Plastome data reveal multiple geographic origins of *Quercus* Group Ilex

Marco Cosimo Simeone, Guido W Grimm, Alessio Papini, Federico Vessella, Simone Cardoni, Enrico Tordoni, Roberta Piredda, Alain Franc, Thomas Denk

Nucleotide sequences from the plastome are currently the main source for assessing taxonomic and phylogenetic relationships in flowering plants and their historical biogeography at all hierarchical levels. One exception is the large and economically important genus *Quercus* (oaks). Whereas differentiation patterns of the nuclear genome are in agreement with morphology and the fossil record, diversity patterns in the plastome are at odds with established taxonomic and phylogenetic relationships. However, the extent and evolutionary implications of this incongruence has yet to be fully uncovered. The DNA sequence divergence of four Euro-Mediterranean Group Ilex oak species (*Quercus ilex Qcoccifera QaucheriQalnifolia* ilex L., Q. coccifera L., Q. aucheri Jaub. & Spach., Q. alnifolia Poech.) was explored at three chloroplast markers (rbcL, trnK-matK, trnH-psbA). Phylogenetic relationships were reconstructed including worldwide members of additional 55 species representing all *Quercus* subgeneric groups. Family and order sequence data were harvested from gene banks to better frame the observed divergence in larger taxonomic contexts. We found a strong geographic sorting in the focal group and the genus in general that is entirely decoupled from species boundaries. Main plastid haplotypes shared by distinct oak lineages from the same geographic region and high plastid diversity in members of Group Ilex are indicative for a polyphyletic origin of their plastomes. The results suggest that incomplete lineage sorting and repeated phases of unidirectional introgression among ancestral lineages of Group Ilex and two other main Groups of Eurasian oaks (Cyclobalanopsis and Cerris) caused this complex pattern. Comparison with the current phylogenetic synthesis also suggests an initial high- versus mid-latitude biogeographic split within *Quercus*. High plastome plasticity of Group Ilex reflects geographic area disruptions, possibly linked with high tectonic activity of past and modern distribution ranges, that did not leave imprints in the nuclear genome of modern species and infrageneric lineages.

1 **Title: Plastome data reveal multiple geographic origins of *Quercus* Group Ilex**

2

3 **Authors**

4 Marco Cosimo Simeone [1,*], Guido W. Grimm [2], Alessio Papini [3], Federico Vessella [1], Simone

5 Cardoni [1], Enrico Tordoni [4], Roberta Piredda [5], Alain Franc [6,7], Thomas Denk [8]

6

7 **Affiliations**

8 [1] = Department of Agricultural and Forestry Science (DAFNE), Università degli Studi della

9 Tuscia, via S. Camillo de Lellis, 01100 Viterbo, Italy

10 [2] = Department of Palaeontology, University of Wien, Althanstrasse 14 (UZA II), 1090 Wien,

11 Austria

12 [3] =Dipartimento di BiologiaVegetale, Università degli studi di Firenze, via La Pira 4, 50121

13 Firenze, Italy

14 [4] =Department of Life Science, Università degli studi di Trieste, via L. Giorgieri 10, 34127

15 Trieste, Italy

16 [5] =Stazione Zoologica Anton Dohrn, Villa Comunale, 8012 Napoli, Italy

17 [6] = INRA, UMR BIOGECO-1202, 69 route d'Arcachon, F-33612 Cestas, France

18 [7] = Université Bordeaux, UMR BIOGECO-1202, Avenue des Facultés, F-33405 Talence, France

19 [8]= Department of Palaeobiology, Swedish Museum of Natural History, Box 50007, 10405

20 Stockholm, Sweden

21

22 **Corresponding Author**

23 Dr. Marco Cosimo Simeone

24   Department of Agricultural and Forestry Science (DAFNE), Università degli Studi della Tuscia,

25   via S. Camillo de Lellis, 01100 Viterbo, Italy

26   E-Mail address: mcsimeone@unitus.it

27 **Abstract**

28 Nucleotide sequences from the plastome are currently the main source for assessing taxonomic

29 and phylogenetic relationships in flowering plants and their historical biogeography at all

30 hierarchical levels. One exception is the large and economically important genus *Quercus* (oaks).

31 Whereas differentiation patterns of the nuclear genome are in agreement with morphology and

32 the fossil record, diversity patterns in the plastome are at odds with established taxonomic and

33 phylogenetic relationships. However, the extent and evolutionary implications of this

34 incongruence has yet to be fully uncovered. The DNA sequence divergence of four Euro-

35 Mediterranean Group Ilex oak species (*Quercus ilex* L., *Q. coccifera* L., *Q. aucheri* Jaub. &

36 Spach., *Q. alnifolia* Poech.) was explored at three chloroplast markers (rbcL, trnK-matK, trnH-

37 psbA). Phylogenetic relationships were reconstructed including worldwide members of

38 additional 55 species representing all *Quercus* subgeneric groups. Family and order sequence

39 data were harvested from gene banks to better frame the observed divergence in larger

40 taxonomic contexts. We found a strong geographic sorting in the focal group and the genus in

41 general that is entirely decoupled from species boundaries. Main plastid haplotypes shared by

42 distinct oak lineages from the same geographic region and high plastid diversity in members of

43 Group Ilex are indicative for a polyphyletic origin of their plastomes. The results suggest that

44 incomplete lineage sorting and repeated phases of unidirectional introgression among ancestral

45 lineages of Group Ilex and two other main Groups of Eurasian oaks (Cyclobalanopsis and Cerris)

46 caused this complex pattern. Comparison with the current phylogenetic synthesis also suggests

47 an initial high- versus mid-latitude biogeographic split within *Quercus*. High plastome plasticity

48 of Group Ilex reflects geographic area disruptions, possibly linked with high tectonic activity of

49 past and modern distribution ranges, that did not leave imprints in the nuclear genome of modern

50 species and infrageneric lineages.

51

52 **Keywords:** Fagaceae, Mediterranean, Plastome polyphyly, Ancient introgression, Incomplete

53 lineage sorting, Decoupled phylogenies

54 **Introduction**

55 *Quercus* L. (oaks) is among the most ecologically diverse and economically important

56 extratropical tree genera in the northern hemisphere (Govaerts and Frodin, 1998). *Quercus* is the

57 largest genus in the order Fagales, comprising ca. 400–500 species. Oaks are concentrated in the

58 Americas (Groups Quercus, Lobatae and Protobalanus; Flora of North America Editorial

59 Committee, 1997) and Southeast Asia and southern China (Group Cyclobalanopsis; Flora of

60 China Editorial Committee, 1999). In contrast, a relatively lower number of species can be found

61 in western Eurasia and the Mediterranean (Groups Ilex and Cerris; Kubitzki, 1993; Menitsky,

62 2005). The six major infrageneric lineages of *Quercus* occur from the tropics to the high

63 mountains of the temperate zone and to the boreal continental, cold temperate regions (Denk and

64 Grimm, 2010). The northern limit of oaks in North America and Eurasia coincides with the

65 border of *Dfb* to *Dfc* and *Dwb* to *Dwc* climates, snow climates with warm versus cool summers

66 (Köppen, 1936; Kottek et al., 2006; Peel et al., 2007).

67 Recent molecular phylogenetic studies at and below the genus level focussed on the nucleome of

68 oaks (Oh and Manos, 2008; Denk and Grimm, 2010; Hipp et al., 2014; Hubert et al., 2014).

69 These studies consistently recovered two main lineages, the 'New World Clade' comprising the

70 white oaks (Group Quercus), red oaks (Group Lobatae) and golden-cup oaks (Group

71 Protobalanus), and the 'Old World Clade' consisting of the cycle-cup oaks (Group

72   Cyclobalanopsis), the Ilex oaks (Group Ilex) and the Cerris oaks (Group Cerris). Evidence from

73   nuclear markers and the fossil record suggests that the initial split in the 'New World Clade' was

74   pre-Oligocene between the lineages leading to Group Lobatae and Group Protobalanus/Quercus

75   (Bouchal et al., 2014; Hubert et al., 2014; Grímsson et al., 2015). This early radiation of the

76   Quercus/Protobalanus lineage left its imprints in the molecular signatures of the few modern

77   species of Group Protobalanus and two narrow endemic white oak species, *Quercus pontica*

78   (north-eastern Turkey, south-western Georgia; Denk and Grimm, 2010) and *Q. sadleriana*

79   (California; Hubert et al., 2014). Within the 'Old World Clade', the major split was established

80   between the evergreen Groups Cyclobalanopsis and Ilex during the Eocene/Oligocene, whereas

81   the chiefly temperate Group Cerris is suggested to have evolved ('budded') from a Group Ilex

82   stock, possibly in Europe, not before the earliest Miocene (Denk and Grimm, 2009; Kmenta,

83   2011; Hubert et al., 2014; Velitzelos et al., 2014).

84   Nuclear amplicon data sets have also contributed to resolve the circumscription of these six

85   groups and to delineate some intergroup and interspecies relationships (López de Heredia et al.,

86   2007; Pearse and Hipp, 2009; Denk and Grimm, 2010; Hubert et al., 2014); well-resolved

87   within-lineage relationships were recently obtained from phylogenomic data in the genetically

88   least-diverged, but species-rich Group Quercus (Hipp et al., 2014). Nucleome-based studies,

89   therefore, clearly indicate a strong correlation between morphology/speciation and nuclear

90   differentiation in oaks. In contrast, oak plastid haplotypes are extensively shared between groups

91   of species (Whittemore and Schaal, 1991; Belahbib et al., 2001; Manos and Stanford, 2001; Petit

92   et al. 2002; Kanno et al., 2004; López de Heredia et al., 2007; Okaura et al., 2007; Neophytou et

93   al., 2010; Gugger and Cavender-Bares 2013). Notably, this was also observed in other genera of

94   Fagaceae such as *Fagus* (Fujii et al. 2002; Lei et al. 2012; Zhang et al. 2013b) and *Lithocarpus*

95   (Cannon and Manos, 2003), and other Fagales such as the northern hemispheric *Carya*

96   (Juglandaceae; Zhang et al., 2013a) and the South American *Nothofagus* (Nothofagaceae; Acosta

97   and Premoli, 2010; Premoli et al., 2012). Plastomes of this large group of long-lived woody

98   plants appear to retain molecular signatures of evolutionary events that cannot be investigated

99   when considering the nuclear DNA alone (e.g., Cavender-Bares et al., 2011; Premoli et al.,

100   2012). As such, they can provide additional information to complement hypotheses on

101   diversification and speciation processes. However, the extent and evolutionary implications of

102   nuclear-plastome incongruence in *Quercus* have yet to be fully uncovered.

103   Testing the potential of DNA barcoding in western Eurasian oaks, Simeone et al. (2013) recently

104   found puzzling diversity in the plastid haplotypes of samples belonging to Group Ilex. In

105   addition to interspecific haplotype sharing, paraphyly to Groups Cerris and Quercus and an

106   underlying geographic partitioning was suggested. In the present study, we increased the

107   geographic coverage and taxon sampling to explore the complex patterns of plastome evolution

108   in *Quercus* Group Ilex. This species group is today confined to extra-tropical regions of Eurasia,

109   spanning from arid Mediterranean maquis to high mountain and sub-alpine Himalayan forests

110   and thickets, and to subtropical forests of SE Asia. Group Ilex includes some 35 evergreen,

111   mostly sclerophyllous taxa, whose taxonomy is still controversial (see Table 1) and

112   biogeographic history is not yet well understood (Menitsky, 2005; Denk and Grimm, 2010). In

113   this work, we compiled plastid sequence data for 81accessions of 20 oak taxa of Group Ilex. The

114   main sampling effort was put into the four species currently occurring in the Mediterranean and

115   adjacent regions in North Africa (Atlas Mountains) and northern Turkey (Black Sea region): the

116   widespread *Quercus ilex* L. and *Q. coccifera* L., and the two East Mediterranean narrow

117   endemics *Q. aucheri* Jaub. & Spach. and *Q. alnifolia* Poech. Data for additional 56 individuals of

118   ca. 40 species were also produced to integrate all subgeneric *Quercus* groups and their

119   worldwide geographic distribution. Additionally, Fagales data sets were harvested from gene

120   banks to allow interpretation of the observed divergence in the plastid markers within a larger

121   taxonomic frame. Our objectives were: (1) to assess the extent of plastome diversity in the Euro-

122   Mediterranean focal group; (2) to outline key phylogeographic patterns within *Quercus* Group

123   Ilex; (3) to establish major evolutionary steps for the differentiation of the 'Old World Clade'.

124

125   **Material and methods**

126

127   *Plant Material, DNA amplification and analyses*

128   Our analysis included 59 individuals of the four Mediterranean *Quercus* Group Ilex species

129   (Table S1) covering their entire range in North Africa and western Eurasia. Additionally, 22

130   individuals of 16 Asian species of Group Ilex were analysed. The final dataset also included all

131   species of the western North American Group Protobalanus (five species, 10 individuals), 16

132   species of Group Quercus (20 individuals, from North America and Eurasia), five species of the

133   East Asian Group Cyclobalanopsis (11 individuals), seven species of the American Group

134   Lobatae (eight individuals), and six species of Group Cerris (seven individuals). The outgroup

135   set was represented by one sample each of the monotypic genera *Notholithocarpus* and

136   *Chrysolepis* (western North America) and one species each of *Castanea* and *Castanopsis* [(NCBI

137   GenBank accessions HQ336406 (complete plastid genome of *C. mollissima*), JN044213,

138   JF941179, FJ185053). Based on their genetic (plastid) signatures these genera are the closest

139   relatives of *Quercus* within the Fagaceae (Manos et al., 2008). For voucher information and

140   accession numbers see Table S1. The molecular analyses included three plastid DNA regions: a

141   part of the *rbc*L gene, the *trn*H-*psb*A intergenic spacer and a portion of the *trn*K/*mat*K region (3'

142   intron and partial gene). These markers were chosen based the variability displayed in previous

143   works (e.g. Manos et al., 2001; Okaura et al., 2007; Simeone et al., 2013) and on the of the high

144   number of their sequences available on GenBank. DNA extractions, primers and PCR protocols

145   were the same as in (Piredda et al., 2011) and (Simeone et al., 2013). Sequencing was performed

146   at Macrogen (http://www.macrogen.com); electropherograms were edited with CHROMAS 2.3

147   (http://www.technelysium.com.au) and checked visually.

148

149   *Statistical tools*

150   The diversity of the investigated regions was evaluated with MEGA 5.2 (Tamura et al., 2011) and

151   DNASP 5.1(Librado and Rozas, 2009). For comparisons of divergence patterns across all

152   Fagales, available data in gene banks were processed using GBK2FAS (Göker et al., 2009);

153   multiple sequence alignments were done with MAFFT v.7 (Katoh and Standley, 2013) using

154   default setting and checked by eye to remove inconsistencies and erroneous sequences (taxa and

155   sequence numbers in Supporting Information). To minimise the effect of alignment gaps, and

156   since we were primarily interested in assessing intra- and intergeneric divergence, alignments

157   included only subsets of the Fagales: 1) Nothofagaceae (data covering all four genera); 2)

158   Fagaceae (10 genera including *Quercus*); 3) Betulaceae-*Ticodendron*-Casuarinaceae (11 genera);

159   4) Juglandaceae (9 genera); 5) Myricaceae (4 genera). Pairwise distance matrices (uncorrected *p*-

160   distance, K2P, HKY, GTR+ Γ) for each marker were calculated with PAUP* 4.0 (Swofford,

161   2002). Minimum intra-specific and minimum/maximum inter-specific distances (calculated with

162   G2CEF; Göker and Grimm, 2008) within and between genera, subgenera in the case of *Fagus*, and

163   infrageneric groups in case of *Quercus*, are listed in Table S2.

164

165   *Phylogenetic analyses*

166   Multiple sequence alignments for the focal group were obtained with CLUSTALW 1.81

167   (Thompson et al., 1994) and checked by eye. The matrices were concatenated with the Python

168   programme COMBINEX2_0.PY (PYTHON v. 2.6.4; BIOPYTHON 1.57).

169   Maximum likelihood trees were inferred with GARLI (Zwickl, 2006; run on the CIPRES portal,

170   http://www.phylo.org/sub_sections/portal/) using four data partitions (*rbc*L and *mat*K codons,

171   *trn*K intron and *trn*H-*psb*A spacer). MRMODELTEST 2.0 (Nylander, 2004) and the Akaike

172   Information Criterion (AIC; Akaike, 1974) were used to decide on the best-fitting substitution

173   model for each partition.

174   MRMODELTEST2.0 results were also used for setting up Bayesian inference, performed with

175   MRBAYES 3.4b4 (Ronquist and Huelsenbeck 2003; Ronquist et al. 2012). RAxML v. 7.0.4

176   (Stamatakis et al., 2008) was used for calculating maximum likelihood bootstrap support (1000

177   replicates). Trees were edited with FIGTREE 1.3.1(Rambaut, 2014) and MESQUITE v. 2.75

178   (Maddison and Maddison, 2011). Median-joining (MJ) haplotype networks were inferred with

179   NETWORK 4.6.1.1 (http://www.fluxus-engineering.com/) for each gene region (*rbc*L, *trn*K/*mat*K,

180   *trn*H-*psb*A), treating gaps either as missing or 5th state. MJ algorithm was invoked with default

181   parameters (equal weight of transversion/transition), in order to handle large datasets and

182   multistate characters.

183   Primary data, analysis and supplementary files (S1–S3) are provided for anonymous download in

184   an online supporting archive hosted at www.palaeogrimm.org/data/Smn15_OSA.zip

185

186   **Results**

187

188   *Levels of intra- and interspecies plastome divergence in Quercus*

189   The entire dataset included 423 plastid DNA sequences (141 samples, three markers each). Table

190   2 shows that *trn*H-*psb*A was the most variable marker region (a 34-bp inversion occurring in

191   approximately 50% of the samples was not considered). The least variable region was *rbc*L, as

192   expected. No indels were found in the *rbc*L and *mat*K coding regions. The combined cpDNA

193   dataset (*trn*H-*psb*A, *trn*K/*mat*K, *rbc*L) resulted in an alignment of 2082 characters (sites), of

194   which122 were variable (thereof 72 parsimony-informative; gaps not considered). The alignment

195   had a nucleotide diversity of 0.0056 and included 74 different haplotypes of which 50 were

196   unique (restricted to a single accession). As a result, the overall haplotype diversity was high (*Hd*

197   = 0.978 ± 0.005). With gaps considered, the number of haplotypes increased to 110, of which 89

198   were unique (*Hd* = 0.9939).

199   In general, the infrageneric divergence calculated in *Quercus* is comparable to that found in other

200   genera of the Fagaceae and Betulaceae, and higher than in Juglandaceae (Table 3). All three gene

201   regions allow distinguishing the generic affinity of an oak individual; the same haplotype may be

202   shared by several or many oak species (usually within the same infrageneric group; Table 3), but

203   not with other genera of the Fagaceae.

204   At the infrageneric level in *Quercus,* minimal inter-species distances can be zero for all three

205   markers and within all infrageneric groups. Notably, maximal inter-species distances within

206   infrageneric groups of *Quercus* can reach or even exceed the level of inter-generic differentiation

207   in Fagaceae (e.g. between *Notholithocarpus, Lithocarpus, Castanopsis, Castanea, Chrysolepis*),

208   Juglandaceae and Myricaceae. The maximum intra-specific distance found in Mediterranean

209   individuals of *Quercus* Group Ilex equals the maximum inter-specific divergence found within

210   this group.

211

212   *Phylogenetic placement of Mediterranean Quercus Group Ilex plastid haplotypes*

213   Individuals of the Mediterranean species of *Quercus* Group Ilex cluster in three well supported

214   distinct clades (Fig. 1). The first clade ('Euro-Med') accommodates most accessions of *Q. ilex*

215   and *Q. coccifera*. In the second clade ('Cerris-Ilex'), accessions of *Q. ilex*, *Q. coccifera,* and one

216   of the five samples of *Q. aucheri* group together with all representatives of *Quercus* Group

217   Cerris and two Himalayan-East Asian species of Group Ilex. Sister to this clade are the three

218   representatives of the single Japanese species of Group Ilex (*Q. phillyraeoides*). In the third clade

219   (West Asia-Himalaya-East Asia; 'WAHEA') the remaining specimens of *Q. aucheri* form a

220   subclade along with the Cypriote endemic *Q. alnifolia,* and several Eastern Mediterranean *Q.*

221   *coccifera.* The second, more divergent and poorly supported subclade comprises two western

222   Himalayan species (*Q. baloot, Q. floribunda*), two individuals of Himalayan-East Asian species

223   of *Quercus* Group Ilex, and one Central China accession of a Cyclobalanopsis member (*Q.*

224   *oxyodon*) sympatric with many group Ilex oaks, including *Q. semecarpifolia*, *Q.*

225   *leucotrichophora*, *Q. floribunda* (Menitsky, 2005). In contrast to Group Ilex, all other

226   infrageneric groups show relatively high chlorotypic coherence, usually forming clades or

227   grouped within the same subtree. The actual root of the tree is obscured; representatives of

228   *Castanea*, *Castanopsis*, and *Notholithocarpus/Chrysolepis* that could be used as putative

229   outgroups are placed in different subtrees.

230

231 *Evolutionary significance of plastid haplotypes in western Mediterranean oaks of Quercus*

232 *Group Ilex*

233 The MJ network for the plastid region with the highest overall variability (*trn*H-*psb*A, only

234 length-homogenous parts considered; Fig. 2) highlights the evolutionary significance of the three

235 main haplotypes, 'Euro-Med', 'Cerris-Ilex', and 'WAHEA'. Three main clusters differ by a

236 minimum of two conserved mutations: 1) Group Quercus, Protobalanus and Lobatae ('New

237 World Oaks'); 2) individuals with 'Euro-Med' haplotypes; 3) individuals with 'Cerris-Ilex' and

238 'WAHEA' haplotypes, representatives of Group Cerris and East Asian species of Group Ilex and

239 Group Cyclobalanopsis ('Old World Oaks'). In general, haplotypes (File S3 includes MJ-

240 networks for the other three regions, *rbc*L gene, *mat*K gene, 3' *trn*K intron,) found in the western

241 Eurasian members of Group Ilex represent unique or ancestral variants. Unique haplotypes of

242 Group Cerris are directly derived from the Group Ilex or shared 'Cerris-Ilex' haplotypes.

243 Haplotypes of Group Cyclobalanopsis are identical to or can be derived from East Asian

244 members of Group Ilex. The graphs further highlight a close relationship of haplotypes of

245 *Chrysolepis* and *Notholithocarpus* with those of the 'New World' oaks; those of *Castanea* and

246 *Castanopsis* can be derived from the 'Old World' oaks basic type.

247 Figures 1 and 2 clearly illustrate that differentiation in the plastid sequences of *Quercus* (and

248 related Fagaceae) is independent from the formation or at least the genetic homogenization

249 (lineage sorting) of the modern clades.

250

251 *Phylogeographic structure in Quercus Group Ilex*

252 Haplotypes forming the 'Euro-Med', 'Cerris-Ilex' and 'WAHEA' lineages are geographically

253 sorted. The phylogenetically isolated 'Euro-Med' haplotypes are encountered in the western

254   Mediterranean populations of *Q. ilex* and *Q. coccifera* (North Africa, Iberia, Southern France,

255   Italy), along the Adriatic coast and into Central Greece (Fig. 3). Also included here are isolated

256   populations of *Q. ilex* from Crete and the southern Black Sea coast. 'Cerris-Ilex' and 'WAHEA'

257   haplotypes are confined to the eastern Mediterranean region. 'Cerris-Ilex' haplotypes are found

258   in the Aegean region (*Q. ilex*, *Q. coccifera* and *Q. aucheri* individuals) and replaced by

259   'WAHEA' haplotypes (*Q. coccifera*, *Q. aucheri*, *Q. alnifolia*) in south-western Turkey and

260   extending to the east (Levant; Fig. 3).The 'Cerris-Ilex' type is also found in the *Q. coccifera*

261   individual from northern Turkey, representing the north-easternmost population of this species.

262

263   **Discussion**

264

265   Despite resolution issues due to weak signals regarding intergeneric relationships, all data on

266   Fagaceae show a deep incongruence between nuclear and plastid data. Nuclear phylogenies

267   unambiguously point towards an inclusive common origin of *all* oaks, i.e. a monophyletic (s.

268   str.) genus *Quercus* (Oh and Manos, 2008; Denk and Grimm, 2010; Hubert et al., 2014). At the

269   same time plastid data repeatedly failed to resolve all oaks as one clade (Manos et al., 2008; this

270   study). Instead, a split emerges (with varying support) between the North American

271   *Notholithocarpus* and North American/northern temperate clade of oaks, the 'New World Oaks',

272   and the Eurasian *Castanea, Castanopsis* and oak lineages, the 'Old World oak' clade; an

273   observation that holds independent from the exact placement of the root in a plastid tree. If we

274   accept the monophyly of the genus *Quercus*, which is backed also by morphology and evidence

275   from the fossil record, haplotypes of *Castanea/Castanopsis* and *Notholithocarpus* that group

276   with the 'New World' and 'Old World' oaks, respectively, can hence only be the result of

277 incomplete lineage sorting during the formation of the modern genera. In addition, the plastid

278 genepool of the earliest oaks must have shown a genetic gradient that was to some extent caused

279 by biogeographic patterns. Although it is impossible to pinpoint the place of origin of oaks, it is

280 clear that the ancient oaks must have been widespread, allowing them to pick up and propagate

281 geographic signatures inherited from their common ancestors with *Notholithocarpus, Castanea*

282 and *Castanopsis*. Geographic signatures in two Mediterranean species of Group Ilex, *Q. ilex* and

283 *Q. coccifera*, are discussed in the following.

284

285 *Major trends of plastome differentiation*

286 The overall low genetic intra- and intertaxonomic (intrageneric lineages, genera) distances

287 suggest low evolutionary rates for the chloroplast genomes of Fagales, at least at the examined

288 loci. However, the data coverage is far from sufficient for most genera and families to precisely

289 evaluate the plastome potential variation within this plant group. In Fagaceae, a comparison with

290 the (genetically) more diverse Nothofagaceae and Betulaceae families reveals that haplotype

291 variation at the *trn*H-*psb*A locus can be sufficiently high to allow inferences at the

292 phylogeographic and systematic level (see Premoli et al., 2012; Grimm and Renner, 2013). In

293 analogy, haplotypes of intrageneric lineages of *Quercus* differ in this marker. Furthermore, a

294 geographic pattern is evident for the most widely sampled groups. Groups Ilex, Lobatae and

295 Quercus appeared the most variable, whereas Group Cerris exhibited the lowest differentiation

296 rates. Interestingly, interclade differentiation among all *Quercus* groups equalled or exceeded the

297 values scored by the four outgroup genera (*Castanea, Castanopsis, Notholithocarpus* and

298 *Chrysolepis*). As a consequence, the outgroup taxa appear scattered across the tree, rather than

299 being culled in a distinct subtree, rendering the plastome of *Quercus* 'non-monophyletic'.

300   Outgroup selection as a potential source of topological ambiguity was already pointed out by

301   Hubert et al. (2014; 108 oak taxa, eight nuclear markers). Ambiguous relationships among this

302   group of genera independently of the strength of the obtained phylogenetic signal were also

303   suggested by a recent study on Fagales (based on molecular, fossil and reproductive syndromes

304   analyses), which resolved the majority of inter-generic relationships in each family except in the

305   Quercoideae group making *Castanopsis* and *Quercus* non-monophyletic (Xiang et al., 2014).

306

307   *Plastid phylogeny does not conform to the current synopsis of oak evolution*

308   Figure 4 highlights the incongruence of the plastid genealogy tree with the current understanding

309   of the evolution of Fagaceae and oaks based on molecular sequence data from the non-coding

310   nuclear gene regions (Manos et al., 2008; Denk and Grimm, 2010), a recent time-calibrated

311   nuclear phylogeny of oaks (Hubert et al. 2014), and the fossil record of modern lineages as

312   documented by pollen investigated under the scanning-electron microscope (Grímsson et al.,

313   2015; see also Denk and Grimm, 2009). Two evolutionary mechanisms (incomplete lineage

314   sorting, reticulation) may account for the observed, highly complex pattern.

315   Firstly, speciation processes in *Quercus* do not immediately leave imprints on the plastome (e.g.

316   Neophytou et al., 2010; Cavender-Bares et al., 2011) as also well documented for *Nothofagus*

317   (Acosta and Premoli, 2010; Premoli et al., 2012). Low mutation rate and long generation time

318   can contribute to slow evolutionary rates and incomplete lineage sorting of organellar genomes

319   (Cavender-Bares et al., 2015; Besnard et al., 2007). In addition, reiterated extinctions and re-

320   colonisations involving bottlenecks, genetic drift, and founder effects may cause random fixation

321   of haplotypes, increasing the probability for retaining ancestral traits. Oaks in general, and

322   especially the Mediterranean taxa, are also characterised by a marked resprouting ability in

323   response to disturbances of varying frequencies and intensities, including the action of man,

324   herds, and wild fires (Barbero et al., 1990). This could also have contributed to clonally preserve

325   and transmit ancestral plastid lineages (maternally inherited) during multiple and reiterated

326   unfavourable conditions since the origin of the Mediterranean region (Blondel and Aronson,

327   1999). At the same time, different environmental selection and adaptation, large population sizes,

328   and long distance pollen dispersal, might have homogenised the nuclear genomes in local

329   populations of a species but not their organelle genomes.

330   Secondly, Fagaceae lineages are susceptible to hybridisation and introgression (Arnold, 2006).

331   This may lead to the formation of morphologically unambiguous individuals of a species with

332   plastid signatures of another (Whittemore and Schaal, 1991; Petit et al., 2004). There is

333   increasing evidence for local introgression in oak communities with morphologically distinct

334   species in the case of European white oaks (Group Quercus; *Q. robur*, *Q. petraea*, *Q. pyrenaica*,

335   *Q. pubescens*, *Q. frainetto*; Curtu et al., 2007; Valbuena-Carabaña et al., 2007; Lepais et al.,

336   2009), as well as in members of *Quercus* subsection *Virentes,* a subgroup of Group Quercus, in

337   North America (Cavender-Bares et al., 2015), and across a wide range of Group Lobatae (Dodd

338   and Afzal-Rafii, 2004; Peñaloza-Ramírez et al., 2010; Moran et al., 2012; Valencia-Cuevas et al..

339   2015). In our focal group, hybrids and different levels of genetic introgression among

340   morphologically pure individuals were molecularly documented in *Q. ilex*/*Q. coccifera* (Ortego

341   and Bonal, 2010) and, to a lesser extent, in *Q. coccifera*/*Q. alnifolia* (Neophytou et al., 2011).

342   Also, the potential for inter-group hybridisation was experimentally demonstrated for *Q. ilex* and

343   *Q. robur* (Group Quercus; Schnitzler et al., 2004), and natural introgression in *Q. ilex*/*Q. suber*

344   was identified in Southern France (Mir et al., 2009) and Iberia (Burgarella et al., 2009).

345   Therefore, it is possible that ancient hybridization and introgression, favoured by the well-known

346   sexual promiscuity between closely related taxa and their ability to disperse pollen over long

347   distances, obscure the true evolutionary origin of an oak species or entire lineage.

348   In the Mediterranean, the dramatic geological and ecological changes during the Neogene

349   (Blondel and Aronson, 1999) likely caused extinction, re-colonisation, range fragmentation and

350   hybridisation linked to secondary contact, especially when species were still young and

351   reproductive barriers likely weaker than today. Taken together, incomplete sorting of ancestral

352   traits and introgression of haplotypes thus appear highly likely mechanisms to decrease inter-

353   species plastid differentiation while at the same time increasing intra-species variation. This

354   provides an explanation why the widespread, morphologically and genetically (at the nuclear

355   level) distinct species *Q. coccifera* and *Q. ilex* have accumulated three distinct main plastid

356   haplotypes, which we discuss in the following.

357

358   *Polyphyletic clues in Quercus Group Ilex*

359   The most striking finding of this study is the plastid polyphyly and a clear geographic pattern

360   displayed by a group of Mediterranean Ilex oaks (*Q. ilex*, *Q. coccifera* and *Q. aucheri*).

361   Phylogeographic patterns reflecting distant vicariant events and a complex history of range

362   expansions and contractions have been previously inferred for other Mediterranean woody

363   species (Besnard et al., 2007; Desamore et al., 2011; Migliore et al., 2012; Chen et al., 2014).

364   Mediterranean *Laurus* (Rodriguez-Sanchez et al., 2009), for instance, also comprises three

365   plastid haplotype lineages roughly corresponding to biogeographic patterns as seen in the

366   Mediterranean Ilex oaks: (1) an eastern lineage in Turkey and the Near East, (2) a second one in

367   the Aegean region, and (3) a probably ancestral lineage of central and western Mediterranean

368   populations. The importance of the Mediterranean basin in shaping the intraspecific divergence

369 of Tertiary plant species is therefore well-acknowledged (Nieto Feliner, 2014). Unfortunately,

370 none of the above studies investigated multiple species, leaving the question whether the

371 observed divergence is restricted to a single species or shared by several species as in the case of

372 *Quercus* Group Ilex.

373 Extensive sampling is more likely to document hidden polyphyly (Wiens and Servedio, 2000);

374 besides a weak phylogenetic signal resulting in poorly resolved and potentially inaccurate gene

375 trees (countered here by using median-joining and bipartition networks), incomplete sorting of

376 ancestral lineages and introgression represent very general sources of polyphyly. However,

377 distinguishing the separate effects of these mechanisms can be very difficult in the absence of

378 nuclear markers and (palaeo-)geography as complementing information (Funk and Omland,

379 2003). As a general rule, the phylogenetically basal position of polyphyly usually hints at

380 retained ancestral polymorphism, while recently introgressed haplotypes may assume a highly

381 derived position in a gene tree. At the same time, incomplete sorting is not predicted to promote

382 the geographic proximity of interspecifically shared haplotypes that may be seen under local

383 introgression (Hare and Avise, 1998, Masta et al., 2002).

384 In our reconstruction, Group Ilex oaks appear to conform to both instances. Nuclear data

385 covering the entire range of *Q. ilex* and *Q. coccifera* in the Mediterranean region unambiguously

386 resolved the two species as close, but mutually monophyletic sister taxa (Denk and Grimm,

387 2010). In the absence of nucleome data for all here included individuals, it is impossible to infer

388 to which degree introgression and incomplete lineage sorting contributed to the plastid gene pool

389 of the Mediterranean species of Group Ilex. Nevertheless, the most straightforward explanation

390 for the observed scenario would be a combined effect of the two mechanisms: asymmetrical

391 introgression of ancestral haplotypes resulting in local genetic clusters decoupled from

392   taxonomic boundaries, in which plastome accessions of species or species complexes may form

393   grades or multiple clades in phylogenetic trees, thus appearing polyphyletic (e.g. Rieseberg and

394   Soltis, 1991; Whittemore and Schaal, 1991).

395   As modelled by Excoffier et al. (2009), interspecific interactions during historical range

396   fluctuations can profoundly affect the observed phylogeographic patterns, and manifest as

397   paraphyly or reticulation. In fact, most range expansions do not occur in completely uninhabited

398   areas, and interbreeding between local and an expanding (invasive) species with subsequent

399   asymmetrical introgression can develop also in absence of selection (Lepais et al., 2009). Plastid

400   haplotypes referring to the original ('lost') species are indeed likely to persist over long

401   evolutionary periods, and may still be found in the invading species. Noteworthy, environmental

402   changes and disturbance of local communities have been shown to increase hybridisation rates

403   (Lagache et al., 2013), hence, the potential for widespread, imbalanced introgression. In Group

404   Ilex oaks, the interspecific capture of plastids among sexually incompletely isolated species

405   likely occurred on the geological timescale, concealing the species relationships at various stages

406   in the history of the genus. In a comprehensive study of the genus *Ilex* (Manen et al., 2010), the

407   high incongruence between a taxonomically compatible nuclear gene tree and a geographically

408   structured plastid tree was explained with extensive extinctions between the Cretaceous and

409   Miocene and multiple hybridization and introgression events between distantly related lineages.

410   This has been documented also for *Platanus* (Grimm and Denk, 2010) and more recently

411   suggested for the evergreen white oaks of *Quercus* subsection Virentes (Eaton et al., 2015).

412   Similar ancient lateral transfers have been also inferred to explain the paraphyly of the

413   maternally inherited mtDNA of *Picea* (Bouillè et al., 2011) and *Pinus* (Tsutsui et al., 2009). As

414   noted above, introgression has been widely demonstrated across a wide range of oaks (Dodd and

415   Afzal-Rafii, 2004; Curtu et al., 2007; Valbuena-Carabaña et al., 2007; Burgarella et al., 2009;

416   Lepais et al., 2009; Mir et al., 2009; Ortego and Bonal, 2010; Peñaloza-Ramírez et al., 2010;

417   Neophytou et al., 2011; Moran et al., 2012; Valencia-Cuevas et al., 2015) and suggested as

418   explanation for the wide haplotype sharing revealed by *Q. suber* and *Q. cerris* in the Italian

419   peninsula (Magri et al., 2007).

420

421   *Temporal and spatial framework of plastome evolution*

422   The three distinct plastid haplotypes observed in modern Mediterranean members of *Quercus*

423   Group Ilex may reflect three radiation phases (range extensions), followed by range disruptions

424   and isolation of plastome lineages within the 'Old World Clade' of *Quercus*. Considering the

425   high diversity of haplotypes in Group Ilex as compared to other major oak lineages (or other

426   genera in the Fagales; see Table 3; Table S2) it can be assumed that the geographical disruptions

427   in the plastome of the ancestors of Group Ilex and interacting lineages predate the manifestation

428   of modern taxa (species and infrageneric groups; Fig. 1). Haplotypes shared between members of

429   Group Ilex and its sister lineages Group Cerris and Group Cyclobalanopsis may indicate

430   common (geographic) origin or may be the result of secondary contact and unidirectional

431   introgression.

432   Independent from the position of the plastid root, the divergence of the 'Euro-Med' haplotype

433   must have coincided with the initial differentiation in *Quercus* (Fig. 1). Oaks had achieved a

434   wide northern hemispheric range by the Eocene. Unequivocal fossils are known from high

435   latitudes (North America, Greenland, North Europe; Crepet and Nixon ,1989; Manchester, 1994;

436   Grímsson et al., 2015) and mid latitudes (Central Europe, South East Asia; Kvaček and Walther,

437   1989; Hofmann, 2010). All major lineages of oaks were established by the end of the Eocene, ca.

438   35 Ma, as evidenced by the fossil record and molecular dating using eight nuclear gene regions

439   (Bouchal et al., 2014: fig. 14; Hubert et al., 2014; Grímsson et al., 2015). During this time, one

440   fraction of oaks must have been geographically and reproductively isolated which would have

441   caused a major split in the plastid gene pool (Fig. 1). Today, the 'Euro-Med' haplotype is the

442   only one exclusively shared by just two, but widespread Mediterranean species of *Quercus*

443   Group Ilex, *Q. ilex* and *Q. coccifera*. Two evolutionary scenarios can explain the establishment

444   of this haplotype in *Q. ilex*-*Q. coccifera* (Fig. 5): (*i*) The 'Euro-Med' haplotype is the remnant of

445   an extinct oak lineage that was intrograded (invaded) and consumed by members of Group Ilex;

446   under this scenario Group Ilex would have migrated into Europe at some point prior to the

447   Miocene where it came into contact with this extinct oak lineage. (*ii*) The 'Euro-Med' haplotype

448   represents the original plastome of Group Ilex; under this scenario, the first split within the

449   modern 'Old World clade' would have been between a western Group Ilex and an eastern Group

450   Cyclobalanopsis (Fig. 5). Fossil evidence and available phylogenies (discussed in the following)

451   lend high credibility to scenario (*i*) as the most plausible explanation.

452   The 'Cerris-Ilex' haplotype is shared between all species of *Quercus* Group Cerris (western

453   Eurasian and East Asian), East Mediterranean (Aegean) individuals and two East Asian species

454   of Group Ilex. This is in agreement with Denk and Grimm (2009) who suggested that *Quercus*

455   Group Cerris evolved from Group Ilex by budding (a hypothesis further confirmed by the 8-

456   nuclear gene data set used by Hubert et al., 2014), and the low support for a Group Ilex clade in

457   an all-Fagaceae (excluding *Fagus*) tree based on over 1000 nuclear ITS sequences (Denk and

458   Grimm, 2010). Hubert et al. (2014) inferred a Miocene age for this budding event, which

459   corresponds to the earliest unequivocal fossil of *Quercus* Group Cerris (Kmenta, 2011) and is

460   younger than the earliest definite fossil record of *Quercus* Group Ilex in Europe (early

461 Oligocene, Cospuden; Denk et al., 2012). Also, dispersed pollen from the Paleogene

462 Changchang Formation, Hainan (Hofmann, 2010), resembles both *Quercus* Group Ilex and

463 Group Cyclobalanopsis; the age of this formation is considered late early to early late Eocene

464 (Lei et al., 1992). The most closely related haplotype to the 'Cerris-Ilex' haplotype is

465 encountered in the widespread East Asian *Q. phillyraeoides*, the only species of Group Ilex

466 extending to Japan (the East Asian members of Group Cerris have a much wider range in north-

467 eastern Asia; Menitsky, 2005). Regarding its phylogenetic position, the emergence of the

468 'Cerris-Ilex' haplotype appears linked with a major taxonomic sorting event in Eurasian

469 Fagaceae, resulting in distinct haplotypes restricted to genera and intrageneric groups of *Quercus*

470 (Fig. 1). Based on the palaeobotanical record, these lineages (*Castanopsis*, *Castanea*, *Quercus*

471 Group Ilex, *Quercus* Group Cyclobalanopsis) were well established at least by the Eocene (Table

472 4, Fig. 5; Grímsson et al., 2015); a deep divergence is reflected by their distinctly different

473 nuclear genomes (Oh and Manos, 2008; Denk and Grimm, 2010; Hubert et al., 2014). Two

474 evolutionary scenarios can explain the occurrence of the 'Cerris-Ilex' haplotype in Aegean

475 individuals of *Q. ilex* and *Q. coccifera* and the westernmost *Q. aucheri*: (*i*) Group Cerris evolved

476 in western Eurasia/Himalaya from an (extinct) subtropical to temperate sublineage of Group Ilex,

477 which left its imprint in the Aegean members of Group Ilex, and *Q. spinosa*, *Q. engleriana* and

478 *Q. phillyraeoides*; (*ii*) Group Cerris shares a common ancestry with the north-east Asian *Q.*

479 *phillyraeoides*. Under this scenario, the budding event of the group took place in north-eastern

480 Asia, from where it migrated into western Eurasia and the Aegean region; in relatively recent

481 times, Group Cerris came into contact with the Mediterranean members of Group Ilex and were

482 locally intrograded.

483   The high similarity of 'Cerris-Ilex' haplotypes lends some credibility to the second scenario.

484   Furthermore, there is evidence for current introgression and occasional hybridization of *Q. suber*

485   (Group Cerris) and *Q. ilex* in the western Mediterranean (Burgarella et al., 2009; Mir et al.,

486   2009). However, it is difficult to explain why *Q. ilex-coccifera* should only intrograde into

487   populations of Cerris oaks at a large scale in the Aegean region. Today, Group Cerris is more

488   diverse than Group Ilex in the East Mediterranean (*Q. brantii*, *Q. cerris*, *Q. ithaburensis, Q.*

489   *macrolepis*, *Q. libani*, *Q. trojana*) with some species adapted to distinctly continental climates

490   (Browicz and Zieliński, 1982; Menitsky, 2005), outside the range of Group Ilex. This diversity

491   and the vast distribution of only two species of Group Cerris in East Asia may point towards a

492   young radiation of the group. During the early and middle Miocene, Cerris did not play an

493   important role in western Eurasia. In contrast, two distinct species complexes of *Quercus* Group

494   Ilex were prominently represented in eastern Mediterranean and Parathethyan plant assemblages,

495   *Q. drymeja* Unger and *Q. mediterranea* Unger (e.g. Velitzelos et al., 2014). Intriguingly, the *Q.*

496   *drymeja* complex includes morphotypes found today in *Q. ilex* as part of its intraspecific

497   variation, and of a range of East Asian species including *Q. engleriana*. *Quercus mediterranea* is

498   the morphological equivalent of *Q. ilex*, *Q. coccifera* and a range of East Asian species including

499   *Q. spinosa*. Hence, the fossil record clearly favours a western Eurasian-Himalayan origin of

500   Group Cerris (scenario *i*).

501   The West Asian-Himalayan-East Asian (WAHEA) haplotype represents Eastern Mediterranean

502   members of *Quercus* Group Ilex and is sister to a clade comprising several Asian species of

503   Group Ilex (Himalayas to the mountains of Southeast Asia). It reflects the second radiation

504   within the Old World Clade and allies after the isolation of the 'Euro-Med' original lineage and

505   prior to the radiation and subsequent sorting within the clade comprising the Cerris-Ilex

506   haplotype (Fig. 1).The modern distribution of species with the WAHEA haplotype follows the

507   Himalayan corridor (Kitamura, 1955; Güner and Denk, 2012). The Himalayan corridor is a

508   narrow band along the southern flanks of the Himalaya with a subtropical to temperate climate

509   (*Cwa, Cwb*; Peel et al. 2007) providing a refuge for plants that were more widespread before the

510   Himalayan uplift. In addition to *Quercus* Group Ilex (Zhou, 1992; Velitzelos et al., 2014),

511   prominent relic taxa include species of *Acer*, *Aesculus*, *Cedrus*, *Cotinus*, *Juglans*, *Platanus*, and

512   *Rhododendron* among others. The 'WAHEA' haplotype represents the western counterpart to the

513   haplotype lineage found in East Asian species of Group Ilex and Cyclobalanopsis. The relic *Q.*

514   *alnifolia*, today preserved only in the mid-montane region of Cyprus (Mt. Troodos), would be a

515   witness of this expansion (Menitsky, 2005).

516

517   *Towards an integrated biogeography of oaks*

518   Plant biogeographic studies at the genus level have commonly relied on few to many chloroplast

519   markers and a single or very few accessions per taxon. In the case of woody angiosperms with a

520   subtropical to temperate distribution such as for example Nothofagaceae (Svenson et al., 2001;

521   Knapp et al., 2005), *Rhus* (Yi et al., 2004), *Cornus* (Xiang et al., 2005), *Carpinus* (Yoo and Wen,

522   2007), *Castanea* (Lang et al., 2007), *Juglans* (Aradhya et al. 2007), and *Carya* (Zhang et al.

523   2013a), such an approach runs the risk of capturing but a limited aspect of the evolutionary

524   history of the focal group. Mere combination with e.g. nuclear ITS data can be problematic, too

525   (compare data shown here with data provided by Denk and Grimm, 2010, on western Eurasian

526   members of Group Ilex). The decoupled evolutionary signals in plastomes and the

527   nucleome/morphology as documented for *Nothofagus* (Acosta and Premoli, 2010; Premoli et al.,

528   2012) and *Quercus* Group Ilex (this study) suggests that the traditional placeholder sampling

529   strategy is not ideal. Signals from few-marker/many-samples data sets are likely to be complex

530   or even puzzling (Figs 1–4), but at the same time provide entirely new perspectives on plant

531   evolution worth exploring. For *Quercus* Group Ilex, our pilot study focussing on Mediterranean

532   species reveals a crucial aspect of oak evolution not seen in the combined nuclear,

533   morphological, and fossil data: large-scale introgression and incomplete lineage sorting among

534   ancestral lineages of modern major groups and species. The new data corroborate hypotheses

535   that Group Cerris evolved ('budded') relatively recent from Group Ilex (over 600 ITS and over

536   900 5S-IGS accessions covering all western Eurasian oak species, Denk and Grimm, 2010;

537   signal from six single-copy nuclear regions, Hubert et al., 2014). *Quercus* Group Cerris probably

538   evolved in western Eurasia and the Himalayas when the then chiefly subtropical low latitude

539   Group Ilex radiated into temperate niches. Within modern members of Group Cerris, a wide

540   spectrum of leaf traits is found from pseudo-evergreen in *Q. suber*, to semi-evergreen in *Q.*

541   *brantii*, *Q. ithaburensis*, *Q. trojana* (partly) and fully deciduous in *Q. acutissima*, *Q.*

542   *castaneifolia*, *Q. cerris*, *Q. libani* and *Q. variabilis*. The conspicuous plastid diversity in the

543   Mediterranean species of Group Ilex and the lineage in general (Figs 1, 2; Table 2) reflects the

544   highly complex geographical history of this group. The 'Euro-Med' haplotype evidences an

545   initial phase of west-east differentiation in low-latitude Eurasian oaks, the 'Old World Clade',

546   probably triggered by the complex topography within its potential range essentially since the

547   Eocene (Fig. 5). During its evolution, Group Ilex was continuously affected by range disruptions

548   caused by tectonic activity south of the Paratethys linked to the collision of Africa and the Indian

549   subcontinent with Eurasia (Fig. 6); progressive rarefaction of the original haplotypes and the

550   occurrence of (repeated) invasion and introgression events that left imprints in the plastome even

551   within the same species is highly likely.

552 Although decoupled from taxonomy, the plastid phylogeny provides important, independent

553 information on the geographic differentiation of *Quercus* prior to the formation of modern

554 species/species groups. The major split within oaks has traditionally been considered between

555 'New World' and 'Old World' oaks (Manos and Stanford, 2001) because of the current

556 distribution of the major lineages of oaks. The plastid data presented here strongly suggest that

557 the early evolution of oaks instead was geographically bound to high latitude Arctic regions and

558 to low latitude subtropical regions (Fig. 5). The high latitude lineages remained genetically

559 homogeneous in the nucleome, but also in the plastome to some degree. Continuous circum-

560 polar distribution prevented pronounced genetic drift in the high latitude lineage, which became

561 the 'New World Clade', and explains low genetic differentiation in deciduous high and mid

562 latitude white oaks until today (Denk and Grimm, 2010). At the same time, the Atlantic, the

563 proto-Mediterranean, and the Paratethys isolated the Eurasian low latitude lineage.

564 Our data should only be viewed as a first step towards a more complete understanding of the

565 biogeography and evolution of oaks. The next step would be to map the plastid variation of

566 *Quercus* Group Ilex across its entire range by sampling multiple stands of the Himalayan and

567 East Asian species to characterise the geographic and taxonomic ranges of the various plastid

568 lineages.

569

570 **Conclusion**

571

572 Taken all evidence together, the first major split of oaks, consistently found in all molecular

573 phylogenies, would have been into two clades. A northern, high-latitude clade leading to white,

574 red and golden-cup oaks, which evolved and diversified in the tectonically quiet parts of the

575   Northern Hemisphere. Its counterpart was a southern, mid-latitude clade made up by Group Ilex

576   and Group Cyclobalanopsis (and later by Group Cerris), in the southern part of Eurasia, and

577   perhaps western North America. Both *Quercus* Group Cyclobalanopsis and Group Ilex were

578   present in southern Eurasia, close to the shores of the Tethys, and western North America by the

579   middle Eocene. The outlined history of further steps in the southern clades appear to consist of

580   two radiations: one lead to the Group Cerris clade (early Miocene) with migration westwards

581   along the southern slopes of the Himalayas, and eastwards over China and Japan. The highly

582   coherent 'Cerris-Ilex' haplotypes are key witnesses of this event indicating that westwards and

583   eastwards migration and radiation of the monophyletic Group Cerris may have been relatively

584   recent. A putative radiation centre is the East-Mediterranean Paratethys region, in which

585   members of Group Ilex and Cerris coexist *and* share highly similar to identical haplotypes. The

586   second radiation is likely to have occurred in the Miocene, too. From a Himalayan Group Ilex

587   stock, the clade with 'WAHEA' haplotype expanded towards the eastern Mediterranean basin.

588   Both the ancient western Eurasian clade, now extinct but evidenced by the 'Euro-Med'

589   haplotype, and the originally Himalayan clade had been invaded by the late Neogene by the

590   direct ancestors of today's *Q. coccifera* and *Q. ilex*. Modern forms of these two Mediterranean

591   oaks (and their two closest relatives, *Q. aucheri* and *Q. alnifolia*) would have preserved the

592   ancestral haplotypes, keeping strong geographic indications of those events along with the

593   formation of the Mediterranean region. Moreover, the Himalayan uplift coincided with the

594   development of the modern monsoon climate; hence adaptations or exaptation to phases of

595   drought may have played an important role at least since 15 million years ago (Wang and Wu

596   2015). Our reconstruction is still speculative, but consistent with (i) all fossil observations, (ii)

597   known molecular phylogenies of plastid haplotypes, and (iii) known processes shaping the

598   diversity of plastid haplotypes by speciation and introgression, linked with invasions. Further

599   detailed studies, especially in the Himalayan and Chinese regions, are needed to validate or

600   disprove this scenario.

601   Because of similar strong correlation between plastome differentiation and geographic

602   distribution at the population level and the species/genus level, similar or identical plastid

603   haplotypes typically shared between co-occurring and often distantly related taxa, polyphyletic

604   signals and reproductive biology, the same processes could have likely played a key role in the

605   evolutionary history of other Fagaceae (e.g., *Fagus*, *Castanea*, *Castanopsis*). Broadening the

606   sampling efforts of phylogenetic analyses of the plastome could help decipher the speciation

607   history of these genera. At the same time, extended nucleome investigations will be obviously

608   necessary to definitely assess a clear molecular phylogeny of Fagaceae.

609 **Acknowledgments**

617

618

619

620 **References**

621 **Acosta MC, Premoli AC.** 2010. Evidence of chloroplast capture in South American *Nothofagus*

622   (subgenus *Nothofagus*, Nothofagaceae). *Molecular Phylogenetics and Evolution* **54**: 235–

623   242.

624 **Akaike H. 1974**. A new look at the statistical model identification. *IEEE Transactions on*

625   *Automatic Control* **19**: 716–723.

626 **Aradhya MK, Potter D, Gao F, Simon CJ.** 2007. Molecular phylogeny of *Juglans*

627   (Juglandaceae): a biogeographic perspective. *Tree Genetics and Genomes* **3**: 363–378.

628 **Arnold ML** (2006) *Evolution through genetic exchange,* vol 3. Oxford: Oxford University Press.

629 **Barbero M, Bonin G, Loisel R, Quezel P.** 1990. Changes and Disturbances of Forest

630   Ecosystems Caused by Human Activities in the Western Part of the Mediterranean Basin.

631   *Vegetatio* **87** : 151-173.

632 **Belahbib N, Pemonge MH, Ouassou A, Sbay H, Kremer A, Petit R.** 2001. Frequent

633   cytoplasmic exchanges between oak species that are not closely related: *Quercus suber*

634   and *Q. ilex* in Morocco. *Molecular Ecology* **10**: 2003-2012.

635 **Besnard G, Rubio de Casas R, Vargars P**. 2007. Plastid and nuclear DNA polymorphism

636   reveals historical processes of isolation and reticulation in the olive tree complex (*Olea*

637   *europaea*). *Journal of Biogeography* **34**: 736-752.

638 **Blondel J, Aronson J.** 1999. *Biology and wildlife of the Mediterranean region*. Oxford: Oxford

639   University Press.

640 **Bouchal J, Zetter R, Grímsson F, Denk T.** 2014. Evolutionary trends and ecological

641   differentiation in early Cenozoic Fagaceae of western North America. *Americal Journal*

642   *of Botany* **101**: 1–18.

643   **Bouillé M, Senneville S, Bousquet J.** 2011. Discordant mtDNA and cpDNA phylogenies

644   indicate geographic speciation and reticulation as driving factors for the diversification of

645   the genus *Picea*. *Tree Genetics and Genomes* 7: 469–484.

646   **Browicz K, Zieliński J.** 1982. *Chorology of trees and shrubs in South-West Asia and adjacent*

647   *regions*, Vol. 1. Warsaw: Polish Scientific Publishers

648   **Burgarella C, Lorenzo Z, Jabbour-Zahab R, Lumaret R, Guichoux E, Petit R, Soto A, Gil**

649   **L.** 2009. Detection of hybrids in nature: application to oaks (*Quercus suber* and *Q. ilex*).

650   *Heredity* **102**: 442-452

651   **Cannon CH, Manos PS.** 2003. Phylogeography of the Southeast Asian stone oaks

652   (*Lithocarpus*). *Journal of Biogeography* **30**: 211-226

653   **Cavender-Bares J, Gonzalez-Rodriguez A, Eaton DAR, Hipp AAL, Beulke A, Manos PS.**

654   2015. Phylogeny and biogeography of the American live oaks (*Quercus* subsection

655   Virentes): A genomic and population genetics approach. *Molecular Ecology* (in press)

656   **Cavender-Bares J, Gonzalez-Rodriguez A, Pahlich A, Koehler K, Deacon N.** 2011.

657   Phylogeography and climatic niche evolution in live oaks (*Quercus* series Virentes) from

658   the tropics to the temperate zone. *Journal of Biogeography* **38**: 962–981

659   **Chen C, Qi ZC, Xu XH, Comes HP, Koch MA, Jin XJ, Fu CX, Qiu YX.** 2014.

660   Understanding the formation of Mediterranean-African-Asian disjunctions: evidence for

661   Miocene climate-driven vicariance and recent long-distance dispersal in the Tertiary

662   relict *Smilax aspera* (Smilacaceae). *New Phytologist* **204**: 243-255.

663   **Crepet WL, Nixon KC.** 1989. Earliest megafossil evidence of Fagaceae: phylogenetic and

664   biogeographic implications. *American Journal of Botany* **76**: 842–855

665   **Curtu AL, Gailing O, Finkeldey R.** 2007. Evidence for hybridization and introgression within

666   a species-rich oak (*Quercus* spp.) community. *Bmc Evolutionary Biology* **7**: 218

667   **Denk T, Grimm GW.** 2009. Significance of pollen characteristics for infrageneric classification

668   and phylogeny in *Quercus* (Fagaceae). *International Journal of Plant Sciences* **170**: 926–

669   940

670   **Denk T, Grimm GW.** 2010. The oaks of western Eurasia: traditional classifications and

671   evidence from two nuclear markers. *Taxon* **59**: 351-366

672   **Denk T, Grímsson F, Zetter R.** 2012. Fagaceae from the early Oligocene of Central Europe:

673   persisting New World and emerging Old World biogeographic links. *Review of*

674   *Palaeobotany and Palynology* **169**: 7–20

675   **Desamore A, Laenen B, Devos N, Popp M, Gonzalez-Mancebo JM, Carine MA,**

676   **Vanderpoorten A.** 2011. Out of Africa: north-westwards Pleistocene expansions of the

677   heather *Erica arborea*. *Journal of Biogeography* **38**:164-176

678   **Dodd RS, Afzal-Rafii Z.** 2004. Selection and dispersal in a multispecies oak hybrid zone.

679   *Evolution* **58**: 261-269

680   **Eaton DAR, Gonzalez-Rodriguez A, Hipp AL, Cavender-Bares J.** 2015. *Introgression*

681   *obscures and reveals historical relationships among the American live oaks*. bioRxiv

682   DOI: 10.1101/016238.

683   **Excoffier L, Foll M, Petit RJ.** 2009. Genetic Consequences of Range Expansions. *Annual*

684   *Review of Ecology, Evolution and Systematics* **40**: 481-501

685   **Flora of China Editorial Committee.** 1999. *Flora of China, Cycadaceae through Fagaceae*,

686   Vol. 4. St. Louis: Missouri Botanical Garden

687 **Flora of North America Editorial Committee.** 1997. *Flora of North America north of Mexico,*

688   Vol. 3. New York: Oxford University Press

689 **Funk D, Omland K.** 2003. Species-level paraphyly and polyphyly: frequency, causes and

690   consequences, with insights from animal mitochondrial DNA. *Annual Reviews of*

691   *Ecology, Evolution and Systematics* **34**: 397–423

692 **Fujii N, Tomaru N, Okuyama K, Koike T, Mikami T, Ueda K.** 2002. Chloroplast DNA

693   phylogeography of Fagus crenata (Fagaceae) in Japan. *Plant Systematics and Evolution*

694   **232**: 21-33

695 **Göker M, García-Blázquez G, Voglmayr H, Tellería MT, Martín MP.** 2009. Molecular

696   taxonomy of phytopathogenic fungi: a case study in *Peronospora*. *Plos One* **4**: e6319

697 **Göker M, Grimm GW.** 2008. General functions to transform associate data to host data, and

698   their use in phylogenetic inference from sequences with intra-individual variability. *Bmc*

699   *Evolutionay Biology* **8**: 86

700 **Govaerts R, Frodin DG.** 1998. *World checklist and bibliography of Fagales (Betulaceae,*

701   *Corylaceae, Fagaceae and Ticodendraceae)*. Royal Botanic Gardens, Kew

702 **Grimm GW, Denk T.** 2010. The reticulate origin of modern plane trees (*Platanus*, Platanaceae)

703   - a nuclear marker puzzle. *Taxon* **59**: 134-147.

704 **Grimm GW, Renner SS.** 2013. Harvesting Betulaceae sequences from GenBank to generate a

705   new chronogram for the family. *Botanical Journal of the Linnean Society* **172**: 465–477

706 **Grímsson F, Zetter R, Grimm GW, Krarup Pedersen G, Pedersen AK, Denk** T. 2015.

707   Fagaceae pollen from the early Cenozoic of West Greenland: revisiting Engler's and

708   Chaney's Arcto-Tertiary hypotheses. *Plant Systematics and Evolution* **301**: 809–832

709   **Gugger PF, Cavender-Bares J.** 2013. Molecular and morphological support for a Florida origin

710   of the Cuban oak. *Journal of Biogeography* **40**: 632-645

711   **Güner TH, Denk T.** 2012. The genus *Mahonia* in the Miocene of Turkey: taxonomy and

712   biogeographic implications. *Review of Palaeobotany and Palynology***175**: 32–46

713   **Hare MP, Avise JC.** 1998. Population structure in the American oyster as inferred by nuclear

714   gene genealogies. *Molecular Biology and Evolution* **15**: 119–28

715   **Hipp AL, Eaton DAR, Cavender-Bares J, Fitzek E, Nipper R, Manos PS.** 2014. A

716   framework phylogeny of the American oak clade based on sequenced RAD data. *Plos*

717   *One* **9**: e93975

718   **Hofmann C-C.** 2010. Microstructure of Fagaceae pollen from Austria (Paleocene/Eocene

719   boundary) and Hainan Island (?middle Eocene). Paper presented at the 8th European

720   Palaeobotany-Palynology Conference, 6-19 July, Budapest, Hungary

721   **Hubert F, Grimm GW, Jousselin E, Berry V, Franc A, Kremer A.** 2014. Multiple nuclear

722   genes stabilize the phylogenetic backbone of the genus *Quercus*. *Systematics and*

723   *Biodiversity* **12**: 405–423.

724   **Huzioka K, Takahasi E.** 1970. *The Eocene flora of the Ube coal-field, southwest Honshu,*

725   *Japan*. Akita University, Mining College, Japan

726   **Kanno M, Yokoyama J, Suyama Y, Ohyama M, Itoh T, Suzuki M.** 2004. Geographical

727   distribution of two haplotypes of chloroplast DNA in four oak species (*Quercus*) in

728   Japan. *Journal of Plant Research* **117**: 311-317

729   **Katoh K, Standley DM.** 2013. MAFFT multiple sequence alignment software version 7:

730   improvements in performance and usability. *Molecular Biology and Evolution* **30**: 772–

731   780

732   **Kitamura S.** 1955. Flowering plants and ferns. In: Kihara H, ed. *Fauna and Flora of Nepal*

733   *Himalaya*. Kyoto: Fauna and Flora Research Society, 73–290

734   **Kmenta M.** 2011. *Die Mikroflora der untermiozänen Fundstelle Altmittweida, Deutschland*.

735   Wien: University of Wien, Austria

736   **Knapp M, Stöckler K, Havell D, Delsuc F, Sebastiani F, Lockhart PJ.** 2005. Relaxed

737   molecular clock provides evidence for long-distance dispersal of *Nothofagus* (Southern

738   Beech). *Plos Biology* **3**: e14

739   **Köppen W.** 1936. Das geographische System der Klimate. In: Köppen W, Geiger R eds.

740   *Handbuch der Klimatologie, Band 1*. Berlin: Teil C. Gebr. Borntraeger, 1–44

741   **Kottek M, Grieser J, Beck C, Rudolf B, Rubel F.** 2006. World map of the Köppen-Geiger

742   climate classification updated. *Meteorologische Zeitschrift* **15**: 259-263

743   **Kubitzki K.** 1993. Fagaceae. In: Kubitzki K, Rohwer JG, Bittrich V, eds. *The families and*

744   *genera of vascular plants*, vol 2. Berlin: Spinger, 301–309

745   **Kvaček Z, Walther H.** 1989. Palaeobotanical studies in Fagaceae of the European Tertiary.

746   Plant *Systematics and Evolution* **162**: 213-229

747   **Lagache L, Klein EK, Guichoux E, Petit RJ.** 2013. Fine-scale environmental control of

748   hybridization in oaks. *Molecular Ecology* **22**: 423-436.

749   **Lang P, Dane F, Kubisiak TL, Huang H.** 2007. Molecular evidence for an Asian origin and a

750   unique westward migration of species in the genus *Castanea* via Europe to North

751   America. *Molecular Phylogenetics and Evolution* **43**: 49-59

752   **Lei M, Wang Q, Wu ZJ, Lopez-Pujol J, Li DZ, Zhang ZY.** 2012. Molecular phylogeography

753   of Fagus engleriana (Fagaceae) in subtropical China: limited admixture among multiple

754   refugia. *Tree Genetics and Genomes* **8**: 1203-1212.

755   **Lei YZ, Zhang QR, He W, Cao XP.** 1992. Tertiary. In: Wang XF,ed. *Geology of Hainan*

756   *Island. I. Stratigraphy and Palaeontology* [in Chinese]. Beijing: Geological Publishing

757   House, 218–266

758   **Lepais O, Petit R, Guichoux E, Lavabre J, Alberto F, Kremer A, Gerber S.** 2009. Species

759   relative abundance and direction of introgression in oaks. *Molecular Ecology* **18**: 2228-

760   2242

761   **Librado P, Rozas J** (2009) DnaSP v5: A software for comprehensive analysis of DNA

762   polymorphism data. *Bioinformatics* **25**: 1451–1452

763   **López de Heredia U, Jiménez P, Collada C, Simeone MC, Bellarosa R, Schirone B, Cervera**

764   **MT, Gil L.** 2007 Multi-marker phylogeny of three evergreen oaks reveals vicariant

765   patterns in the Western Mediterranean. *Taxon* **56**: 1209-1209

766   **Maddison WP, Maddison DR.** 2011. *Mesquite: a modular system for evolutionary analysis.*

767   *Version 2.75*. http://mesquiteproject.org.

768   **Magri D, Fineschi S, Bellarosa R, Buonamici A, Sebastiani F, Schirone B, Simeone MC,**

769   **Vendramin GG.** 2007. The distribution of *Quercus suber* chloroplast haplotypes

770   matches the palaeogeographical history of the western Mediterranean. *Molecular Ecology*

771   **16**: 5259-5266

772   **Manchester SR.** 1994. Fruits and seeds of the Middle Eocene nut beds flora, Clarno Formation,

773   Oregon. *Palaeontographica Americana* **58**: 1–205

774   **Manos PS, Cannon CH, Oh S-H.** 2008. Phylogenetic relationships and taxonomic status of the

775   paleoendemic Fagaceae of Western North America: recognition of a new genus,

776   *Notholithocarpus*. *Madroño* **55**: 181–190

777 **Manos PS, Stanford AM.** 2001. The historical biogeography of Fagaceae: tracking the tertiary

778   history of temperate and subtropical forests of the Northern Hemisphere. *International*

779   *Journal of Plant Sciences* **162**: S77-S93

780 **Masta SE, Sullivan B, Lamb T, Routman EJ.** 2002. Phylogeography, species boundaries, and

781   hybridization among toads of the *Bufo americanus* group. *Molecular Phylogenetics and*

782   *Evolution* **24**: 302–14

783 **McIntyre DJ.** 1991. Pollen and spore flora of an Eocene forest, eastern Axel Heiberg Island,

784   N.W.T. In: Christie RL, McMillan NJ, eds. *Tertiary Fossil Forest of the Geodetic Hills,*

785   *Axel Heiberg Island, Arctic Archipelago*. Geological Survey of Canada, Bulletin, vol 403,

786   83–97

787 **McIver EE, Basinger JF.** 1999. Early Tertiary floral evolution in the Canadian high arctic.

788   *Annals of the Missouri Botanical Garden* **86**: 523–545

789 **Menitsky YL.** 2005. *Oaks of Asia*. Science Publishers, Enflied, New Hampshire, USA

790 **Migliore J, Baumel A, Juin M, Medail F.** 2012. From Mediterranean shores to central Saharan

791   mountains: key phylogeographical insights from the genus *Myrtus*. *Journal of*

792   *Biogeography* **39**: 942-956

793 **Mir C, Jarne P, Sarda V, Bonin A, Lumaret R.** 2009. Contrasting nuclear and cytoplasmic

794   exchanges between phylogenetically distant oak species (*Quercus suber* L. and *Q. ilex*

795   L.) in Southern France: inferring crosses and dynamics. *Plant Biology* **11**: 213-226

796 **Moran EV, Willis J, Clark JS.** 2012. Genetic evidence for hybridization in red oaks (*Quercus*

797   sect. Lobatae, Fagaceae). *American Journal of Botany* **99**: 92-100

798   **Neophytou C, Aravanopoulos FA, Fink S, Dounavi A.** 2011. Interfertile oaks in an island

799   environment: II. Limited hybridization *Quercus alnifolia* Poech and *Q. coccifera* L. in a

800   mixed stand. *European Journal of Forest Research* **130**: 623–635

801   **Neophytou C, Dounavi A, Fink S, Aravanopoulos FA.** 2010. Interfertile oaks in an island

802   environment: I. High nuclear genetic differentiation and high degree of chloroplast DNA

803   sharing between *Q. alnifolia* and *Q. coccifera* in Cyprus. A multipopulation study.

804   *European Journal of Forest Research* **130**: 543–555

805   **Nieto Feliner G.** 2014. Patterns and processes in plant phylogeography in the Mediterranean

806   Basin. A review. *Perspectives in Plant Ecology, Evolution and Systematics* **16**: 265-278.

807   **Nylander JAA.** 2004. *MrModeltest. 2.1*. Program distributed by the author. Evolutionary

808   Biology Centre, Uppsala University, Uppsala, Sweden

809   **Oh S-H, Manos PS.** 2008. Molecular phylogenetics and cupule evolution in Fagaceae as

810   inferred from nuclear CRABS CLAW sequences. *Taxon* **57**: 434–451

811   **Okaura T, Quang ND, Ubukata M, Harada K.** 2007. Phylogeographic structure and late

812   Quaternary population history of the Japanese oak *Quercus mongolica* var. crispula and

813   related species revealed by chloroplast DNA variation. *Genes and Genetics Systems* **82**:

814   465-477

815   **Ortego J, Bonal R.** 2010. Natural hybridisation between kermes (*Quercus coccifera* L.) and

816   holm oaks (*Q. ilex* L.) revealed by microsatellite markers. *Plant Biology* **12**: 234-238

817   **Pearse IS, Hipp AL.** 2009. Phylogenetic and trait similarity to a native species predict herbivory

818   on non-native oaks. *Proceedings of the National Academy of Sciences of the USA* **106**:

819   18097-18102

820   **Peel MC, Finlayson BL, McMahon TA.** 2007. Updated world map of the Köppen-Geiger

821   climate classification. *Hydrology and Earth System Sciences* **11**: 1633-1644

822   **Peñaloza-Ramírez JM, González-Rodríguez A, Mendoza-Cuenca L, Caron H, Kremer A,**

823   **Oyama K.** 2010. Interspecific gene flow in a multispecies oak hybrid zone in the Sierra

824   Tarahumara of Mexico. *Annals of Botany* **105**: 389-399

825   **Petit RJ, Bodénès C, Ducousso A, Roussel G, Kremer A.** 2004. Hybridization as a mechanism

826   of invasion in oaks. *New Phytologist* **161**: 151-164

827   **Petit RJ, Csaikl UM, Bordács S, Burg K, Coart E, Cottrell J, van Dam B, Deans JD,**

828   **Dumolin-Lapègue S, Fineschi S.** 2002. Chloroplast DNA variation in European white

829   oaks: phylogeography and patterns of diversity based on data from over 2600

830   populations. *Forest Ecology and Management* **156**: 5-26

831   **Piredda R, Simeone MC, Attimonelli M, Bellarosa R, Schirone** B. 2011. Prospects of

832   barcoding the Italian wild dendroflora: oaks reveal severe limitations to tracking species

833   identity. *Molecular Ecology Resources* **11**: 72-83

834   **Premoli AC, Mathiasen P, Acosta MC, Ramos VA.** 2012. Phylogeographically concordant

835   chloroplast DNA divergence in sympatric *Nothofagus* s.s. How deep can it be? *New*

836   *Phytologist* **193**: 261–275

837   **Rambaut A.** 2014. FigTree, a graphical viewer of phylogenetic trees, v1.4.2.. Institute of

838   Evolutionary Biology University of Edinburgh. http://tree.bio.ed.ac.uk/software/figtree/.

839   **Rieseberg LH, Soltis DE.** 1991. Phylogenetic Consequences of Cytoplasmic Gene Flow in

840   Plants. *Evolutionary Trends in Plants* **5**: 65-84

841 **Rodriguez-Sanchez F, Guzman B, Valido A, Vargas P, Arroyo J.** 2009. Late Neogene history

842 of the laurel tree (*Laurus* L., Lauraceae) based on phylogeographical analyses of

843 Mediterranean and Macaronesian populations. *Journal of Biogeography* **36**: 1270-1281

844 **Ronquist F, Huelsenbeck JP.** 2003. MrBayes 3: Bayesian phylogenetic inference under mixed

845 models. *Bioinformatics* **19**: 1572-1574

846 **Ronquist F, Klopfstein S, Vilhelmsen L, Schulmeister S, Murray DL, Rasnitsyn AP.** 2012.

847 A total-evidence approach to dating with fossils, applied to the early radiation of the

848 hymenoptera. *Systematics Biology* **61**: 973–999

849 **Schnitzler JP, Steinbrecher R, Zimmer I, Steigner D, Fladung M.** 2004. Hybridization of

850 European oaks (Quercus ilex× Q. robur) results in a mixed isoprenoid emitter type. *Plant*

851 *Cell and Environment* **27**: 585-593

852 **Simeone MC, Piredda R, Papini A, Vessella F, Schirone B.** 2013. Application of plastid and

853 nuclear markers to DNA barcoding of Euro–Mediterranean oaks (*Quercus*, Fagaceae):

854 problems, prospects and phylogenetic implications. *Botanical Journal of the Linnean*

855 *Society* **172**: 478–499

856 **Stamatakis A, Hoover P, Rougemont J.** 2008. A rapid bootstrap algorithm for the RAxML

857 web servers. *Systematics Biology* **57**: 758–771

858 **Svenson U, Backlund A, McLoughlin S, Hill RS.** 2001. *Nothofagus* biogeography revisited

859 with special emphasis on the enigmatic distribution of subgenus *Brassospora* in New

860 Caledonia. *Cladistics* **17**: 28–47

861 **Swofford DL.** 2002. PAUP*: *Phylogenetic Analysis Using Parsimony (and Other Methods) 4.0*

862 *Beta*. Sunderland, MA: Sinauer Associates

863   **Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S** (2011) MEGA5: molecular

864   evolutionary genetics analysis using maximum likelihood, evolutionary distance, and

865   maximum parsimony methods. *Molecular Biology and Evolution* **28**: 2731–2739

866   **Thompson JD, Higgins DG, Gibson TJ.** 1994. CLUSTAL W: improving the sensitivity of

867   progressive multiple sequence alignment through sequence weighting, position specific

868   gap penalties and weight matrix choice. *Nucleic Acids Research* **22**: 4673-4680

869   **Tsutsui K, Suwa A, Sawada K, Kato T, Ohsawa TA, Watano Y.** 2009. Incongruence among

870   mitochondrial, chloroplast and nuclear gene trees in Pinus subgenus Strobus (Pinaceae).

871   *Journal of Plant Research* **122**: 509–521

872   **Valbuena-Carabaña M, González-Martínez SC, Hardy OJ, Gil .L** 2007. Fine-scale spatial

873   genetic structure in mixed oak stands with different levels of hybridization. *Molecular*

874   *Ecology* **16**: 1207-1219

875   **Valencia-Cuevas L, Mussali-Galante P, Piñero D, Castillo-Mendoza E, Rangel-Altamirano**

876   **G, Tovar-Sánchez E.** 2015. Hybridization of *Quercus castanea* (Fagaceae) across a red

877   oak species gradient in Mexico. *Plant Systematics and Evolution* **301**:1085-1097

878   **Velitzelos D, Bouchal JM, Denk T.** 2014. Review of the Cenozoic floras and vegetation of

879   Greece. *Review of Palaeobotany and Palynology* **204**: 56–117.

880   **Wang N, Wu FX.** 2015. New Oligocene cyprinid in the central Tibetan Plateau documents the

881   pre-uplift tropical lowlands. *Ichthyological Research* **62**: 274-285

882   **Weitschat W, Wichard W.** 2003. *Atlas of Plants and Animals in Baltic Amber*. Munchen: Pfeil

883   Verlag

884   **Whittemore AT, Schaal BA** (1991) Interspecific gene flow in sympatric oaks. *Proceedings of*

885   *the National Academy of Sciences of the USA* **88**: 2540-2544

886   **Wiens JJ, Servedio MR.** 2000. Species delimitation in systematics: inferring diagnostic

887   differences between species. *Proceedings of the Royal Society of London Series B* **267**:

888   631–36

889   **Xiang Q-Y, Manchester SR, Thomas DT, Zhang W, Fan C.** 2005. Phylogeny, biogeography,

890   and molecular dating of Cornelian cherries (*Cornus,* Cornaceae): Tracking Tertiary plant

891   migration. *Evolution* **59**: 1685-1700

892   **Xiang X-G, Wang W, Li R-Q, Lin L, Liu Y, Zhou Z-K, Li Z-Y, Chen Z-D.** 2014. Large-

893   scale phylogenetic analyses reveal fagalean diversification promoted by the interplay of

894   diaspores and environments in the Paleogene. *Perspectives in Plant Ecology, Evolution*

895   *and Systematics* **16**: 101–110

896   **Yi T, Miller AJ, Wen J.** 2004. Phylogenetic and biogeographic diversification of *Rhus*

897   (Anacardiaceae) in the Northern Hemisphere. *Molecular Phylogenetics and Evolution* **33**:

898   861-879

899   **Yoo K-O, Wen J.** 2007. Phylogeny of *Carpinus* and subfamily Coryloideae (Betulaceae) based

900   on chloroplast and nuclear ribosomal sequence data. *Plant Systematics and Evolution*

901   **267**: 25–35

902   **Zhang J-B, Li R-Q, Xiang X-G, Manchester SR, Lin L, Wang W, Wen J, Chen Z-D.** 2013a.

903   Integrated fossil and molecular data reveal the biogeographic diversification of the

904   eastern Asian-eastern North American disjunct hickory genus (*Carya* Nutt.). *Plos One* **8**:

905   e70449

906   **Zhang ZY, Wu R, Wang Q, Zhang ZR, Lopez-Pujol J, Fan DM, Li DZ.** 2013b. Comparative

907   phylogeography of two sympatric beeches in subtropical China: Species-specific

908   geographic mosaic of lineages. *Ecology and Evolution* **3**: 4461-4472.

909   **Zhou Z.** 1992. The fossil history of *Quercus*. *Acta Botanica Yunnanica* **15**: 21-33

910   **Zwickl DJ.** 2006. *Genetic algorithm approaches for the phylogenetic analysis of large*

911   *biological sequence datasets under the Maximum Likelihood criterion*. Ph.D. thesis,

912   University of Texas, Austin, USA

913

914   **Figure captions**

915

916   **Figure 1** ML tree of plastid accessions; tentatively rooted with the *Notholithocarpus-Chrysolepis*

917   subtree. Stars indicate subtrees comprising accessions of Mediterranean members of *Quercus*

918   Group Ilex. Number at branches indicate non-parametric bootstrap support under maximum

919   likelihood using two different implementations and posterior probabilities calculated using

920   Bayesian inference

921

922   **Figure 2** Haplotype network based on length-conserved portions of the *trn*H-*psb*A spacer.

923   Colouration refers to the taxonomic affiliation of specimens

924

925   **Figure 3** Geographic pattern of plastid haplotype variation in Mediterranean members of

926   *Quercus* Group Ilex. (A) Map showing the taxonomic identity of sampled specimens. (B) Map

927   showing the plastid haplotypes of sampled specimens

928

929   **Figure 4** Mapping of chloroplast evolution in oaks (using the same rooting scenario as in Fig. 1)

930   on current evolutionary synopsis (based on nuclear sequence data, morphology, and the fossil

931   record; modified after (Grímsson et al., 2015, fig. 16). Colouring of the plastid lineages refers to

932   branches/subclades in Fig. 1: bluish, common (ancestral) and 'New World' oak/castanoids

933   plastid haplotype lineages; green, lineages of the unique 'Euro-Med' plastid haplotype found

934   only in Mediterranean members of Group Ilex; reddish, lineages of 'Old World' oaks and

935   Eurasian castanoids. Note that members of Group Ilex keep plastid haplotypes of five different

936   evolutionary sources/systematic affinities. Abbreviations: C = Cretaceous, Pa = Paleocene, E =

937   Eocene, O = Oligocene, M = Miocene, Pl = Plio-/Pleistocene

938

939   **Figure 5** Eocene set-up and the origin of the 'Euro-Med' haplotype (palaeogeographic map ©

940   Ron Blakey, Colorado Plateau Geosystems). **A** Unequivocal fossil record of oaks in the Eocene

941   mapped on a palaeotopographic map (© Ron Blakey, Colorado Plateau Geosystems)

942   highlighting a primary split into a high-latitude and mid-latitude lineage that likely correspond to

943   the deep phylogenetic split seen in nuclear and plastid sequence data of modern oaks between the

944   'New World Clade' (Groups Protobalanus, Quercus and Lobatae) and the 'Old World Clade'

945   (Groups Cyclobalanopsis, Ilex, Cerris). **B-C** Scenarios that can explain the occurrence of the

946   unique 'Euro-Med' haplotype in westernmost members of *Quercus* Group Ilex. **B** The 'Euro-

947   Med' haplotype belonged to an extinct oak lineage geographically/biologically separated from

948   both the ancestors of the New World and Old World Clade. Westward expansion of Himalayan

949   members of Group Ilex and subsequent large-scale introgression/hybridisation homogenised the

950   western members of Group Ilex and the extinct oak lineage, retaining and evolving the original

951   haplotype in the Mediterranean region. **C** The 'Euro-Med' haplotype reflects geographic

952   fragmentation within the Paleogene range of the Old World Clade that was overprinted to some

953   degree after later radiation phases of Group Ilex

954

955   **Figure 6** Tectonic activity during the Eocene (palaeogeographic map © Ron Blakey, Colorado

956   Plateau Geosystems) and past and modern distribution of the New World (white) and Old World

957   (yellow) groups within *Quercus*. Black lines indicate major subduction zones, red lines major

958   orogenies. Note that the high latitude lineage of oaks (*Quercus* Group Lobatae, Group Quercus,

959   Group Protobalanus) evolved in tectonically stable regions, whereas the low latitude lineage

960   (*Quercus* Group Ilex, Group Cyclobalanopsis, Group Cerris) evolved in tectonically unstable

961   regions. Uppercase and lowercase letters refer to extant and extinct distribution areas of major

962   oak lineages: P,p = Group Protobalanus; Q,q = Group Quercus; L,l = Group Lobatae; I,i = Group

963   Ilex; C,c = Group Cerris; Y,y = Group Cyclobalanopsis

964

965

966   **Supporting Information**

967   **Table S1** List of specimens and accessions included in this study

968   **Table S2** Intra- and intertaxonomic minimum and maximum pairwise genetic distances

969   **File S3** MJ networks based on different plastid gene regions

970

971   **Data Archival Location**

972   All sequence data generated as part of this study are available on GenBank

973   (http://www.ncbi.nlm.nih.gov/genbank/) under accession numbers LM652969-LM653098; other

974   relevant data are within the paper and its Supplementary files. Primary data and analyses are

975   provided for anonymous download at www.palaeogrimm.org/data/Smn15_OSA.zip

976

Figure 1(on next page)

ML tree of the investigated oak accessions

ML tree of plastid accessions; tentatively rooted with the Notholithocarpus-Chrysolepis subtree. Stars indicate subtrees comprising accessions of Mediterranean members of Quercus Group Ilex. Number at branches indicate non-parametric bootstrap support under maximum likelihood using two different implementations and posterior probabilities calculated using Bayesian inference

0.1 expected subst./site

ML-BS (Garli, 100 replicates)/ML-BS (RAxML, 650 replicates)/PP (MrBayes)
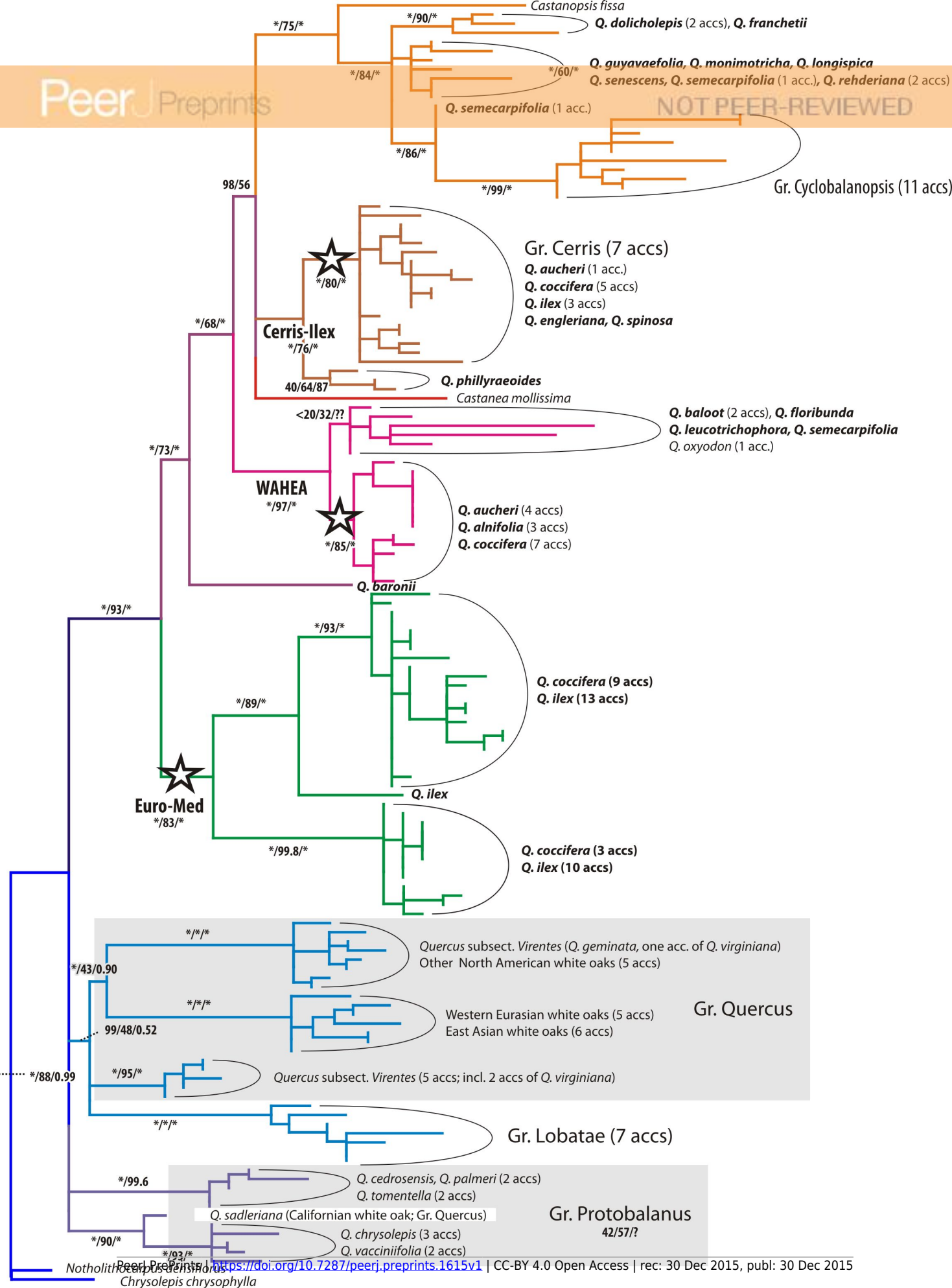
Figure 2(on next page)

Haplotype network based on the trnH-psbA spacer

Haplotype network based on length-conserved portions of the trnH-psbA spacer. Colouration refers to the taxonomic affiliation of specimens
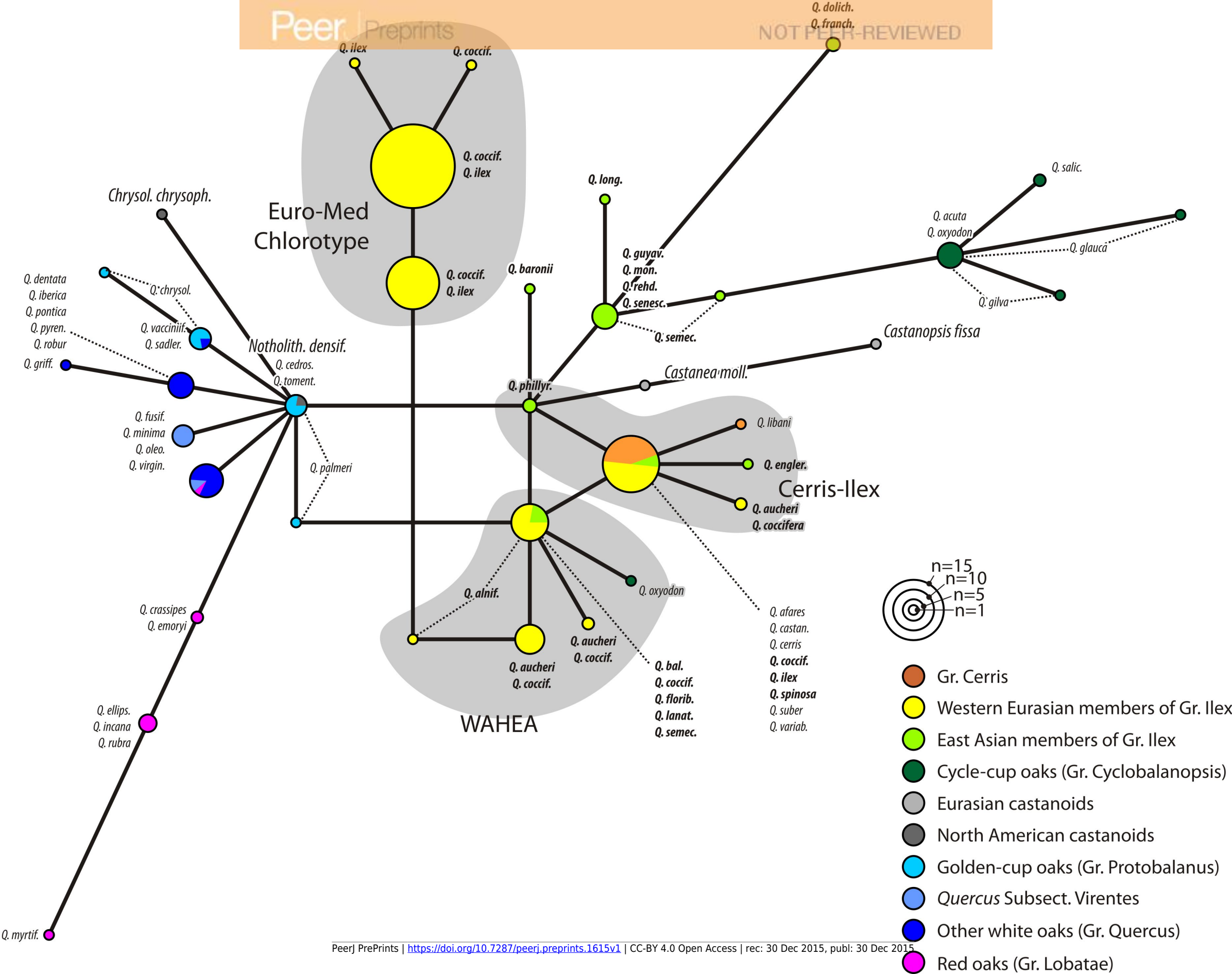
Q. dolich.
Q. franch.

Q. ilex   *Q. coccif.*

Q. coccif.
Q. ilex

Euro-Med Chlorotype

Chrysol. chrysoph.

Q. long.

Q. salic.

Q. acuta
Q. oxyodon

Q. dentata
Q. iberica
Q. pontica
Q. pyren.
Q. robur

Q. chrysol.

Q. vacciniif.
Q. sadler.

Q. griff.

Notholith. densif.
Q. cedros.
Q. toment.

Q. coccif.
Q. ilex

Q. baronii

Q. guyav.
Q. mon.
Q. rehd.
Q. senesc.

Q. glauca

Q. gilva

Q. semec.

Castanopsis fissa

Q. fusif.
Q. minima
Q. oleo.
Q. virgin.

Castanea moll.

Q. phillyr.

Q. libani

Cerris-Ilex

Q. palmeri

Q. engler.

Q. aucheri
Q. coccifera

n=15
n=10
n=5
n=1

Q. crassipes
Q. emoryi

Q. alnif.

Q. oxyodon

Q. afares
Q. castan.
Q. cerris
Q. coccif.
Q. ilex
Q. spinosa
Q. suber
Q. variab.

Q. ellips.
Q. incana
Q. rubra

Q. aucheri
Q. coccif.

Q. aucheri
Q. coccif.

WAHEA

Q. bal.
Q. coccif.
Q. florib.
Q. lanat.
Q. semec.

Q. myrtif.

Gr. Cerris

Western Eurasian members of Gr. Ilex

East Asian members of Gr. Ilex

Cycle-cup oaks (Gr. Cyclobalanopsis)

Eurasian castanoids

North American castanoids

Golden-cup oaks (Gr. Protobalanus)

Quercus Subsect. Virentes

Other white oaks (Gr. Quercus)

Red oaks (Gr. Lobatae)

Figure 3 (on next page)

Plastid haplotype variation in Mediterranean members of Quercus Group Ilex

Geographic pattern of plastid haplotype variation in Mediterranean members of *Quercus* Group Ilex. (A) Map showing the taxonomic identity of sampled specimens. (B) Map showing the plastid haplotypes of sampled specimens
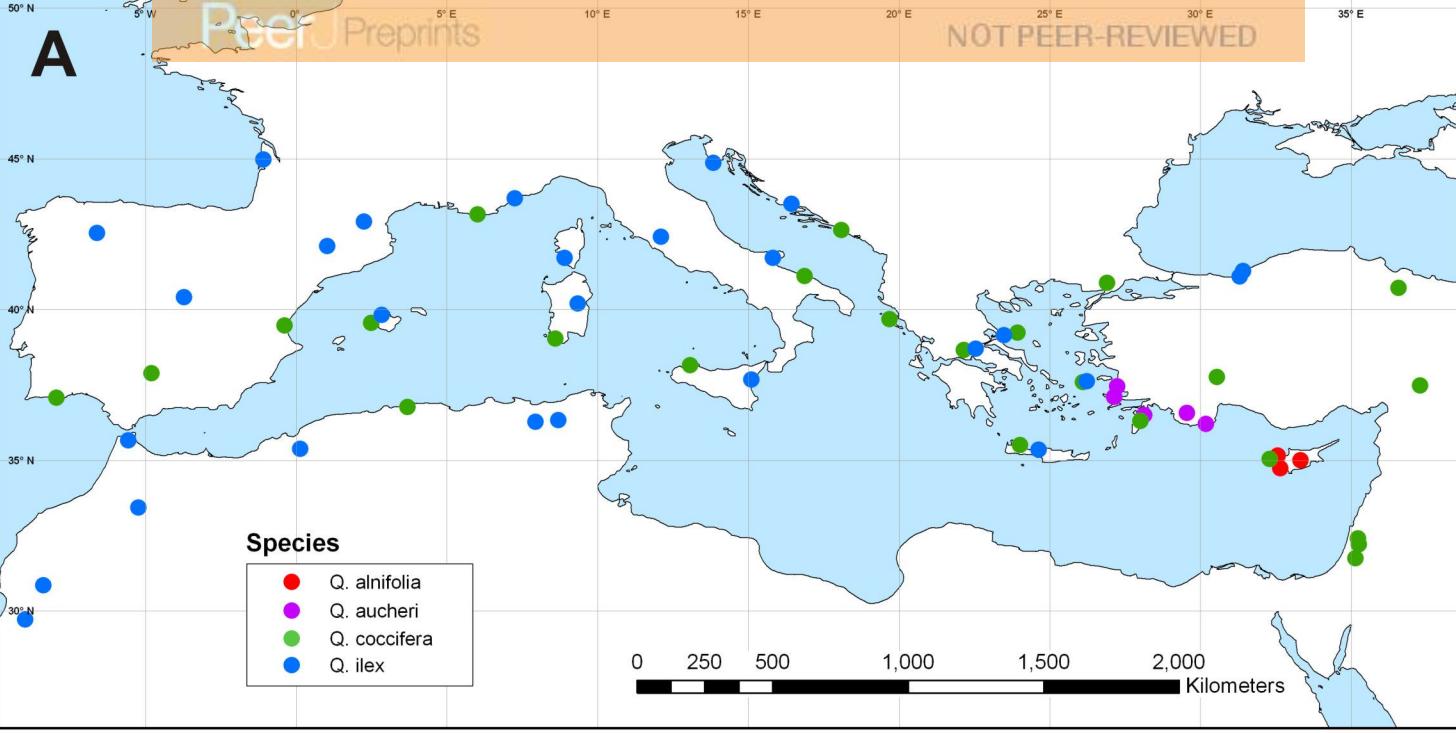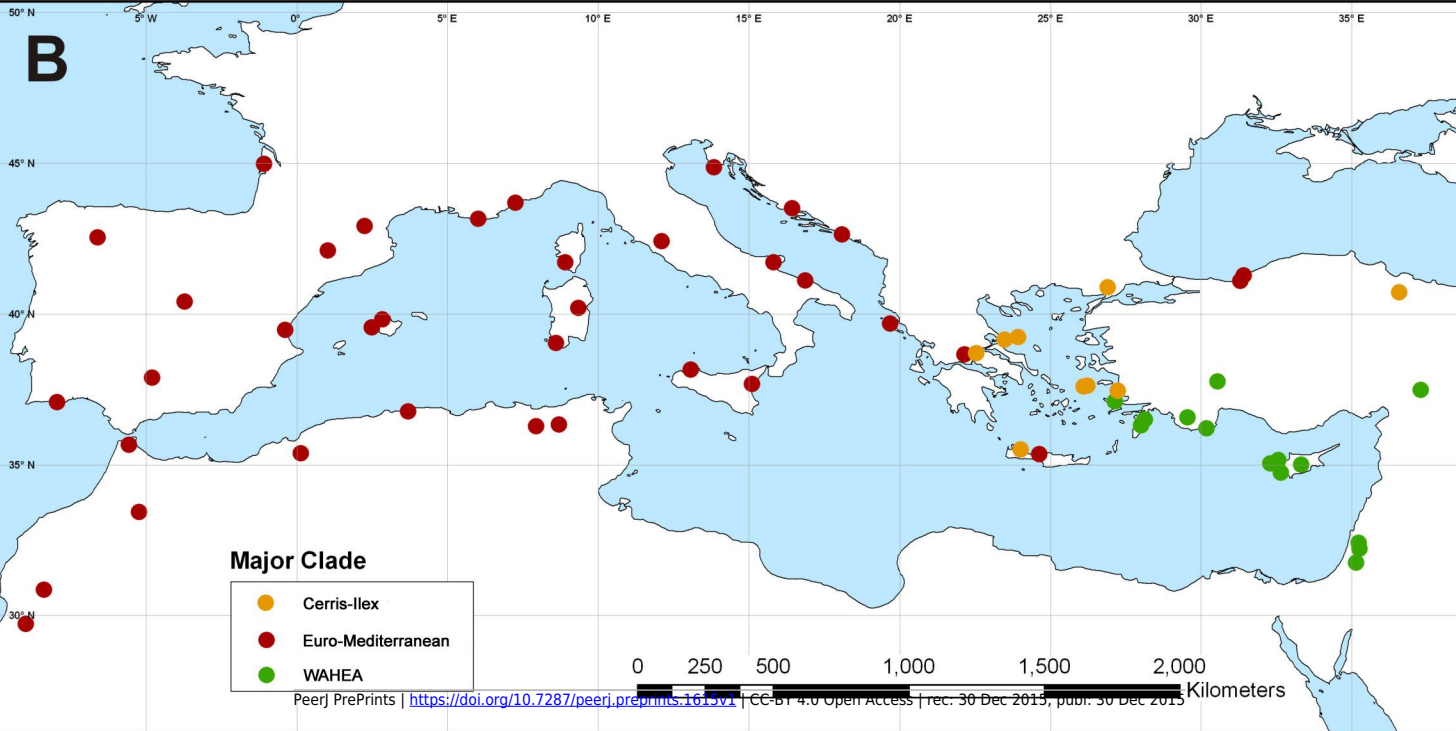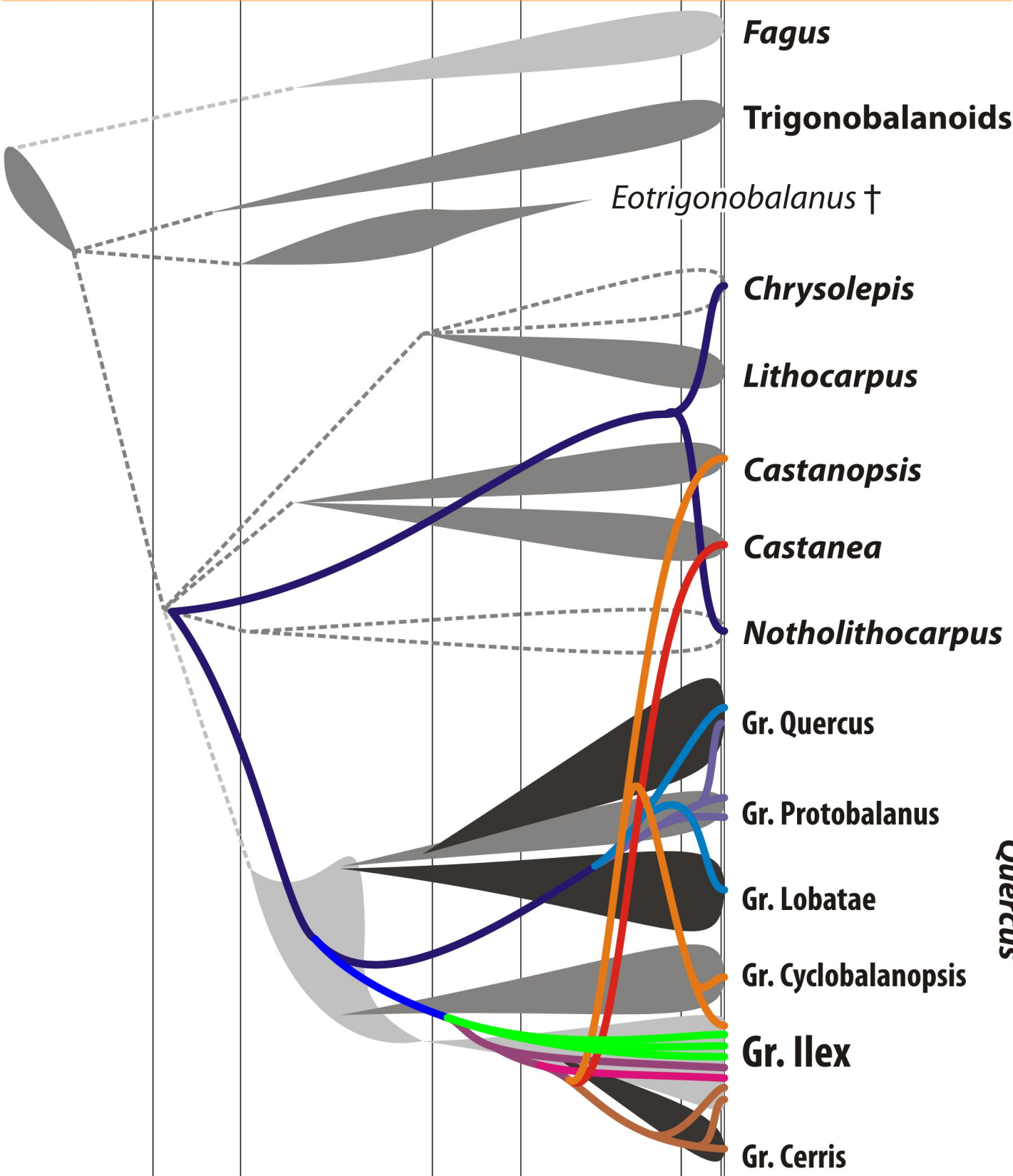
Figure 4(on next page)

Map of chloroplast evolution in oaks

Mapping of chloroplast evolution in oaks (using the same rooting scenario as in Fig. 1) on current evolutionary synopsis (based on nuclear sequence data, morphology, and the fossil record; modified after (Grímsson et al., 2015, fig. 16). Colouring of the plastid lineages refers to branches/subclades in Fig. 1: bluish, common (ancestral) and 'New World' oak/castanoids plastid haplotype lineages; green, lineages of the unique 'Euro-Med' plastid haplotype found only in Mediterranean members of Group Ilex; reddish, lineages of 'Old World' oaks and Eurasian castanoids. Note that members of Group Ilex keep plastid haplotypes of five different evolutionary sources/systematic affinities. Abbreviations: C = Cretaceous, Pa = Paleocene, E = Eocene, O = Oligocene, M = Miocene, Pl = Plio-/Pleistocene

Fagus

Trigonobalanoids

Eotrigonobalanus †

Chrysolepis

Lithocarpus

Castanopsis

Castanea

Notholithocarpus

Gr. Quercus

Gr. Protobalanus

Gr. Lobatae

Gr. Cyclobalanopsis

Gr. Ilex

Gr. Cerris

Quercus

C   Pa   E   O   M   Pl

66   56   34   23   5 Ma

Figure 5(on next page)

Origin of the 'Euro-Med' haplotype

Eocene set-up and the origin of the 'Euro-Med' haplotype (palaeogeographic map © Ron Blakey, Colorado Plateau Geosystems). A Unequivocal fossil record of oaks in the Eocene mapped on a palaeotopographic map (© Ron Blakey, Colorado Plateau Geosystems) highlighting a primary split into a high-latitude and mid-latitude lineage that likely correspond to the deep phylogenetic split seen in nuclear and plastid sequence data of modern oaks between the 'New World Clade' (Groups Protobalanus, Quercus and Lobatae) and the 'Old World Clade' (Groups Cyclobalanopsis, Ilex, Cerris). B-C Scenarios that can explain the occurrence of the unique 'Euro-Med' haplotype in westernmost members of *Quercus* Group Ilex. B The 'Euro-Med' haplotype belonged to an extinct oak lineage geographically/biologically separated from both the ancestors of the New World and Old World Clade. Westward expansion of Himalayan members of Group Ilex and subsequent large-scale introgression/hybridisation homogenised the western members of Group Ilex and the extinct oak lineage, retaining and evolving the original haplotype in the Mediterranean region. C The 'Euro-Med' haplotype reflects geographic fragmentation within the Paleogene range of the Old World Clade that was overprinted to some degree after later radiation phases of Group Ilex

A

High-latitude lineage:
including ancestors of New World Clade

Affinity with East Asia
(*Quercus* Gr. Cyclobalanopsis, *Lithocarpus*)

Modern distribution
of Euro-Med chlorotype

Mid-latitude lineage
including ancestors of
Old World Clade

Legend:
- ⭐ (white) Oak of unknown affinity/exinct lineage
- ⭐ (pink) New World Clade
- ⭐ (blue) Group Quercus
- ⭐ (red) Group Lobatae
- ⭐ (yellow) Old World Clade

B

Turgai Street closure

Extinct oak lineage

Westward expansion of Group Ilex

C

Turgai Street closure

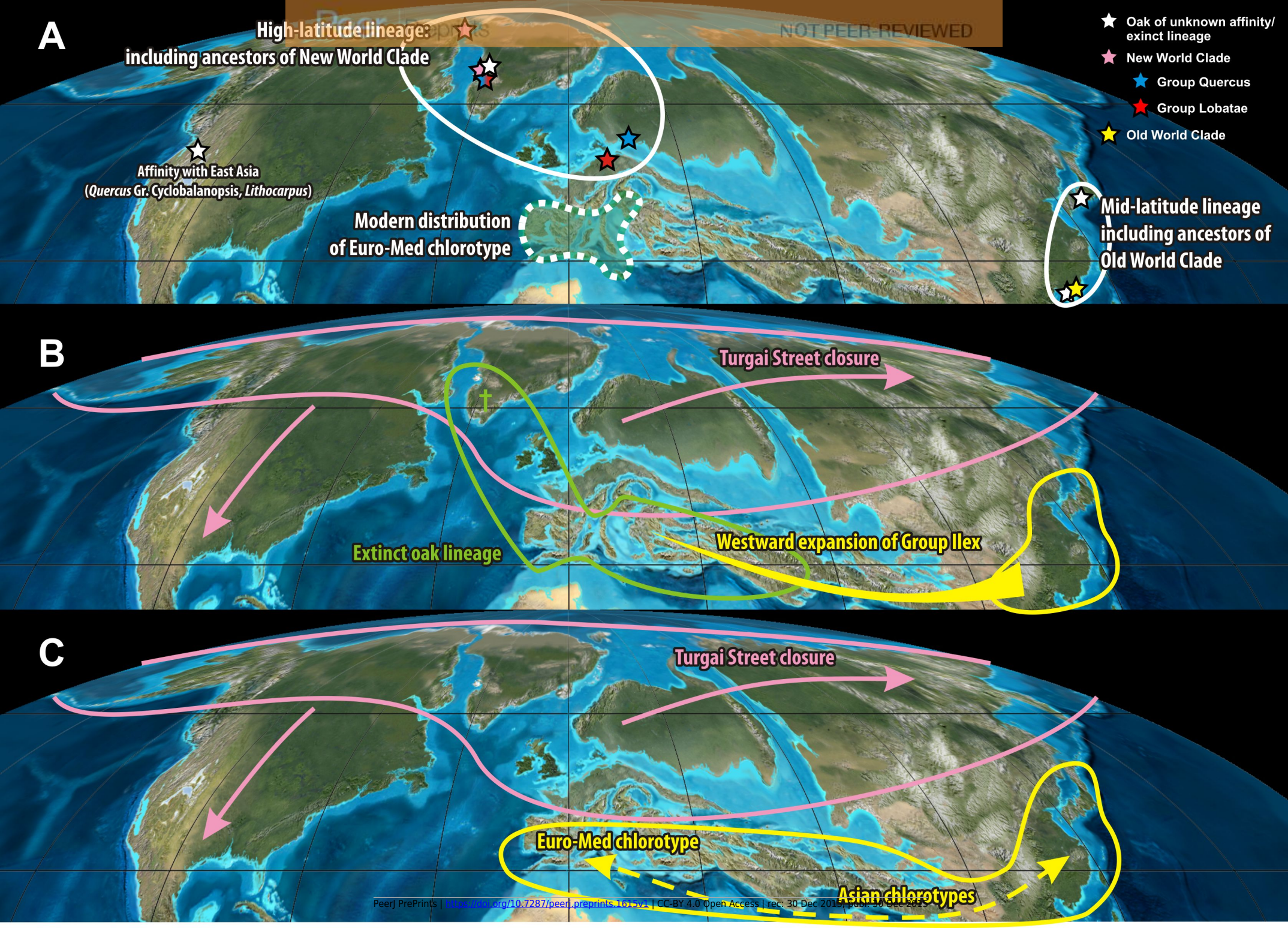Euro-Med chlorotype

Asian chlorotypes

Figure 6(on next page)

Tectonic activity during the Eocene

Tectonic activity during the Eocene (palaeogeographic map © Ron Blakey, Colorado Plateau Geosystems) and past and modern distribution of the New World (white) and Old World (yellow) groups within *Quercus*. Black lines indicate major subduction zones, red lines major orogenies. Note that the high latitude lineage of oaks (*Quercus* Group Lobatae, Group Quercus, Group Protobalanus) evolved in tectonically stable regions, whereas the low latitude lineage (*Quercus* Group Ilex, Group Cyclobalanopsis, Group Cerris) evolved in tectonically unstable regions. Uppercase and lowercase letters refer to extant and extinct distribution areas of major oak lineages: P,p = Group Protobalanus; Q,q = Group Quercus; L,l = Group Lobatae; I,i = Group Ilex; C,c = Group Cerris; Y,y = Group Cyclobalanopsis

'New World Clade'

'Old World Clade'

Table 1(on next page)

Species list

Species included in *Quercus* Group Ilex according to Denk and Grimm (2010); nomenclature followed Govaerts and Frodin (1998); species investigated in the present study are bolded. Taxonomic remarks and species distributions according to Govaerts and Frodin (1998), *Menitsky (2005) and **Flora of China Editorial Committee (1999).

1

2 **Table 1** – Species included in *Quercus* Group Ilex according to Denk and Grimm (Denk and Grimm 2010); nomenclature followed

3 Govaerts and Frodin (1998); species investigated in the present study are bolded. Taxonomic remarks and species distributions

4 according to Govaerts and Frodin (1998), *Menitsky (2005) and **Flora of China Editorial Committee (1999).

5

Species	Taxonomic remarks	Distribution
Q. acrodonta Seemen	Includes *Q. handeliana* A. Camus*/**	C, E and S China
***Q. alnifolia* Poech**		Cyprus
Q. aquifolioides Rehder & E.H.Wilson	Includes *Q. semecarpifolia* subsp. *glabra**	Tibet, C and SW China to Myanmar
***Q. aucheri* Jaub. &Spach**		SW Anatolia
***Q. baloot* Griff.**		Pakistan, Afghanistan
***Q. baronii* Skan**	Numerous morpho-ecological traits in common with members of Group Cerris*	NC and SW China
Q. bawanglingensis C.C. Huang, Ze X. Li & F.W. Xing	Poorly known; uncertain status, related to *Q. phillyreoides**	SE China
***Q. coccifera* L.**	Includes *Q. calliprinos* Webb*	Mediterranean
Q. cocciferoides Hand.-Mazz.	Includes *Q. taliensis* A. Camus**	CS China
***Q. floribunda* Lindl. ex A. Camus**	Basionym: *Q. dilatata* Lindl. ex A.DC. *nom. illegit.*	Pakistan, Afghanistan, Nepal
***Q. dolicholepis* A. Camus**	Includes *Q. fimbriata**	CW to SW China
***Q. engleriana* Seemen**		Tibet to E China, Myanmar
Q. fimbriata Y.C. Hsu & H. Wei Jen	Included in *Q. semecarpifolia** or *Q. dolicholepis**	C China
***Q. franchetii* Skan**		C China to N Vietnam
Q. gilliana Rehder & E.H. Wilson	Included in *Q. spinosa**	Tibet, C and S China
***Q. guyavifolia* H. Lév.**	Included in *Q. semecarpifolia* subsp. *glabra**; *Q.*	C and S China

	pannosa or *Q. aquifolioides* var. *rufescens***	
Q. ilex **L.**		Mediterranean
Q. kingiana Craib		C China to N Thailand
Q. lanata Sm.	Included in *Q. leucotrichophora***	Buthan to Vietnam
Q. leucotrichophora **A. Camus**	Basionym: *Q. incana* Roxb. *nom. illegit.*	N Pakistan, N India, Nepal to N Vietnam
Q. lodicosa O.E. Warb. & E.F. Warb.		SE Tibet to Myanmar
Q. longispica **(Hand.-Mazz.) A. Camus**	Includes *Q. semecarpifolia* subsp. *glabra**; *Q. rehderiana***	C and S China
Q. marlipoensis Hu & Cheng	Poorly known*; very close to *Q. engleriana***	C China
Q. monimotricha **Hand.-Mazz.**		C and S China
Q. oxyphylla Hand.-Mazz.	Includes *Q. spathulata* Seemen; included in *Q. dolicholepis***	C and S China
Q. pannosa **Hand.-Mazz.**	Possibly conspecific with *Q. semecarpifolia**	C and S China
Q. phillyreoides **A. Gray**	Includes *Q. utilis**	C China to Japan
Q. pseudosemecarpifolia **A. Camus**	Includes *Q. semecarpifolia* subsp. *glabra**, *Q. rehderiana***	Tibet to CS China
Q. rehderiana Hand.-Mazz.	Included in *Q. semecarpifolia* subsp. *glabra**; includes *Q. longispica* and *Q. pseudosemecarpifolia***	Tibet to C and S China
Q. semecarpifolia **Sm.**	Includes: *Q. fimbriata, Q. gujavifolia, Q. aquifolioides, Q. rehderiana, Q. longispica, Q. pseudosemecarpifolia**	Afghanistan to Myanmar
Q. senescens **Hand.-Mazz.**		E Himalaya, Tibet, C and S China
Q. setulosa Hickel & A. Camus		C China to Vietnam
Q. spinosa **David**		NC and SW China to Taiwan
Q. tarokoensis Hayata		E Taiwan
Q. utilis Hu & Cheng	Included in *Q. phillyreoides* subsp. *fokiensis**	C China

6

Table 2(on next page)

Diversity values of the DNA fragments used for the analyses

1

2　**Table 2.** Diversity values and models of DNA evolution of the fragments used for the analyses in 59 *Quercus* species (137

3　individuals) and 4 outgroup taxa. AL: Aligned length (bp); P: uncorrected p-distance (min. − max.); N hap: Number of identified

4　haplotypes, brackets: with gaps considered; Hd: Haplotype diversity; S: Number of polymorphic sites; θ: Nucleotide polymorphism;

5　π: Nucleotide diversity; PIC: Number of Parsimony Informative Characters; ME: Model of evolution.

6

Markers	AL	P	Nhap	Hd	S	PICs	θw	π	ME
*rbc*L	743	0.00 − 0.008	28	0.846±0.027	26	18	0.0063	0.0027	HKY+I
*trn*H-*psb*A	634	0.00 − 0.035	37 (84)	0.944±0.008	38	23	0.0159	0.009	GTR+G
*trn*K/*mat*K	705	0.00 − 0.022	49 (51)	0.952±0.008	59	31	0.0156	0.0064	n.d.
*trn*K (intron)	401	n.d.	32 (34)	0.821±0.028	36	16	0.0169	0.0048	GTR+G
*mat*K (codons)	304	n.d.	25	0.925±0.008	23	15	0.0146	0.0086	HKY
*rbc*L+*trn*K/*mat*K+*trn*H-*psb*A	2082	0.00 − 0.014	74 (110)	0.978±0.005	122	72	0.0119	0.0056	**Combined**
*rbc*L + *mat*K	1047	n.d.	49	0.965±0.006	49	33	0.0085	0.0044	n.d.
*trn*H-*psb*A + *trn*K	1035	n.d.	57 (103)	0.954±0.008	69	34	0.0161	0.0067	n.d.
*trn*H-*psb*A + *trn*K-*mat*K	1339	n.d.	65 (110)	0.970±0.006	92	49	0.0155	0.0072	n.d.

7

Table 3 (on next page)

Divergence patterns in other Fagaceae and Fagales

Divergence patterns in *Quercus* compared to other Fagaceae and Fagales based inter-species pair wise uncorrected p-distances of sequences retrieved from GenBank and produced in this study; Nt = number of taxa; Ns = number of sequences.

1   **Table 3.** Divergence patterns in *Quercus* compared to other Fagaceae and Fagales based inter-species pair wise uncorrected *p*-

2   distances of sequences retrieved from GenBank and produced in this study; Nt = number of taxa; Ns = number of sequences.

Taxon	Nt	Ns	Intrageneric divergence *rbc*L Min.	Max.	*mat*K Min.	Max.	*trn*H-*psb*A Min.	Max.	Mean intergeneric divergence at family level[‡] *rbc*L Min.	Max.	*mat*K Min.	Max.	*trn*H-*psb*A Min.	Max.
Quercus	87/87/86	219/255/382	0.000	0.010	0.000	0.023	0.000	0.042	0.003	0.011	0.006	0.015	0.006	0.042
Fagus	8/9/6	33/30/19	0.000	0.014	0.000	0.007	0.000	0.042	0.024	0.036	0.091	0.098	0.120	0.147
Other Fagaceae	37/21/28	102/32/86	0.000	0.020	0.000	0.021	0.000	0.020	0.003	0.013	0.000	0.019	0.011	0.032
Nothofagaceae*	23/†/14	35/†/53	0.000	0.027	–[†]	–[†]	0.000	0.017	0.012	0.023	–[†]	–[†]	0.017	0.041
Betulaceae	55/19/77	131/34/247	0.000	0.011	0.000	0.006	0.000	0.069	0.006	0.024	0.014	0.033	0.011	0.079
Juglandaceae	18/†/21	23/†/28	0.000	0.005	–[†]	–[†]	0.000	0.007	0.000	0.021	–[†]	–[†]	0.006	0.034

* Values for *rbc*L may be over-estimated (data usually older than 15 years; sequences show features characteristic for sequencing and editing artifacts)

[†] Insufficient data.

[‡] Values for *Quercus* and other Fagaceae not including *Fagus* (see *Fagus* for max. inter-generic divergence in Fagaceae)

3

Table 4(on next page)

Eocene fossils of Quercus

1 **Table 4.** Eocene fossil record of *Quercus*.

Locality, site, age	Reference	Taxon, organ	Affinity
Clarno Fm., Oregon, western U.S.A.; ~48 Ma	Manchester 1994	"*Quercus*" *paleocarpa* Manchester; cupules and acorns	Group Cyclobalanopsis (? *Lithocarpus*)
Axel-Heiberg Island, Canadian Arctic; ~45 Ma	McIntyre 1991 McIver and Basinger 1999	Pollen and leaves	New World Clade; *Quercus* Group Quercus/Lobatae
Hareøn, western Greenland; ~42 Ma	Grímsson et al. 2015	*Quercus* sp. 4–5; pollen	Extinct/ancestral type
		Quercus sp. 6–7; pollen	New World Clade (aff. Group Protobalanus)
		Quercus sp. 1–3; pollen	*Quercus* Group Quercus and/or Lobatae
Baltic amber, northern Europe; ~45 Ma	Crepet and Nixon 1989; Weitschat and Wichard 2003	Flower and *in situ* pollen	*Quercus* Group Quercus
Königsaue, near Aschersleben, Germany; middle Eocene (48–38 Ma)	Kvaček and Walther 1989	*Quercus subhercynica* H. Walther & Kvaček; leaf	*Quercus* Group Lobatae
Ube coal-field, southwestern Honshu, Japan; middle Eocene (48–38 Ma)	Huzioka and Takahasi 1970	*Cyclobalanopsis nathoi* Huzioka & Takahashi; leaf, acorn (?), cupule (?)	*Quercus,* affinity unclear
Changchang, Hainan; middle (?) Eocene (50–35 Ma)	Hofmann 2010	*Quercus* pollen types 2-8, 10; pollen	*Quercus,* affinity unclear (extinct, Group Quercus/Lobatae?, Group Protobalanus?)
		Quercus pollen type 1; pollen	*Quercus* Group Ilex
		Quercus pollen type 9; pollen	*Quercus* Group Cyclobalanopsis

2