Molecular Phylogenetics and Evolution 63 (2012) 230–243

Contents lists available at SciVerse ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Expanded phylogenetic and dating analyses of the apples and their relatives (Pyreae, Rosaceae)

Eugenia Y.Y. Lo *, Michael J. Donoghue

Department of Ecology and Evolutionary Biology, Yale University, 21 Sachem Street, New Haven, CT 06520, USA

ARTICLE INFO

Article history:
Received 1 March 2011
Revised 30 September 2011
Accepted 6 October 2011
Available online 25 January 2012

Keywords:
Pyreae
Maloideae
Rosaceae
Sorbus
Divergence times
Hybridization
Northern hemisphere disjunctions

ABSTRACT

Despite previous efforts to elucidate relationships within the Pyreae (Rosaceae), relationships among the major sub-lineages, generic limits, and divergence times have remained uncertain. The present study greatly expands phylogenetic analyses of the Pyreae by using a combination of 11 chloroplast regions plus nuclear ribosomal ITS sequences from 486 individuals representing 331 species and 27 genera. Maximum likelihood and Bayesian analyses generally support existing generic boundary, although *Sorbus*, as previously circumscribed, is clearly non-monophyletic. Two significant conflicts were detected between the chloroplast and ITS phylogenies, suggesting that hybridization played a role in the origins of *Micromeles* and *Pseudocydonia*. In addition, we provide estimates of the divergence times of the major lineages. Our findings support the view that the major Pyreae lineages were established during the Eocene–Oligocene period, but that most of the modern diversity did not originate until the Miocene. At least five major, early Old World-New World disjunctions were detected and these vicariance events are generally most consistent with movement through the Beringia.

1. Introduction

The Rosaceae is a moderately large angiosperm lineage, with approximately 3000 species in 100 genera (Kalkman, 2004), including a clade of mostly fleshy-fruited genera some of which are widely cultivated and of considerable economic importance (e.g., apple (*Malus*), chokeberry (*Aronia*), loquat (*Eriobotrya*), pear (*Pyrus*), quince (*Cydonia*), and serviceberry (*Amelanchier*)). A new classification of the Rosaceae based on molecular data delimited members into three subfamilies Dryadoideae, Rosoideae, and Spiraeoideae (Potter et al., 2007). Within the Spiraeoideae, the genus *Gillenia* is included in the clade Pyrodae; *Lindleya*, *Kageneckia*, and *Vauquelinia* are included in the clade Pyreae; and members of the long-recognized subfamily Maloideae, of which fruit-type is generally a pome, in the clade Pyrinae (Potter et al., 2007). Pyreae are widespread in northern temperate regions and several lineages have radiated into the far north and achieved circumboreal distributions. There are two competing hypotheses concerning the origin of the Pyreae. One is that the Pyreae could be the product of wide hybridization between the ancestors of the Spiraeoideae ($n = 9$) and the Amygdaloideae ($n = 8$) (Phipps et al., 1991, and references therein). However, based on phylogenetic, morphological, and fossil evidence, Evans and Campbell (2002) and Evans and Dickinson (2005) suggested an aneuploid origin of Pyreae that involved an initial chromosome doubling (or polyploidization) event in the *Gillenia* ($n = 9$) lineage, followed by an aneuploid loss of one pair of homologous chromosomes. In Pyrinae there are approximately 950 species (Campbell et al., 2007), and they share a base chromosome number of $n = 17$ (with the exception of *Vauquelinia*, with $n = 15$).

Several attempts have been made to resolve relationships among the recognized genera; however, to date only a limited number of species (ca. 50 of 950) have been included in phylogenetic studies at this level. Genetic diversity in the larger groups, such as *Malus*, *Cotoneaster*, *Sorbus*, and *Crataegus* has not been well represented in previous analyses, which has hindered a critical evaluation of the monophyly of these genera. One example concerns the group *Sorbus*, which was previously circumscribed to include both the pinnate-leaved species (*Sorbus s.s.* and *Cormus*) and the simple-leaved species (*Aria*, *Micromeles*, *Chamaemespilus*, and *Torminalis*) (Table 1; Rehder, 1940; Yu, 1974; Phipps et al., 1990; Aldasoro et al., 1998). Albeit limited samples in previous studies, both morphological (e.g., Fig. 11 in Phipps et al., 1991) and molecular data (Campbell et al., 2007) indicate non-monophyletic relationships among species of the group. In addition, although fossil information is available for several lineages of Rosaceae (Wolfe and Wehr, 1988; DeVore and Pigg, 2007), the lack of a well-sampled and robust phylogeny has hindered an assessment of the ages of lineages and their biogeographic histories.

* Corresponding author. Address: Department of Ecology and Evolutionary Biology, Yale University, P.O. Box 208105, New Haven, CT 06520, USA. Fax: +1 203 432 7909.
E-mail address: eugenia.loyy@gmail.com (E.Y.Y. Lo).

1055-7903/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2011.10.005

Table 1
Summary of Pyreae samples included in this study. Number in parentheses indicates the estimated total number of species described in each group. Information on voucher specimens is provided in Supplementary Appendix 1. Biogeographic regions: AS = Asia; EU = Europe; LA = Latin America; NA = North America.

Genus	Included species	No. of individuals	Geography
Amelanchier Medik.	14 (25)	21	AS; NA; EU
Aronia Medik.	3 (3)	5	NA
Chaenomeles Lindl.	4 (5)	8	AS
Cotoneaster Medik.	40 (70)	56	AS; NA; EU
Crataegus L.	60 (~150)	62	AS; NA; EU
Cydonia Mill.	1 (1)	3	AS
Dichotomanthes Kurz	1 (1)	3	AS
Docynia Decne.	2 (2)	5	AS
Docyniopsis Koidz.	1 (1)	1	AS
Eriobotrya Lindl.	2 (18)	5	AS
Eriolobus (DC.) Roem.	1 (1)	1	AS
Gillenia Moench	2 (1)	2	NA
Heteromeles Roem.	1 (1)	3	NA
Kageneckia Ruiz & Pavon	2 (3)	2	LA
Lindleya H.B.K.	1 (1)	1	LA
Malacomeles (Decne.) Engler	1 (1)	1	LA
Malus Mill.	29 (40)	41	AS; NA; EU
Mespilus L.	1 (2)	2	EU
Osteomeles Lindl.	2 (2)	4	AS; NA
Peraphyllum Nutt.	1 (1)	2	NA
Photinia Lindl. sensu stricto	1 (40)	2	AS
Pourthiaea Decne.	2 (25)	3	AS
Pyracantha Roem.	3 (3)	5	AS; EU
Pseudocydonia C.K. Schneid	1 (1)	4	AS
Pyrus L.	16 (~20)	22	AS; EU; AF
Rhaphiolepis Lindl.	2 (5)	5	AS
Sorbus subgen. *Aria* (Pers.) G. Beck	36 (~50)	54	EU
Sorbus subgen. *Chamaemespilus* (Medik.) K. Koch.	1 (1)	2	EU
Sorbus subgen. *Cormus* (Spach) Duch.	1 (1)	2	EU
Sorbus subgen. *Micromeles* Decne.	15 (20)	19	AS
Sorbus subgen. *Sorbus*	78 (80)	130	AS; NA; EU
Sorbus subgen. *Torminalis* (DC.) K. Koch	1 (1)	2	EU
Stranvaesia Lindl.	1 (5)	3	AS
Vauquelinia Humb. & Bonpl.	1 (3)	1	NA
Prunus L. (outgroup)	4	4	AS; NA
Total	331	486	

Lindleya, *Kageneckia*, and *Vauquelinia* resemble the pome-bearing members of Pyreae in their base chromosome number ($x = 17$, with the exception of *Vauquelinia* with $x = 15$; Goldblatt, 1976), the symbiotic association with the fungal pathogen *Gymnosporagium* (*Vauquelinia* and several members of the Pyrinae; Savile, 1979), and some floral characteristics such as connate carpels and presence of two basal collateral ovules with funicular obturators in each ovary (*Lindleya*, *Vauquelinia*, and several members of the Pyrinae; Robertson et al., 1991; Rohrer et al., 1994). Recent analyses of molecular data support a sister group relationships of these taxa with members of the Pyrinae (Morgan et al., 1994; Evans et al., 2000; Evans and Campbell, 2002; Campbell et al., 2007; Potter et al., 2007). However, relationships among the major lineages of the group are still not confidently resolved. Reasons cited for poor phylogenetic resolution include hybridization, gene paralogy, rapid radiation, and/or slow divergence at the molecular level (Campbell et al., 2007). Pyreae are notorious among botanists for weak reproductive barriers, both among closely related species as well as among members of different genera (Phipps et al., 1991; Robertson et al., 1991). Hybridization creates intermediate phenotypes and genotypes and allows the introgression of maternally inherited plastid genomes from one species into another. This may result in strongly contrasting phylogenetic signals between plastid and nuclear genes (Pamilo and Nei, 1988), which can be useful in identifying potential hybrid lineages and tracing their parentage. Polyploidy is also common in Pyreae, and the recurrent gain of nuclear gene copies via genome multiplications and/or the incomplete sorting of gene lineages could lead to phylogenetic incongruence among nuclear genes (e.g., topological differences reported between nuclear ribosomal ITS and the four GBSSI (waxy) gene copies; Campbell et al., 2007). Finally, a failure to resolve relationships among deep branches has suggested the possibility of an ancient rapid radiation within Pyreae (Campbell et al., 1995, 2007).

With the goal of obtaining a robust phylogeny that allows us to critically evaluate generic limits and identify potential hybrid lineages, we inferred relationships among a greatly expanded sample of Pyreae species based on a combination of slowly and rapidly evolving chloroplast regions. We compared these results to similarly broadly sampled phylogenetic trees based on nuclear ribosomal ITS sequences. These analyses not only help us to understand the evolutionary history of the Pyreae, but also allow us to estimate divergence times for the major lineages and to identify possible morphological and biogeographic factors underlying the diversification of the group.

2. Materials and methods

2.1. Taxon sampling and gene regions

Our sampling attempted to maximize the taxonomic and geographical coverage of each previously recognized genus within Pyreae. Data were obtained from a total of 486 individuals representing 331 species and 27 previously recognized genera (Table 1; Supplementary Appendix 1). In some cases up to three individuals per species were examined. For most of the larger genera, including *Cotoneaster, Crataegus, Malus*, and *Sorbus*, at least half of the described species were included. Species of *Prunus* (*P. hortulana*, *P. nigra*, *P. persica*, and *P. virginiana*) were included for rooting purposes (Morgan et al., 1994; Potter et al., 2007). *Prunus* is also well known in the fossil record back to the Middle Eocene (Cevallos-Ferriz and Stockey, 1991; Manchester, 1994; DeVore and Pigg, 2007), which is useful in calibrating divergence time analyses. Samples were either collected in the field or in botanical gardens, or obtained from herbarium specimens. Except as noted in Supplementary Appendix 1, voucher specimens are deposited in the Yale Herbarium (YU) or in the Arnold Arboretum (AA) Herbarium or the Gray Herbarium (GH) of the Harvard University Herbaria (HUH). Total DNA was extracted from dried plant tissues using the QIAGEN DNA extraction kit following the manufacturer's protocol.

Eleven coding and non-coding regions of the chloroplast genome were sequenced. Six of these regions – *trn*L–*trn*F, *trn*K + *mat*K, *rpl*16 intron, *rps*16 intron, *atp*B–*rbc*L, and *rbc*L – were amplified using the primers published in Campbell et al. (2007). For the remaining five regions – *trn*G–*trn*S, *rpl*20–*rps*12, *trn*C–*ycf*6, *psb*A–*trn*H, and *trn*H–*rpl*2 – primer information can be found in Shaw et al. (2005) and Lo et al. (2009). For each of the gene regions, many new sequences were generated in this study: *trn*L–*trn*F: 311, *trn*K + *mat*K: 296, *rpl*16: 299, *rps*16: 278, *atp*B–*rbc*L: 301, *rbc*L: 272, *trn*G–*trn*S: 355, *rpl*20–*rps*12: 355, *trn*C–*ycf*6: 302, *psb*A–*trn*H: 360, and *trn*H–*rpl*2: 352. These sequences were added to published sequences for Pyreae species obtained from GenBank, and together these yielded an alignment of 11,056 bp. In addition to the chloroplast regions, we amplified the internal transcribed spacer (ITS) of the nuclear ribosomal DNA regions using primers in White et al. (1990). A total of 258 new ITS sequences were added to the 159 published sequences obtained from GenBank, yielding an alignment of 657 bp.

Chloroplast sequences were obtained by direct sequencing whereas PCR products of the ITS sequences were cloned using pDrive vector (QIAGEN) when ambiguous nucleotides were detected with direct sequencing. Plasmids and PCR products were sequenced in both forward and reverse directions on an ABI 3730*xl* (Applied Biosystems) automated DNA sequencer with BigDye terminator cycle sequencing kits. When multiple sequences were found to be identical among individuals of the same species, a single sequence was used to represent the species. Sequences for each gene region were deposited in GenBank with the accession numbers presented in Suppl. Appendix 2. The data matrices underlying the published trees are available in TreeBASE (www.treebase.org).

2.2. Phylogenetic analyses

Sequences of each gene region were aligned with MUSCLE (version 4.0; Edgar, 2004) and manually adjusted with the Sequence Alignment Editor version 1.d1 (SE-AL; Rambaut, 2002). The nucleotide substitution model was determined by the Akaike Information Criterion (AIC) method using Modeltest (version 3.06; Posada and Crandall, 1998). The best-fitting model and related parameters of the datasets were used in maximum likelihood (ML) analyses conducted using RAxML (v7.0.4; Stamatakis et al., 2005) and in Bayesian inference (BI) using Mr. Bayes (v3.0b4; Huelsenbeck and Ronquist, 2001). For ML analyses, bootstrap support (BS) was assessed with 1000 replicates with the rapid bootstrap algorithm implemented in the RAxML (Stamatakis et al., 2008). Bayesian analyses were performed with four Markov chains each initiated with a random tree and two independent runs each for 10,000,000 generations, sampling every 100th generation. In preliminary analyses using the full dataset (containing over 10 Kbps for nearly 400 terminals), posterior probabilities and other parameters were not converging, even when the number of generations was increased to 20 million. We therefore reduced the number of terminals as follows: for genera containing <10 species, all species were included; for genera containing 10–30 species, half of the species in each genus were included (species were selected to represent each resolved subclade based on the RAxML results); for genera containing >30 species, one-third of the species in each genus were included (species again represent the RAxML subclades). This sampling strategy resulted in a reduced dataset with a total of 158 terminals for Bayesian analyses and posterior probabilities and other parameters were shown to converge after 10 million generations with this reduced dataset. Likelihood values were monitored for stationarity with Tracer (v1.4.1; Rambaut and Drummond, 2003). Trees and other sampling points prior to the burn-in cut-off (i.e., 25,000 out of 100,000 trees) were discarded and the remaining trees were imported into Phyutility (v2.2; Smith and Dunn, 2008) to generate a majority-rule consensus. Posterior probability values (PP; Ronquist and Huelsenbeck, 2003) were used to evaluate node support in the Bayesian trees.

Tree topologies and bootstrap values were inspected to identify possible cases of incongruence between the cpDNA and ITS datasets. To further assess such incompatibilities and to test specific hypotheses, we employed the one-tailed non-parametric Shimodaira–Hasegawa test (SH; Shimodaira and Hasegawa, 1999) as implemented in PAUP* (version 4b10; Swofford, 2002). This test compares the likelihood scores of the best ML trees obtained from either the chloroplast or the nuclear data with trees resulting from analyses in which topologies were variously constrained. Four specific questions were addressed: (1) whether *Sorbus* and its allies (*Aria*, *Micromeles*, *Chamaemespilus*, *Torminalis*, and *Cormus*) are monophyletic; (2) whether there is a significant conflict in the placement of *Micromeles* between chloroplast and nuclear data; (3) whether there is a significant conflict in the placement of *Pseudocydonia* between chloroplast and nuclear data, and; (4) whether the ITS data fit the "three-clade" topology seen in the chloroplast tree after the removal of conflicting taxa (see below). To test each of these hypotheses, the taxa of interest were constrained to be monophyletic using Mesquite (version 2.71; Maddison and Maddison, 2008); all other branches were unconstrained and only those trees compatible with the constraint were retained in the analyses (Suppl. Appendix 2a–d). Substitution models and ML parameters for these SH tests were obtained as outlined above. We used resampling estimated log-likelihood (RELL) optimization and 1000 bootstrap replicates.

2.3. Divergence time estimations and fossils

Divergence times for the major lineages of Pyreae were estimated using the penalized likelihood (PL) method implemented in r8s version 1.71 (Sanderson, 2003) and the Bayesian method in BEAST (Drummond and Rambaut, 2007). PL is a semi-parametric rate-smoothing approach that allows rate heterogeneity among branches when estimating node ages in the phylogenetic trees (Sanderson, 2002). The RAxML topology was used for calibration and the rate smoothing parameter (λ) was determined by cross-validation analysis. Confidence intervals around the divergence times were estimated by the non-parametric bootstrap procedure (Baldwin and Sanderson, 1998; Sanderson and Doyle, 2001). One hundred bootstrap matrices were simulated using Mesquite (Maddison and Maddison, 2008) under the maximum likelihood criterion and specified substitution model. For each matrix, trees of the same topology but with different branch lengths were generated from ML heuristic searches in PAUP* (Swofford, 2002) and these were then used for age estimation with the same parameters. The central 95% of the age distribution provides the confidence interval.

Unlike PL, BEAST incorporates uncertainty in phylogenetic trees by estimating prior probability distributions of parameters such as tree topology and branching rate (Drummond and Rambaut, 2007). Also, an uncorrelated lognormal relaxed-clock model allows rate variation across branches. Two independent MCMC runs were performed with the best-fit RAxML tree as the starting tree for 10,000,000 generations, sampling every 100th generation. The outgroup *Prunus* species were constrained to be sister to all of the Pyreae taxa; all other relationships were unconstrained. A Yule tree prior was specified to model speciation. We estimated divergence times using the combined chloroplast and ITS data, but with the taxa that showed conflicting phylogenetic positions removed, specifically *Pseudocydonia sinensis* and the species of *Micromeles*.

In the PL analyses, we used two different maximum root age constraints chosen to bracket previously inferred ages for the Rosaceae: (1) 73 MY, the youngest credible age of crown Rosaceae inferred in a recent molecular dating analysis (Forest and Chase, 2009); and (2) 104 MY, the oldest credible age of crown Rosales inferred in recent dating analyses (which is equivalent to the age of stem Rosaceae given that Rosaceae is inferred to be sister to the remainder of the Rosales) (Wikstrom et al., 2001; Magallón and Castillo, 2009; Bell et al., 2010). In BEAST, we modeled the root of crown Rosaceae as a lognormal distribution in two separate runs, one with the offset value equal to 73 MY and the other with offset value equal to 104 MY. Values for the mean and standard deviation were set to 1.5 and 1, respectively, to encapsulate the Upper (70–95 MY) and Lower Cretaceous (104–130 MY) in the 95% quantiles of the lognormal prior distribution given the uncertainty in the absolute root age of Rosaceae and the assumption that the age of stem Rosaceae cannot be older than the age of the earliest angiosperms. In addition, we constrained the minimum age of two internal nodes. One constraint was based on documentation of *Prunus*-type fruit (seed and endocarp) and leaf fossil from the Middle Eocene (approximately 40 MY) at the Okanogan Highlands

Princeton chert of southern British Columbia (Cevallos-Ferriz and Stockey, 1991) and in the coeval Clarno Nut Beds of Oregon (Manchester, 1994). The other constraint was based on the oldest fossil record of leaves of *Amelanchier* from the Middle Eocene at the One Mile Creek locality, Princeton, British Columbia (approximately 40 MY; Wolfe and Wehr, 1988). These fossils were identified as species of *Prunus* and *Amelanchier*, but their relationships to the extant species within these groups are uncertain. We assumed that these fossils are members of their respective crown groups, and assigned the fossil ages of *Prunus* and *Amelanchier* as minimum ages for their respective most recent common ancestors. For the internal node calibrations, we used a lognormal distribution with offset values of 40, a mean of 1.5, and a standard deviation of 1.

3. Results

3.1. Relationships among lineages

Consistent with previous findings (Morgan et al., 1994; Campbell et al., 1995, 2007; Potter et al., 2007), both the cpDNA (Fig. 1a) and ITS (Fig. 1b) trees indicate that the root of the Pyreae is situated among *Kageneckia*, *Lindleya*, and *Vauquelinia* and that the remainder of the Pyreae forms a well-supported clade (PP > 90%; BS 87%). In the cpDNA tree, the root falls along the branch connecting a *Lindleya* plus *Kageneckia* clade to a *Vauquelinia* plus Pyrinae clade (Fig. 1a). In the ITS tree, relationships among *Lindleya–Kageneckia*, *Vauquelinia*, and Pyrinae are unresolved (Fig. 1b).

Within the large Pyrinae clade, cpDNA data support three major clades, labeled A, B, and C in Fig. 1a. Clade A (PP 100%; BS 83%) contains species of *Malacomeles*, *Peraphyllum*, *Amelanchier*, *Crataegus*, and *Mespilus*, with the first three sister to the latter two. Each of these genera was resolved as monophyletic. Basal relationships within clade B (PP 84%; BS 82%) in the cpDNA tree are poorly resolved, but this clade contains a *Chaenomeles–Pseudocydonia* clade, a *Malus–Eriolobus–Docynopsis–Docynia* clade, and an *Aria–Aronia–Pourthiaea* clade. The relationships of *Dichotomanthes* and *Cydonia* within clade B are not clearly resolved. Clade C (PP 70%; BS 64%) contains two major subclades: *Cotoneaster* (*Eriobotrya–Rhaphiolepis*) (PP 90%; BS 71%), and *Pyrus* (*Cormus* (*Micromeles*, *Sorbus sensu stricto*)) (PP 97%; BS 78%; Fig. 1a). Species of *Photinia davidiana* (previously named as *Stranvaesia davidiana*; Guo et al., 2011) and *Heteromeles* are nested in clade C but their relationships within the clade are not well resolved.

Clade A was also recovered in the ITS tree (PP 100%; BS 86%), with matching relations among the genera (Fig. 1b). Clade B does not appear in the ITS trees, although one component – the *Malus–Eriolobus–Docynopsis–Docynia* clade – is supported (PP 100%; BS 98%). It is noteworthy that these are the only Pyreae that produce dihydrochalcones (Challice, 1973). Likewise, clade C does not appear in the ITS tree, though several elements are similar, such as a strongly supported *Eriobotrya–Rhaphiolepis* clade (PP 100%; BS 98%), marked also by fruits that lack a conspicuous core and contain 1–3 large seeds, as well as by not being hosts for *Gymnosporangium* fungi.

Despite limited agreement between the cpDNA and ITS trees with respect to broader relationships among the genera, it is important to note that both datasets generally strongly support the monophyly of the individual genera (Fig. 1). One interesting exception to this rule is *Sorbus* in its broad sense. In this case, cpDNA and ITS analyses agree that there are two separate, distantly related clades. One of these clades corresponds to *Sorbus sensu stricto*, with pinnately compound leaves (in clade C; PP 87%; BS 77%), and the other corresponds to *Aria*, which contains mostly European species with simple leaves (in clade B; PP 88%; BS 72%). For both cpDNA and ITS, constraining the monophyly of *Sorbus* (*sensu lato*) significantly decreased the likelihood of the resulted trees in the SH analyses ($P < 0.01$; Table 1; Appendix 2a).

3.2. Conflicting positions and tests of incongruence

Several apparently major differences between the cpDNA and ITS trees are not well supported in one or both datasets. An example concerns the placement of *Cotoneaster*, *Eriobotrya + Rhaphiolepis*, and their possible relatives *Heteromeles* and *Stranvaesia*. While these taxa appear in clade C in the cpDNA tree, along with *Pyrus*, *Cormus*, and *Sorbus sensu stricto* (Fig. 1a), they are instead united in the ITS tree with *Malus*, *Aria*, and other members of cpDNA clade B (Fig. 1b). However, the support for the ITS placement is relatively low (PP and BS < 50%; Fig. 1b) and SH tests do not reject the cpDNA-based topology for the ITS data (Table 1; Appendix 2d). We therefore interpret this as a lack of resolving power in the ITS data, not as a strong conflict in need of further explanation.

Between the cpDNA and ITS data there do appear to be two strongly supported phylogenetic inconsistencies (Fig. 1). The first is the placement of the *Micromeles* species, which have previously been recognized as part of *Sorbus sensu lato*. *Micromeles* appears as sister to pinnate-leaved *Sorbus sensu stricto* in clade C in the cpDNA tree (PP 90%; BS 70%; Fig. 1a). In contrast, *Micromeles* appears as sister to simple-leaved *Aria* in the ITS data (PP 75%; BS 62%; Fig. 1b). Although support for this relationship in the ITS tree is relatively low, SH tests clearly reject the cpDNA topology for the ITS data, as well as the ITS topology for the cpDNA data ($P < 0.01$; Table 1; Appendix 2b).

A second well-supported incongruence concerns the placement of the monotypic *Pseudocydonia* (Fig. 1). In cpDNA tree, *P. sinensis* is sister to *Chaenomeles* (PP 90%; BS 78%; clade B in Fig. 1a), whereas in ITS tree it is sister to *Cydonia* (PP 100%; BS 100%; Fig. 1b). Constraining the monophyly of *Pseudocydonia* plus *Cydonia* in cpDNA analysis, or the monophyly of *Pseudocydonia* plus *Chaenomeles* in ITS analysis, yields significantly worse results (SH test; $P < 0.01$; Table 1; Appendix 2c).

3.3. Combined analyses

Because strongly supported conflicts between cpDNA and ITS data are limited to the placement of *Micromeles* and *Pseudocydonia*, we removed these two groups and merged the remaining data to carry out a combined analysis. The resulting ML tree, as shown in Fig. 2, gives better resolution and stronger support to relationships among the genera. The three major clades (A, B, and C) obtained in the cpDNA analyses (Fig. 1) were recovered in both the ML and Bayesian trees. Bootstrap and posterior probability values are generally higher for these clades in the combined than in the separate analyses. Similar to the cpDNA and ITS results, clade A (PP 100%; BS 100%) contains species of (*Malacomeles* (*Peraphyllum*, *Amelanchier*)) and (*Crataegus*, *Mespilus*). In the *Crataegus* clade (PP 93%; BS 100%), the species are further divided into three well-supported subclades, which correspond to the sectional classification of the group as well as to geographic distributions (see Lo et al., 2007, 2009 for details).

In clade B the relationships of *Pyracantha* and *Stranvaesia* are poorly resolved. However, the *Malus–Eriolobus–Docynopsis–Docynia* clade is strongly supported (PP 98%; BS 83%). Species of *Torminalis* and *Chamaemespilus* are nested in the *Aria* subclade (PP 100%; BS 90%), and species of *Photinia* and *Aronia* are shown to be closely related (PP 98%; BS 88%). *Dichotomanthes* and *Cydonia* are nested in the *Aria–Aronia–Photinia* clade of the combined data (PP 78%; BS 65%), but relationships among them are not well resolved. While the clade containing all *Malus* species is poorly supported, a

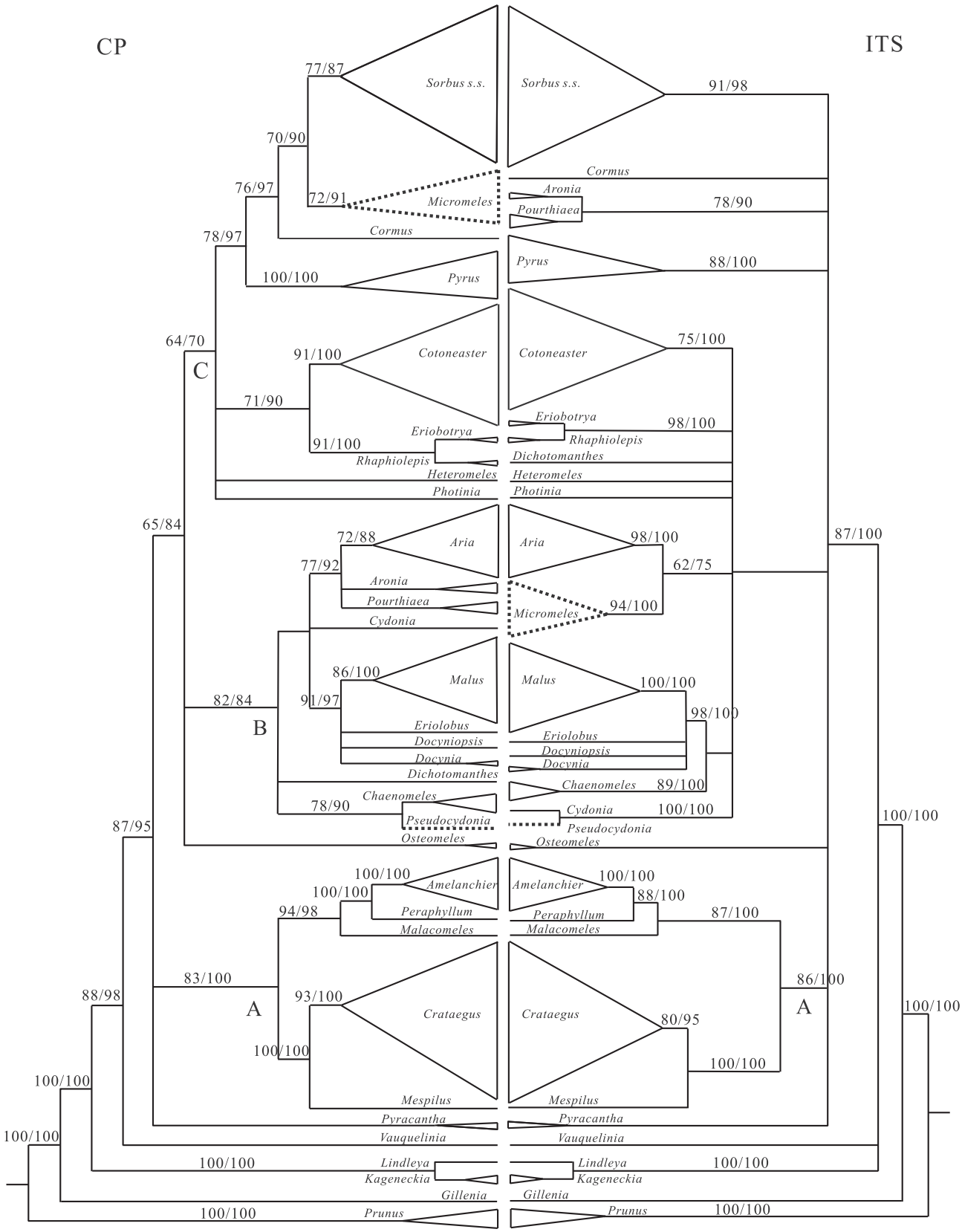


Fig. 1. Summary trees from maximum likelihood (ML) analyses of (a) combined chloroplast DNA sequence data, and (b) nuclear ribosomal ITS DNA sequence data for Pyreae. Species of *Prunus* were included for rooting. Species names appear in Fig. 2 and voucher specimen information is provided in Appendix 1 (Supplementary material). The three major clades resolved in the chloroplast-based tree are labeled A (*Amelanchier–Crataegus*), B (*Aria–Malus*), and C (*Cotoneaster–Sorbus*). Nodes with bootstrap values (BS; left) and posterior probabilities (PP; right) >50% are indicated. Dotted lines mark taxa showing significant conflict between the chloroplast and ITS trees.

subclade that contains most but not all of the European and Asian species is supported (PP 87%; BS 73%).

Clade C in the combined tree (PP 95%; BS 88%; Fig. 2) contains two major subclades: (*Cotoneaster*, *Heteromeles* (*Eriobotrya*, *Rhaphiolepis*)) (PP 92%; BS 80%) and *Pyrus* (*Cormus*, *Sorbus sensu stricto*) (PP 97%; BS 85%). Consistent with the results of the separate analyses (Fig. 1), *Eriobotrya* and *Rhaphiolepis* are closely related with one another (PP 100%; BS 92%), but their connection to the *Cotoneaster–Heteromeles* clade lacks support (Fig. 2). Within the *Cotoneaster* clade (PP 100%; BS 100%), the species are divided into two well-supported subclades, which both contain species assigned to the two subgenera, *Chaenopetalum* and *Cotoneaster* (Fryer and Hylmö, 2009), implying that neither one is monophyletic. Within the *Sorbus sensu stricto* clade (PP 100%; BS 100%), the species are also divided into two well-supported subclades, which correspond to the subgeneric classification as well as to fruit color differences (McAllister, 2005). One subclade (PP 94%; BS 85%), corresponding to subgenus *Sorbus* (including *S. californica*, *S. tianshanica*, and their relatives), contains mostly red-fruited species from Europe, Asia, and North America. The other subclade (PP 100%; BS 95%), corresponding to subgenus *Albocashmarianae* (including *S. parvifructa*, *S. gonggashanica*, and their relatives), contains white-fruited species exclusively from Asia. Intrageneric relationships in *Sorbus*, as well as character evolution and biogeography, are the subject of a separate paper (Lo, McAllister, and Donoghue, in prep).

3.4. Divergence time among lineages

The confidence interval of the ages obtained from PL and Bayesian analyses are largely overlapped with one another, with the greatest differences being in the earliest divergence events, especially the split between (*Kageneckia*, *Lindleya*) and the remaining Pyreae, which includes *Vauquelinia* and the Pyrinae, which in turn includes clades A, B, and C (node 1 in Table 2; Fig. 3). Analyses using the two root age constraints (73 and 104 MY) show a similar pattern: PL tends to give younger ages than BEAST for the earliest divergences but consistently older ages for more nested events closer to the present (starting at around node 6). As expected the 73 and 104 MY constraints yielded consistently younger and older divergence time estimates, respectively. The greatest differences between these analyses in absolute terms concern the very first splitting events (i.e., nodes 1 and 2 in Table 2; Fig. 3), where there is limited overlap in the ranges of the two estimates. Moving closer to the present, the estimates from the two methods, under either constraint, tend to converge (Table 2).

When the root age for Rosaceae was set to 73 MY, PL analyses estimated that the split between *Vauquelinia* and the large Pyrinae clade occurred at around 53 ± 10 MY, during the early Eocene (node 2 in Table 2; Figs. 2 and 3). BEAST estimated that this divergence occurred earlier, during the Paleocene (58 ± 6 MY). Splits at nodes 3–11 are inferred to have occurred during the Eocene (or possibly in the early Oligocene), nodes 12–19 are centered in the Oligocene (but might extend into the early Miocene), and nodes 20–23 are squarely in the Miocene to the present. We note that divergences between a number of the major genera are inferred during the late Eocene, including the splits between *Amelanchier* and its relatives and *Crataegus*, between *Malus* and *Aria*, between *Cotoneaster* and its relatives and (*Sorbus sensu stricto*, *Pyrus*), and between *Sorbus sensu stricto* and *Pyrus* (Table 2).

When the root age for Rosaceae was set to 104 MY, PL analyses estimated that the split between *Vauquelinia* and the Pyrinae occurred at around 70 ± 12 MY, near the end of the Cretaceous into the Paleocene (node 2 in Table 2; Figs. 2 and 3). BEAST estimated that this divergence occurred much earlier, within the late Cretaceous (81 ± 10 MY). Splits at nodes 3–11 are inferred to have occurred during the Paleocene and to the middle Eocene, nodes 12–19 from the late Eocene through the Oligocene, and nodes 20–23 from the Miocene to the present. Divergences between the major genera are correspondingly older, mostly between 50 and 60 MY (Table 2).

Crown clade *Sorbus sensu stricto* could be as young as 28 ± 10 MY, and crown *Aria* could be as young as 20 ± 7 MY (Table 2). This suggests that hybridization occurred between the ancestral lineages of *Sorbus* and *Aria* (the probable parents of *Micromeles*) before the early Miocene. Likewise, the crown age of *Chaenomeles* (clade B) was estimated to be as recent as 20 ± 9 MY, and that of *Cydonia* (clade B) could be as recent as 36 ± 4 MY (Table 2; Fig. 2). Therefore, hybridization between members of the ancestral lineages of *Chaenomeles* and *Cydonia* (the probable parents of *Pseudocydonia*) could have happened before the early Miocene.

4. Discussion

The occurrence of complicating phenomena such as hybridization, polyploidy, and apomixis has, until recently, vitiated attempts to infer the evolutionary history of the Pyreae. Nevertheless, appreciable progress has been made over the last two decades in understanding the phylogeny, morphology, fossil history, and reproductive biology of the group (e.g., Robertson et al., 1991, 1992; Rohrer et al., 1994; Campbell et al., 1995; Evans and Dickinson, 1999; Aldasoro et al., 2004; DeVore and Pigg, 2007; Dickinson et al., 2007). More specifically, Campbell et al. (2007) and Potter et al. (2007) have conducted phylogenetic analyses of the Pyreae, and have discussed the evolution of several key morphological and biochemical features. However, previous studies have been limited in both taxon and gene sampling, and it has been our primary goal to, on the one hand, include additional and variable chloroplast gene regions (i.e., intergenic regions including *atp*B–*rbc*L, *trn*G–*trn*S, *psb*A–*trn*H, *trn*H–*rpl*2, *trn*C–*ycf*6, *rpl*20–*rpl*12; see Section 2). These regions evolve faster than coding genes such as *rbc*L and *ndh*F, and thus provide better phylogenetic resolution to our studied taxa. On the other hand, we aim to greatly expand the sampling in order to more confidently establish phylogenetic relationships among and within the major Pyreae lineages. These together can provide a broad framework for the analysis of evolutionary patterns in the group and for an improved classification system.

4.1. Generic limits and hybridization between ancestral lineages

The circumscription of genera within the Pyreae has been controversial for more than a century (see Table 1 in Robertson et al., 1991). The disagreement of defining genera in a narrow sense stems in part from weak reproductive barriers between species belonging to different generic and subgeneric groups and, consequently, the existence of differentiated phenotypes that have blurred taxonomic boundaries at several levels. The most extreme opinion on the classification of the Pyreae was that of Sax (1931) who proposed merging all the existing genera into a single genus. This suggestion was unacceptable to many workers who noted on the evident morphological diversity within the group, and among many of the previously recognized genera (Phipps et al., 1991; Robertson et al., 1991). A number of studies based on morphological, anatomical, and phytochemical data resulted in the recognition of 26 genera within the Pyreae (Kovanda, 1965; Kalkman, 2004). Our molecular data, which represent nearly half of the described species, are consistent with the recognition of many of the smaller (e.g., *Chaenomeles*, *Pyrus*, and *Amelanchier*) and larger (e.g., *Crataegus*, *Malus*, and *Cotoneaster*) genera. In general, these

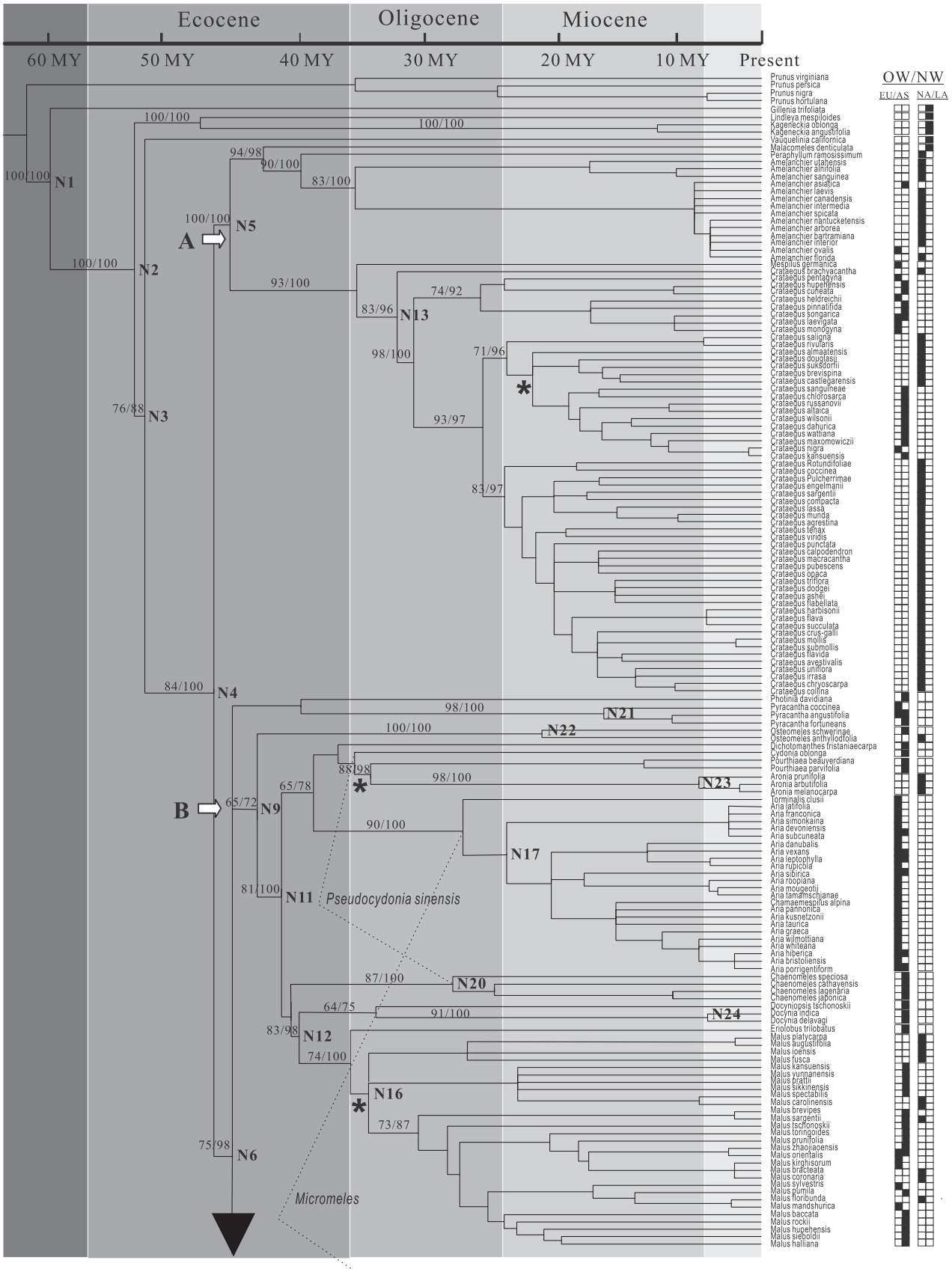


Fig. 2. Chronogram of Pyreae based on penalized likelihood analyses of the combined chloroplast and ITS data. Boxes on the right indicate the biogeographic distribution of the examined taxa: EU = Europe, AS = Asia, NA = North America, and LA = Latin America. Nodes with bootstrap values (BS; left) and posterior probabilities (PP; right) >50%, which are estimated to be older than 25 MY are indicated. For this analysis (see text) the root node of Pyreae and *Prunus* was set at 73 MY; in addition, two internal minimum age constraints were used, at the root of *Prunus* (40 MY) and at the root of *Amelanchier* (40 MY). Dotted lines indicate possible hybridization events giving rise to putative hybrid taxa, *Pseudocydonia* and *Micromeles*.

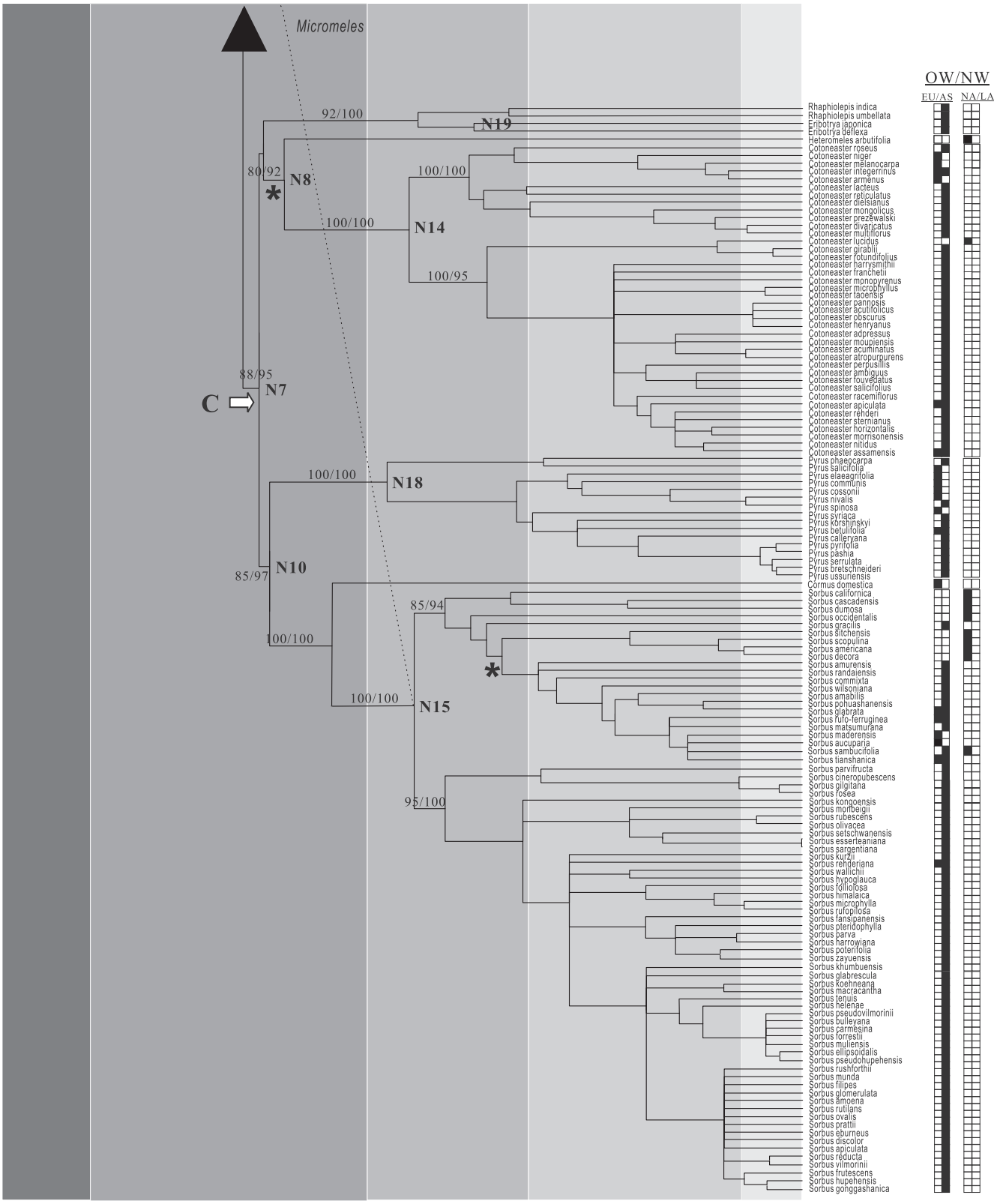


Fig. 2 (*continued*)

are supported as being monophyletic (Figs. 1 and 2). The one major exception concerns *Sorbus*.

Sorbus L. has previously been circumscribed to include both the pinnate-leaved species (*Sorbus s.s.* and *Cormus*) and the simple-leaved species (*Aria*, *Micromeles*, *Chamaemespilus*, and *Torminalis*) (Table 1; Rehder, 1940; Yu, 1974; Phipps et al., 1990; Aldasoro et al., 1998). In large part, this delimitation reflected the occurrence in Europe and western Asia of numerous apomictic microspecies of apparent hybrid origin involving species of *Aria, Sorbus s.s.*, and *Torminalis* (Aas et al., 1994; Nelson-Jones et al., 2002).

Table 2
Results of Shimodaira–Hasegawa tests for the four hypotheses tested. Detailed information on topological constraints is provided in Supplementary Fig. 1a–d. The likelihood score of the best ML tree obtained from the data is compared with trees resulting from analyses that are compatible with the constraint. *P*-values indicate whether the two resulting trees are significantly different from one another.

Hypothesis	Dataset	Topological treatment in constraint analyses	Likelihoods	*P*-value
(A) Monophyly of *Sorbus s.l.*	cpDNA	No constraint	−23798.03	
		Aria, *Sorbus*, *Chamaemespilus*, *Cormus*, *Micromeles*, and *Torminalis* constrained to be monophyletic	−23917.38	<0.001
	ITS	No constraint	−6712.11	
		Aria, *Sorbus*, *Chamaemespilus*, *Cormus*, *Micromeles*, and *Torminalis* constrained to be monophyletic	−6758.37	0.01
(B) Hybrid origin of *Micromeles*	cpDNA	No constraint	−19776.28	
		Micromeles constrained as sister to *Aria*	−19885.63	<0.001
	ITS	No constraint	−5619.02	
		Micromeles constrained as sister to *Sorbus*	−5647.42	0.03
(C) Hybrid origin of *Pseudocydonia*	cpDNA	No constraint	−17526.82	
		Pseudocydonia constrained as sister to *Cydonia*	−17601.73	0.01
	ITS	No constraint	−3232.64	
		Pseudocydonia constrained as sister to *Chaenomeles*	−3279.50	0.02
(D) Three-clade topology of Pyreae	cpDNA	No constraint	−29760.94	
		–	–	–
	ITS	No constraint	−13125.12	
		Constrained into three clades	−13130.05	>0.05

Table 3
Divergence time estimates for Pyreae clades from penalized likelihood (PL) and Bayesain (BI) analyses. The divergence of Pyreae and *Prunus* was set using two maximum age constrains 73 and 104 MY (see Section 2). Node numbers refer to those marked in Fig. 2.

Node #		PL-73 (CI)	BI-73 (CI)	PL-104 (CI)	BI-104 (CI)
Age of split between clades					
1	*Gillenia* and Pyreae	60 (56–65)	68 (63–73)	78 (64–90)	92 (84–100)
2	(*Kageneckia–Lindleya*) and the remaining Pyreae	53 (44–64)	66 (58–72)	71 (54–86)	86 (70–96)
3	*Vauquelinia* and clades ABC	53 (40–63)	58 (52–64)	70 (46–82)	81 (64–91)
4	Clade A and (B–C)	48 (40–60)	55 (47–63)	61 (42–76)	65 (51–79)
5	*Amelanchier* and *Crataegus*	47 (40–60)	50 (44–53)	59 (41–74)	58 (46–71)
6	Clade B and C	45 (39–52)	47 (40–52)	58 (47–66)	59 (47–74)
7	*Cotoneaster* and (*Sorbus–Pyrus*)	43 (26–50)	43 (34–50)	56 (49–61)	49 (32–61)
8	*Cotoneaster* and (*Eriobotrya–Rhaphiolepis*)	43 (25–47)	40 (31–49)	55 (48–61)	47 (30–59)
9	*Osteomeles* and the remaining clade B	42 (27–47)	40 (31–47)	54 (40–60)	51 (38–65)
10	*Pyrus* and *Sorbus*	42 (27–49)	38 (30–48)	54 (41–60)	45 (29–57)
11	*Malus* and *Aria*	41 (34–45)	40 (29–48)	46 (31–57)	40 (28–51)
12	*Chaenomeles* and *Malus*	39 (26–44)	36 (26–45)	50 (41–57)	44 (32–58)
Age of the first split within clades (by clade size)					
13	*Crataegus* (ca. 150 species)	33 (28–40)	32 (25–41)	39 (28–53)	33 (25–47)
14	*Cotoneaster* (ca. 70 species)	30 (24–36)	22 (15–32)	39 (30–46)	25 (11–42)
15	*Sorbus sensu stricto* (ca. 80 species)	29 (23–35)	28 (18–38)	36 (27–44)	35 (20–50)
16	*Malus* (ca. 40 species)	36 (28–44)	30 (20–39)	46 (38–56)	34 (22–45)
17	*Aria* (ca. 50 species)	29 (22–36)	20 (15–28)	37 (28–47)	25 (14–37)
18	*Pyrus* (ca. 20 species)	33 (29–38)	27 (21–40)	42 (24–48)	33 (18–48)
19	*Eriobotrya* (ca. 18 species)	25 (21–29)	21 (14–25)	33 (15–40)	20 (12–36)
20	*Chaenomeles* (5 species)	25 (19–30)	20 (12–31)	32 (22–42)	24 (13–37)
21	*Pyracantha* (3 species)	14 (11–18)	7 (3–15)	17 (5–24)	10 (5–16)
22	*Osteomeles* (2 species)	20 (15–25)	14 (5–28)	24 (10–37)	17 (5–36)
23	*Aronia* (3 species)	4 (2–6)	3 (0.4–8)	5 (2–12)	3 (0.5–8)
24	*Docynia* (2 species)	4 (2–6)	2.5 (0.2–8)	6 (1–10)	3 (0.3–9)

These microspecies appeared to span the major groups, making it at least inconvenient to recognize a number of separate genera (McAllister, 2005). Another factor concerned fruit traits. Although some of these traits appeared to link segregate groups (e.g., heterogeneous flesh unites *Aria* and *Micromeles*), the distribution of others (e.g., starch composition, the shape and distribution of sclereids, and the structure of the seed coat and endosperm) variously overlapped among the segregate groups (Kovanda, 1961; Gabrielian, 1978). Leaf structure (pinnate versus simple) and fruit pulp texture (homogeneous versus heterogeneous) were the main characters cited in separating the species into two major genus-level groups: "*Sorbus s.s.*" and *Aria*.

Our chloroplast and ITS results clearly indicate that the pinnate- and simple-leaved species do not form a single monophyletic group. Instead, all of the pinnate-leaved species (*Sorbus s.s.* and *Cormus*) are strongly united within clade C, while most of the simple-leaved species (*Aria*, *Chamaemespilus*, and *Torminalis*) are united within clade B. The major problem concerns the placement of the simple-leaved *Micromeles* species (Fig. 1). In the ITS data *Micromeles* appears to be closely related to the other simple-leaved species (*Aria, Chamaemespilus*, and *Torminalis*), but in the chloroplast data *Micromeles* is associated with the pinnate-leaved ones (*Sorbus s.s* and *Cormus*). One possible explanation for this conflict is that the ancestral lineage of *Micromeles* originated via hybridization between members of the ancestral lineages of the pinnate- and the simple-leaved groups. In this case, given our data, it would appear that an ancestor of a pinnate-leaved species served as the maternal parent, and the chloroplast genome was introgressed into

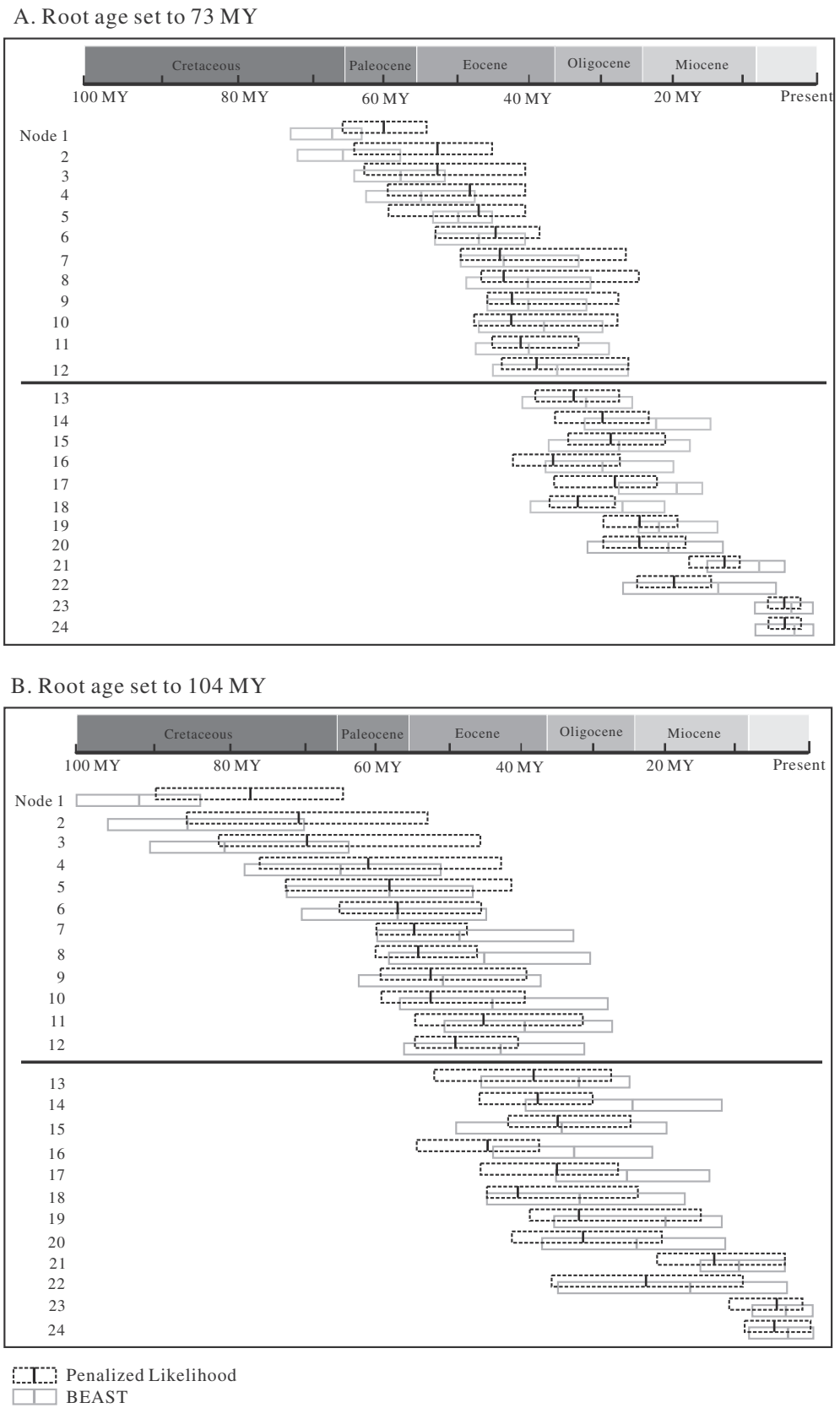
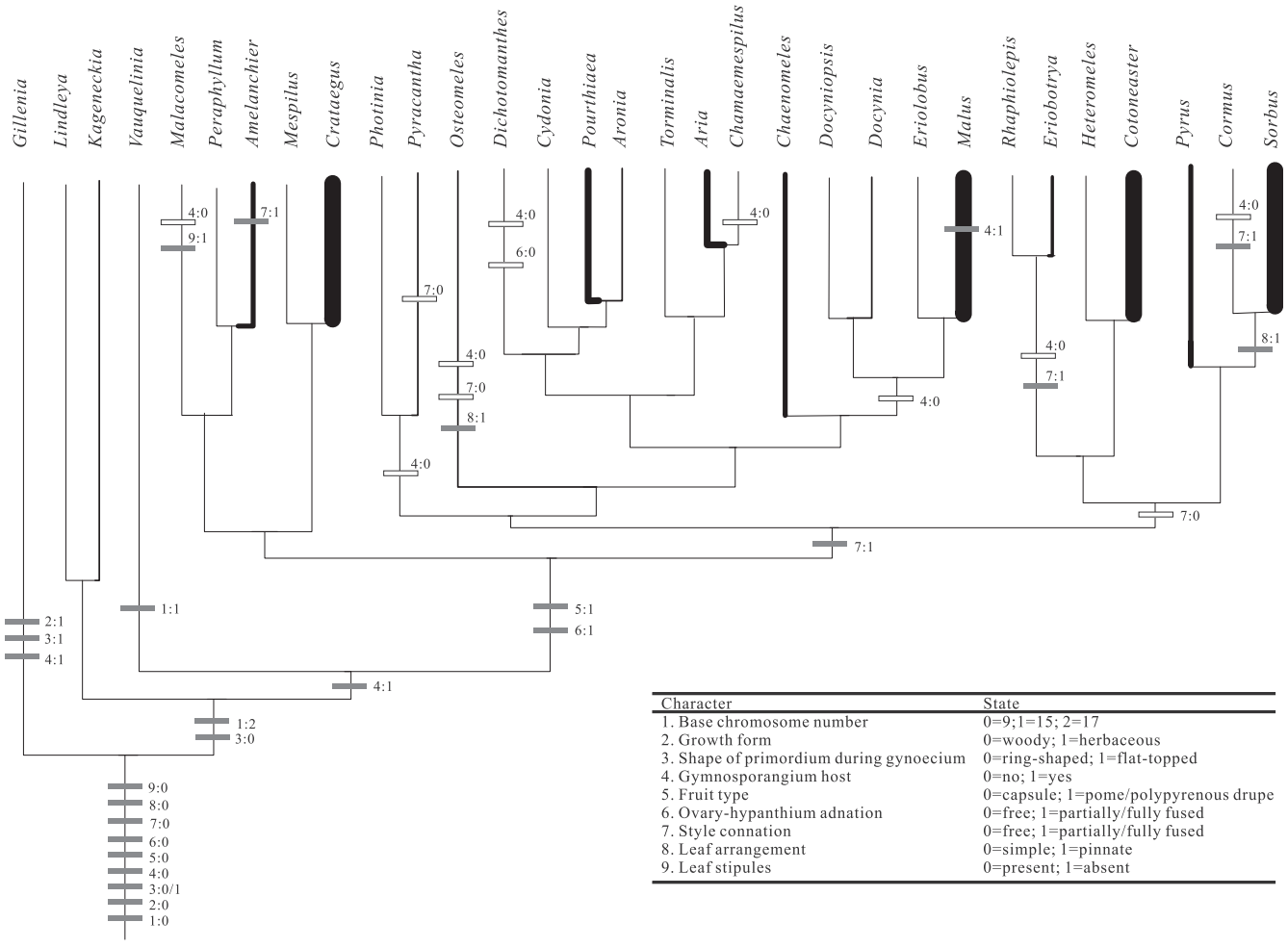


Fig. 3. Divergence time estimates within the Pyreae based on the penalized likelihood (PL) and BEAST (BI) analyses (see text). For both PL and BI analyses, the root age was set at both 73 and 104 MY, bracketing the inferred maximum and minimum ages for crown Roscaceae. Nodes 1–11 mark estimated divergences among major (above genus) lineages; nodes 12–23 mark estimated divergences within genera (see Fig. 2, Table 3).

the ancestor of *Micromeles*. Based on our nuclear data, the paternal parent would have been an ancestor of a simple-leaved species.

Other lines of evidence support a hybrid origin hypothesis for *Micromeles*. Species of *Micromeles* show some morphological resemblance to both of the putative parental taxa. Apart from their simple, unlobed leaves, *Micromeles*, *Aria*, and *Chamaemespilus* share pomes with heterogeneous flesh due to the presence of clusters of large tanniferous cells in the fleshy tissues (Kovanda, 1961; Iketani and Ohashi, 1991). On the other hand, in species of *Micromeles* the lower leaf surfaces are more or less glabrous, resembling most of the eastern Asian and North American *Sorbus* species, but differing from all *Aria* species in which the leaves are white-tomentose beneath. In addition, *Micromeles* and most of the white-fruited *Sorbus* species (e.g., section *Multijugae*) have fused carpel apices, whereas *Aria* species have free carpel apices (Aldasoro et al., 2004; McAllister, 2005). Geographically, species of *Aria* and *Chamaemespilus* are found only from West Europe to the Caucasus region. There is a slight overlap in Iran, but otherwise they do not overlap in distribution with the *Micromeles* species, which are currently found in eastern Asia, including China, Korea, Japan, and Malaysia (Aldasoro et al., 2004). On the other hand, species of the pinnate-leaved *Sorbus* clade occur throughout the moist, cool temperate regions of the northern hemisphere, and they overlap with *Micromeles* in eastern Asia and with simple-leaved *Aria* species in western Europe (McAllister, 2005). For example, *S. aucuparia* alone spans from Madeira in the west to the Soviet Far East and into North America. Based on these distribution patterns, there are two feasible hypotheses concerning the geographic origin of *Micromeles*. One is that hybridization took place in Europe where *Aria* and *Sorbus* overlapped. In this case, the hybrid lineage dispersed into central and eastern Asia subsequent to its formation, followed by the extinction of *Micromeles* species in Europe, probably during the late Cenozoic glacial period. An alternative scenario is that hybridization took place in Asia, but this would necessitate that *Aria* previously co-existed with *Sorbus* in Asia and subsequently went extinct there. The reason why only *Aria* went extinct in Asia but not all *Sorbus s.s.* and its relatives is unclear.

A second major incongruence between the chloroplast and nuclear data is the placement of the monotypic *Pseudocydonia* (Figs. 1 and 2). This conflict was also noted by Campbell et al. (2007, p. 138): "Nuclear and morphological data, but not cpDNA, support the *Cydonia–Pseudocydonia* clade." One potential explanation is that the *Pseudocydonia* lineage originated via hybridization between members of the ancestral lineages of *Cydonia* and *Chaenomeles*. In this case, an ancestor of *Chaenomeles* served as the maternal donor and its chloroplast genome was introgressed into *Pseudocydonia*. *Pseudocydonia sinensis* is a diploid species and is native to China, where species of *Cydonia* and *Chaenomeles* co-occur. There are several morphological characters shared between *Pseudocydonia* and its putative parents. For example, *Pseudocydonia*, *Cydonia*, and *Chaenomeles* share similar fruit feature (egg-shaped, yellowish pomes of medium size, ca. 5–7 in.). *Cydonia* flowers bear



Character	State
1. Base chromosome number	0=9;1=15; 2=17
2. Growth form	0=woody; 1=herbaceous
3. Shape of primordium during gynoecium	0=ring-shaped; 1=flat-topped
4. Gymnosporangium host	0=no; 1=yes
5. Fruit type	0=capsule; 1=pome/polypyrenous drupe
6. Ovary-hypanthium adnation	0=free; 1=partially/fully fused
7. Style connation	0=free; 1=partially/fully fused
8. Leaf arrangement	0=simple; 1=pinnate
9. Leaf stipules	0=present; 1=absent

Fig. 4. Mapping of nine non-molecular characters (Savile, 1979; Robertson et al., 1991; Rohrer et al., 1994; Kalkman, 2004; Evans and Dickinson, 2005) on the combined chloroplast- and nuclear ITS-based ML tree. Character state changes are labeled along branches. Gray boxes represent state change from 0 to 1 and white boxes represent reversals. Thickness of the terminal branches is proportional to the number of species included in each of the groups included in this study (Table 1).

persistent, recurved calyx lobes and free styles, while *Pseudocydonia* resembles *Chaenomeles* in having a deciduous calyx and fused style (Robertson et al., 1991; Kalkman, 2004). Our phylogenetic analyses suggest that *P. sinensis* may have originated as early as the late Miocene.

4.2. The timing of lineage divergences

The origin of the Pyreae appears to have involved three major changes (Fig. 4; Evans and Campbell, 2002; Evans and Dickinson, 2005): (1) an initial chromosome doubling (or polyploidization) event followed by an aneuploid reduction that resulted in an increase in chromosome number from $n = 9$ (both *Gillenia stipulata* and *G. trifoliata*) to $n = 15$ (*Vauquelinia*) or 17 (the rest of Pyreae); (2) a shift from herbaceous (*Gillenia*) to woody (Pyreae) growth form or vice versa (i.e., *Gillenia* evolved the herbaceous habit); and (3) a change from a flat-topped primordium (*Gillenia*) to a ring-shaped primordium (Pyreae) on the floral apex during gynoecium development (Evans and Dickinson, 2005) or vice versa. Subsequent to these changes, there was another major transition that marked the origin of the Pyrinae, possibly near the onset of the Eocene. This was a shift from dry capsular fruits that rely on wind dispersal (which characterize the early-diverging lineages, *Kageneckia, Lindleya*, and *Vauquelinia*, in which the carpels are free from the hypanthium and the ovaries are superior) to fleshy fruits (pomes or polypyrenous drupes, derived from carpels that are adnate to the fleshy hypanthium through various extent of intercalary growth resulting in an inferior ovary; Evans and Dickinson, 2005) that rely on animal dispersal (Rohrer et al., 1994). This change in fruit morphology also appears to have been accompanied by a geographical range shift, again possibly during the early Eocene, from a more limited distribution in southwestern US and northern Mexico (*Vauquelinia*) and Central and South America (*Kageneckia* and *Lindleya*) to a widespread northern temperate distribution in both the Old and New World.

Between and within Pyreae genera there appear to be at least five major, early Old World-New World disjunctions (labeled by asterisks in Fig. 2). A number of other intercontinental disjunctions have occurred more recently within several of the major lineages, but a better understanding of these cases will require more detailed analyses within these clades. One of the early disjunctions is within *Crataegus* in clade A (Fig. 2), reflected by the connection between the North American and Asian species (BS 71%; PP 96%). This geographic split within *Crataegus* may have taken place by the early Miocene (22 ± 3 MY), much earlier than the time estimated in a previous study (4.6 ± 0.9 MY; Lo et al., 2009). This discrepancy in divergence times reflects differences between these studies in sampling and in the fossils used for calibration. Within clade B two early Old World-New World disjunctions are evident. One is between the American *Aronia* Medik. and the Asian *Pourthiaea* Decne. and the other is within *Malus* (where there have also been several other, more recent disjunctions). These splits are estimated to have occurred as recently as the early Oligocene. *Aronia* and *Pourthiaea* were previously described in *Photinia* Lindl. (Robertson et al., 1991). However, recent molecular data indicated that *Photinia* sensu lato is not a monophyletic group, and that *Aronia* and *Pourthiaea* could be recognized as distinct genera (Guo et al., 2011). Despite limited sampling of *Photinia* (*P. davidiana*) and *Pourthiaea* (*P. beauverdiana* and *P. parviflora*) in this study, our findings are consistent with those of Guo et al. (2011) showing that *Aronia* and *Pourthiaea* are closely related to one another and that *P. davidiana* is distant from the *Aronia–Pourthiaea* clade (Figs. 1 and 2). Given that *Aronia* species are found exclusively in the New World, whereas *Pourthiaea* species in the Old World at present, the disjunction between the two groups suggests that their species could be previously widespread, but geological and climatic changes during the Oligocene may have resulted in the extinction or differentiation of several lineages within the groups. Within clade C two additional early disjunctions are detected. One is between the monotypic North American *Heteromeles* and the mainly Asian *Cotoneaster*, which may date back to the Eocene. The other is within *Sorbus*, probably during the late Oligocene (Fig. 2). In view of their inferred timing, we suspect that most of these intercontinental vicariance events are most consistent with movement through the Bering Land Bridge, though we cannot strongly rule out movement in several of the older cases across the North Atlantic (Tiffney, 1985a,b; Donoghue et al., 2001; Milne, 2006). Beringia would certainly have been the likely route for the several more recent disjunctions within the major lineages.

Several of the genus-level clades appear to have originated around the Eocene–Oligocene boundary (Table 2; Fig. 3). For instance, crown *Crataegus* (node 12) within clade A may have started to diversify at around 32 ± 8 MYA. The *Sorbus sensu stricto* (node 14) and *Cotoneaster* (node 13) clades within clade C may have started to diversify during the mid-Oligocene (Table 2). According to our estimates, *Malus* (node 15) of clade B may have started to diversify by the early Oligocene, somewhat earlier than the diversification of crown *Aria* (node 16). We note that the crown ages of the various genera do not appear to be correlated with their species richness. For example, smaller genera such as *Eriobotrya* (18 species), *Chaenomeles* (5 species), *Osteomeles* (2 species), and *Pyrus* (20 species) also appear to date back to the Oligocene (20–33 MY), with the exception of crown *Aronia* (3 species) and *Docynia* (2 species), which may have begun to diversify some 6 MY ago (Table 2). Although the stem lineages of most of the major lineages seem to have been established in the Eocene, it was not until much later, probably mostly in the Miocene, that much of the modern species diversity originated (Fig. 2). One possibility is that the origin of the several major Pyreae lineages during the Eocene–Oligocene transition may have been related to cooling and increased aridity at northern temperate latitudes (Liu et al., 2009). Possible increases in diversification rate during the Miocene might relate to even more pronounced aridification and to the spread of several lineages around the northern hemisphere (cf. Moore and Donoghue, 2007).

While the precise location and the causes of shifts in the diversification rate warrant much additional attention, we are struck by the pattern seen in Fig. 2 in which rather large clades (with some 40–100 species) are sister to much small ones (<10 species). Given that these paired lineages have been diverging for the same length of time, it would appear that there have been significant shifts (upward or downward) in the rate of diversification. In future studies it will be important to explore correlations with geographic shifts, with morphological trait evolution, and with change in ploidy and reproductive system. As always, however, it will be difficulty to tease apart the effects of increased speciation versus increased extinction in these lineages (Ricklefs, 2007; Rabosky, 2009).

4.3. Summary and future directions

Our analyses of a much broader sample of Pyreae diversity have identified three major clades within the Pyrinae (above the early diverging *Kaganeckia*, *Lindleya*, and *Vauquilinia* lineages) that probably originated in the early to mid-Eocene. Within these three clades, many of the previously recognized genera appear to be monophyletic with the major exception of *Sorbus sensu lato*, which is clearly demonstrated to form two distantly related clades. One of these clades, *Sorbus sensu stricto* (Robertson et al., 1991; McAllister, 2005), is characterized by pinnately compound leaves, and the other one, *Aria*, has simple leaves. Within the three major clades we find a repeated pattern in which species-rich and species-poor lineages are sister groups, suggesting that there have been re-

peated shifts in diversification rate. We also have identified a number of inter-continental vicariance events, which may have come about via migration, possibly through Beringia, at several different times. We also note that the origin of several major clades may have been related to the onset of colder conditions during the Oligocene, and that their further diversification may have correlated with the spread of colder and more arid climates during the Miocene. Although our chloroplast and ITS datasets yield generally similar trees, several strong conflicts suggest the possibility that two lineages – *Micromeles* and *Pseudocydonia* – may have originated following hybridization events. We are hopeful that nuclear genes can be useful in further resolving relationships within the Pyreae. Thus far, however, analyses of nuclear genes (e.g., the four copies of GBSSI; Evans et al., 2000) have yielded inconsistent phylogenetic results, perhaps as a function of lineage sorting events and/or the loss of gene copies subsequent to duplication (Campbell et al., 2007). The application of coalescent-based methods (see Edwards, 2009; Knowles, 2009) may prove useful as studies of nuclear genes progress.

Acknowledgments

We thank Hugh McAllister, Natacha Frachon, David Boufford, Knud IB Christensen, Tim Baxter, Michael Dosmann, Kathryn Richardson, Melinda Peters, Xue-Jun Ge, Maria Kuzmina, Tim Dickinson, and Nadia Talent for providing materials as well as assisting in plant collection and identification; Patrick Sweeney for his help with voucher specimens; the Yale Molecular Systematics and Conservation Genetics Laboratory and the W.M. Keck Foundation Biotechnology Resources Laboratory for providing sequencing services; and the Harvard University Herbaria, the New York Botanical Garden, the Ness Garden (Liverpool), the Jardin Botanique de Montréal, the Royal Botanical Garden (Edinburgh), the Missouri Botanical Garden, the South China Botanical Garden, the Singapore Botanical Garden, and the Yale Herbarium for granting access to dried specimens and/or living collections. We are especially grateful to Hugh McAllister for sharing his knowledge and for his valuable comments on an early version of this manuscript; to Dan Potter, Elizabeth Anne Zimmer, and an anonymous reviewer for their comments and suggestions to improve the manuscript. This work was supported by Yale University funds provided to M.J.D.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ympev.2011.10.005.

References

Aas, G., Maier, J., Baltisberger, M., Merzger, S., 1994. Morphology, isozyme variation, cytology, and reproduction of hybrids between *Sorbus aria* (L.) Crantz and *S. torminalis* (L.). Crantz. Bot. Helv. 104, 195–214.

Aldasoro, J.J., Aedo, C., Navarro, C., Garmendia, F.M., 1998. The genus *Sorbus* (Maloideae, Rosaceae) in Europe and in North Africa: morphological analysis and systematics. Syst. Bot. 23, 189–212.

Aldasoro, J.J., Aedo, C., Garmendia, F.M., de la Hoz, F.P., Navarro, C., 2004. Revision of *Sorbus* subgenera *Aria* and *Torminaria* (Rosaceae–Maloideae). Syst. Bot. Monogr. 69, 1–148.

Baldwin, B.G., Sanderson, M.J., 1998. Age and rate of diversification of the Hawaiian silversword alliance (Compositae). Proc. Natl. Acad. Sci. USA 95, 9402–9406.

Bell, C.D., Soltis, D.E., Soltis, P.S., 2010. The age and diversification of the angiosperms re-revisited. Am. J. Bot. 97, 1296–1303.

Campbell, C.S., Baldwin, B.G., Donoghue, M.J., Wojciechowski, M.F., 1995. A phylogeny of the genera of Maloideae (Rosaceae): evidence from Internal Transcribed Spacers of nuclear ribosomal DNA sequences and congruence with morphology. Am. J. Bot. 82, 903–918.

Campbell, C.S., Evans, R.C., Morgan, D.R., Dickinson, T.A., Arsenault, M.P., 2007. Phylogeny of subtribe Pyrinae (formerly the Maloideae, Rosaceae): limited resolution of a complex evolutionary history. Plant Syst. Evol. 266, 119–145.

Cevallos-Ferriz, S.R.S., Stockey, R.A., 1991. Fruits and seeds from the Princeton chert (Middle Eocene) of British Columbia: Rosaceae (Prunoideae). Bot. Gaz. 152, 369–379.

Challice, J., 1973. Phenolic compounds of the subfamily Pomoideae: a chemotaxonomic survey. Phytochemistry 12, 1095–1101.

DeVore, M.L., Pigg, K.B., 2007. A brief review of the fossil history of the family Rosaceae with a focus on the Eocene Okanogan Highlands of eastern Washington State, USA, and British Columbia, Canada. Plant Syst. Evol. 266, 45–57.

Dickinson, T.A., Lo, E.Y.Y., Talent, N., 2007. Polyploidy, reproductive biology, and Rosaceae: understanding evolution and making classifications. Plant Syst. Evol. 266, 59–78.

Donoghue, M.J., Bell, C.D., Li, J.H., 2001. Phylogenetic patterns in northern hemisphere plant geography. Int. J. Plant Sci. 162, S41–S52.

Drummond, A.J., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. 7, 214–223.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughout. Nucl. Acids Res. 32, 1792–1797.

Edwards, S.V., 2009. Is a new and general theory of molecular systematics emerging? Evolution 63, 1–19.

Evans, R.C., Dickinson, T.A., 1999. Floral ontogeny and morphology in subfamily Spiraeoideae Endl. (Rosaceae). Int. J. Plant Sci. 160, 981–1012.

Evans, R.C., Alice, L.A., Campbell, C.S., Kellogg, E.A., Dickinson, T.A., 2000. The granule bound starch synthase (*GBSSI*) gene in Rosaceae: multiple putative loci and phylogenetic utility. Mol. Phylogenet. Evol. 17, 388–400.

Evans, R.C., Campbell, C.S., 2002. The origin of the apple subfamily (Rosaceae: Maloideae) is clarified by DNA sequence data from duplicated GBSSI genes. Am. J. Bot. 89, 1478–1484.

Evans, R.C., Dickinson, T.A., 2005. Floral ontogeny and morphology in *Gillenia* (''Spiraeoideae'') and subfamily Maloideae C. Weber (Rosaceae). Int. J. Plant Sci. 166, 427–447.

Forest, F., Chase, M.W., 2009. Eurosid I. In: Hedges, S.B., Kumar, S. (Eds.), The Timtree of Life. Oxford University Press, pp.188–196.

Fryer, J., Hylmö, B., 2009. Cotoneasters: A Comprehensive Guide to Shrubs for Flowers, Fruits, and Foliage. Timber Press Inc..

Gabrielian, E., 1978. Rjabiny (*Sorbus* L.) Zapadnoj Azii i Gimalaev, Yerevan.

Goldblatt, P., 1976. Cytotaxonomic studies in the tribe Quillajeae (Rosaceae). Ann. Mol. Bot. Gard. 63, 200–206.

Guo, W., Yu, Y., Shen, R.J., Liao, W.B., Chin, S.W., Potter, D., 2011. A phylogeny of *Photinia* sensu lato (Rosaceae) and related genera based on nrITS and cpDNA analysis. Plant Syst. Evol. 291, 91–102.

Huelsenbeck, J.P., Ronquist, F., 2001. Mr. Bayes: a program for the Bayesian inference of phylogeny. Bioinformatics 17, 754–755.

Iketani, H., Ohashi, H., 1991. Anatomical structure of fruits and evolution of the tribe Sorbeae in the subfamily Maloideae (Rosaceae). J. Jpn. Bot. 66, 319–351.

Kalkman, C., 2004. Rosaceae. In: Kubitzki, K. (Ed.), Flowering Plants – Dicotyledons: Celastrales, Oxalidales, Rosales, Cornales, Ericales. Springer, Berlin, pp. 343–386.

Knowles, L.L., 2009. Statistical phylogeography. Ann. Rev. Ecol. Evol. Syst. 40, 593–612.

Kovanda, M., 1961. Flower and fruit morphology of *Sorbus* in correlation to the taxonomy of the genus. Preslia 7, 1–16.

Kovanda, M., 1965. On the generic limits in the Maloideae. Preslia 37, 27–34.

Liu, Z., Pagani, M., Zinniker, D., DeConto, R., Huber, M., Brinkhuis, H., Sunita, R., Shah, R., Leckie, M., Pearson, A., 2009. Global cooling during the Eocene–Oligocene climate transition. Science 27, 1187–1190.

Lo, E.Y.Y., Stefanović, S., Dickinson, T.A., 2007. Molecular reappraisal of relationships between *Crataegus* and *Mespilus* (Rosaceae, Pyreae) – two genera or one? Syst. Bot. 32, 596–616.

Lo, E.Y.Y., Stefanović, S., Christensen, K.I., Dickinson, T.A., 2009. Evidences for genetic association between East Asian and Western North American *Crataegus* L. (Rosaceae) and rapid divergence of the Eastern North American lineages based on multiple DNA sequences. Mol. Phylogenet. Evol. 51, 157–168.

Maddison, W.P., Maddison, D.R., 2008. Mesquite: A Modular System for Evolutionary Analysis. Version 2.71. <http://mesquiteproject.org>.

Magallón, S., Castillo, A., 2009. Angiosperm diversification through time. Am. J. Bot. 96, 349–365.

Manchester, S.R., 1994. Fruits and seeds of the Middle Eocene Nut Beds flora, Clarno Formation, North Central Oregon. Palaeontogr. Am. 58, 1–205.

McAllister, H., 2005. The Genus *Sorbus* – Mountain Ash and Other Rowans. Royal Botanical Gardens, Kew.

Milne, R.I., 2006. Northern hemisphere plant disjunctions: a window on Tertiary land bridges and climate change. Ann. Bot. 98, 465–472.

Moore, B.R., Donoghue, M.J., 2007. Correlates of diversification in the plant clade Dipsacales: geographic movement and evolutionary innovations. Am. Nat., S28–S55.

Morgan, D.R., Soltis, D.E., Robertson, K.R., 1994. Systematic and evolutionary implications of *rbc* L sequence variation in Rosaceae. Am. J. Bot. 81, 890–903.

Nelson-Jones, E.B., Briggs, D., Smith, A.G., 2002. The origin of intermediate species of the genus *Sorbus*. Theor. Appl. Genet. 105, 953–963.

Pamilo, P., Nei, M., 1988. Relationships between gene trees and species trees. Mol. Biol. Evol. 5, 568–583.

Phipps, J.B., Robertson, K.R., Smith, P.G., Rohrer, J.R., 1990. A checklist of the subfamily Maloideae (Rosaceae). Can. J. Bot. 68, 2209–2269.

Phipps, J.B., Robertson, K.R., Rohrer, J.R., Smith, P.G., 1991. Origins and evolution of subfamily Maloideae (Rosaceae). Syst. Bot. 16, 303–332.

Posada, D., Crandall, K.A., 1998. Modeltest: testing the model of DNA substitution. Bioinformatics 14, 817–818.
Potter, D., Eriksson, T., Evans, R.C., Oh, S.H., Smedmark, J.E.E., Morgan, D.R., Kerr, M., Robertson, K.R., Arsenault, M.P., Dickinson, T.A., Campbell, C.S., 2007. Phylogeny and classification of Rosaceae. Plant Syst. Evol. 266, 5–43.
Rabosky, D.L., 2009. Extinction rates should not be estimated from molecular phylogenies. Evolution 64, 1816–1824.
Rambaut, A., 2002. Se–Al Sequence Alignment Editor v2.0a11. University of Oxford, Oxford.
Rambaut, A., Drummond, A.J., 2003. Tracer. <http://beast.bio.ed.ac.uk/tracer>.
Rehder, A., 1940. Manual of Cultivated Trees and Shrubs Hardy in North America Exclusive of the Subtropical and Warmer Temperate Regions, second ed. MacMillan, New York.
Ricklefs, R.E., 2007. Estimating diversification rates from phylogenetic information. Trends Ecol. Evol. 22, 601–610.
Robertson, K.R., Phipps, J.B., Rohrer, J.R., 1991. A synopsis of genera in Maloideae (Rosaceae). Syst. Bot. 16, 376–394.
Robertson, K.R., Phipps, J.B., Rohrer, J.R., 1992. Summary of leaves in the genera of Maloideae (Rosaceae). Ann. Mol. Bot. Gard. 79, 81–94.
Rohrer, J.R., Robertson, K.R., Phipps, J.B., 1994. Floral morphology of Maloideae (Rosaceae) and its systematic relevance. Am. J. Bot. 81, 574–581.
Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574.
Sanderson, M.J., 2002. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. Mol. Biol. Evol. 19, 101–109.
Sanderson, M.J., 2003. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. Bioinformatics 19, 301–302.
Sanderson, M.J., Doyle, J.A., 2001. Sources of error and confidence intervals in estimating the age of angiosperms from *rbc* L and 18S rDNA data. Am. J. Bot. 88, 1499–1516.
Savile, D.B.O., 1979. Fungi as aids in higher plant classification. Bot. Rev. 45, 380–495.
Sax, K., 1931. The origin and relationships of the Pomoideae. J. Arnold Arbor. 12, 3–22.
Shaw, J., Lickey, E.B., Beck, J.T., Farmer, S.B., Liu, W.S., Miller, J., Siripun, K.C., Winder, C.T., Schilling, E.E., Small, R.L., 2005. The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. Am. J. Bot. 92, 142–166.
Shimodaira, H., Hasegawa, M., 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol. Biol. Evol. 16, 1114–1116.
Smith, S.A., Dunn, C.W., 2008. Phyutility: a phyloinformatics tool for trees, alignments and molecular data. Bioinformatics 24, 715–716.
Stamatakis, A., Ludwig, T., Meier, H., 2005. RAxML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees. Bioinformatics 21, 456–463.
Stamatakis, A., Hoover, P., Rougemont, J., 2008. A rapid bootstrap algorithm for the RAxML web servers. Syst. Biol. 57, 758–771.
Swofford, D.L., 2002. PAUP*: Phylogenetic Analysis Using Parsimony (and Other Methods), Version 40b10. Sinauer Associates, Sunderland.
Tiffney, B.H., 1985a. Perspectives on the origin of the floristic similarity between eastern Asia and eastern North America. J. Arnold Arbor. 66, 73–94.
Tiffney, B.H., 1985b. The Eocene North Atlantic land bridge: its importance in Tertiary and modern phylogeography of the Northern Hemisphere. J. Arnold Arbor. 66, 243–273.
Wikstrom, N., Savolainen, V., Chase, M.W., 2001. Evolution of the angiosperms: calibrating the family tree. Proc. Roy. Soc. Biol. Sci. B 268, 2211–2220.
Wolfe, J.A., Wehr, W., 1988. Rosaceous *Chamaebatiaria*-like foliage from the paleogene of western North America. Aliso 12, 177–200.
Yu, T., 1974. Rosaceae–Prunoideae and Connaraceae – Flora Republicae Popularis Sinicae, vol. 36. Science, Beijing.
White, T.J., Bruns, T., Lee, S., Taylor, J., 1990. Amplification and direct sequencing of fungal ribosomal genes for phylogenies. In: Innis, M., Gelfand, D., Sninsky, J., White, T. (Eds.), PCR Protocols: A Guide to Methods and Applications. Academic Press, San Diego, pp. 315–322.