individuals of 16 Asian species of Group Ilex were analysed. The final dataset also included all species of the western North American Group Protobalanus (five species, 10 individuals), 16 species of Group Quercus (20 individuals, from North America and Eurasia), five species of the East Asian Group Cyclobalanopsis (11 individuals), seven species of the American Group Lobatae (eight individuals), and six species of Group Cerris (seven individuals). The outgroup set was represented by one sample each of the monotypic genera *Notholithocarpus* and *Chrysolepis* (western North America) and one species each of *Castanea* and *Castanopsis* [(NCBI GenBank accessions HQ336406 (complete plastid genome of *C. mollissima*), JN044213, JF941179, FJ185053)]. Based on their genetic (plastid) signatures these genera are the closest relatives of *Quercus* within the Fagaceae (Manos et al., 2008). For voucher information and accession numbers see ~~Table~~ File S1. The molecular analyses included three plastid DNA regions: a part of the *rbc*L gene, the *trn*H-*psb*A intergenic spacer and a portion of the *trn*K/*mat*K region (3' intron and partial gene). These markers were chosen based on the variability displayed in previous ~~works~~ analyses (e.g. Manos et al., 2001; Okaura et al., 2007; Simeone et al., 2013) and ~~on the of~~ the high number of their sequences available on in? GenBank. Primer sequences for the three regions were obtained from Kress et al. (2007), Shaw et al. (2005) and Piredda et al. (2011), respectively. DNA extractions~~,~~ ~~primers~~ and PCR protocols were the same as in ~~(~~Piredda et al.~~,~~ (2011)~~ and (Simeone et al., 2013)~~. Sequencing of both DNA strands was performed at Macrogen (http://www.macrogen.com), using the forward and reverse PCR primers,~~at Macrogen (http://www.macrogen.com)~~; electropherograms were edited with CHROMAS 2.3 (http://www.technelysium.com.au) and checked visually.

not italics

~~*Statistical tools*~~*Assessment of overall diversity in* Quercus *and Fagaceae*

The diversity of the investigated regions was evaluated with MEGA 5.2 (Tamura et al., 2011) and DNASP 5.1(Librado and Rozas, 2009). For comparisons of divergence patterns across all Fagales, available data in gene banks were processed using GBK2FAS (Göker et al., 2009); multiple sequence alignments were done with MAFFT v.7 (Katoh and Standley, 2013) using default settings and treating gaps either as missing or 5th state. The MJ algorithm was invoked with default parameters (equal weight of transversion/transition), in order to handle large datasets and multistate characters. Primary data, analyse~~is~~, results and supplementary files (S1–S3) are provided ~~for anonymous download~~ in ~~an~~ the online supporting archive ~~hosted at~~ (OSA; mirrored at www.palaeogrimm.org/data/Smn15_OSA.zip).

Results

Levels of intra- and interspecies plastome divergence in Quercus

The entire dataset included 423 plastid DNA sequences (141 samples, three markers each). The sequence quality was high for all the marker regions: 100% of unambiguous full-length electropherograms with 100% overlap between complementary sequences were recovered across all taxa. Table 2 shows that *trn*H-*psb*A was the most variable marker region (a 34-bp inversion occurring in approximately 50% of the samples was ??? reverse-complemeted and a binary character was inserted to keep record of it~~not considered~~). As expected, t~~T~~he least variable region was *rbc*L~~;~~ ~~as expected~~. No indels were found in the *rbc*L and *mat*K coding regions. The combined cpDNA dataset (*trn*H-*psb*A, *trn*K/*mat*K, *rbc*L) resulted in an alignment of 2082 characters (sites), of which 122 were variable (thereof 72 parsimony-informative; gaps not considered). The ~~alignment~~ combined regions had a nucleotide diversity of 0.~~0056~~006 and included 74 different haplotypes of which 50 were unique (restricted to a single accession). As a result, the overall haplotype diversity was high (*Hd* = 0.978 ± 0.005). With gaps considered, the number of haplotypes increased to 110, of which 89 were unique (*Hd* = 0.~~9939~~994).

obtained with forward and reverse primer ? as sequencing primer

Not clear. Please insert a new sentence.

In general, the infrageneric divergence calculated in *Quercus* is comparable to that found in other genera of the Fagaceae and Betulaceae, and higher than in Juglandaceae (Table 3). All three gene regions allow distinguishing the generic affinity of an oak individual; the same haplotype may be the western Mediterranean populations of *Q. ilex* and *Q. coccifera* (North Africa, Iberia, Southern France, Italy), along the Adriatic coast and into Central Greece (Fig. 3). Also included here are isolated populations of *Q. ilex* from Crete and the southern Black Sea coast. 'Cerris-Ilex' and 'WAHEA' haplotypes are confined to the eastern Mediterranean region. 'Cerris-Ilex' haplotypes are found in the Aegean region (*Q. ilex*, *Q. coccifera* and *Q. aucheri* individuals) and replaced by 'WAHEA' haplotypes (*Q. coccifera*, *Q. aucheri*, *Q. alnifolia*) in south-western Turkey and extending to the east (Levante region; Fig. 3). The 'Cerris-Ilex' type is also found in the *Q. coccifera* individual from northern Turkey, representing the north-easternmost population of this species.

Discussion

All currently available molecular data on Fagaceae show a deep incongruence between nuclear and plastid data. Nuclear phylogenies unambiguously point towards an inclusive common origin of all oaks, i.e. a monophyletic (s. str.) genus *Quercus* (Oh and Manos, 2008; Denk and Grimm, 2010; Hubert et al., 2014). At the same time plastid data repeatedly failed to resolve all oaks as one clade (Manos et al., 2008; this study). Instead, a split emerges (with varying support) between the North American *Notholithocarpus* and a North American/northern temperate clade of oaks, the 'New World Oaks', and the Eurasian *Castanea, Castanopsis* and oak lineages, the 'Old World oak' clade. If we accept the monophyly of the genus *Quercus*, which is further supported by morphology and evidence from the fossil record, haplotypes of *Castanea/Castanopsis* and *Notholithocarpus* that group with the 'New World' and 'Old World' oaks, respectively, can only be the result of incomplete lineage sorting during the formation of the modern genera and/or ancestral gene flow between early diverged lineages. In addition, the plastid gene pool genepool of the earliest oaks must have shown a genetic gradient that reflected to some extent a biogeographic pattern. Although it is impossible to pinpoint the place of origin of oaks, it is clear that paleoclimatic and paleogeographic conditions during the

Eocene facilitated a rapid spread over the Holarctic region, allowing them to pick up and propagate geographic signatures inherited from their common ancestors with *Notholithocarpus*, *Castanea* and *Castanopsis*.

Major trends of plastome differentiation

Overall low genetic intra- and intertaxonomic (intrageneric lineages, genera) distances suggest low evolutionary rates for the chloroplast genomes of Fagales at the examined loci. However, the data coverage is far from sufficient for most genera and families to precisely assess the potential variation within the plastome of this plant group. In Fagaceae, a comparison with the (genetically) more diverse Nothofagaceae and Betulaceae families shows that haplotype variation at the *trn*H-*psb*A locus can be sufficiently high to allow phylogeographic and systematic inferences (see Premoli et al., 2012; Grimm and Renner, 2013). We observed similar levels of variation for haplotypes of intrageneric lineages of *Quercus* at this marker. Furthermore, a geographic pattern is evident for the most widely sampled groups. Groups Ilex, Lobatae and Quercus were the most variable groups, whereas Group Cerris exhibited the lowest differentiation (variation?) rates among species. Interclade differentiation among all *Quercus* groups equalled or exceeded that found in the four outgroup genera (*Castanea*, *Castanopsis*, *Notholithocarpus* and *Chrysolepis*). As a consequence, the outgroup taxa appear scattered across the tree, rather than being culled in a distinct subtree, rendering the plastome of *Quercus* 'non-monophyletic'. Outgroup selection as a potential source of topological ambiguity has previously been pointed out by Hubert et al. (2014)~~:~~ who analyzed 108 oak taxa~~,~~ at eight nuclear markers~~)~~. Ambiguous relationships within Fagales independently of the strength of the obtained phylogenetic signal were also suggested in a recent study based on plastid DNA, fossil and reproductive syndromes ~~analyses~~, which resolved the majority of inter-generic relationships in each family except for in the Quercoideae group making *Castanopsis* and *Quercus* "non-monophyletic" (Xiang et al., 2014).

How polyphyletic is Quercus Group Ilex?

Figure 4 highlights the incongruence of the plastid genealogy tree with the current understanding of the evolution of Fagaceae and oaks based on molecular sequence data from non-coding nuclear gene regions (Manos et al., 2001; Denk and Grimm, 2010), a recent time-calibrated nuclear phylogeny of oaks (Hubert et al. 2014), and the fossil record of modern lineages as documented by pollen investigated under the scanning-electron microscope (Grímsson et al., 2015; see also Denk and Grimm, 2009). Extensive sampling is more likely to reveal polyphyly (Wiens and Servedio, 2000), but caution is needed when trees are interpreted regarding single (monophyly, paraphyly) or multiple (polyphyly) origins of species. Indeed, polyphyly as expressed in phylogenetic trees, 'non-monophyly', is more likely due to homoplasy between distant branches in a tree. Additional reasons for topological polyphyly is that we work with gene trees, and not species trees: in this case, topological polyphyly ('non-monophyly') may be due to ancestral polymorphisms or introgression causing incongruences between the trees. However, distinguishing the effects of these mechanisms may be very difficult in the absence of nuclear markers and (palaeo-)geography as complementary information (Funk and Omland, 2003). As a general rule, deep putative polyphyly or paraphyly (lineages resolved as a "basal" grade) hints at retained ancestral polymorphism, i.e. a lineage sorting phenomenon, while recently introgressed haplotypes may assume a highly derived position in a gene tree. At the same time, incomplete lineage sorting is not predicted to promote the geographic proximity of interspecifically shared haplotypes that may be seen under local introgression (Hare and Avise, 1998, Masta et al., 2002).

this is available for the Fagaceae

A strictly polyphyletic origin of *Quercus* Group Ilex or its Mediterranean members is unlikely. Nuclear data covering the entire range of *Q. ilex* and *Q. coccifera* in the Mediterranean region unambiguously resolved the two species as close, but mutually monophyletic sister taxa (Denk and Grimm, 2010). The 'non-monophyly' of Group Ilex plastomes seen in the tree (Fig. 1), including haplotypes shared with Group Cerris or closely related to Group Cyclobalanopsis therefore reflects either incomplete lineage sorting or introgression or both. In the absence of nucleome data for all here included individuals, it is impossible to infer to which degree introgression and incomplete lineage sorting contributed to the plastid gene pool of the Mediterranean species of Group Ilex. Nevertheless, the most straightforward explanation for the observed scenario would be a combined effect: asymmetrical introgression of ancestral haplotypes resulting in local genetic clusters decoupled from taxonomic boundaries, in which plastome accessions of species or species complexes may form grades or multiple clades in phylogenetic trees, thus appearing para- or polyphyletic (e.g. Rieseberg and Soltis, 1991; Whittemore and Schaal, 1991).

As modelled by Excoffier et al. (2009), interspecific gene flow? interactions during historical range fluctuations can profoundly affect the observed phylogeographic patterns, and manifest as paraphyly or reticulation (polyphyly in a broad sense). In fact, most range expansions do not occur in completely uninhabited areas, and interbreeding between local and an expanding (invasive) species with subsequent asymmetrical introgression can develop also in absence of selection, e.g. when one species is dominant and most abundant (Lepais et al., 2009). Plastid haplotypes referring to the original ('lost') species are indeed likely to persist over long evolutionary periods, and may still be found in the invading species. Noteworthy, environmental changes and disturbance of local communities have been shown to increase hybridisation rates (Lagache et al., 2013), hence, the potential for widespread, imbalanced introgression. In Group Ilex oaks, the interspecific capture of plastids among sexually incompletely isolated species likely occurred on the geological timescale, concealing the species relationships at various stages in the history of the genus. In a comprehensive study of the genus *Ilex* (Manen et al., 2010), the high incongruence between a taxonomically compatible nuclear gene tree and a geographically structured plastid tree was explained with extensive extinctions between the Cretaceous and Miocene and multiple hybridization and introgression events between distantly related lineages. This has been documented also for *Platanus* (Grimm and Denk, 2010) and more recently suggested for the evergreen white oaks of *Quercus* subsection *Virentes* (Eaton et al., 2015). Similar ancient lateral transfers have been also inferred to explain the paraphyly of the maternally inherited mtDNA of *Picea* species (Bouillè et al., 2011) and *Pinus* species (Tsutsui et al., 2009).

Decoupling of plastid signatures and taxonomy in oaks

Speciation processes in *Quercus* do not immediately leave imprints in the plastome (e.g. Neophytou et al., 2010; Cavender-Bares et al., 2011) as also well documented for *Nothofagus* (Acosta and Premoli, 2010; Premoli et al., 2012). Low mutation rates and long generation times can contribute to slow evolutionary rates and incomplete lineage sorting of organellar genomes (Cavender-Bares et al., 2015; Besnard et al., 2007). In addition, reiterated extinction and re-colonisation involving bottlenecks, genetic drift, and founder effects may cause random fixation of haplotypes, increasing the probability for retaining ancestral traits. Oaks in general, and especially the Mediterranean taxa, are also characterised by a marked resprouting ability in response to environmental disturbances (Barbero et al., 1990). This could have contributed to clonally preserve and transmit ancestral plastid lineages (maternally inherited) during multiple unfavourable conditions since the origin of the Mediterranean region (Blondel and Aronson, 1999). At the same time, large population sizes and long distance pollen dispersal might have contributed to homogenise the nuclear genomes in local populations of a species but not their organelle genomes. Topographic barriers hindering seed dispersal and balancing selection on the organelle genomes could have preserved ancestral polymorphisms within the species' gene pools, also (Funk and Homland, 2003). However, although there is some evidence for balancing selection on the mtDNA in plants (Städler & Delph 2002) it has not yet been suggested as an explanation of intraspecific plastid heterogeneity.

Additionally, Fagaceae lineages are susceptible to hybridisation and introgression (Arnold, 2006). This may lead to the formation of morphologically unambiguous individuals of a species with plastid signatures of another (Whittemore and Schaal, 1991; Petit et al., 2004). There is strong evidence for local introgression in oak communities with morphologically distinct species in the case of European white oaks (Group Quercus: *Q. robur*, *Q. petraea*, *Q. pyrenaica*, *Q. pubescens*, *Q.*

page 37-38. There is little information on Fagaceae in this publication *frainetto*; Curtu et al., 2007; Valbuena-Carabaña et al., 2007; Lepais et al., 2009), as well as in members of *Quercus* subsection *Virentes*, a subgroup of Group Quercus, in North America (Cavender-Bares et al., 2015), and across a wide range of Group Lobatae (Dodd and Afzal-Rafii, 2004; Peñaloza-Ramírez et al., 2010; Moran et al., 2012; Valencia-Cuevas et al., 2015). In our focal group, hybrids and different levels of genetic introgression among morphologically pure individuals were molecularly documented via genetic assignment analysis in *Q. ilex*/*Q. coccifera* (Ortego and Bonal, 2010) and, to a lesser extent, in *Q. coccifera*/*Q. alnifolia* (Neophytou et al., 2011). Also, the potential for inter-group hybridisation was experimentally demonstrated for *Q. ilex* and *Q. robur* (Group Quercus; Schnitzler et al., 2004), and natural introgression in *Q. ilex* and *Q. suber* (Group Cerris) was identified in southern France (Mir et al., 2009) and the Iberian Peninsula (Burgarella et al., 2009). Therefore, it is possible that ancient hybridization and introgression, favoured by the well-known sexual promiscuity between closely related taxa and their ability to disperse pollen over long distances, obscure the true evolutionary origin of an oak species or entire lineage. In the Mediterranean, the profound geological and ecological changes during the Neogene (Blondel and Aronson, 1999) likely caused extinction, re-colonisation, range fragmentation and hybridisation linked to secondary contact, especially when species were still young and reproductive barriers likely weaker than today. Taken together, incomplete sorting of ancestral traits and introgression of haplotypes thus appear highly likely mechanisms to decrease inter-species plastid differentiation while at the same time increasing intra-species variation.

?! "introgression" is part of the at cp evol. history cpDNA markers.

Temporal and spatial framework of plastome evolution

The three distinct plastid haplotypes observed in modern Mediterranean members of *Quercus* Group Ilex may reflect three radiation phases (range extensions), followed by range disruptions and isolation of plastome lineages within the 'Old World Clade' of *Quercus*. Considering the high diversity of haplotypes in Group Ilex as compared to other major oak lineages (or other genera in the Fagales; see Table 3; Table S2) it can be hypothesized that the geographical disruptions in the plastome of the ancestors of Group Ilex and interacting lineages predate the manifestation of modern taxa (species and infrageneric groups; Fig. 1). Haplotypes shared between members of Group Ilex and its sister lineages Group Cerris and Group Cyclobalanopsis may indicate the same geographic origin or may be the result of secondary contact and asymmetrical introgression.

New paragraph.

Evolutionary hypotheses concerning the unique 'Euro-Med' haplotype — Independent from the actual position of the plastid root (note the scattered placement of putative outgroups in Fig. 1), the divergence of the 'Euro-Med' haplotype must have coincided with the initial differentiation in *Quercus* (Fig. 1). Oaks had achieved a wide northern hemispheric range by the Eocene. Unequivocal fossils are known from high latitudes (North America, Greenland, North Europe; Crepet and Nixon, 1989; Manchester, 1994; Grímsson et al., 2015) and mid-latitudes (Central Europe, South East Asia; Kvaček and Walther, 1989; Hofmann, 2010). All major lineages of oaks were established by the end of the Eocene, ca. 35 Ma, as evidenced by the fossil record and molecular dating using eight nuclear gene regions (Bouchal et al., 2014: fig. 14; Hubert et al., 2014; Grímsson et al., 2015). During this time, one fraction of oaks, represented by the 'Euro-Med' plastids, must have been geographically and reproductively isolated which would have caused a major split in the plastid gene pool (Fig. 1). Today, the 'Euro-Med' haplotype within *Quercus* Group Ilex is the only one exclusively shared by just two, but widespread Mediterranean species of Group Ilex, *Q. ilex* and *Q. coccifera*. This haplotype is markedly distinct from haplotypes in other oaks or Fagaceae (Fig. 1). Two evolutionary hypotheses could explain the establishment of this unique haplotype in *Q. ilex*-*Q. coccifera* (Fig. 5): (*i*) The 'Euro-Med' haplotype is the remnant of an extinct oak lineage that was intrograded (invaded) and consumed by members of Group Ilex (Fig. 5b); under this scenario Group Ilex would have migrated into Europe at some point prior to the Miocene where it came into contact with this extinct oak lineage. (*ii*) The 'Euro-Med' haplotype represents an original plastome of Group Ilex; under this scenario, the first split within the modern 'Old World clade' would have been between western members of Group Ilex and an eastern cluster comprising precursors, members of Group Ilex and Group Cyclobalanopsis (Fig. 5c). At the moment we can only speculate which hypothesis applies.

During the Eocene and Oligocene a number of extinct genera of Fagaceae (e.g. *Eotrigonobalanus*, *Trigonobalanopsis*) and extinct lineages of *Quercus* were present in the northern hemisphere (Kohlman-Adamska, 2000, 2001, Stuchlik et al., 2007; Denk et al., 2012; Grímsson et al., 2015). It is possible that the plastid of one of these lineages was captured by members of *Quercus* Group Ilex when expanding westwards into the modern area of the Euro-Med haplotype. This would fit with the first hypothesis.

Scenario *ii* would be in agreement with geographic fragmentation and subsequent incomplete lineage sorting across the former Paratethyan and current Himalayan arcs. During its evolution, Group Ilex must have been continuously affected by range disruptions caused by tectonic activity south of the Paratethys linked to the collision of Africa and the Indian subcontinent with Eurasia (Fig. 6); progressive rarefaction of the original haplotypes and the occurrence of (repeated) invasion and introgression events that left imprints in the plastome even within the same species is highly likely in such a topographic setting. Nevertheless, under scenario *ii* the diversity and distinctness of the 'Euro-Med' haplotype would require an initial geographic separation between the westernmost (European) and other members of Group Ilex, predating the formation of the modern lineages and surviving later (Miocene) migration waves. Therefore, scenario *ii* appears less probable than scenario *i*.

Evolutionary significance of the 'Cerris-Ilex' haplotype and the origin of Group Cerris—The 'Cerris-Ilex' haplotype is shared between all species of *Quercus* Group Cerris (western Eurasian and East Asian), East Mediterranean (Aegean) individuals and two East Asian species of Group Ilex. This is in agreement with Denk and Grimm (2010) who suggested that *Quercus* Group Cerris evolved from Group Ilex by budding (a hypothesis further confirmed by the 8-nuclear gene data set used by Hubert et al., 2014), and the low support for a Group Ilex clade in an all-Fagaceae (excluding *Fagus*) tree based on over 1000 nuclear ITS sequences (Denk and Grimm, 2010). Hubert et al. (2014) inferred a Miocene age for this budding event, which corresponds to the earliest unequivocal fossil of *Quercus* Group Cerris (Kmenta, 2011) and is younger than the earliest definite fossil record of *Quercus* Group Ilex in Europe (early Oligocene, Cospuden; Denk et al., 2012). Also, dispersed pollen from the Paleogene Changchang Formation, Hainan (Hofmann, 2010), resembles both *Quercus* Group Ilex and Group Cyclobalanopsis; the age of this formation is considered late early to early late Eocene (Lei et al., 1992). The haplotype most closely related to the 'Cerris-Ilex' haplotype is encountered in the widespread East Asian *Q. phillyraeoides*, the only species of Group Ilex extending to Japan (in contrast, the East Asian members of Group Cerris have a much wider range in north-eastern Asia: Menitsky, 2005). Regarding its phylogenetic position, the emergence of the 'Cerris-Ilex' haplotype appears to have been linked with a major taxonomic sorting event in Eurasian Fagaceae, resulting in distinct haplotypes restricted to genera and intrageneric groups of *Quercus* (Fig. 1). Based on the palaeobotanical record, these lineages (*Castanopsis*, *Castanea*, *Quercus* Group Ilex, *Quercus* Group Cyclobalanopsis) were well established at least by the Eocene (Table 4, Fig. 5; Grímsson et al., 2015); a deep divergence is reflected by their distinctly different nuclear genomes (Oh and Manos, 2008; Denk and Grimm, 2010; Hubert et al., 2014). Again, two evolutionary scenarios could explain the occurrence of the 'Cerris-Ilex' haplotype in Aegean individuals of *Q. ilex* and *Q. coccifera* and the westernmost *Q. aucheri*: (*i*) Group Cerris evolved in western Eurasia/Himalaya from an (extinct) subtropical to temperate sublineage of Group Ilex, which left its imprint in the Aegean members of Group Ilex, and *Q. spinosa*, *Q. engleriana* and *Q. phillyraeoides*. (*ii*) Group Cerris shares a common ancestry with the north-east Asian *Q. phillyraeoides* of Group Ilex. Under the latter scenario, the budding event of the group would have taken place in north-eastern Asia, from where it migrated into western Eurasia and the Aegean region. In relatively recent times, Group Cerris came into contact with the Mediterranean members of Group Ilex and was locally introgressed (e.g. Burgarella et al., 2009; Mir et al., 2009). However, it is difficult to explain why *Q. ilex-coccifera* should only intrograde into populations of Cerris oaks at a large scale in the Aegean region. Therefore, scenario *i* appears to be more plausible. The fossil record of Group Cerris, in particular the Central and East Asian one, could potentially shed further light on this hypothesis. Unfortunately, it is currently not very well understood.

Origin and evolutionary significance of the 'WAHEA' haplotype of Group Ilex—The West Asian-Himalayan-East Asian (WAHEA) haplotype represents Eastern Mediterranean members of *Quercus* Group Ilex and is sister to a clade comprising several Asian species of Group Ilex (Himalayas to the mountains of Southeast Asia). Based on its position in the plastid tree, this haplotype seems to reflect a second radiation within the Old World Clade and allies after the isolation of the 'Euro-Med' lineage and prior to the radiation and subsequent sorting within the clade comprising the 'Cerris-Ilex' haplotypes (Fig. 1). The modern distribution of species with the WAHEA haplotype follows the Himalayan corridor (Kitamura, 1955; Güner and Denk, 2012). The Himalayan corridor is a narrow band along the southern flanks of the Himalaya with a subtropical to temperate climate (*Cwa, Cwb*; Peel et al. 2007) providing a pathway for plants originating from humid temperate Cenozoic laurel and mixed broad-leaved deciduous and evergreen forests. In addition to *Quercus* Group Ilex (Zhou, 1992; Velitzelos et al., 2014), prominent relic taxa include species of *Acer, Aesculus, Cedrus, Cotinus, Juglans, Platanus*, and *Rhododendron* among others. The 'WAHEA' haplotype represents the western counterpart to the haplotype lineage found in East Asian species of Group Ilex and Cyclobalanopsis (Fig. 1). The relic *Q. alnifolia*, today only found in the mid-montane region of Cyprus (Mt. Troodos), would be a witness of this expansion (Menitsky, 2005).

Conclusion

Plant biogeographic studies at the genus level have commonly relied on few to many chloroplast markers and a single or very few accessions per taxon. In the case of woody angiosperms with a subtropical to temperate distribution such as for example Nothofagaceae (Svenson et al., 2001; Knapp et al., 2005), *Rhus* (Yi et al., 2004), *Cornus* (Xiang et al., 2005), *Carpinus* (Yoo and Wen, Check for consistent formatting of citations

2007), *Castanea* (Lang et al., 2007), *Juglans* (Aradhya et al., 2007), and *Carya* (Zhang et al., 2013a), such an approach runs the risk of capturing but a limited aspect of the evolutionary history of the focal group. Mere combination with e.g. nuclear ITS data can be problematic, too (compare data shown here with data provided by Denk and Grimm, 2010, on western Eurasian members of Group Ilex). The decoupled evolutionary signals in plastomes and the nucleome/morphology as documented for *Nothofagus* (Acosta and Premoli, 2010; Premoli et al., 2012) and *Quercus* Group Ilex (this study) suggest that the traditional placeholder sampling strategy is not ideal. Signals from few-marker/many-samples data sets are likely to be complex or even puzzling (Figs 1–4), but at the same time provide entirely new perspectives on plant evolution worth exploring. For *Quercus* Group Ilex, our pilot study focussing on Mediterranean species reveals a crucial aspect of oak evolution not seen in the combined nuclear, morphological, and fossil data: large-scale introgression and/or incomplete lineage sorting among ancestral lineages of modern major groups and species. The new data corroborate hypotheses that Group Cerris evolved ('budded') relatively recent from Group Ilex as discussed by Denk and Grimm (2010, using over 600 ITS and over 900 5S-IGS accessions covering all western Eurasian oak species) and inferred by Hubert et al. (2014, using dated phylogenies based on seven single-copy nuclear regions and the ITS region). Our preferred hypothesis is that *Quercus* Group Cerris evolved in western Eurasia and the Himalayas when the then chiefly subtropical low latitude Group Ilex radiated into temperate niches. Accordingly, within modern members of Group Cerris, a wide spectrum of leaf traits is found from pseudo-evergreen in *Q. suber*, to semi-evergreen in *Q. brantii*, *Q. ithaburensis*, *Q. trojana* (partly) and fully deciduous in *Q. acutissima*, *Q. castaneifolia*, *Q. cerris*, *Q. libani* and *Q. variabilis*. The conspicuous plastid diversity in the Mediterranean species of Group Ilex and the lineage in general (Figs 1, 2; Table 2) can only be interpreted as reflecting the highly complex geographical history of this group, due to growing geographic isolation between clades from the Oligocene onwards. The 'Euro-Med' haplotype evidences an initial phase of geographic differentiation predating the formation of modern lineages, but its origin and evolutionary significance remain enigmatic (Fig. 5).

Although decoupled from taxonomy, the plastid phylogeny provides important, independent information on the geographic differentiation of *Quercus* prior to the formation of modern species/species groups. So far, the major split within oaks has been between 'New World' and 'Old World' oaks (following Manos et al., 2001) because of the current distribution of the major, molecular-defined lineages of oaks. This view replaced traditional concepts (reviewed in Denk and Grimm, 2010) recognising two subgenera/genera, one in subtropical to tropical East Asia (subgenus/genus *Cyclobalanopsis* = Group Cyclobalanopsis) and the other ubiquitous on the northern hemisphere (subgenus/genus *Quercus*, including all other infrageneric groups). The plastid data presented here suggest that the early evolution of oaks instead was geographically bound to high latitude Arctic regions and to low latitude subtropical regions (Fig. 5). The high latitude lineages remained genetically homogeneous in the nucleome, but also in the plastome to some degree. Continuous circum-polar distribution could have prevented pronounced genetic drift in the high latitude lineage, which became the 'New World Clade', and explains low genetic differentiation in deciduous high and mid latitude white oaks until today (Denk and Grimm, 2010). At the same time, the Atlantic, the proto-Mediterranean, the Paratethys and the rise of the Himalaya isolated the Eurasian low latitude lineage, and may explain the differentiation between plastome clades, known to be mirror of geographic isolation rather than more than species taxonomy in case of introgression. The high diversity of southern plastomes, the low diversity of northern plastomes, and the propensity of oaks to introgress, altogether built a consistent framework for deciphering the historical biogeography of this group.

Our data should only be viewed as a first step towards a more complete understanding of the biogeography and evolution of oaks. The next step would be to map the plastid variation of *Quercus* Group Ilex across its entire range by sampling multiple stands of the Himalayan and East Asian species to characterise the geographic and taxonomic ranges of the various plastid lineages. Finally, because of similar strong correlation between plastome differentiation and geographic distribution at the population level and the species/genus level, similar or identical plastid haplotypes typically