

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

Authors: Guido W. Grimm¹ and Susanne S. Renner²

¹*Department of Palaeobotany, Swedish Museum of Natural History, P.O. Box 50007, 10405 Stockholm*

²*Systematic Botany and Mycology, University of Munich (LMU), Menzinger Strasse 67, 80638 Munich, Germany*

Corresponding author's e-mail: guido.grimm@nrm.se

Running head: A new chronogram for the Betulaceae

Abstract

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 relatively small amounts 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 Paleocene, and the most recent common ancestors of each of the living genera to the middle to late Miocene. A substitution rate shift within the Coryloideae between 25 and 15 million years ago precedes the mid-Miocene climatic optimum and may be linked to temperate niches that became available following the mid-Miocene.

Keywords: Fossil calibrations, consensus sequences, GenBank data curation, molecular clocks, substitution rate shift

INTRODUCTION

The Fagales family Betulaceae comprises 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 the island of 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 of overlapping geographic distributions (: *Alnus* Mill. (29–35) and *Betula* (42–50) make up the Betuloideae, and *Carpinus* L. (26–35), *Corylus* L. (16), *Ostrya* Scop. (5–9), and *Ostryopsis* Decne. (3) make up the Coryloideae. Several species of *Alnus*, 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.

In spite of its relatively small size and several molecular-phylogenetic studies, relationships among the subfamilies have remained unclear (Table 1). The first plastid *rbcL* phylogeny for the 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 species of *Ostryopsis* 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 species of *Betula*; Li, 2008: 22 species of Coryloideae). Those of these studies that included outgroups mostly placed the root between the two subfamilies of Betulaceae (Table 1). A study focusing on the entire order Fagales, however, found *Alnus* and *Betula* as 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, Yoo & Wen, 2007, Li, 2008). No study so far has sampled both species of *Ostryopsis* (a third species, *O. intermedia* B.Tian & J.Q.Liu, was described recently Tian *et al.*, 2010).

Molecular data show that the closest relatives of the Betulaceae are the Ticodendraceae, a monotypic family ranging from Panama to Mexico, and the 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* from the late Paleocene (56–59.2) of southern England and North Dakota (Crane, 1981, Crane, Manchester & Dilcher, 1990). These nutlets resemble both *Ostrya* and *Corylus*, and do not represent any extant genus (Sun & Stockey, 1992). An extinct lineage close to the root of the

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

clade (formed by Betulaceae, Casuarinaceae, and Ticodendraceae) may be represented by *Endressianthus*, from the late Cretaceous of Portugal (Santonian/ Campanian, 71 Myr; Friis, Pedersen & Schönenberger, 2003).

The relatively good fossil record of Betulaceae has encouraged molecular-clock studies comparing the clade's plastid substitution rates to 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 Myr or 45 Myr, Bousquet *et al.* inferred *rbcL* rates of 0.37 or 0.67×10^{-4} substitution per site per Myr (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 million years 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 10 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 10 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 found in each genus (judged from hundreds of sequences), we hoped to better resolve genus and subfamily relationships in the 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 larger taxon and gene sampling and if there were any drastic shifts in substitution rates (i.e., branch length differences).

MATERIALS 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 www.palaeogrimm.org/data. The sources of our new sequences (accession numbers KC412166–

1
2 KC412181) are *Ostryopsis davidiana* Decne. are Huhehaote, Inner Mongolia, 40°54'/111°34', alt.
3 1280 m, Liu Jianquan 152-3 (LZU) and *Ostryopsis nobilis* Balf.f. & W.W.Sm., Daju, Yunnan,
4 27°16'/100°13', alt. 1910 m, Liu Jianquan 53-7 (LZU).
5

6
7 GenBank data were harvested and processed using GBK2FAS (Göker *et al.*, 2009), and
8 alignments were done with MAFFT v. 5 (standard settings; Katoh *et al.*, 2005), followed by visual
9 checked for inconsistencies or erroneous sequences (details are described in the supplementary
10 information [SI] 1). The full alignments comprised 230 sequences of *atpB-rbcL* and *rbcL*, 49 of
11 granule-bound starch synthase (GBSSI), 502 of ITS (with many Betulaceae represented by multiple
12 sequences), 252 of *trnH-psbA*, 75 of *rpl16*, and 395 of *trnK/matK*, 146 of the *trnL* region. **There**
13 **were no statistically supported (>80% ML bootstrap support) conflicting topological placements in**
14 **trees generated from the individual markers.** We also aligned and studied sequences of the nuclear-
15 encoded 18S rDNA, nitrate reductase (*NIA*), mitochondrial *matR*, and the plastid *atpB* gene, but did
16 no use them in the final analyses because they contained no genus-level signal or in the case of *NIA*
17 were too variable to be aligned between the genera. Instead of choosing a single placeholder
18 accession for each species, species consensus sequences were generated with G2CEF (Göker &
19 Grimm, 2008), using the option “strict” and gaps treated as missing data. We then built a 46-taxon-
20 9321-nucleotide matrix that includes five species of *Alnus*, nine of *Betula*, nine of *Corylus*, five of
21 *Ostrya*, two of *Ostryopsis*, and 14 of *Carpinus* for a total of 44 Betulaceae, which strikes a balance
22 between species sampling and nucleotide sampling: The matrix contained 32.37% empty cells (gaps
23 or missing data). The matrix Delta Value (mDV; Holland *et al.*, 2002; calculated using DIST_STATS
24 (Auch *et al.*, 2006) of the concatenated 46-species matrix is low (0.18), indicating a high tree-
25 likeness of the signal contained in the matrix (mDV).
26
27

28
29 For dating, we built a reduced matrix of 29 species by selecting up to five species per genus
30 based on their individual Delta Values (iDV; Auch *et al.*, 2006, Göker & Grimm, 2008; SI 2),
31 computed from a pairwise model-based (HKY + Γ) distance matrix (see dating). Sequences with
32 low iDV behave more tree-like, whereas sequences with high iDV produce topological
33 incongruence or indecisiveness. Earlier studies have established the monophyly of *Alnus* (Chen &
34 Li, 2004: 34 of the estimated 35 species sampled for ITS); *Corylus* (Forest & Bruneau, 2000: 15
35 species sampled for 5S; Erdogan & Mehlenbacher, 2000: 12 species samples for nuclear and plastid
36 markers; Whitcher & J., 2001: 13 species sampled for ITS); and *Betula* (Järvinen *et al.*, 2004: 16
37 species sampled for a nuclear and a plastid locus; Li *et al.*, 2005: 34 species sampled for ITS).
38
39

PHYLOGENETIC ANALYSES

40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
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

1
2 previous section). Phylogenetic trees were estimated using ML optimization in RAxML-HPC
3 version 7.2.6 (Stamatakis, 2006b) and Bayesian optimization (BI) in BEAST version 1.74
4 (Drummond and Rambaut, 2012). The ML analyses used the per site rate model (originally labeled
5 “CAT” model), an approximation of the GTR + Γ model (Stamatakis, 2006a) with 25 rate
6 categories, independent models for each data partition, and model parameters estimated over the
7 duration of specified runs. Final model parameters and likelihood were optimized under a GTR + Γ
8 model.
9

10
11
12
13 Statistical support came from the fast implementation (Stamatakis, Hoover & Rougemont,
14 2008) of nonparametric bootstrapping (Felsenstein, 1985) in RAxML with the number of replicates
15 determined by the extended majority rule consensus bootstrap criterion (Pattengale *et al.*, 2009).
16 Competing bootstrap support for alternate phylogenetic splits was investigated with the consensus
17 network module implemented in SplitsTree 4 (Holland & Moulton, 2003, Huson & Bryant, 2006),
18 with edge weights set to “COUNT” (‘bipartition networks’, Grimm *et al.*, 2006). Alternative
19 topologies, generated using Mesquite 2.75 (Maddison & Maddison, 2011) were tested using the
20 Shimodaira-Hasegawa test (Shimodaira & Hasegawa, 1999) implemented in RAxML.
21

22
23
24
25
26
27 Bayesian tree searches relied on the uncorrelated lognormal relaxed clock, the HKY + Γ
28 substitution model with four rate categories, and a Yule tree prior. Monte Carlo Markov chains
29 (MCMC) were run for 10 million generations, with parameters sampled every 1000 generations.
30 Log files were then analyzed with Tracer version 1.5 (<http://beast.bio.ed.ac.uk/>) to assess
31 convergence and to confirm that the effective sample sizes for all parameters were larger than 200,
32 indicating that MCMC chains were run long enough to reach stationarity. After discarding c. 25%
33 of the saved trees as burn-in, maximum clade credibility trees with median branch lengths based on
34 the remaining trees were produced using TreeAnnotator (part of the BEAST package) and FigTree
35 version 1.3.1 (<http://tree.bio.ed.ac.uk/>).
36
37
38
39
40
41
42
43

44 MOLECULAR CLOCK CALIBRATIONS

45
46 Three fossil constraints were used in alternative runs and under strict or relaxed clock models, in
47 each case using gamma prior distributions, with the fossil age as the offset and a shape and scale of
48 1, which allowed some proportion of nodes to be 4–5 Myr older than the offset. First, 71 Myr-old
49 flower fossils (*Endressianthus miraensis* and *E. foveocarpus*) from Portugal, which are
50 “particularly close to members of the Betulaceae and may represent an extinct lineage at the root of
51 the Betulaceae” (Friis *et al.*, 2003: S201), were used as a minimum constraint on the
52 Tiodendraceae/ Betulaceae node. These fossils are of Campanian-Maastrichtian age (Friis *et al.*,
53 2003). Second, we constrained the split between *Alnus* and *Betula* to minimally late Paleocene (58
54 Myr), based on the earliest *Alnus*-type pollen from North Bohemia (Konzalova, 1971). Fruiting
55
56
57
58
59
60

1
2 *Alnus* and *Betula* material is not known until the middle Eocene (48 million years ago; Chen *et al.*,
3 1999). Third, the crown age of Coryloideae was constrained to minimally 56 Myr old based on the
4 extinct genus *Cranea* known from late Paleocene (56–59.2 Myr) localities of the Fort Union
5 Formation in Wyoming and from a lower Eocene (41–56 Myr) locality in the Big-Horn Basin
6 (Manchester & Chen, 1998). Parsimony analyses of morphological characters cited in Forest *et al.*
7 (2005) placed *Cranea* as sister to *Carpinus*, *Ostrya*, and *Ostryopsis*. Another extinct genus of
8 Coryloideae is *Palaecarpinus*, known from Paleocene of North America, Southern England, and
9 China (Sun & Stockey, 1992, Manchester & Guo, 1996, Manchester *et al.*, 2004). The involucre
10 and nuts of *Palaecarpinus* are similar to those of extant *Ostryopsis*, and the male catkins have
11 triporate pollen so that “the reconstructed plant conforms to the subfamily Coryloideae but cannot
12 be placed in a modern genus” (Manchester & Chen, 1998: 522).
13
14
15
16
17
18
19
20
21

22 RESULTS

23 RELATIONSHIPS, DIVERGENCE TIMES AND SUBSTITUTIONS RATES IN THE BETULACEAE

24 A neighbor-net shows the genetic distinctness of five of the six genera (Fig. 1). Edge lengths mirror
25 the ML bootstrap and PP values obtained in tree inference. A long, tree-like portion separates the
26 Betuloideae and Coryloideae (Figs. 1, 2; SI 3). Nuclear GBSSI sequences, and a ML tree from the
27 plastid data (Fig. 3), support the genera as mutually monophyletic with a bootstrap support of 92%
28 for *Ostrya* (5 species of 5–9 species sampled) and 65% for *Carpinus* (14 of 26–35 species sampled;
29 Fig. 4). The best-scoring ML trees from the 29- and 46-taxon matrices (SI 3) placed *Alnus* as sister
30 to all remaining Betulaceae and *Corylus* as sister to *Ostryopsis*, but with low bootstrap support.
31 Investigation of the bootstrap support for alternative placements (Table 2) revealed low support for
32 *Alnus* and *Betula* as sister groups from the concatenated data, but high (83/99.8) support for
33 *Ostryopsis* as sister clade to *Carpinus/Ostrya*. A Shimodaira-Hasegawa test showed that the
34 placement of *Alnus*, *Betula*, and *Ostryopsis* preferred by the Bayesian analysis (below) was not
35 significantly worse than the topology of the best-scoring ML tree (Table 3).
36
37
38
39
40
41
42
43
44

45 Divergence times and their 95% CIs for important nodes under strict and relaxed clock
46 models with alternative fossil calibrations are shown in Table 4. Figure 2 shows the chronogram
47 obtained under a relaxed clock calibrated with *Endressianthus* (71 Myr at the node marked in Fig.
48 2). The ages of the most recent common ancestors of the extant species of the six genera range from
49 the middle to late Miocene (Fig. 2, Table 4). The split between the two subfamilies is inferred as
50 Paleocene/Eocene (63–43 Myr), and the divergence between *Alnus* and *Betula* occurred soon
51 thereafter (60–38 Myr). The divergence between *Corylus* and the remaining Coryloideae is much
52 younger, dating to the early Oligocene (39–22 Myr). The two species of *Ostryopsis* diverged from
53 each other c. 5.6 Myr ago. In relaxed clock runs, the ucl.d.stdev and coefficient of variation
54
55
56
57
58
59
60

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 coloring of branches in Figure 3. There is an increase along the stem lineage of the Coryloideae followed by a slowdown in their crown group to the level observed in the 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} subs/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; SI 3; OSA), and distance optimization (Fig. 1; SI 2) support the mutual monophyly of the two subfamilies more clearly than did earlier studies (summarized in Table 1). An analysis of relationships in the order Fagales, however, had found *Alnus* and *Betula* as successive sisters to the remaining genera, instead of as sister groups to each other (Li *et al.*, 2004). We the current amount of DNA data, 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 (fig. 23 in Manchester & Crane, 1987), in *Ostrya* of an utriculate envelope that completely encloses the nutlet (fig. 24 in the same study). The nuclear GBSSI sequences support the mutual monophyly of the two genera (Fig. 4), while nuclear nitrate reductase gene sequences indicate that *Ostrya* may be nested within *Carpinus* (Li, 2008). We found that the latter marker cannot unambiguously be aligned across genera of Betulaceae (see NEXUS file; Supplementary Information).

RECONCILING A DEEP FOSSIL RECORD AND RELATIVELY YOUNG

MOLECULAR DIVERGENCE TIMES OF EXTANT BETULACEAE

Inferring divergence times in the Betulaceae has proven 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 Myr ago (Crane *et al.*, 1990, Scriven & Hill, 1995, Chen *et al.*, 1999, Friis *et al.*, 2003, Manchester, 2011). Both clock studies used nuclear 18S rDNA

1 and plastid *rbcL* and *atpB* genes, with the first relying on a single calibration point (the divergence
2 of Cucurbitales and Fagales), the second on 36 calibration points. The reason for the too young ages
3 probably lies in the near absence of phylogenetic signal in the 18S data for Fagales, meaning this
4 marker cannot be used to infer divergence times from genetic branch lengths, and the short branch
5 lengths from *rbcL* and *atpB*. The former shows 16, the second 14 substitutions between the
6 outgroup *Ticodendron* and the ten species of *Betula* and *Alnus* in our matrix, but Wikström *et al.*
7 (2001) and Bell *et al.* (2010) included a single species from each of these genera, resulting in their
8 young inferred divergence times.
9

10 Compared to the (too) young ages inferred in these angiosperm-wide studies, the ages
11 inferred in the Betulaceae-focused work of Forest *et al.* (2005) were surprisingly old, in several
12 cases exceeding earliest paleobotanical occurrences of Betulaceae. The number of species sampled
13 by Forest *et al.* was similar to ours (26 vs. 27 Betulaceae), but only 462 nucleotides were used
14 compared to the 9321 used here, and fossils were applied as single fixed calibration points, rather
15 than as prior probability distributions (such as the gamma distributions used here; *Material and*
16 *Methods*). Forest *et al.* also focused on median ages obtained with 10 alternative placements of five
17 fossils, meaning that each fossil's effect entered the median twice, once when that fossil was
18 assigned to the crown, once when it was assigned to the stem. Of the five fossil records used to fix
19 the crown or stem ages (*Alnus* 65 Myr; *Betula* 49 Myr; *Corylus* to 49 Myr, *Carpinus* 49 Myr,
20 *Ostrya* 33.7 Myr), the *Alnus* constraint is also used here. Our other constraints were *Endressianthus*
21 (Friis *et al.*, 2003) for the Betulaceae stem lineage and *Cranea* (Manchester and Chen, 1996, 1998)
22 for the Coryloideae stem lineage because we wanted to use *Corylus*, *Carpinus*, and *Ostrya* fossils to
23 cross-validate our DNA-based estimates.
24

25 Cross validation (against other fossil evidence) shows that the inferred divergence times
26 match the fossil record well (Fig. 3). Thus, fruits and leaves assignable to *Ticodendron*
27 (*Ferrigocarpus*) have been found in early mid-Eocene deposits from Oregon (48.32 Myr) and in the
28 London Clay (ca. 50 Myr old; Manchester, 2011). Since *Ticodendron* consists of a single surviving
29 species, these fossils cannot be assigned in a DNA phylogeny. Nevertheless, they indicate that the
30 split between Ticodendraceae and Betulaceae must be at least 50 Myr old, which does not conflict
31 with our constraint of this split to 71 Myr, based on the *Endressianthus* fossil (in one run). Also, the
32 age of 49 (38–60) Myr that we inferred for the *Alnus/Betula* split using the *Endressianthus*
33 calibration fits reasonably well with the earliest *Alnus* foliage and ovulate cones from the 49–52
34 Myr-old McAbee flora of British Columbia, a site that also yielded *Betula leopoldae* leaves, catkins
35 and pollen (Crane & Stockey, 1987, Dillhoff *et al.*, 2005). Third, the age of 53 (43–63) Myr we
36 inferred for the Coryloideae stem lineage (when this was not itself constrained) overlaps with the
37 earliest fossils of Coryloideae from late Paleocene/Thanetian (56–59.2 Myr). The earliest fossils
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1 assigned to the genera *Corylus* and *Carpinus* appear in the early Eocene Klondike Mt. Formation
2 dated to 50–49 Myr (Pigg, Manchester & Wehr, 2003); they were used by Forest et al. (2005) to fix
3 crown or stem ages of these genera to 49 Myr. The ages inferred here for the *Corylus* and
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

The main earlier effort to date the Betulaceae (Forest *et al.*, 2005) could not use the
Endressianthus fossil (Friis et al., 2003) as a constraint because of the study's taxon sampling,
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*, *Paraalnipollenites*) also used here to fix the age of the *Alnus* stem lineage to 65
Myr (while 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 ACCTTRAN 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 41.7–57.9 Myr (Forest et al.'s table 3), in agreement with the Coryloideae fossil record, for
example, *Cranea* and *Palaeocarpinus* known from late Paleocene (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*
species 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 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, very 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 and Donoghue (1993), fossils should be assigned
conservatively, that is, as 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, also must be
diffusing any correct temporal signal contained in the genetic branch lengths.

SUBSTITUTION RATE SHIFTS WITHIN 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 the Coryloideae followed by a slowdown in their crown group to the level observed in the 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* are 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; 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, Skj  th *et al.*, 2007), and backward trajectory analyses suggest pollen dispersal over long distances in northeastern (  auliene & Veriankaite, 2006) and northwestern Europe (Skj  th *et al.*, 2008). *Betula* is also the most cold-tolerant genus among the northern hemisphere Fagales, with some species extending into high-alpine and arctic environments (Dfc, Dfd, ET climates according to K  ppen-Geiger; Kottek *et al.*, 2006; SI 4). *Alnus* resembles *Betula* in its pollen and seed dispersal (Ridley, 1930), and its species also cover a wide range of climates (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 (Eocene; Mai, 1995).

By contrast, *Corylus*, just like *Fagus* and *Quercus*, relies on jays for dispersal (*Garrulus*, Corvidae; Haffer & Bauer, 1993) as well as other stationary birds and small mammals (Ridley, 1930). By 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 the Coryloideae between 25 and 15 million years ago (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 and Sanderson, 2006; Chatterjee *et al.*, 2009). Starting from more than 900 available Betulaceae sequences 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

1
2 criterion of “tree-like signal” to select the most suitable of available Betulaceae species. The
3 resulting networks and phylogenies suggest that *Ostryopsis* is indeed the sister to *Carpinus/Ostrya*
4 and that the Betuloideae and Coryloideae are mutually monophyletic, which was not clear before.
5 We used relatively few of the many Betulaceae fossils as constraints, preferring to instead use them
6 to cross-validate inferred node ages. Inferred substitution rate shifts in a few places in the
7 phylogeny may be linked to different rates of allopatric species formation, but this needs to be
8 tested with a denser phylogenetic, geographic, and ecological species sample.
9
10
11
12

13 14 15 **ACKNOWLEDGMENTS**

16
17 We thank Jianquan Liu for sharing his unpublished sequences of *Ostryopsis*. GWG acknowledges
18 funding by the Vetenskåpradet (Swedish research council), project no. 2008-3726.
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

REFERENCES

- 1
2
3
4
5 **Albert, VA, Backlund, A, Bremer, K, Chase, MW, Manhart, JR, Mishler, BD, Nixon, KC.**
6 **1994.** Functional constraints and *rbcL* evidence for land plant phylogeny. *Annals of the*
7 *Missouri Botanical Garden* **81**: 534–567.
8
9
10 **Auch, AF, Henz, SR, Holland, BR, Göker, M. 2006.** Genome BLAST distance phylogenies
11 inferred from whole plastid and whole mitochondrion genome sequences. *BMC*
12 *Bioinformatics* **7**: 350.
13
14
15 **Bell, CD, Soltis, DE, Soltis, PS. 2010.** The age and diversification of the angiosperms re-revisited.
16 *American Journal of Botany* **97**: 1296–1303.
17
18 **Bininda-Emonds, ORP. 2004.** Trees versus characters and the supertree/supermatrix "paradox".
19 *Systematic Biology* **53**: 356–359.
20
21
22 **Bousquet, J, Strauss, SH, Li, P. 1992.** Complete congruence between morphological and *rbcL*-
23 based molecular phylogenies in birches and related species (Betulaceae). *Molecular Biology*
24 *and Evolution* **9**: 1076–1088.
25
26
27 **Chatterjee H, Ho S, Barnes I, Groves C. 2009.** Estimating the phylogeny and divergence times of
28 primates using a supermatrix approach. *BMC Evolutionary Biology* **9**: 259.
29
30 **Chen, Z-D, Li, J-H. 2004.** Phylogenetics and biogeography of *Alnus* (Betulaceae) inferred from
31 sequences of nuclear ribosomal DNA its region. *International Journal of Plant Sciences*
32 **165**: 325–335.
33
34
35 **Chen, Z-D, Manchester, SR, Sun, H-Y. 1999.** Phylogeny and evolution of the Betulaceae as
36 inferred from DNA sequences, morphology and palaeobotany. *American Journal of Botany*
37 **86**: 1168–1181.
38
39
40 **Cohen, KM, Finney, S, Gibbard, PL. 2012.** International chronostratigraphic chart, IUGS, Aug.
41 2012 (<http://www.stratigraphy.org>) Accessed 11/11/2012.
42
43
44 **Crane, PR. 1981.** Betulaceous leaves and fruits from the British Upper Palaeocene. *Botanical*
45 *Journal of the Linnean Society* **83**: 103–136.
46
47 **Crane, PR. 1989.** Early fossil history and evolution of the Betulaceae. In: Crane, PR, Blackmore, S
48 eds. *Evolution, systematics, and fossil history of the Hamamelidae. Vol. 2: 'Higher'*
49 *Hamamelidae*. Oxford: Clarendon Press.
50
51
52 **Crane, PR, Manchester, SR, Dilcher, D. 1990.** A preliminary survey of fossil leaves and well-
53 preserved reproductive structures from the Sentinel Butte Formation (Paleocene) near
54 Almont, North Dakota. *Fieldiana Geology, New Series* **20**: 1-63.
55
56
57 **Crane, PR, Stockey, RA. 1987.** *Betula* leaves and reproductive structures from the Middle Eocene
58 of British Columbia, Canada. *Canadian Journal of Botany* **65**: 2490–2500.
59
60

- 1
2 **Denk, T, Grimm, GW. 2009.** The biogeographic history of beech trees. *Review of Palaeobotany*
3 *and Palynology* **158**: 83–100.
- 4
5 **Dillhoff, RM, Leopold, EB, Manchester, SR. 2005.** The McAbee flora of British Columbia and
6 its relation to the early-middle Eocene Okanagan Highlands flora of the Pacific Northwest.
7 *Canadian Journal of Earth Sciences* **42**: 151–166.
- 8
9
10 **Doyle, JA, Donoghue, MJ. 1993.** Phylogenies and angiosperm diversification. *Paleobiology* **19**:
11 141–167.
- 12
13 **Drummond, AJ, Suchard, MA, Xie, D, Rambaut A. 2012.** Bayesian phylogenetics with BEAUti
14 and the BEAST 1.7. *Molecular Biology and Evolution* **29**: 1969–1973.
- 15
16 **Erdogan, V, Mehlenbacher, SA. 2000.** Phylogenetic relationships of *Corylus* species (Betulaceae)
17 based on nuclear ribosomal DNA ITS region and chloroplast matK gene sequences.
18 *Systematic Botany* **25**: 727–737.
- 19
20
21 **Felsenstein, J. 1985.** Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*
22 **39**: 783–791.
- 23
24 **Ford, RH, Sharik, TL, Feret, PP. 1983.** Seed dispersal of the endangered virginia round-leaf birch
25 (*Betula uber*). *Forest Ecology and Management* **6**: 115–128.
- 26
27
28 **Forest, F, Bruneau, A. 2000.** Phylogenetic analysis, organization, and molecular evolution of the
29 non-transcribed spacer of 5S ribosomal RNA genes in *Corylus* (Betulaceae). *International*
30 *Journal of Plant Science* **161**: 793–806.
- 31
32
33 **Forest, F, Savolainen, V, Chase, MW, Lupia, R, Bruneau, A, Crane, PR. 2005.** Teasing apart
34 molecular- versus fossil-based error estimates when dating phylogenetic trees: a case study
35 in the birch family (Betulaceae). *Systematic Botany* **30**: 118–133.
- 36
37
38 **Friis, EM, Pedersen, KR, Schönenberger, J. 2003.** *Endressianthus*, a new Normapolles-
39 producing plant genus of Fagalean affinity from the Late Cretaceous of Portugal.
40 *International Journal of Plant Sciences* **164**: S201–S223.
- 41
42
43 **Grimm, GW, Renner, SS, Stamatakis, A, Hemleben, V. 2006.** A nuclear ribosomal DNA
44 phylogeny of *Acer* inferred with maximum likelihood, splits graphs, and motif analyses of
45 606 sequences. *Evolutionary Bioinformatics* **2**: 279–294.
- 46
47
48 **Göker, M, García-Blázquez, G, Voglmayr, H, Tellería, MT, Martín, MP. 2009.** Molecular
49 taxonomy of phytopathogenic fungi: a case study in *Peronospora*. *PLoS ONE* **4**: e6319.
- 50
51
52 **Göker, M, Grimm, GW. 2008.** General functions to transform associate data to host data, and their
53 use in phylogenetic inference from sequences with intra-individual variability. *BMC*
54 *Evolutionary Biology* **8**: 86.
- 55
56
57 **Haffer, J, Bauer, KM. 1993.** Corvidae – Rabenvögel. In: Glutz von Blotzheim, UN, Bauer, KM
58 eds. *Handbuch der Vögel Mitteleuropas. Band 13-III. Passeriformes (4. Teil)*.
- 59
60

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
- 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.
- 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, O., ML, 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. *Vestnik Ustredniho Geologickeho* **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.
- 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, DE, Soltis, PS, 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.

- 1
2 **Maddison, WP, Maddison, DR. 2011.** Mesquite: a modular system for evolutionary analysis, v.
3 2.75 (<http://mesquiteproject.org>).
4
- 5 **Mai, DH. 1981.** Entwicklung und klimatische Differenzierung der Laubwaldflora Mitteleuropas im
6 Tertiär. *Flora* **171**: 525–582.
7
- 8 **Mai, DH 1995.** *Tertiäre Vegetationsgeschichte Europas* Jena, Stuttgart, New York: Gustav Fischer
9 Verlag.
10
- 11 **Manchester, SR. 2011.** Fruits of Ticodendraceae (Fagales) from the Eocene of Europe and North
12 America. *International Journal of Plant Sciences* **172**: 1179–1187.
13
- 14 **Manchester, SR, Chen, Z-D. 1998.** A new genus of Coryloideae (Betulaceae) from the Paleocene
15 of North America. *International Journal of Plant Sciences* **159**: 522–532.
16
- 17 **Manchester, SR, Crane, PR. 1987.** A new genus of Betulaceae from the Oligocene of western
18 North America. *Botanical Gazette* **148**: 263–273.
19
- 20 **Manchester, SR, Guo, S-X. 1996.** *Paleocarpinus* (extinct Betulaceae) from northwestern China:
21 new evidence for Paleocene floristic continuity between Asia, North America, and Europe.
22 *International Journal of Plant Sciences* **157**: 240–246.
23
- 24 **Manchester, SR, Pigg, KB, Crane, PR. 2004.** *Palaeocarpinus dakotensis* sp. n. (Betulaceae:
25 Coryloideae) and associated staminate catkins, pollen, and leaves from the Paleocene of
26 North Dakota. *International Journal of Plant Sciences* **165**: 1135–1148.
27
- 28 **Matlack, GR. 1989.** Secondary dispersal of seed across snow in *Betula lenta*, a gap-colonizing tree
29 species. *Journal of Ecology* **77**: 853–869.
30
- 31 **Matlack, GR. 1991.** Influence of fruit size and weight on wind dispersal in *Betula lenta*, a gap-
32 colonizing tree species. *American Midland Naturalist* **128**: 30–39.
33
- 34 **Pattengale, ND, Masoud, A, Bininda-Emonds, ORP, Moret, BME, Stamatakis, A. 2009.** How
35 many bootstrap replicates are necessary? In: Batzoglou, S ed *RECOMB 2009*. Berlin,
36 Heidelberg: Springer-Verlag
37
- 38 **Pigg, KB, Manchester, SR, Wehr, WC. 2003.** *Corylus*, *Carpinus*, and *Palaeocarpinus*
39 (Betulaceae) from the middle Eocene Klondike Mountain and Allenby Formations of
40 northwestern North America. *International Journal of Plant Science* **164**: 807–822.
41
- 42 **Prieditis, N. 1997.** *Alnus glutinosa* – dominated wetland forests of the Baltic Region: community
43 structure, syntaxonomy, and conservation. *Plant Ecology* **129**: 49–94.
44
- 45 **Ridley, HN 1930.** *The dispersal of plants throughout the world*. Ashford: L. Reeve & Co. Ltd.
46
- 47 **Sanderson, MJ. 1997.** A nonparametric approach to estimating divergence times in the absence of
48 rate constancy. *Molecular Biology and Evolution* **14**: 1218–1232.
49
50
51
52
53
54
55
56
57
58
59
60

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
- Šauliene, 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. 2008.** Are the birch trees in Southern England a source of *Betula* pollen for North London? *International Journal of Biometeorology* DOI: 10.1007/s00484-00008-00192-00481 [e-pub].
- 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, J., W. 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.

- 1
2 **Yoo, K-O, Wen, J. 2002.** Phylogeny and biogeography of *Carpinus* and subfamily Coryloideae
3 (Betulaceae). *International Journal of Plant Sciences* **163**: 641–650.
4
5 **Yoo, K-O, Wen, J. 2007.** Phylogeny of *Carpinus* and subfamily Coryloideae (Betulaceae) based on
6 chloroplast and nuclear ribosomal sequence data. *Plant Systematics and Evolution* **267**: 25–
7 35.
8
9
10 **Zachos, JC, Pagani, M, Sloan, L, Thomas, E, Billups, K. 2001.** Trends, rhythms, and aberrations
11 in global climate 65 Ma to present. *Science* **292**: 686-693.
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Figures and Tables

Figure 1. Neighbor Net splits graph from the 46-taxon matrix of 9321 aligned nucleotides of nuclear and plastid sequences. Values at edges refer to 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). Abbreviations: NA, North American spp. of *Ostrya*; EA, Eurasian species of *Ostrya*.

Figure 2. A chronogram for the Betulaceae obtained under a relaxed clock model applied to an alignment of 29 species and 9321 aligned nucleotides (*Materials and Methods*). Labels above or below branches indicate support from, in the following order, Bayesian analysis of the 29-taxon matrix, and 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 shown by gray bars. The geographic origin of each accession is shown after the species name: (1), Australasia; (2), Central America (incl. Mexico); (3), eastern N. America; (4) (cool) temperate northeastern Asia (incl. Siberia, Russian Far East, Mongolia, N.E. China, N. Korea, and N. Japan); (5), Himalayas foothills and adjacent mountain areas in S.E. Asia and Southwest China; (6) (temperate) western Eurasia; (7), Andes; (8), S. W. China (Yunnan, Sichuan, Guizhou); (9) Mountainous areas in N.C. China; (10) Rio Grande valley, N. America; (11) Mountainous areas in N. Vietnam and/or S. China (incl. islands of Hainan and Taiwan); (12) (warm) temperate East Asia (C. and E. China, Korea, Japan); (13) Eastern Mediterranean hinterland.

Figure 3. Unrooted maximum likelihood tree based on all available nuclear GBSSI sequences (matrix included in 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 high support. Edge thickness represents low (<50, thin), moderate (50–90, medium) and high BS support (>90; thick lines). Exact values indicated for interior edges. The taxon sampling is comparable to that of the concatenated 46-taxon matrix. Species not covered by the 46-taxon matrix are marked in gray. Two ingroup species (*Carpinus tschonoskii* and *Ostryopsis davidiana*) and the two outgroups in the 46-taxon matrix lack GBSSI sequences. “A” indicates the outgroup-inferred root from ML tree searches and favored by the Fagales dataset of Li *et al.* (2004); “B” indicates the root here inferred under Bayesian inference and favored by most other studies with low to moderate support (compare Fig. 2; Tables 1, 2).

Table 1. Support for genus relationships within the Betulaceae based on different taxon and gene samples. Numbers in square brackets refer to the number of species (according to the original papers; see SI 1 for identification or sequencing errors). ‘NA’, clade not represented in tree; ‘~’, according bipartition not testable. Clades that received moderate to high support independent of gene- and taxon-sampling highlighted in bold: ‘BS_{NJ/P}’, Bootstrap support under NJ or MP; PP, Bayesian posterior probabilities.

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 _{NJ} /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 (Co.)</i>	~ [1]	~ [1]	100/1.0 [5]	~ [1]	64/65 [4], 100/1.0 [3]	100/99 [6]
<i>C.-O.-Os.</i>	NA	NA	72/0.94	48	<50/<50, <50/0.5	86/NA**
<i>Ostryopsis (Os.)</i>	~ [1]	~ [1]	~ [1]	~ [1]	~/~, ~/~ [1]	100/~ [1]
<i>Co.-Os.</i>	99	88/1.0	NA	NA	NA, NA	NA
<i>C.-O.</i>	100	100/1.0	95/1.0	82	93/<50, 79/0.97	98/91
<i>Carpinus (C.)</i>	~ [1]	~ [1]	65/0.97 [5]	Grade [2]	NA/55 [18], 63/0.97 [13]	Grade [8]
<i>Ostrya (O.)</i>	~ [1]	~ [1]	93/1.0 [2]	89 [2]	88/Grade [5], Grade [5]	95/99 [3]
Betuloideae Coryloideae split*	~	~	~	99	<50/<50	99

*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 the remaining Betulaceae, while *matK* indicating an *Alnus/Betula* sister relationship (Kato *et al.*, 1998; G. Grimm, pers. obs.).

‡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.

Table 2. ML bootstrap support for Betulaceae generic relationships in the 46-taxon, and 29-taxon data sets.

Phylogenetic split ^a	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									
ML <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
BI Betuloideae clade	29	11	<10	~	60	83	~	32	38
Position of <i>Ostryopsis</i> within Coryloideae									
ML <i>Ostryopsis</i> sister to <i>Corylus</i>	<1	13	<10	<1	31	13	17	<5	<10
BI <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

^aFlag (**ML**,**BI**) indicates ML-preferred (best-scoring 29-/46-taxon trees) and Bayesian-preferred (Fig. 3) phylogenetic splits.

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

Topology	ln L (GTR+□)	D(L)	Standard deviation	Significantly worse (p<0.01) than best-scoring ML tree
Best-scoring ML 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 & B)	-23299.279698	-1.838041	5.250697	No

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	<i>Corylus</i> stem 56	<i>Corylus</i> stem 56	<i>Alnus</i> stem 58	<i>Alnus</i> stem 58
	<i>Strict</i>	<i>Relaxed</i> (Fig. 3)	<i>Strict</i>	<i>Relaxed</i>	<i>Strict</i>	<i>Relaxed</i>
<i>Betulaceae</i> Stem	Constrained	Constrained	154 (178-133)	133 (175-101)	92 (105-80)	85 (102-69)
<i>Betulaceae</i> crown group	52 (58-46)	53 (63-43)	113 (126-99)	98 (124-77)	67 (74-61)	63 (71-58)
<i>Alnus/Betula</i> Crown	46 (52-39)	49 (60-38)	99 (114-84)	91 (118-68)	Constrained	Constrained
Coryloideae Crown	26 (30-22)	29 (39-22)	Constrained	Constrained	33 (39-28)	35 (45-26)
<i>Ostryopsis</i> stem	23 (27-19)	25 (34-18)	50 (55-45)	49 (56-41)	30 (35-25)	30 (39-22)

ONLINE SUPPORTING MATERIAL

SI 1. Notes on data curation, including notes on apparently erroneous sequences in GenBank

SI 2. Distance statistics

SI 3. The best-scoring ML trees inferred from 29- and 46-taxon matrices.

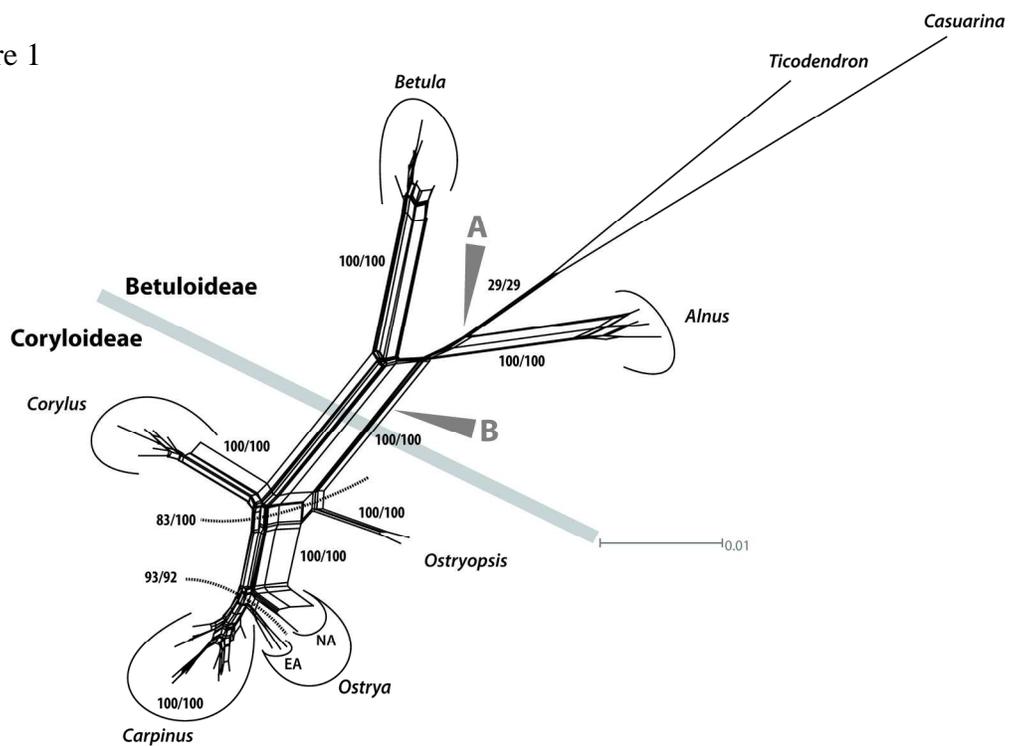
SI 4. Geographic and climatic distribution of Betulaceae with a graphical overview (Fig. S1) and tabulated data (File S1).

Figure S1. Map of the geographical and climatic distribution of Betulaceae. The pie charts show the number of species and proportion of genera divided by large scale (bio)geographic unit and climate zones (details provided in File S1). Base map from Kottek et al. (2006; freely available via <http://koeppen-geiger.vu-wien.ac.at/>): red tones: equatorial (tropical) climates (A-climates); yellow tones: Dry climates (deserts, BW, and steppe, BS); light green tones: warm temperate, summer-dry climates (Cs), lush green tones: warm temperate, fully humid climates (Cf), brown tones, warm temperate, winter-dry climates (Cw); pink/purple tones: snow climates (D); blue tones: polar climates (tundra, ET, and frost climate, EF).

File S1. An XLSX file with tabulated distribution and climate data for Betulaceae species, compiled from various data sources as indicated. See sheet “Readme!” for further content and instructions.

PDF Proof

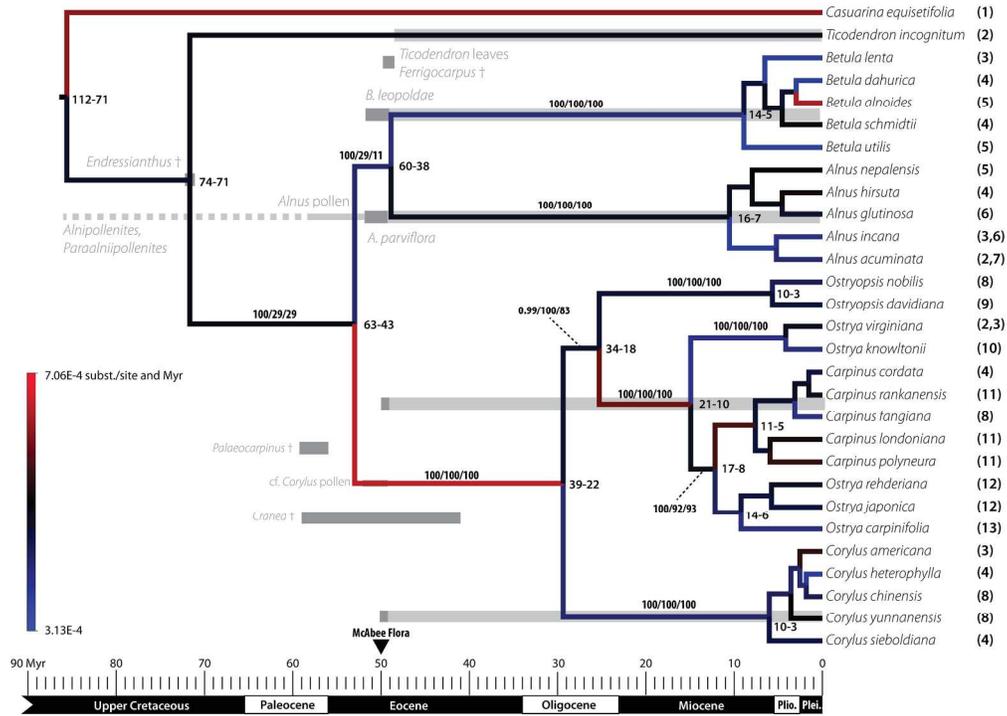
Figure 1



190x140mm (300 x 300 DPI)

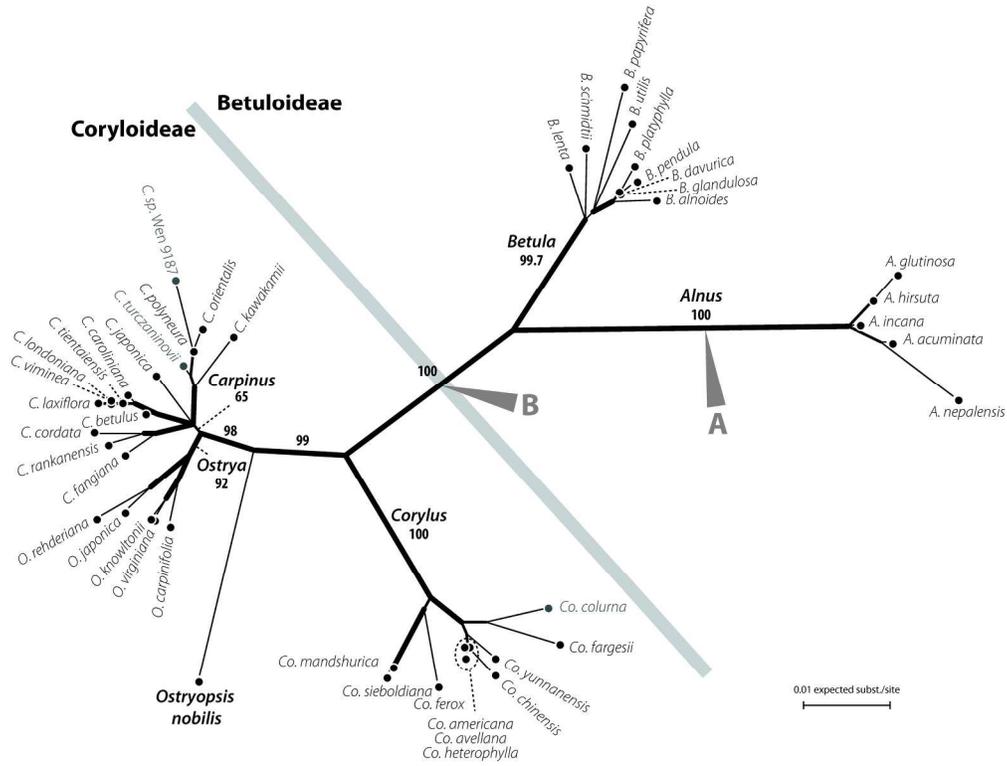
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Figure 2



202x143mm (300 x 300 DPI)

Figure 3



207x157mm (300 x 300 DPI)