 2, 7–22.

HALL, T.A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series 41, 95–98.

HUELSBENCKE, J.P. & ROJOUX, F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17, 754–755.

HUSON, D.H. & BRYANT, D. 2006. Application of phylogenetic networks in evolutionary studies. Molecular Biology and Evolution 23, 254–267.

JAHREN, A. 2007. The Arctic forest of the middle Eocene. Annual Review of Earth and Planetary Sciences 35, 509–540.

KMENTA, M. 2011. Die Mikroflora der untertertiären Fundstelle Altmitwiede. Master thesis, University of Vienna, Vienna, Austria.

KREMER, A. & PETIT, R. 1993. Gene diversity in natural populations of oak species. Annals of Forest Science 50, 186–202.

KREMER, A., ABBOTT, A.G., CARLSON, J.E., MANOS, P.S., PLOMION, C., SISCO, P., STATON, M.E., UENO, S. & VENDRAMIN, G.G. 2012. Genomics of Fagaceae. Tree Genetics and Genomes 8, 583–610.

KVACEK, Z. & WALThHER, H. 1989. Palaeobotanical studies in Fagaceae of the European Tertiary. Plant Systematics and Evolution 162, 213–229.

LUMARET, R. & JABBOUR-ZAHAB, R. 2009. Ancient and current gene flow between two distantly related Mediterranean oak species, Quercus suber and Q. ilex. Annals of Botany 104, 725–736.

MANCHESTER, S. 1994. Fruits and seeds of the Middle Eocene nut beds flora, Clarno Formation, Oregon. Palaeoentographica Americana 58, 1–205.

MANOS, P.S. & STANFORD, A.M. 2001. The historical biogeography of Fagaceae: tracking the tertiary history of temperate and subtropical forests of the northern hemisphere. International Journal of Plant Sciences 162, 77–93.

MANOS, P.S., DOYLE, J.J. & NIXON, K.C. 1999. Phylogeny, biogeography, and processes of molecular differentiation in Quercus subgenus Quercus (Fagaceae). Molecular Phylogenetics and Evolution 12, 333–349.

MANOS, P.S., ZHOU, Z.K. & CANNON, C.H. 2001. Systematics of Fagaceae: phylogenetic tests of reproductive trait evolution. International Journal of Plant Sciences 162, 1361–1379.

MANOS, P.S., CANNON, C.H. & OUI, S.-H. 2008. Phylogenetic relationships and taxonomic status of the paleoendemic Fagaceae of Western North America: recognition of a new genus, Notholithocarpus. Madroño 55, 181–190.

MCINTYRE, D. 1991. Pollen and spore flora of an Eocene forest, eastern Axel Heiberg Island. In: CHRISTIE, R. & MCILLAN, N., Eds., Tertiary Fossil Forests of the Geodetic Hills, Axel Heiberg Island, Arctic Archipelago. Geological Survey of Canada Bulletin 403, 83–97.

MCIVER, E. & BASINGER, J. 1999. Early Tertiary floral evolution in the Canadian high arctic. Annals of the Missouri Botanical Garden 86, 523–545.

MENTISKY, Y.L. 1984. Doby Asii. Nauka, Leningrad, Soviet Union.

MENTISKY, Y.L. 2005. Oaks of Asia. Science Publishers, Enfield, NH, USA.

MOR, C., TOUMI, L., JANNE, P., SADA, V., DI GIUSTO, F. & LUMARET, R. 2006. Endemic North African Notholithocarpus. Madroño 55, 181–190.

MORAN, E.V., WILSON, J. & CLARK, J.S. 2012. Genetic evidence for hybridization in red oaks (Quercus sect. Lobatae, Fagaceae). American Journal of Botany 99, 92–100.
Phylogenetics of Quercus

Nixon, K. 1993. Infrageneric classification of Quercus (Fagaceae) and typification of sectional names. *Annals of Forest Science* **50**, 25–34.

Ohi, S.H. & Manos, P.S. 2008. Molecular phylogenetics and cupule evolution in Fagaceae as inferred from nuclear CRABS CLAW sequences. *Taxon* **57**, 434–451.

Ørsted, A.S. 1867. *Bidrag til Egeskægtenes Systematik*. Bianco Luno, Kjøbenhavn, Denmark.

Premoli, A.C., Mathiasen, P., Cristina Acosta, M. & Ramos, V.A. 2012. Phylogeographically concordant chloroplast DNA divergence in sympatric Nothofagus s.s. How deep can it be? *New Phytologist* **193**, 261–275.

Rambaut, A. & Drummond, A. 2007. *Tracer v1.4*. http://beast.bio.ed.ac.uk/Tracer (accessed 13 June 2014).

Rokas, A., Kruger, D. & Carroll, S.B. 2005. Animal evolution and the molecular signature of radiations compressed in time. *Science* **310**, 1933–1938.

Rushiton, B. 1993. Natural hybridization within the genus *Quercus* L. *Annals of Forest Science* **50**, 73–90.

Samuel, R., Bacherbier, A., Jobst, J. & Ehrendorfer, F. 1998. ITS sequences from nuclear rDNA suggest unexpected phylogenetic relationships between Euro-Mediterranean, East Asian and North American taxa of *Quercus* (Fagaceae). *Plant Systematics and Evolution* **211**, 129–139.

Schnitzler, J.P., Steinbrecher, R., Zimmer, I., Steigner, D. & Fladung, M. 2004. Hybridization of European oaks (*Quercus ilex* x *Q-robur*) results in a mixed isoprenoid emitter type. *Plant Cell and Environment* **27**, 585–593.

Schwarz, O. 1936. Entwurf zu einem natürlichen System der Cupuliferen und derGattung Quercus L. *Notizblätter des Botanischen Garten zu Berlin-Dahlem* **13**, 1–22.

Schwarz, O. 1936–39. *Monographie der Eichen Europas und des Mittelmeergebietes*. Series: Repertorium specierum nov. regni vegetabilis, Sonderhefte, Selbstverlag Friedrich Fedde, Dahlem-Berlin, Germany.

Schwarz, O. 1993. *Quercus* L. In: Tutin, T.G., Heywood, V.H., Burges, N.A., Valentine, D.H., Walters, S.M. & Webb, D.A., Eds., *Flora Europaea*. Cambridge University Press, Cambridge, UK.

Scornavacca, C., Berry, V., Lefort, V., Dougey, E.J.P. & Ranwez, V. 2008. PhySIC_IST: cleaning source trees to infer more informative supertrees. *BioMed Central Bioinformatics* **9**, 413.

Simone, M.C., Prezza, R., Papini, A., Vessella, F. & Schibone, B. 2013. Application of plastid and nuclear markers to DNA barcoding of Euro–Mediterranean oaks (*Quercus*, Fagaceae): problems, prospects and phylogenetic implications. *Botanical Journal of the Linnean Society* **172**, 478–499.

Smiley, C.J. & Huggins, L.M. 1981. *Pseudofagus idahoensis*, n. gen. et sp. (Fagaceae) from the Miocene Clarkia Flora of Idaho. *American Journal of Botany* **58**, 741–761.

Solomon, A. 1983a. Pollen morphology and plant taxonomy of white oaks in eastern North America. *American Journal of Botany* **70**, 481–494.

Solomon, A. 1983b. Pollen morphology and plant taxonomy of red oaks in eastern North America. *American Journal of Botany* **70**, 495–507.

Stamatakis, A. 2006. *raxml-vi-hpc*: Maximum-Likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688–2690.

Stamatakis, A., Hoover, P. & Rougemont, J. 2008. A rapid bootstrap algorithm for the RAxML web servers. *Systematic Biology* **57**, 758–771.

Tamura, K., Dudley, J., Nei, M. & Kumar, S. 2007. *MEGA4*: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular Biology and Evolution* **24**, 1596–1599.

Ueno, S., Le Provost, G., Leger, V., Klopp, C., Norot, C., Fugiero, J.M., Salin, F., Salse, J., Abrouk, M., Murat, F., Brendel, O., Derory, J., Abadie, P., Leger, P., Cabane, C., Barre, A., De Daruvar, A., Couloux, A., Wincker, P., Reviron, M.P., Kremer, A. & Plomion, C. 2010. Bioinformatic analysis of ESTs collected by Sanger and pyrosequencing methods for a keystone forest tree species: oak. *BioMed Central Genomics* **11**, 650.

Valencia, S.A. 2004. Diversidad del genero Quercus (Fagaceae) en Mexico. *Boletín de la Sociedad Botánica de México* **75**, 33–54.

Velitzeiros, D., Bouchar, J.M. & Denk, T. 2014. Review of the Cenozoic floras and vegetation of Greece. *Review of Palaeobotany and Palynology* **204**, 86–117.

Whitefield, J.B. & Kier, K.M. 2008. Ancient rapid radiations of insects: challenges for phylogenetic analysis. *Annual Review of Entomology* **53**, 449–472.

Xu, L., 2004. *Diversité de l’ADN chloroplastique et relations phylogénétiques au sein des Fagacées et du genre Quercus*. PhD thesis, University of Henri Poincaré, Nancy, France.

Xu, L., Kremer, A., Bodenes, C. & Huang, M.R. 2005. Chloroplastic DNA diversity and phylogenetic relationships in botanical sections of the genus Quercus L. In: Rogers, R., Ducousso, A. & Kanazashi, A., Eds., *Integration of silviculture and genetics in creating and sustaining of oak forests*. Proceedings of the Joint Meeting of IUFRO Working Groups Genetics of Quercus and Improvement and Silviculture of Oakes. FFPRI (Forestry and Forest Products Research Institute), pp. 113–121.

Zhang, D.X. & Hewitt, G.M. 1996. Nuclear integrations: challenges for mitochondrial DNA markers. *Trends in Ecology and Evolution* **11**, 247–251.

Associate Editor: Nadia Bystriakova