



Harvesting Betulaceae sequences from GenBank to generate a new chronogram for the family

GUIDO W. GRIMM^{1*} and SUSANNE S. RENNER²

¹Department of Palaeobotany, Swedish Museum of Natural History, PO Box 50007, 10405 Stockholm, Sweden

²Systematic Botany and Mycology, University of Munich (LMU), 80638 Munich, Germany

Received 15 December 2012; revised 6 April 2013; accepted for publication 29 April 2013

Betulaceae, with 120–150 species in six genera, are a family of Fagales that occurs mainly in the Northern Hemisphere. Previous studies of the evolution of *Alnus*, *Betula*, *Carpinus*, *Corylus*, *Ostrya* and *Ostryopsis* have relied on a relatively small number of sequence data and molecular clocks with fixed-point calibrations. We exploited GenBank to construct Betulaceae matrices of up to 900 sequence accessions and 9300 nucleotides of nuclear and plastid DNA; we also computed species consensus sequences to build 46- and 29-species matrices that strike a balance between species sampling and nucleotide sampling. Trees were rooted on Ticodendraceae and Casuarinaceae, and divergence times were inferred under relaxed and strict molecular clocks, using alternative fossil constraints. The data support the traditional two subfamilies, Betuloideae (*Alnus*, *Betula*) and Coryloideae, and show that *Ostryopsis* is sister to *Ostrya/Carpinus*. The fossil record and molecular clocks calibrated with alternating fossils indicate that the stem lineage of Betulaceae dates back to the Upper Cretaceous, the two subfamilies to the Palaeocene and the most recent common ancestors of each of the living genera to the mid- to late Miocene. A substitution rate shift in Coryloideae between 25 and 15 Mya preceded the mid-Miocene climatic optimum and may be linked to temperate niches that became available following the mid-Miocene. © 2013 The Linnean Society of London, *Botanical Journal of the Linnean Society*, 2013, **172**, 465–477.

ADDITIONAL KEYWORDS: consensus sequences – fossil calibrations – GenBank data curation – molecular clocks – substitution rate shift.

INTRODUCTION

Betulaceae (Fagales) comprise some 120–150 species of trees or shrubs, occurring mostly in the northern temperate zone. A few species occur as far south as the Andes and Sumatra (Kubitzki, 1993). Betulaceae can be recognized by their stipulate, doubly serrate leaves, catkins and small winged fruits or nuts associated with leafy bracts. There are six genera in two subfamilies with overlapping distributions: *Alnus* Mill. (29–35 species) and *Betula* L. (42–50) make up Betuloideae and *Carpinus* L. (26–35), *Corylus* L. (16), *Ostrya* Scop. (5–9) and *Ostryopsis* Decne. (3) make up Coryloideae. Several *Alnus* species, especially *A. rubra* Bong., are used in reforestation because of their fast growth and nitrogen fixation; European hazelnut, *Corylus avellana* L., is an economically important food crop.

Despite the relatively small size and several molecular phylogenetic studies, relationships among subfamilies have remained unclear (Table 1). The first phylogenetic analysis, based on plastid *rbcL*, for Betulaceae was published > 20 years ago (Bousquet, Strauss & Li, 1992), but did not include *Ostryopsis*. Subsequent studies added plastid *matK* sequences (Kato *et al.*, 1998: one *Ostryopsis* sp. sampled), nuclear internal transcribed spacer (ITS) sequences (Chen, Manchester & Sun, 1999; Yoo & Wen, 2002; Forest *et al.*, 2005; Li, Shoup & Chen, 2005), further plastid genes and spacers (Yoo & Wen, 2007: combined 2906 nucleotides) and nuclear nitrate reductase sequences (*NIA*, 3rd intron; Li, Shoup & Chen, 2007: 11 diploid *Betula* spp.; Li, 2008: 22 species of Coryloideae). The studies that included outgroups mostly placed the root between the two subfamilies of Betulaceae (Table 1). A study focusing on the order Fagales, however, found *Alnus* and *Betula* as

*Corresponding author. E-mail: guido.grimm@nrm.se

Table 1. Support for relationships in Betulaceae based on different taxon and gene samples. Numbers in square brackets refer to the number of species [according to the original papers (see also Supporting Information, S1) for identification or sequencing errors]

Clade	Kato <i>et al.</i> (1998) <i>matK</i> BS _P	Li <i>et al.</i> (2004) Multigene† BS _P /PP	Forest <i>et al.</i> (2005)‡ 5S IGS + ITS BS _P /PP	Li <i>et al.</i> (2005) ITS BS _P	Yoo & Wen (2002, 2007)‡ ITS/morph, multigene§ BS _P /BS _P , BS _P /PP	Li (2008) ‡ <i>NIA</i> BS _{NIP} /BS _P
Betulaceae	100	100/1.0	~	~	~ / ~, ~ / ~	~
Betuloideae*	64	NA	79/1.0	~	~ / ~, ~ / ~	~
<i>Alnus</i>	~ [1]	~ [1]	100/1.0 [6]	100 [3]	~ / ~, ~ / ~ [1]	~ [1]
<i>Betula</i>	~ [1]	~ [1]	100/1.0 [7]	100 [34]	~ / ~, ~ / ~ [1]	100/100 [3]
Coryloideae*	100	100/1.0	96/1.0	~	~ / ~, ~ / ~	~
<i>Corylus</i>	~ [1]	~ [1]	100/1.0 [5]	~ [1]	64/65 [4], 100/1.0 [3]	100/99 [6]
<i>Carpinus-Ostrya-Ostryopsis</i>	NA	NA	72/0.94	48	< 50/< 50, < 50/0.5	86/NA¶
<i>Ostryopsis</i>	~ [1]	~ [1]	~ [1]	~ [1]	~ / ~, ~ / ~ [1]	100/~ [1]
<i>Corylus-Ostryopsis</i>	99	88/1.0	NA	NA	NA, NA	NA
<i>Carpinus-Ostrya</i>	100	100/1.0	95/1.0	82	93/<50, 79/0.97	98/91
<i>Carpinus</i>	~ [1]	~ [1]	65/0.97 [5]	Grade [2]	NA/55 [18], 63/0.97 [13]	Grade [8]
<i>Ostrya</i>	~ [1]	~ [1]	93/1.0 [2]	89 [2]	88/Grade [5], Grade [5]	95/99 [3]
Betuloideae/Coryloideae split*	~	~	~	99	<50/<50	99

NA, clade not represented in tree; ~, thus bipartition not testable. Clades that received moderate to strong support independent of gene and taxon sampling, highlighted in bold; BS_{NIP}, bootstrap support under neighbor joining (NJ) or maximum parsimony (MP); PP, Bayesian posterior probabilities.

*The mutual monophyly of Betuloideae/Coryloideae can only be tested in rooted trees.

†Three plastid genes (*atpB*, *matK*, *rbcL*); one plastid intron (*trnL*); one nuclear ribosomal RNA gene (18S rDNA) and one mitochondrial gene (*matR*). The signal stems mostly from the *matK* partition, with the *trnL* data forcing *Alnus* as sister to remaining Betulaceae and *matK* indicating an *Alnus/Betula* sister relationship (Kato *et al.*, 1998; G. W. Grimm, pers. observ.).

‡Sampling included multiple accessions from the same species or individuals.

§One plastid gene and two intergenic spacer regions (*matK*, *trnL-trnF*, *trnH-psbA*).

¶In the parsimony tree shown in Li (2008), *Corylus* and *Ostryopsis* are swapped, the resulting *Corylus/Carpinus/Ostrya* clade receiving a BS_P of 56.

successive sisters to the remaining genera (Li *et al.*, 2004: nuclear 18S rDNA, mitochondrial *matR* and four plastid markers; Table 1). Species of *Ostrya* and *Carpinus* were variously intermixed in nuclear and plastid trees (Yoo & Wen, 2002, 2007; Li, 2008). No study so far has sampled more than one species of *Ostryopsis* (a third species, *O. intermedia* B.Tian & J.Q.Liu, was described by Tian, Liu & Liu, 2010).

Molecular data show that the closest relatives of Betulaceae are Ticodendraceae, a monotypic family ranging from Panama to Mexico, and Casuarinaceae, with 69 species in Australia and 27 in Malesia (Li *et al.*, 2004). All three families have good fossil records (Crane, 1989; Scriven & Hill, 1995; Chen *et al.*, 1999; Manchester, Pigg & Crane, 2004; Manchester, 2011). The first *Alnus*-type pollen is from the Upper Cretaceous (Coniacian, 86.3–89.8 Myr; Konzalova, 1971; Forest *et al.*, 2005; absolute ages from the chronostratigraphic chart of Cohen, Finney & Gibbard, 2012) and the oldest macrofossils assignable to the family are nutlets of *Palaeocarpinus* Crane from the late Palaeocene (56–59.2) of southern England and North Dakota (Crane, 1981; Crane, Manchester & Dilcher, 1990). These nutlets resemble *Ostrya* and *Corylus*, and do not represent any extant genus (Sun & Stockey, 1992). An extinct lineage close to the root of the clade (formed by Betulaceae, Casuarinaceae and Ticodendraceae) may be represented by *Endressianthus* Friis, Pedersen & Schoenenberger, from the Late Cretaceous of Portugal (Santonian/Campanian, 71 Myr; Friis, Pedersen & Schöenenberger, 2003).

The relatively good fossil record for Betulaceae has encouraged molecular-clock studies comparing the plastid substitution rates in the clade with those in other angiosperms (Bousquet *et al.*, 1992) and studies that inferred divergence times among the genera (Forest *et al.*, 2005). Assuming a split of *Alnus* and *Betula* at either 80 or 45 Myr, Bousquet *et al.* inferred *rbcL* rates of 0.37 or 0.67×10^{-4} substitution per site per million years (and similar rates when they used coryloid fossils as calibrations). Forest *et al.* applied non-parametric rate smoothing (Sanderson, 1997) and five Betulaceae fossils as fixed calibrations, using one at a time, either placed at crown group nodes or at stem lineage nodes. Many of the inferred ages exceeded oldest fossil occurrences of the relevant clades. For instance, the Betulaceae crown group was estimated at 25 Myr older than the first Betulaceae-type pollen from the Coniacian (86.3–89.8 Myr). However, these estimates were based on just 462 nucleotides of ITS and the age for each node was the median of the ages obtained from ten alternative fossil constraints, including five where ancient fossils were assigned to the crown groups of the living genera *Alnus*, *Betula*, *Carpinus*, *Corylus* and *Ostrya*. The combination of a small amount of sequence data

with so-called ‘consensus estimates’ from ten vastly divergent crown and stem constraints throws doubt on the inferred ages.

Taking advantage of the Betulaceae DNA sequences now available in GenBank, we compiled a matrix of up to 900 sequence accessions and 9300 nucleotides of nuclear and plastid DNA. Species consensus sequences were then computed to obtain less patchy matrices that strike a balance between species sampling and nucleotide sampling. We also added new sequences for *Ostryopsis*. By representing the genetic variation in each genus (judged from hundreds of sequences), we hoped to better resolve genus and subfamily relationships in Betulaceae. To infer divergence times, we used relaxed and strict clock methods with prior age probability distributions on stem lineage fossils, rather than fixed calibrations on either stem or crown groups. We also wanted to test whether plastid substitution rates inferred 20 years ago (Bousquet *et al.*, 1992) would hold up with more sophisticated clock approaches and greater taxon and gene sampling and if there were any drastic shifts in substitution rates (i.e. branch length differences).

MATERIAL AND METHODS

SAMPLING OF TAXA AND GENETIC MARKERS, DNA ALIGNMENT

Species names, GenBank accession numbers and vouchers (where available) are listed in the nexus-formatted single-partition matrices included in the online supporting archive (OSA) hosted at <http://www.palaeogrimm.org/data>. The sources of our new sequences (accession numbers KC412166–KC412181) are *Ostryopsis davidiana* Decne. from Huhehaote, Inner Mongolia, 40°54′/111°34′, alt. 1280 m, Liu Jianquan 152-3 (LZU) and *Ostryopsis nobilis* Balf.f. & W.W.Sm., Daju, Yunnan, 27°16′/100°13′, alt. 1910 m, Liu Jianquan 53-7 (LZU).

GenBank data were harvested and processed using GBK2FAS (Göker *et al.*, 2009) and alignments were carried out with MAFFT ver. 5 (standard settings; Katoh *et al.*, 2005), followed by a visual check for inconsistencies or erroneous sequences (details are described in the Supporting Information, Appendix S1). The full alignments comprised 230 sequences for *atpB-rbcL* and *rbcL*, 49 for granule-bound starch synthase (GBSSI), 502 for ITS (with many Betulaceae represented by multiple sequences), 252 for *trnH-psbA*, 75 for *rpl16*, 395 for *trnK/matK* and 146 for the *trnL* region. There were no statistically supported [$> 80\%$ maximum likelihood (ML) bootstrap support] conflicting topological placements in trees generated from the individual markers. We also aligned and studied sequences of the nuclear-encoded 18S rDNA

and nitrate reductase (*NIA*), mitochondrial *matR* and the plastid *atpB* gene, but did not use them in the final analyses because they contained no genus-level signal or, in the case of *NIA*, were too variable to be aligned between the genera. Instead of choosing a single placeholder accession for each species, species consensus sequences were generated with G2CEF (Göker & Grimm, 2008), using the option 'strict' and gaps treated as missing data. We then built a 46-taxon, 9321-nucleotide matrix that includes five *Alnus* spp., nine *Betula* spp., nine *Corylus* spp., five *Ostrya* spp., two *Ostryopsis* spp. and 14 *Carpinus* spp., giving a total of 44 Betulaceae, which strikes a balance between species sampling and nucleotide sampling: The matrix contained 32.37% empty cells (gaps or missing data). The matrix Delta Value (mDV; Holland *et al.*, 2002; calculated using DISTSTATS (Auch *et al.*, 2006) of the concatenated 46-species matrix is low (0.18), indicating a high tree-likeness of the signal contained in the matrix (mDV).

For dating, we built a reduced matrix of 29 species by selecting up to five species per genus based on their individual Delta Values (iDV; Auch *et al.*, 2006; Göker & Grimm, 2008; see also Supporting Information, Appendix S2), computed from a pairwise model-based (HKY + Γ) distance matrix (see dating). Sequences with low iDV behave in a more tree-like fashion, whereas sequences with high iDV produce topological incongruence or indecisiveness. Earlier studies have established the monophyly of *Alnus* (Chen & Li, 2004: 34 of the estimated 35 species sampled for ITS), *Corylus* (Forest & Bruneau, 2000: 15 species sampled for 5S ribosomal DNA; Erdogan & Mehlenbacher, 2000: 12 species sampled for nuclear and plastid markers; Whitcher & Wen, 2001: 13 species sampled for ITS) and *Betula* (Järvinen *et al.*, 2004: 16 species sampled for a nuclear and a plastid locus; Li *et al.*, 2005: 34 species sampled for ITS).

PHYLOGENETIC ANALYSES

Trees were rooted on Casuarinaceae and Ticodendraceae. The individual data partitions did not yield genus-level topological contradictions and the markers were therefore concatenated (see previous section). Phylogenetic trees were estimated using ML optimization in RAxML-HPC ver. 7.2.6 (Stamatakis, 2006b) and Bayesian optimization (BI) in BEAST ver. 1.74 (Drummond *et al.*, 2012). The ML analyses used the per site rate model (originally labelled 'CAT' model), an approximation of the GTR + Γ model (Stamatakis, 2006a) with 25 rate categories, independent models for each data partition and model parameters estimated over the duration of specified runs. Final model parameters and likelihood were optimized under a GTR + Γ model.

Statistical support came from the fast implementation (Stamatakis, Hoover & Rougemont, 2008) of non-parametric bootstrapping (Felsenstein, 1985) in RAxML, with the number of replicates determined by the extended majority rule consensus bootstrap criterion (Pattengale *et al.*, 2009). Competing bootstrap support for alternate phylogenetic splits was investigated with the consensus network module implemented in SplitsTree 4 (Holland & Moulton, 2003; Huson & Bryant, 2006), with edge weights set to 'COUNT' ('bipartition networks', Grimm *et al.*, 2006). Alternative topologies, generated using Mesquite 2.75 (Maddison & Maddison, 2011) were tested using the Shimodaira–Hasegawa test (Shimodaira & Hasegawa, 1999) implemented in RAxML.

Bayesian tree searches relied on the uncorrelated lognormal relaxed clock, the HKY + Γ substitution model with four rate categories, and a Yule tree prior. Markov chain Monte Carlo (MCMC) were run for 10 million generations, with parameters sampled every 1000 generations. Log files were then analysed with Tracer ver. 1.5 (<http://beast.bio.ed.ac.uk/>) to assess convergence and to confirm that the effective sample sizes for all parameters were > 200, indicating that MCMC chains were run long enough to reach stationarity. After discarding *c.* 25% of the saved trees as burn-in, maximum clade credibility trees with median branch lengths based on the remaining trees were produced using TreeAnnotator (part of the BEAST package) and FigTree ver. 1.3.1 (<http://tree.bio.ed.ac.uk/>).

MOLECULAR CLOCK CALIBRATIONS

Three fossil constraints were used in alternative runs and under strict or relaxed clock models, in each case using gamma prior distributions, with the fossil age as the offset and a shape and scale of 1, which allowed some proportion of nodes to be 4–5 Myr older than the offset. First, 71 Myr-old flower fossils (*Endressianthus miraensis* Friis, Pedersen & Schoenenberger and *E. foveocarpus* Friis, Pedersen & Schoenenberger) from Portugal, which are 'particularly close to members of the Betulaceae and may represent an extinct lineage at the root of the Betulaceae' (Friis *et al.*, 2003: S201), were used as a minimum constraint on the Ticodendraceae/ Betulaceae node. These fossils are of Campanian–Maastrichtian age (Friis *et al.*, 2003). Second, we constrained the split between *Alnus* and *Betula* to minimally late Palaeocene (58 Myr), based on the earliest *Alnus*-type pollen from northern Bohemia (Konzalova, 1971). Fruiting *Alnus* and *Betula* material is not known until the mid-Eocene (48 Mya; Chen *et al.*, 1999). Third, the crown age of Coryloideae was constrained to a minimal 56 Myr old based on the extinct genus

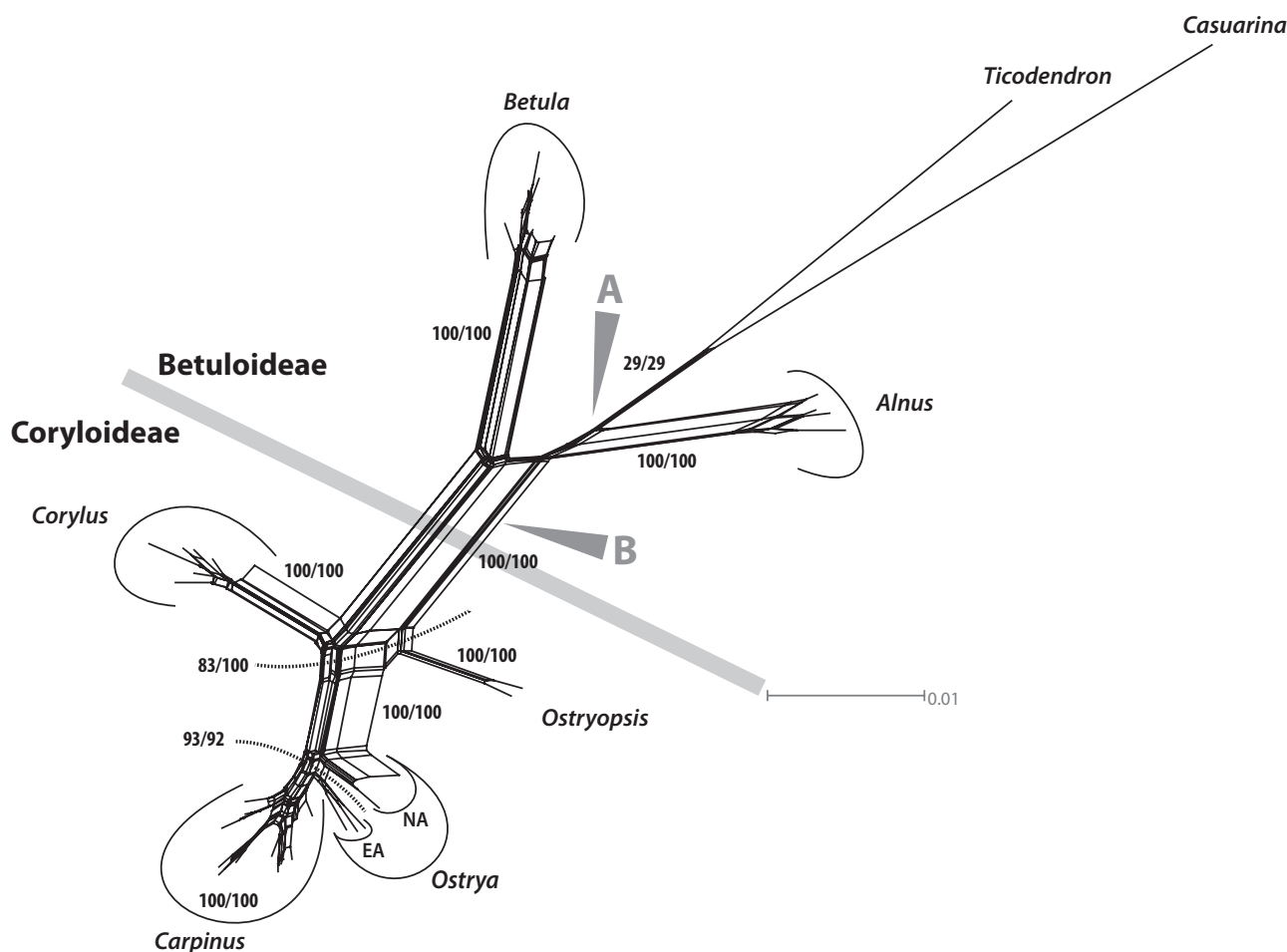


Figure 1. NeighborNet splits graph from the 46-taxon matrix of 9321 aligned nucleotides of nuclear and plastid sequences. Values at edges refer to maximum likelihood (ML) bootstrap support from 350/1000 replicates, with the first value obtained from the 46-taxon matrix, the second from the 29-taxon matrix. 'A' indicates the outgroup-inferred root from ML tree searches and 'B' that from the Bayesian inference (compare Fig. 3). NA, North American spp. of *Ostrya*; EA, Eurasian *Ostrya* spp.

Cranea Manchester & Chen, known from late Palaeocene (56–59.2 Myr) localities of the Fort Union Formation in Wyoming and from a lower Eocene (41–56 Myr) locality in the Big-Horn Basin (Manchester & Chen, 1998). Parsimony analyses of morphological characters cited in Forest *et al.* (2005) placed *Cranea* as sister to *Carpinus*, *Ostrya* and *Ostryopsis*. Another extinct genus of Coryloideae is *Palaeocarpinus* Crane, known from the Palaeocene of North America, southern England and China (Sun & Stockey, 1992; Manchester & Guo, 1996; Manchester *et al.*, 2004). The involucre and nuts of *Palaeocarpinus* are similar to those of extant *Ostryopsis*, and the male catkins have triporate pollen so that 'the reconstructed plant conforms to the subfamily Coryloideae but cannot be placed in a modern genus' (Manchester & Chen, 1998: 522).

RESULTS

RELATIONSHIPS, DIVERGENCE TIMES AND SUBSTITUTION RATES IN BETULACEAE

A NeighborNet shows the genetic distinctness of five of the six genera (Fig. 1). Edge lengths mirror the ML bootstrap and posterior probability (PP) values obtained in tree inference. A long, tree-like portion separates Betuloideae and Coryloideae (Figs 1, 2; see also Supporting Information, Appendix S3). Nuclear GBSSI sequences (Fig. 3) and an ML tree from the plastid data support the genera as mutually monophyletic, with bootstrap support of 92% for *Ostrya* (five of 5–9 species sampled) and 65% for *Carpinus* (14 of 26–35 species sampled; Appendix S4a). The best-scoring ML trees from the 29- and 46-taxon matrices (see also Supporting Information, Appen-

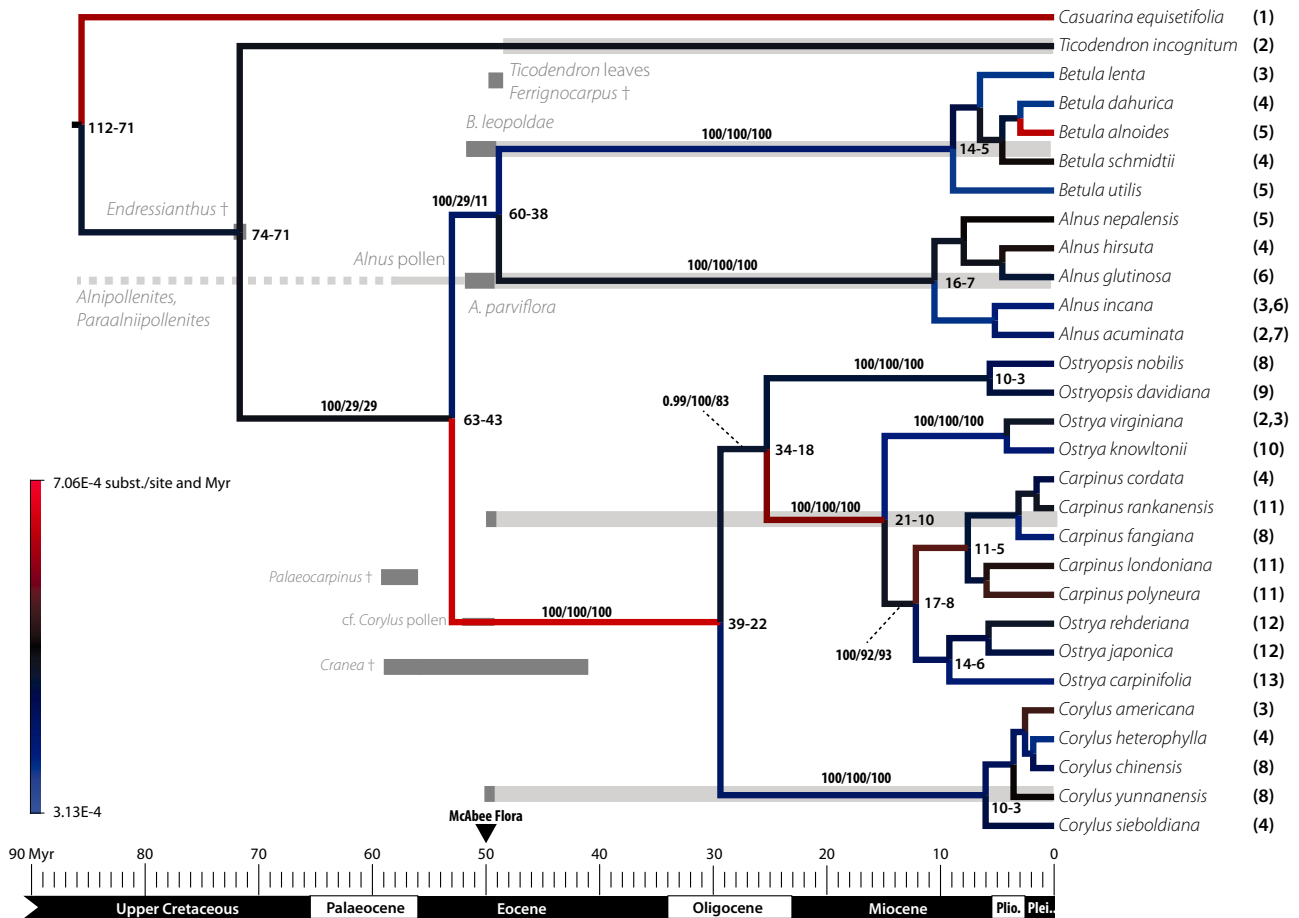


Figure 2. A chronogram for Betulaceae obtained under a relaxed clock model applied to an alignment of 29 species and 9321 aligned nucleotides (MATERIAL AND METHODS). Labels above or below branches indicate support from, in the following order, Bayesian analysis of the 29-taxon matrix and maximum likelihood (ML) bootstrapping of the 29- and 46-taxon matrix. Labels at nodes indicate 95% posterior probability intervals. The geological time scale is in million years and follows Cohen *et al.* (2012); ranges of modern genera and extinct fossil lineages are shown by grey bars. The geographical origin of each accession is shown after the species name: 1, Australasia; 2, Central America (including Mexico); 3, eastern North America; 4, (cool) temperate north-eastern Asia (including Siberia, Russian Far East, Mongolia, north-eastern China, North Korea, and northern Japan); 5, Himalayan foothills and adjacent mountain areas in South-East Asia and south-western China; 6, (temperate) western Eurasia; 7, Andes; 8, south-western China (Yunnan, Sichuan, Guizhou); 9, mountainous areas in north central China; 10, Rio Grande valley, North America; 11, mountainous areas in northern Vietnam and/or southern China (including Hainan and Taiwan); 12, (warm) temperate East Asia (central and eastern China, Korea, Japan); 13, Eastern Mediterranean hinterland.

dix S3) placed *Alnus* as sister to all remaining Betulaceae and *Corylus* as sister to *Ostryopsis*, but with low bootstrap support. Investigation of the bootstrap support for alternative placements (Table 2) revealed weak support for *Alnus* and *Betula* as sister groups from the concatenated data, but strong (83/99.8) support for *Ostryopsis* as sister clade to *Carpinus/Ostrya*. A Shimodaira–Hasegawa test showed that the placement of *Alnus*, *Betula* and *Ostryopsis* preferred by the Bayesian analysis (below) was not significantly worse than the topology of the best-scoring ML tree (Table 3).

Divergence times and their 95% confidence intervals for important nodes under strict and relaxed clock models with alternative fossil calibrations are shown in Table 4. Figure 2 shows the chronogram obtained under a relaxed clock calibrated with *Endressianthus* (71 Myr at the node marked in Fig. 2). The ages of the most recent common ancestors of the extant species of the six genera range from the mid- to late Miocene (Fig. 2, Table 4). The split between the two subfamilies is inferred as Palaeocene/Eocene (63–43 Myr) and the divergence between *Alnus* and *Betula* occurred soon thereafter (60–38 Myr). The divergence between

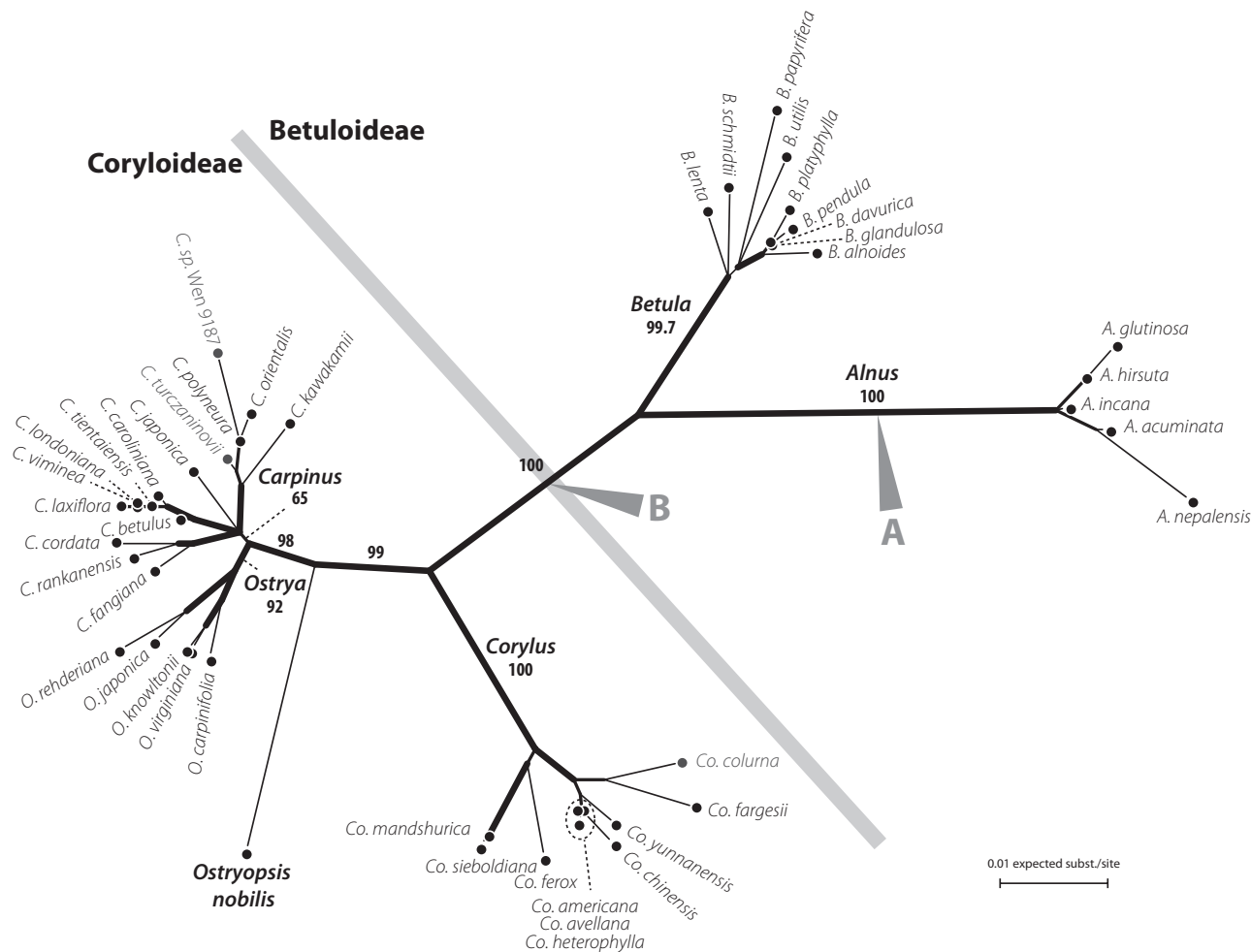


Figure 3. Unrooted maximum likelihood (ML) tree based on all available nuclear GBSSI sequences [matrix included in online supporting archive (OSA); referenced as K.-O. Yoo and J. Wen, unpublished (submitted 2008) in GenBank]. Species of *Carpinus* and *Ostrya* form sister clades, with moderate to strong support. Edge thickness represents low (< 50, thin), moderate (50–90, medium) and strong bootstrap support (> 90; thick lines). Exact values indicated for interior edges. The taxon sampling is comparable with that of the concatenated 46-taxon matrix. Species not covered by the 46-taxon matrix are marked in grey. Two in-group species (*Carpinus tschonoskii* and *Ostryopsis davidiana*) and the two out-groups in the 46-taxon matrix lack GBSSI sequences. 'A' indicates the outgroup-inferred root from ML tree searches and favoured by the Fagales data set of Li *et al.* (2004); 'B' indicates the root here inferred under Bayesian inference and favoured by most other studies with low to moderate support (compare Fig. 2; Tables 1 and 2).

Corylus and the remaining Coryloideae is much younger, dating to the early Oligocene (39–22 Myr). The two species of *Ostryopsis* diverged from each other *c.* 5.6 Myr ago. In relaxed clock runs, the ucd.stdev and coefficient of variation parameter both were 0.34 or 0.35, indicating that the 29-taxon–9321-nucleotide matrix is relatively clock-like. Substitution rates are illustrated by the blue and red colouring of branches in Figure 3. There is an increase along the stem lineage of Coryloideae followed by a slowdown in their crown group to the level observed in Betuloideae. Median substitution rates range from 0.7×10^{-5} at the coryloid root to 0.3×10^{-5} in some branches of *Alnus*,

Betula and *Corylus*, similar to plastid substitution rates of 0.67 or 0.37×10^{-5} substitutions/site/Myr inferred by Bousquet *et al.*, (1992).

DISCUSSION

RELATIONSHIPS IN BETULACEAE

The relationships among the six genera of Betulaceae found under Bayesian (Fig. 2), maximum likelihood [Fig. 2; Supporting Information, Appendix S3; online supporting archive (OSA)] and distance optimization (Fig. 1; Supporting Information, Appendix S2)

Table 2. Maximum likelihood (ML) bootstrap support for Betulaceae generic relationships in the 46-taxon and 29-taxon data sets

Phylogenetic split*	29-taxon set	46-taxon set	<i>atpB-rbcL</i> <i>rbcL</i> gene	GBSSI	ITS	<i>trnH-psbA</i>	<i>rpl16</i>	<i>trnK/matK</i>	<i>trnL/trnL-trnF</i>
Mutual monophyly of Betuloideae and Coryloideae									
Maximum likelihood									
<i>Alnus</i> sister to other Betulaceae	< 5	< 1	35	~	18	10	~	40	27
<i>Betula</i> sister to other Betulaceae	48	61	< 10	~	22	< 5	~	25	10
Bayesian inference									
Betuloideae clade	29	11	< 10	~	60	83	~	32	38
Position of <i>Ostryopsis</i> in Coryloideae									
Maximum likelihood									
<i>Ostryopsis</i> sister to <i>Corylus</i>	< 1	13	< 10	< 1	31	13	17	< 5	< 10
Bayesian inference									
<i>Ostryopsis</i> sister to <i>Carpinus/Ostrya</i>	99.8	83	53	99	45	19	58	95	79
<i>Ostryopsis</i> sister to other Coryloideae	< 1	< 5	27	< 1	23	45	15	< 5	< 10

*Flag (maximum likelihood, Bayesian inference) indicates maximum likelihood-preferred (best-scoring 29-/46-taxon trees) and Bayesian-preferred (Fig. 3) phylogenetic splits.

GBSSI, granule-bound starch synthase; ITS, internal transcribed spacer.

support the mutual monophyly of the two subfamilies more clearly than did earlier studies (summarized in Table 1). An analysis of relationships in Fagales, however, had found *Alnus* and *Betula* as successive sisters to the remaining genera, instead of as sister groups (Li *et al.*, 2004). With the current level of DNA sampling, this question may not be resolvable, especially as long-branch attraction from the outgroup *Ticodendron* may be affecting the placement of *Alnus* (Table 2; Fig. 1).

The slightly contradictory signal coming from the plastid genes, non-coding plastid regions and nuclear spacers and introns can be seen by comparing the placements of *Ostryopsis* relative to *Corylus* and *Carpinus/Ostrya* in Figures 1, 2 and 3. The latter two genera differ mainly in their involucre morphology, which in *Carpinus* consists of a prominent unilateral wing (Manchester & Crane, 1987: fig. 23), in *Ostrya* of an utriculate envelope that completely encloses the nutlet (Manchester & Crane, 1987: fig. 24). The nuclear GBSSI sequences support the mutual monophyly of the two genera (see also Supporting Information, Appendix S4a,b), whereas nuclear nitrate reductase gene sequences indicate that *Ostrya* may be nested in *Carpinus* (Li, 2008). We found that the latter marker cannot unambiguously be aligned across genera of Betulaceae (see also NEXUS file in OSA, <http://www.palaeogrimm.org/data>).

RECONCILING A DEEP FOSSIL RECORD AND RELATIVELY YOUNG MOLECULAR DIVERGENCE TIMES OF EXTANT BETULACEAE

Inferring divergence times in Betulaceae has proved more challenging than one would expect from their good fossil record. Angiosperm-wide molecular dating efforts placed the split between Betulaceae and Casuarinaceae at 35–37 or 27–29 Myr and that between *Alnus* and *Betula* at 19–25 or 18–20 Myr (Wikström, Savolainen & Chase, 2001; Bell, Soltis & Soltis, 2010), dates that are much younger than those inferred from fossils, which show that Casuarinaceae, Betulaceae and Ticodendraceae had differentiated by 86 and 71 Mya (Crane *et al.*, 1990; Scriven & Hill, 1995; Chen *et al.*, 1999; Friis *et al.*, 2003; Manchester, 2011). Both clock studies used nuclear 18S rDNA and plastid *rbcL* and *atpB* genes, with the first relying on a single calibration point (the divergence of Cucurbitales and Fagales) and the second on 36 calibration points. The reason for the too-young ages probably lies in the short branch lengths from *rbcL* and *atpB*, and the near absence of a phylogenetic signal in the 18S data for Fagales, meaning this marker cannot be used to infer divergence times from genetic branch lengths. *RbcL* shows 16, *atpB* 14 substitutions between the outgroup *Ticodendron* and the

Table 3. Results of the Shimodaira–Hasegawa test based on the 46-taxon data set. The Bayesian topology is shown in Fig. 2

Topology	ln L (GTR + Γ)	D(L)	Standard deviation	Significantly worse ($P < 0.01$) than best-scoring maximum likelihood tree
Best-scoring maximum likelihood tree	-23297.441657	~	~	~
Alternative A: Betuloideae clade	-23299.413332	-1.971675	2.680129	No
Alternative B: <i>Ostryopsis</i> sister to <i>Carpinus/Ostrya</i>	-23301.058958	-3.617301	5.783962	No
Bayesian-preferred topology (= A and B)	-23299.279698	-1.838041	5.250697	No

ten species of *Betula* and *Alnus* in our matrix, but Wikström *et al.* (2001) and Bell *et al.* (2010) included a single species from each of these genera, resulting in their young inferred divergence times.

Compared with the (too) young ages inferred in these angiosperm-wide studies, the ages inferred in the Betulaceae-focused work of Forest *et al.* (2005) were surprisingly old, in several cases exceeding earliest palaeobotanical occurrences of Betulaceae. The number of species sampled by Forest *et al.* was similar to ours (26 vs. 27 Betulaceae), but only 462 nucleotides were used compared with the 9321 used here, and fossils were applied as single fixed calibration points, rather than as prior probability distributions (such as the gamma distributions used here; MATERIAL AND METHODS). Forest *et al.* also focused on median ages obtained with ten alternative placements of five fossils, meaning that the effect of each fossil entered the median twice, once when that fossil was assigned to the crown and once when it was assigned to the stem. Of the five fossil records these authors used to fix the crown or stem ages (*Alnus* 65.0 Myr; *Betula* 49.0 Myr; *Corylus* 49.0 Myr, *Carpinus* 49.0 Myr, *Ostrya* 33.7 Myr), only the *Alnus* constraint is also used here. Our other constraints were *Endressianthus* (Friis *et al.*, 2003) for the Betulaceae stem lineage and *Cranea* (Manchester & Chen, 1996, Manchester & Chen, 1998) for the Coryloideae stem lineage, because we wanted to use *Corylus*, *Carpinus* and *Ostrya* fossils to cross-validate our DNA-based estimates.

Cross-validation (against other fossil evidence) shows that the inferred divergence times match the fossil record well (Fig. 3). Thus, fruits and leaves assignable to *Ticodendron* Gómez-Laur. & L.D.Gómez (*Ferrignocarpus* Manchester) have been found in early mid-Eocene deposits from Oregon (48.32 Myr) and in the London Clay (c. 50 Myr old; Manchester, 2011). As *Ticodendron* consists of a single surviving species, these fossils cannot be assigned in a DNA phylogenetic tree. Nevertheless, they indicate that the split between Ticodendraceae and Betulaceae must be at least 50 Myr old, which does not conflict with our constraint

of this split to 71 Myr, based on the *Endressianthus* fossil (in one run). Also, the age of 49 (38–60) Myr that we inferred for the *Alnus/Betula* split using the *Endressianthus* calibration fits reasonably well with the earliest *Alnus* foliage and ovulate cones from the 49–52 Myr-old McAbee flora of British Columbia, a site that also yielded leaves, catkins and pollen of *Betula leopoldae* Wolfe & Wehr (Crane & Stockey, 1987; Dillhoff, Leopold & Manchester, 2005). Third, the age of 53 (43–63) Myr we inferred for the Coryloideae stem lineage (when this was not itself constrained) overlaps with the earliest fossils of Coryloideae from the late Palaeocene/Thanetian (56–59.2 Myr). The earliest fossils assigned to *Corylus* and *Carpinus* appear in the early Eocene Klondike Mountain Formation dated to 50–49 Myr (Pigg, Manchester & Wehr, 2003); they were used by Forest *et al.* (2005) to fix crown or stem ages of these genera to 49 Myr. The ages inferred here for the *Corylus* and *Carpinus/Ostrya* stem lineages are 29 (22–39) Myr and 15 (10–21) Myr (Fig. 2, Table 4), 20 and 34 Myr younger than the Klondike fossils, suggesting that these fossils represent extinct precursors.

The main earlier effort to date Betulaceae (Forest *et al.*, 2005) could not use the *Endressianthus* fossil (Friis *et al.*, 2003) as a constraint because of the taxon sampling in that study, which precluded assignment of this fossil (non-parametric rate smoothing does not permit constraining the root itself). The study did, however, use the earliest *Alnus*-type pollen (*Alnipollenites* R.Potonié and *Paraalnipollenites* L.S.Hills & S.Wallace) also used here to fix the age of the *Alnus* stem lineage to 65 Myr (whereas we assigned a gamma prior with an offset at 58 Myr, allowing a proportion of the inferred ages to be 62 Myr or older). With this calibration, Forest *et al.* inferred an age for the Ticodendraceae/ Betulaceae split of 70.8 Myr (their Table 3, using ML branch lengths; 86.5 Myr using ACCTRAN branch lengths and 81.2 Myr using DELTRAN branch lengths), nicely matching the age of *Endressianthus* (71 Myr). The same calibration gave an age for the Coryloideae crown group of 41.7–

Table 4. Estimated node ages in million years (Myr) for selected divergence events under a strict clock model and a relaxed clock model using different calibration fossils as indicated above columns. Ages are in million years and the values in brackets are the 95% posterior probability intervals

Nodes of interest	Molecular clock model									
	Betulaceae stem 71		Betulaceae stem 71		Corylus stem 56		Corylus stem 56		Alnus stem 58	
Betulaceae stem	Strict	Constrained	Relaxed (Fig. 3)	Constrained	Strict	Relaxed	Strict	Relaxed	Strict	Relaxed
Betulaceae crown group	52 (58–46)	46 (52–39)	53 (63–43)	49 (60–38)	154 (178–133)	133 (175–101)	92 (105–80)	85 (102–69)	92 (105–80)	85 (102–69)
<i>Alnus/Betula</i> crown	26 (30–22)	23 (27–19)	49 (60–38)	29 (39–22)	113 (126–99)	98 (124–77)	67 (74–61)	63 (71–58)	67 (74–61)	63 (71–58)
Coryloideae crown			29 (39–22)	25 (34–18)	99 (114–84)	91 (118–68)	Constrained	Constrained	Constrained	Constrained
<i>Ostryopsis</i> stem			25 (34–18)		50 (55–45)	49 (56–41)	30 (35–25)	30 (39–22)	30 (35–25)	30 (39–22)

57.9 Myr (Forest *et al.*; Table 3), in agreement with the Coryloideae fossil record; for example, *Cranea* and *Palaeocarpinus* known from the late Palaeocene (56–59 Myr) and early Eocene (41–56 Myr; Sun & Stockey, 1992; Manchester & Guo, 1996; Manchester *et al.*, 2004). When Forest *et al.* (2005) instead fixed the age of the most recent common ancestor of the six *Alnus* spp. in their 462-nucleotide long ITS tree to 65 Myr, they obtained ages of 130.5 Myr for the Ticodendraceae/ Betulaceae node and 76.9 Myr for the Coryloideae crown group, exceeding the oldest fossils of these clades. Conversely, when they did not fix the ages of *Corylus* and *Carpinus*, but only used their *Alnus/Betula* stem constraint of 65 Myr, they inferred ages of 15.1 Myr for the *Corylus* crown and 24.3 for the *Carpinus* crown, similar to our results. Their oldest inferred ages always resulted from fixing the ages of genus-level crown groups to Eocene times.

Placing loose prior probabilities on stem lineage ages, not crown ages, is usually preferable based on two arguments. First, it is implausible that the most recent common ancestor of a handful of living species that barely differ in >9000 nucleotides of plastid, mitochondrial and nuclear DNA lived 65 Myr ago. Clades that survive that long are either extremely species-rich or show long genetic branches. Second, following Doyle & Donoghue (1993), fossils should be assigned conservatively, i.e. constraining the ages of stem lineages, not crown groups. Placing five fossils at ten nodes and then focusing on the median age obtained with all of them must also be diffusing any correct temporal signal contained in the genetic branch lengths.

SUBSTITUTION RATE SHIFTS IN BETULACEAE AND THEIR RELATION WITH ECOLOGY

Although the inferred median substitution rates (from 0.7×10^{-5} at the coryloid root to 0.3×10^{-5} in some branches of *Alnus*, *Betula* and *Corylus*) are normal for woody angiosperms (Albert *et al.*, 1994), there clearly is an increase along the stem lineage of Coryloideae, followed by a slowdown in their crown group to the level observed in Betuloideae (Fig. 2). The branch length differences (and rate shifts) may have to do with average seed dispersal and gene flow distances that could influence species formation. Pollen and seeds of *Betula* can be carried over significant distances (e.g. Ford, Sharik & Feret, 1983; Hjelmroos, 1991; Matlack, 1991), partly facilitated by climates with continuous ice/snow covers during winter (D- and E-type climates; see also Supporting Information, SI 4), allowing for secondary dispersal (Matlack, 1989). Birch pollen can be detected across Scandinavia before the flowering season of local birch populations (Hjelmroos, 1991; Skjøth *et al.*, 2007) and

backward trajectory analyses suggest pollen dispersal over long distances in north-eastern Europe (Šauliėne & Veriankaite, 2006) and north-western Europe (Skjōth *et al.*, 2008). *Betula* is also the most cold-tolerant genus of Northern Hemisphere Fagales, with some species extending into high-alpine and arctic environments (Dfc, Dfd, ET climates according to Köppen–Geiger; Kottek *et al.*, 2006; see also Supporting Information, SI 4). *Alnus* resembles *Betula* in its pollen and seed dispersal (Ridley, 1930) and *Alnus* spp. also occur in a wide range of climates (see also Supporting Information, SI 4), although they all depend on ready access to ground water (Tallantire, 1992; Prieditis, 1997). These ecological traits may explain why *Alnus* and *Betula* were able to diversify into new habitats (Mai, 1995; Denk & Grimm, 2009).

Corylus, like *Fagus* L. and *Quercus* L., relies on jays for dispersal (*Garrulus*, Corvidae; Haffer & Bauer, 1993) and other birds and small mammals (Ridley, 1930). In contrast, the winged seeds of *Carpinus*, *Ostrya* and *Ostryopsis* are wind-dispersed (Ridley, 1930), although they may not cover the same distances as *Betula* and *Alnus* seeds. Our sparse species sampling, however, precludes a more in-depth analysis. The inferred substitution rate shift in Coryloideae between 25 and 15 Mya (Fig. 2; Table 4) precedes the mid-Miocene climatic optimum (Zachos *et al.*, 2001) and may relate to new niches that became available after the mid-Miocene optimum.

CONCLUSIONS

Using curated GenBank data is becoming ever more important as researchers construct huge matrices from public data repositories (Bininda-Emonds, 2004; McMahon & Sanderson, 2006; Chatterjee *et al.*, 2009). Starting from >900 available sequences for Betulaceae in GenBank, we built 46- and 29-species matrices of 9321 aligned nucleotides that balanced the trade-off between taxon and DNA sampling. However, instead of choosing a single placeholder accession for each species, as is usually done, we computed species consensus sequences and then used the criterion of ‘tree-like signal’ to select the most suitable available species of Betulaceae. The resulting networks and phylogenetic trees suggest that *Ostryopsis* is indeed sister to *Carpinus/Ostrya* and that Betuloideae and Coryloideae are mutually monophyletic, which was not clear before. We used relatively few of the many Betulaceae fossils as constraints, preferring to instead use them to cross-validate inferred node ages. Inferred substitution rate shifts in a few places in the phylogenetic trees may be linked to different rates of allopatric species formation, but this needs to be tested with denser phylogenetic, geographical and ecological species sampling.

ACKNOWLEDGEMENTS

We thank Jianquan Liu for sharing his unpublished sequences of *Ostryopsis*. G.W.G. acknowledges funding by the Vetenskapsrådet (Swedish Research Council), project no. 2008–3726.

REFERENCES

- Albert VA, Backlund A, Bremer K, Chase MW, Manhart JR, Mishler BD, Nixon KC. 1994. Functional constraints and *rbcL* evidence for land plant phylogeny. *Annals of the Missouri Botanical Garden* **81**: 534–567.
- Auch AF, Henz SR, Holland BR, Göker M. 2006. Genome BLAST distance phylogenies inferred from whole plastid and whole mitochondrion genome sequences. *BMC Bioinformatics* **7**: 350.
- Bell CD, Soltis DE, Soltis PS. 2010. The age and diversification of the angiosperms re-revisited. *American Journal of Botany* **97**: 1296–1303.
- Bininda-Emonds ORP. 2004. Trees versus characters and the supertree/supermatrix ‘paradox’. *Systematic Biology* **53**: 356–359.
- Bousquet J, Strauss SH, Li P. 1992. Complete congruence between morphological and *rbcL*-based molecular phylogenies in birches and related species (Betulaceae). *Molecular Biology and Evolution* **9**: 1076–1088.
- Chatterjee H, Ho S, Barnes I, Groves C. 2009. Estimating the phylogeny and divergence times of primates using a supermatrix approach. *BMC Evolutionary Biology* **9**: 259.
- Chen Z-D, Li J-H. 2004. Phylogenetics and biogeography of *Alnus* (Betulaceae) inferred from sequences of nuclear ribosomal DNA its region. *International Journal of Plant Sciences* **165**: 325–335.
- Chen Z-D, Manchester SR, Sun H-Y. 1999. Phylogeny and evolution of the Betulaceae as inferred from DNA sequences, morphology and palaeobotany. *American Journal of Botany* **86**: 1168–1181.
- Cohen KM, Finney S, Gibbard PL. 2012. International chronostratigraphic chart, IUGS, Aug. 2012. Available at: <http://www.stratigraphy.org> Accessed 11 November 2012.
- Crane PR. 1981. Betulaceous leaves and fruits from the British Upper Palaeocene. *Botanical Journal of the Linnean Society* **83**: 103–136.
- Crane PR. 1989. Early fossil history and evolution of the Betulaceae. In: Crane PR, Blackmore S, eds. *Evolution, systematics, and fossil history of the Hamamelidae. Vol. 2: ‘Higher’ Hamamelidae*. Oxford: Clarendon Press, 87–116.
- Crane PR, Manchester SR, Dilcher D. 1990. A preliminary survey of fossil leaves and well-preserved reproductive structures from the Sentinel Butte Formation (Paleocene) near Almont, North Dakota. *Fieldiana Geology, New Series* **20**: 1–63.
- Crane PR, Stockey RA. 1987. *Betula* leaves and reproductive structures from the Middle Eocene of British Columbia, Canada. *Canadian Journal of Botany* **65**: 2490–2500.
- Denk T, Grimm GW. 2009. The biogeographic history of beech trees. *Review of Palaeobotany and Palynology* **158**: 83–100.

- Dillhoff RM, Leopold EB, Manchester SR. 2005.** The McAbee flora of British Columbia and its relation to the early–middle Eocene Okanagan Highlands flora of the Pacific Northwest. *Canadian Journal of Earth Sciences* **42**: 151–166.
- Doyle JA, Donoghue MJ. 1993.** Phylogenies and angiosperm diversification. *Paleobiology* **19**: 141–167.
- Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012.** Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution* **29**: 1969–1973.
- Erdogan V, Mehlenbacher SA. 2000.** Phylogenetic relationships of *Corylus* species (Betulaceae) based on nuclear ribosomal DNA ITS region and chloroplast *matK* gene sequences. *Systematic Botany* **25**: 727–737.
- Felsenstein J. 1985.** Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**: 783–791.
- Ford RH, Sharik TL, Feret PP. 1983.** Seed dispersal of the endangered Virginia round-leaf birch (*Betula uber*). *Forest Ecology and Management* **6**: 115–128.
- Forest F, Bruneau A. 2000.** Phylogenetic analysis, organization, and molecular evolution of the non-transcribed spacer of 5S ribosomal RNA genes in *Corylus* (Betulaceae). *International Journal of Plant Science* **161**: 793–806.
- Forest F, Savolainen V, Chase MW, Lupia R, Bruneau A, Crane PR. 2005.** Teasing apart molecular- versus fossil-based error estimates when dating phylogenetic trees: a case study in the birch family (Betulaceae). *Systematic Botany* **30**: 118–133.
- Friis EM, Pedersen KR, Schönenberger J. 2003.** *Endresianthus*, a new Normapolles-producing plant genus of fagalean affinity from the Late Cretaceous of Portugal. *International Journal of Plant Sciences* **164**: S201–S223.
- Göker M, García-Blázquez G, Voglmayr H, Tellería MT, Martín MP. 2009.** Molecular taxonomy of phytopathogenic fungi: a case study in *Peronospora*. *PLoS ONE* **4**: e6319.
- Göker M, Grimm GW. 2008.** General functions to transform associate data to host data, and their use in phylogenetic inference from sequences with intra-individual variability. *BMC Evolutionary Biology* **8**: 86.
- Grimm GW, Renner SS, Stamatakis A, Hemleben V. 2006.** A nuclear ribosomal DNA phylogeny of *Acer* inferred with maximum likelihood, splits graphs, and motif analyses of 606 sequences. *Evolutionary Bioinformatics* **2**: 279–294.
- Haffer J, Bauer KM. 1993.** Corvidae–Rabenvögel. In: Glutz von Blotzheim UN, Bauer KM, eds. *Handbuch der Vögel Mitteleuropas. Band 13-III. Passeriformes (4. Teil)*. Wiesbaden: Aula, 1375–2022.
- Hjelmroos M. 1991.** Evidence for long-distance transport of *Betula* pollen. *Grana* **30**: 215–228.
- Holland B, Moulton V. 2003.** Consensus networks: a method for visualising incompatibilities in collections of trees. In: Benson G, Page R, eds. *Algorithms in bioinformatics: Third International Workshop, WABI, Budapest, Hungary. Proceedings*. Berlin, Heidelberg, Stuttgart: Springer Verlag, 165–176.
- Holland BR, Huber KT, Dress A, Moulton V. 2002.** Delta Plots: a tool for analyzing phylogenetic distance data. *Molecular Biology and Evolution* **19**: 2051–2059.
- Huson DH, Bryant D. 2006.** Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* **23**: 254–267.
- Järvinen P, Palme A, Morales LO, Lannenpaa M, Keinanen M, Sopanen T, Lascoux M. 2004.** Phylogenetic relationships of *Betula* species (Betulaceae) based on nuclear ADH and chloroplast *matK* sequences. *American Journal of Botany* **91**: 1834–1845.
- Kato H, Oginuma K, Gu Z, Hammel B, Tobe H. 1998.** Phylogenetic relationships of Betulaceae based on *matK* sequences, with particular reference to the position of *Ostryopsis*. *Acta Phytotaxonomica et Geobotanica* **49**: 89–97.
- Katoh K, Kuma K, Toh H, Miyata T. 2005.** MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Research* **33**: 511–518.
- Konzalova M. 1971.** *Paraalnipollenites* Hills and Wallace 1969, in the Turonian of the upper Cretaceous of North Bohemia. *Vestník Ustředního Geologického ústavu* **46**: 39–40.
- Kottek M, Grieser J, Beck C, Rudolf B, Rubel F. 2006.** World map of the Köppen–Geiger climate classification updated. *Meteorologische Zeitschrift* **15**: 259–263.
- Kubitzki K. 1993.** Betulaceae. In: Kubitzki K, ed. *The families and genera of vascular plants*. Berlin: Springer, 152–157.
- Li J. 2008.** Sequences of low-copy nuclear gene support the monophyly of *Ostrya* and paraphyly of *Carpinus* (Betulaceae). *Journal of Systematics and Evolution* **46**: 333–340.
- Li J, Shoup S, Chen Z. 2005.** Phylogenetics of *Betula* (Betulaceae) inferred from sequences of nuclear ribosomal DNA. *Rhodora* **107**: 69–86.
- Li J, Shoup S, Chen Z. 2007.** Phylogenetic relationships of diploid species of *Betula* (Betulaceae) inferred from DNA sequences of nuclear nitrate reductase. *Systematic Botany* **32**: 357–365.
- Li R-Q, Chen Z-D, Lu A-M, Soltis S, Soltis DE, Manos PS. 2004.** Phylogenetic relationships in Fagales based on DNA sequences from three genomes. *International Journal of Plant Science* **165**: 311–324.
- McMahon MM, Sanderson MJ. 2006.** Phylogenetic supermatrix analysis of GenBank sequences from 2228 papilionoid legumes. *Systematic Biology* **55**: 818–836.
- Maddison WP, Maddison DR. 2011.** Mesquite: a modular system for evolutionary analysis, v. 2.75. Available at: <http://mesquiteproject.org>.
- Mai DH. 1995.** Tertiäre Vegetationsgeschichte Europas *Jena, Stuttgart*. New York: Gustav Fischer Verlag.
- Manchester SR. 2011.** Fruits of Ticodendraceae (Fagales) from the Eocene of Europe and North America. *International Journal of Plant Sciences* **172**: 1179–1187.
- Manchester SR, Chen Z-D. 1996.** *Palaeocarpinus aspinosa* sp. nov. (Betulaceae) from the Paleocene of Wyoming, USA. *International Journal of Plant Sciences* **157**: 644–655.
- Manchester SR, Chen Z-D. 1998.** A new genus of Coryloideae (Betulaceae) from the Paleocene of North America. *International Journal of Plant Sciences* **159**: 522–532.
- Manchester SR, Crane PR. 1987.** A new genus of Betulaceae from the Oligocene of western North America. *Botanical Gazette* **148**: 263–273.

- Manchester SR, Guo S-X. 1996.** *Paleocarpinus* (extinct Betulaceae) from northwestern China: new evidence for Paleocene floristic continuity between Asia, North America, and Europe. *International Journal of Plant Sciences* **157**: 240–246.
- Manchester SR, Pigg KB, Crane PR. 2004.** *Palaeocarpinus dakotensis* sp. n. (Betulaceae: Coryloideae) and associated staminate catkins, pollen, and leaves from the Paleocene of North Dakota. *International Journal of Plant Sciences* **165**: 1135–1148.
- Matlack GR. 1989.** Secondary dispersal of seed across snow in *Betula lenta*, a gap-colonizing tree species. *Journal of Ecology* **77**: 853–869.
- Matlack GR. 1991.** Influence of fruit size and weight on wind dispersal in *Betula lenta*, a gap-colonizing tree species. *American Midland Naturalist* **128**: 30–39.
- Pattengale ND, Masoud A, Bininda-Emonds ORP, Moret BME, Stamatakis A. 2009.** How many bootstrap replicates are necessary? In: Batzoglou S, ed. *RECOMB 2009*. Berlin, Heidelberg: Springer-Verlag, 184–200.
- Pigg KB, Manchester SR, Wehr WC. 2003.** *Corylus*, *Carpinus*, and *Palaeocarpinus* (Betulaceae) from the middle Eocene Klondike Mountain and Allenby formations of northwestern North America. *International Journal of Plant Science* **164**: 807–822.
- Prieditis N. 1997.** *Alnus glutinosa* – dominated wetland forests of the Baltic Region: community structure, syntaxonomy, and conservation. *Plant Ecology* **129**: 49–94.
- Ridley HN. 1930.** *The dispersal of plants throughout the world*. Ashford: L. Reeve & Co. Ltd..
- Sanderson MJ. 1997.** A nonparametric approach to estimating divergence times in the absence of rate constancy. *Molecular Biology and Evolution* **14**: 1218–1232.
- Šauliėnė I, Veriankaite L. 2006.** Application of backward air mass trajectory analysis in evaluation airborne pollen dispersion. *Journal of Environmental Engineering and Landscape Management* **14**: 113–120.
- Scriven LJ, Hill RS. 1995.** Macrofossil Casuarinaceae: their identification and the oldest macrofossil record, *Gymnostoma antiquum* sp. nov., from the late Paleocene of New South Wales, Australia. *Australian Systematic Botany* **8**: 1035–1053.
- Shimodaira H, Hasegawa M. 1999.** Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Molecular Biology and Evolution* **16**: 1114–1116.
- Skjøth CA, Smith M, Brandt J, Emberlin J. 2009.** Are the birch trees in southern England a source of *Betula* pollen for North London? *International Journal of Biometeorology* **53**: 75–86.
- Skjøth CA, Sommer J, Stach A, Smith M, Brandt J. 2007.** The long-range transport of birch (*Betula*) pollen from Poland and Germany causes significant pre-season concentrations in Denmark. *Clinical and Experimental Allergy* **37**: 1204–1212.
- Stamatakis A. 2006a.** Phylogenetic models of rate heterogeneity: a high performance computing perspective. *Proceedings of 20th IEEE/ACM International Parallel and Distributed Processing Symposium (IPDPS2006), High Performance Computational Biology Workshop*. Rhodos, Greece, April 2006.
- Stamatakis A. 2006b.** RAxML-VI-HPC: maximum-likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**: 2688–2690.
- Stamatakis A, Hoover P, Rougemont J. 2008.** A rapid bootstrap algorithm for the RAxML web servers. *Systematic Biology* **57**: 758–771.
- Sun F, Stockey RA. 1992.** A new species of *Paleocarpinus* (Betulaceae) based on infructescences, fruits, and associated staminate inflorescences and leaves from the Paleocene of Alberta, Canada. *International Journal of Plant Sciences* **153**: 136–146.
- Tallantire PA. 1992.** The alder [*Alnus glutinosa* (L.) Gaertn.] problem in the British Isles: a third approach to its palaeohistory. *New Phytologist* **122**: 717–731.
- Tian B, Liu T-L, Liu J-Q. 2010.** *Ostryopsis intermedia*, a new species of Betulaceae from Yunnan, China. *Botanical Studies* **51**: 257–262.
- Whitcher IN, Wen J. 2001.** Phylogeny and biogeography of *Corylus* (Betulaceae): inference from ITS sequences. *Systematic Botany* **26**: 283–298.
- Wikström N, Savolainen V, Chase MW. 2001.** Evolution of the angiosperms: calibrating the family tree. *Proceedings of the Royal Society London: B, Biological Sciences* **268**: 2211–2220.
- Yoo K-O, Wen J. 2002.** Phylogeny and biogeography of *Carpinus* and subfamily Coryloideae (Betulaceae). *International Journal of Plant Sciences* **163**: 641–650.
- Yoo K-O, Wen J. 2007.** Phylogeny of *Carpinus* and subfamily Coryloideae (Betulaceae) based on chloroplast and nuclear ribosomal sequence data. *Plant Systematics and Evolution* **267**: 25–35.
- Zachos JC, Pagani M, Sloan L, Thomas E, Billups K. 2001.** Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science* **292**: 686–693.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

- Appendix S1.** Notes on data curation, including notes on apparently erroneous sequences in GenBank.
- Appendix S2.** Distance statistics.
- Appendix S3.** The best-scoring maximum likelihood (ML) trees inferred from 29- and 46-taxon matrices.
- Appendix S4.** Geographic and climatic distribution of Betulaceae with a graphical overview (**S4a**) and tabulated data (**S4b**).