Journal of Systematics and Evolution 46 (3): 349–374 (2008) doi: 10.3724/SP.J.1002.2008.08056
(formerly Acta Phytotaxonomica Sinica) http://www.plantsystematics.com

Tracking character evolution and biogeographic history through time in Cornaceae—Does choice of methods matter?

Qiu-Yun (Jenny) XIANG* David T. THOMAS

(*Department of Plant Biology, North Carolina State University,* Raleigh, NC 27695, USA)

Abstract This study compares results on reconstructing the ancestral state of characters and ancestral areas of distribution in Cornaceae to gain insights into the impact of using different analytical methods. Ancestral character state reconstructions were compared among three methods (parsimony, maximum likelihood, and stochastic character mapping) using MESQUITE and a full Bayesian method in BAYESTRAITS and inferences of ancestral area distribution were compared between the parsimony-based dispersal-vicariance analysis (DIVA) and a newly developed maximum likelihood (ML) method. Results indicated that among the six inflorescence and fruit characters examined, "perfect" binary characters (no homoplasy, no polymorphism within terminals, and no missing data) are little affected by choice of method, while homoplasious characters and missing data are sensitive to methods used. Ancestral areas at deep nodes of the phylogeny are substantially different between DIVA and ML and strikingly different between analyses including and excluding fossils at three deepest nodes. These results, while raising caution in making conclusions on trait evolution and historical biogeography using conventional methods, demonstrate a limitation in our current understanding of character evolution and biogeography. The biogeographic history favored by the ML analyses including fossils suggested the origin and early radiation of *Cornus* likely occurred in the late Cretaceous and earliest Tertiary in Europe and intercontinental disjunctions in three lineages involved movements across the North Atlantic Land Bridge (BLB) in the early and mid Tertiary. This result is congruent with the role of NALB for post-Eocene migration and in connecting tropical floras in North America and Africa, and in eastern Asia and South America. However, alternative hypotheses with an origin in eastern Asia and early Trans-Beringia migrations of the genus cannot be ruled out.

Key words AReA, BAYESTRAITS, BEAST, biogeography, chromosome evolution, Cornaceae, fruit and inflorescence evolution, LAGRANGE, MESQUITE, divergence time.

Reconstructing ancestral states is a powerful method to understand the pathways and pattern of character evolution, which holds the key to our understanding of evolutionary processes (Cunningham, 1999; Maddison & Maddison, 1992, 2001; Ronquist, 2004). During the past decade, character mapping on a phylogeny has become more and more a common component in research articles to test evolutionary hypotheses, as well as to understand evolution of organism traits and behaviors, and more recently of genes, genomes, and gene functions (Blanchette et al., 2004; Ouzounis, 2005; e.g., Rossnes et al., 2005; Renner et al., 2007; Kondo & Omland, 2007; Zhang et al., 2008; Ekman et al., 2008; also see Ronquist, 2004). A number of analytical methods have been developed for reconstructing ancestral states of characters, which may be divided into four major categories: parsimony (MP) (Maddison & Maddison, 1992, 2001, 2007; Rossnes et al., 2005), maximum likelihood (ML) (Harvey & Pagel, 1991; Schluter et al., 1997; Pagel, 1999), Bayesian inference (Huelsenbeck et al., 2000, 2001; Ronquist 2004; Pagel et al., 2004), and stochastic character mapping (SCM) (Nielsen, 2002; Huelsenbeck et al., 2003). Until approximately five years ago, the easy use of the parsimony method implemented in MacClade 3.0 and 4.03 (Maddison & Maddison, 1992, 2001) had made it the most useful approach for tracking evolutionary history of morphological characters (see Huelsenbeck et al., 2003; Ronquist, 2004). The recent development of other methods and their automation in computer packages provided alternative methods that are more robust and realistic for users to study character evolution and overcome drawbacks of the parsimony method [e.g., ML and stochastic character mapping in MESQUITE 2.01 (Maddison & Maddison, 2007), the ML method in BAYESTRAITS 1.0 (Pagel & Meade, 2006), the Bayesian methods in BAYESTRAITS 1.0 (Pagel & Meade, 2006) and SIMMAP (Bollback, 2006)]. Both ML and SCM in MESQUITE 2.01 apply stochastic models of character state change and can explicitly accommodate uncertainty in ancestral states. The

Received: 17 April 2008 Accepted: 4 May 2008
* Author for correspondence. E-mail: jenny_xiang@ncsu.edu;
Tel.: 919-515-2728; Fax: 919-515-3436.

Bayesian approach in BAYESTRAITS and SIMMAP estimates the instantaneous rates of character change using maximum likelihood and accommodates phylogenetic uncertainty by evaluating the ancestral character state on trees sampled from the posterior distribution. In previous practice, most authors applied one or the other method in their studies, and few used multiple approaches to explore the differences of results derived from different methods. Recently, Pedersen et al. (2007) employed maximum parsimony in MacClade 4.03 (Maddison & Maddison, 2001) and Bayesian construction with BAYESTRAITS 1.0 (Pagel & Meade, 2006) in the moss family Bryaceae and found largely congruent results. In contrast, Ekman et al. (2008) found striking differences among parsimony, maximum likelihood, stochastic character mapping, and Bayesian reconstructions in their study of the fungi group *Ascus* (Lecanorales). It is unclear to what extent the congruence or incongruence is affected by phylogenetic uncertainty and the nature of the characters under consideration. Explicit knowledge of how reconstructions change under different methods (and how this depends on the type of character) would be useful in guiding the choice of methods for a study. It is also particularly important to re-evaluate our previous understanding of character evolution based on a single method, especially using parsimony alone. Inconsistent results from different methods would suggest that many previous studies on character evolution based on a single method may need to be reevaluated. This is especially relevant to studies published before the newer methods (e.g., ML, Bayesian, and SCM) became available. In the present study, we used BAYESTRAITS v 1.0 (Pagel & Meade, 2006) to reconstruct evolutionary histories of six key discrete morphological traits in Cornaceae and compared the results with those from MP, ML, and SCM implemented in MESQUITE 2.01(Maddison & Maddison, 2007) to gain more insight into the sensitivity of "character mapping" to the choice of methods.

Inferring the biogeographic histories (e.g., origin, persistence, dispersal, and extinction) of lineages is analogous to reconstruction of character evolutionary histories in many ways (but see below) and is fundamental to understanding the origin and evolution of the modern distribution of biodiversity. However, comparing to the study of character histories, there are fewer quantitative statistical methods available for historical biogeography. The present practice of tracking biogeographic histories on phylogeny is, in this regard, facing more challenges than studying character evolution. During the past decade, the parsimony-based dispersal-vicariance analysis (DIVA; Ronquist, 1996, 1997) has been the most widely used approach in phylogenetic biogeography (e.g., Xiang & Soltis, 2001; Donoghue & Smith, 2004; Sanmartín & Ronquist, 2004; Xiang et al., 2006; Hines, 2008) due to many advantages of the method, e.g., permitting inference of explicit vicariant and dispersal events, their relative timing, and dispersal directions, as well as the availability of fast and user-friendly software (DIVA 1.1, Ronquist, 1997). However, the method inherits the intrinsic principle of parsimony, and therefore it may underestimate character transformation events (i.e., dispersal). Although some studies applied the character model of likelihood methods for biogeographic studies to estimate the likelihood of ancestral areas (Nepokroeff et al., 2003), the models of character evolution are not appropriate for geographic ranges (see Ree & Smith, 2008). As a response to the need of more and better analytical biogeographic methods, Ree and colleagues recently developed a likelihood-based approach for inference of ancestral geographic areas (Ree et al., 2005; Ree & Smith, 2008). The newly developed method represents a significant advance in biogeographic methodology by implementing a maximum likelihood statistical model that includes information from biological and abiotic factors in calculating the likelihood of biogeographic pathways for a given phylogenetic tree and the distributions of taxa. For example, the rate of dispersal and local extinction, lineage surviving time, and the probabilities of dispersal between geographic ranges at different geological times (Ree et al., 2005; Moore et al., 2006) can all be used in the estimation. Their methods model geographic range evolution by stochastic dispersal, local extinction, and speciation in a set of discrete areas in continuous time. Unlike the character models, a single taxon is allowed to have more than one character state (e.g., a wide distribution in two or more areas), as a result of direct dispersal (Ree et al., 2005). A recent version of the method of Ree and Smith (2008) uses instantaneous rates of dispersal and local extinction to significantly improve the computation time for parameter estimation previously done using a simulation approach in Ree et al. (2005). This makes it feasible for ancestral area reconstruction using ML optimization analogous to character mapping. In the present study, we apply the methods of Ree et al. (2005) and Ree & Smith (2008) to Cornaceae and compare the results from those derived from DIVA. Knowledge of congruences and discordance between the results from DIVA and the

ML analysis is particularly useful for evaluating our previous understanding of lineage biogeographic histories and synthesis on global biogeography based on DIVA (Donoghue & Smith, 2004; Sanmartín et al., 2001; Sanmartín & Ronquist, 2004).

The dogwood family Cornaceae is the key member of Cornales, a lineage occupying a special position on the tree of life of angiosperms for being basal in the Asteridae clade (APG, 2003; Soltis et al., 2005). The family has been circumscribed differently among authors (Eyde, 1988; Takhtajan, 1987; see Xiang et al., 1993; APG, 2003; Fan & Xiang, 2003; Soltis et al., 2005). Here we take the narrow concept proposed by Takhtajan (1987) that Cornaceae contains only *Cornus* L. *s. l.*, based on the following reasons. First, the closest genus to *Cornus* is *Alangium* Craib, based on results of molecular phylogenetic studies (Chase et al., 1993; Xiang et al., 1998, 2002; Fan & Xiang, 2003). Second, *Alangium* has long been recognized as a family, Alangiaceae DC. (de Candolle, 1828). Third, the name of Alangiaceae has been proposed to be conserved (Bullock, 1959; APG, 2003). Cornaceae s.s. consists of approximately 50–55 species that are morphologically diverse. Species of the genus exhibit striking variation in inflorescence architecture and display, fruit type, structure, and color, among other attributes (see Xiang et al., 2006). As a consequence of the morphological heterogeneity, the species have been grouped in various ways, in multiple genera, subgenera, or sections during the past century (see Eyde, 1988; Murrell, 1993, 1996; Xiang et al., 1996; Fan & Xiang, 2001; Xiang & Boufford, 2005). Recent molecular phylogenetic studies reconstructed a well resolved, robust species phylogeny for *Cornus* and resolved the species into four major clades (Xiang et al., 2006; Xiang et al., in review, available upon request). These are: (1) the blue-or-white fruited group (BW); (2) the cornelian cherries (CC), (3) the big-bracted group (BB), and (4) the dwarf dogwoods (DW). Relationships within and among each group are also well resolved, and mostly strongly supported (Xiang et al., 2006; Xiang et al., in review). This provides a solid basis to track the evolutionary histories of morphological characters in the genus and help to understand the tempo and pattern of key morphological changes associated with the dogwoods radiation, such as changes in inflorescence and fruit architecture and morphology that must have played important roles in the genus diversification.

With its highest species diversity in eastern Asia and eastern North America, the dogwood genus also exhibits several historically important biogeographic patterns involving intercontinental disjunctions in eastern Asia, North America, Europe, tropical eastern Africa, and South America (Fig. 1; Thorne, 1972; Boufford & Spongberg, 1983; Wu, 1983; Milne & Abbott, 2002; Donoghue & Smith, 2004; also see Xiang et al., 2006; Xiang et al., in review). With an excellent fossil record that includes fruit stones that can be reliably identified into subgroups (Eyde, 1988; Manchester, 1994; Crane et al., 1990; Manchester et al., in press) and a well-resolved, robust phylogeny, the genus is ideal for a comprehensive biogeographic analysis and provides a good model for exploring the impacts of different analytical methods and fossils in phylogenetic biogeography. Xiang et al. (2006) proposed, based on results of DIVA and fossils, that the genus evolved and early diversified into the four major clades in Europe followed by mutiple intercontinental migrations mostly across the North Atlantic Land Bridge. The migration routes in three of the four disjunct lineages however remained to be evaluated with divergence time information not available in Xiang et al. (2006). In this study, the biogeographic history of *Cornus* is reconstructed using the likelihood-based method of Ree et al. (2005) and Ree & Smith (2008). The ancestral areas at deep nodes are compared to those inferred from DIVA (Ronquist, 1997). The impacts of fossils and different constraints on maximum areas are also explored for each method. Furthermore, the results from DIVA and ML are also compared with an analysis treating distributions as a single, multistate character in BAYESTRAITS.

The major goal of this study is to understand the morphological evolution of *Cornus* in space and time using the most recent and best method available. A second goal of the study is to gain new insights on the influences of method choice in reconstructing the ancestral state of characters and ancestral areas of biogeographic distributions.

1 Material and methods

1.1 Estimation of ancestral character states

Sampling of characters and phylogeny: Six characters mostly from inflorescence architecture and fruit morphology were chosen for the analyses (Table 1), including (1) chromosome numbers, (2) inflorescence bracts, (3) inflorescence type, (4) development of inflorescence bud, (5) fruit type, and (6) fruit color. Five of the six characters are multistate and one is binary (character 5). For all of these characters, all taxa have a single state except for character 6, which is polymorphic within one terminal taxon (represented

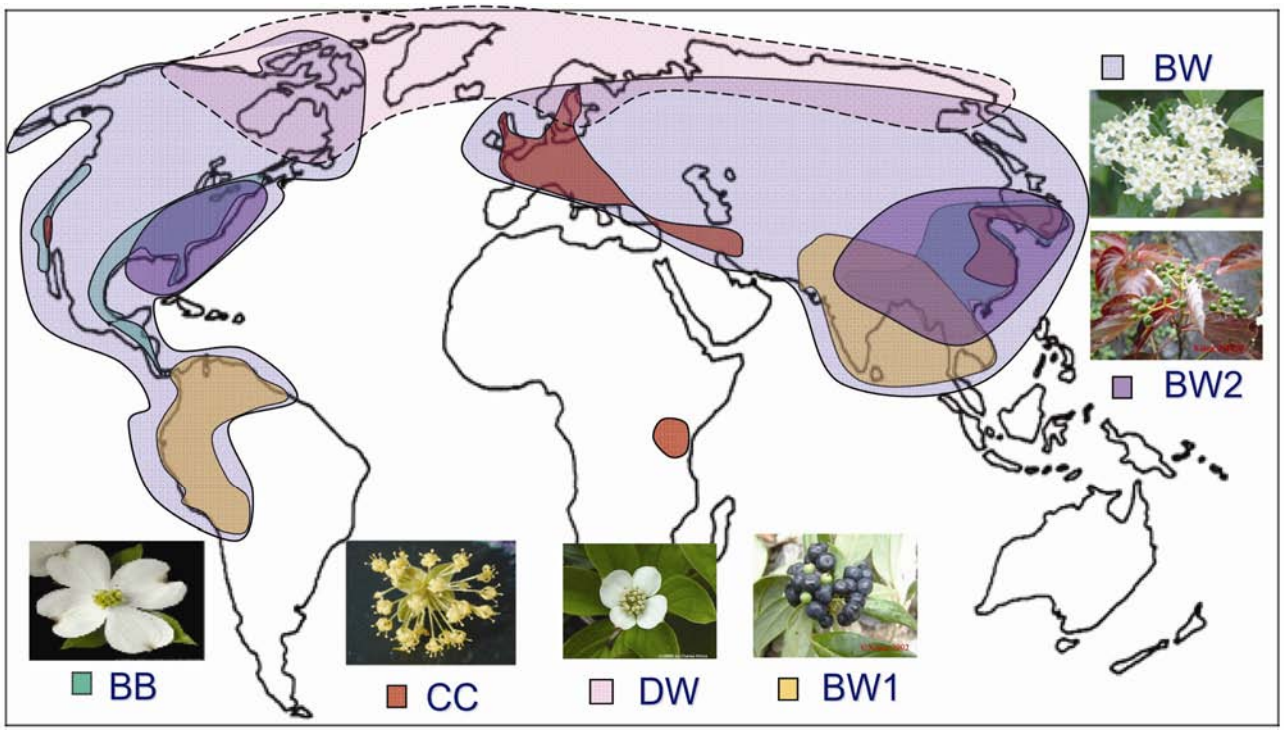

Fig. 1. Approximate geographic distributions of *Cornus* major clades. BB, big-bracted dogwoods; CC, cornelian cherries; DW, dwarf dogwoods; BW, blue- or white-fruited dogwoods; BW1, *C. oblonga* and *C. peruviana*; BW2, *C. alternifolia* and *C. controversa*.

by *C. racemosa* Lam.), of the blue-fruited or white-fruited clade (BW) and the outgroups. The six characters represent a range of character conditions that are commonly encountered by investigators, e.g., characters without polymorphism and homoplasy (character 5), characters with missing data and homoplasy (character 1), characters with homoplasy, but no missing data (characters 2, 3, 4), and characters with polymorphism and homoplasy (character 6). The well-resolved molecular phylogenetic trees of *Cornus* derived from six gene regions (*rbc*L, *mat*K, *ndh*F, *atp*B, ITS, and 26S rDNA; Xiang et al., in review) were used as the basis for reconstruction of ancestral states. These trees contain 22 *Cornus* species and two outgroup taxa, *Alangium* and *Diplopanax* Hand.-Mazz., sister and close relative of *Cornus*, respectively (Xiang et al., 1998, 2002; Fan & Xiang, 2003). The 22 *Cornus* species represent all the subgenera recognized so far by different authors at different times and all the species from each subgenus (see Xiang et al., 2006), with the exception for subgen. *Syncarpea* from the Big-bracted group (BB) and subgen. *Kraniopsis* from the BW group. The former is missing *C. multinervosa*, closely related to *C. kousa* Hance, and *C. elliptica* (Chun) Q. Y. Xiang & Boufford, closely related to *C. capitata* Wall. Subgenus *Kraniopsis* consists of ~30 species and was represented by three species, each representing one of the three subclades indentified by the ITS-*mat*K species phylogeny that included all of the species (Xiang et al., 2006). Species within each of the subclades are morphologically highly similar (not variable for the morphological characters under study), thus missing these species will not affect tracking the evolutionary histories of these characters in the genus.

1.2 Optimization of ancestral character states

Among the various methods, the Bayesian approach implemented in BAYESTRAITS 1.0 (Pagel & Meade, 2006) stands out by not only taking into account both branch length and phylogenetic uncertainty, but also allowing the user to explore a variety of models for character transition, defining nodes of interest, and a reversible-jump Markov Chain Monte Carlo (Pagel & Meade, 2006). We used 1,000 trees with branch lengths randomly selected from the tree pool after burn-in from the Bayesian analysis of the six-gene data of *Cornus* (Xiang et al., in review). The 50% majority rule tree of these 1000 trees is identical to that from all the Bayesian trees after burn-in, suggesting that these 1000 trees are likely representative

Table 1 Character state of six morphological characters in *Cornus* (character 1–6) and distribution of species (character 7)

Characters	1	2	3	4	5	6	7
C. nuttallii Audubon	0	2	2	1	0	4	3
C. florida L.	0	2	2	2	0	4	2
C. kousa Hance	0	2	2	3	1	4	1
C. disciflora Moc. & Sesse ex DC.	0	1	2	2	0	2	34
C. capitata Wall.	0	2	2	1	1	4	1
C. oligophlebia Merr.	0	0	0	0	0	1	1235
C. hongkongensis Hemsley	0	2	2	1	1	4	1
C. alternifolia L. f.	1	0	0	3	0	1	2
C. controversa Hemsley	1	0	0	3	0	1	1
C. peruviana J. F. Macbr.	0	0	0	0	0	1	4
C. walteri Wangerin	0	0	0	0	0	1	1235
C. racemosa Lam.	0	0	0	0	0	01	23
C. oblonga Wall.	0	0	0	0	0	1	1
C. chinensis Wangerin	2	1	1	2	0	2	1
C. sessilis Torr. ex Durand	1	1	1	3	0	2	3
C. eydeana QY Xiang & YM Shui	2	1	1	2	0	3	1
C. mas L.	2	1	1	2	0	4	5
C. officinalis Seib. & Zucc.	2	1	1	2	0	4	1
C. volkensii Harms	?	1	1	2	0	2	7
C. suecica L.	0	2	0	0	0	4	1235
C. canadensis L. f.	0	2	0	0	0	4	123
C. unalaschkensis Ledeb.	0	2	0	0	0	4	3
Alangium	0	0	0	0	0	01234	1 or 123567
Diplopanax	0	0	0	0	0	124	12345

Character: character state, score:

1. Chromosome number: n=11, 0; n=10, 1; n=9, 2.
2. Bracts: rudimentary (minute, early deciduous), distal to inflorescence branches, 0; 4, nonpetaloid at the base of inflorescence, 1; 4 or 6, petaloid at the base of inflorescence, 2.
3. Inflorescence type: branched compound cymes, 0; umbel (cymose), 1; glomerule (capitate cyme), 2.
4. Developmental state of winter inflorescence bud: floral buds undeveloped, 0; floral buds developed and unprotected, 1; floral buds developed and protected by inflorescence bracts, 2; floral buds developed and protected by scales of a mixed bud, 3.
5. Fruit type: simple, 0; compound, 1.
6. Fruit color: white, 0, blue or black, 1; red then black, 2; dark purple 3; red, 4.
7. Distributions: eastern Asia, 1; eastern North America, 2; western North America, 3; Central and South America, 4; Europe, 5; Australia, 6; Africa, 7.

Cornus racemosa represents a subclade in areas 2 and 3, and *C. oligophlebia* and *C. walteri* represent a subclade occurring in areas of 1, 2, 3, and 5 (Xiang et al., 2006). These areas were coded for these three species accordingly in the analysis. *Diplopanax* was coded as 1, 2, 3, 4, and 5 to incorporate distributions of taxa in the clade it represents (nysssoids, mastixioids, *Grubbia* and *Curtisia*) in the Cornales.

of the entire tree pool. The model of "multistates", MCMC mode, an exponential prior seeded from a uniform on the interval 0 to 10 (rjhp exp 0.010), a burn-in of 10,000 and sampling of the chain every 300 generations were applied. These parameters were recommended by the program manual. Two separate analyses with a rate parameter of 2 and 5, respectively were performed. Both analyses were commanded to estimate the ancestral state for each character at each node of the phylogeny. The nodes were specified using the command "addMRCA". A total of 10,000,000 iterations were run for each analysis.

The evolutionary history of each of the six characters was also traced over the 50% majority rule tree from Bayesian analysis of Xiang et al. (in review) using parsimony, maximum likelihood (ML), and Stochastic character mapping (SCM) available in MESQUITE 2.01 (Maddison & Maddison, 2007). For ML and SCM analyses, both the "best" tree topology and a chronogram (the "best" tree with branch length in time) were used for a comparison. The chronogram was derived from the divergence time analysis using the program Multidivtime (Thorne et al., 1998; Kishino et al., 2001) from Xiang et al. (in review). In parsimony reconstruction, the character states were treated as "unordered (i.e., allow free transformation of a character state to any other states). The ML reconstructions were conducted using the MK1 model of evolution (Schluter et al., 1997; Pagel, 1999). The Mk1 (Markov k-state 1 parameter model) is a k-state generalization of the Jukes-Cantor model, corresponding to Lewis's (2001) Mk model, which gives equal probability (or rate) for changes between any two character states. The rate of change can be estimated for each individual character based on the data and branch length by MESQUITE. The stochastic character mapping (Nielsen, 2002; Huelsenbeck et al., 2003) in MESQUITE uses continuous-time Markov models to derive posterior distributions of ancestral state. It simulates realizations of precise histories of character evolution in a way consistent with Mk1 model and ancestral states (Maddison & Maddison, 2007). Reconstructions of ancestral states for character 6, which has polymorphism for some taxa, were not allowed for ML and SCM on MESQUITE 2.01. Therefore, only analyses with MP and BAYESTRAIT were conducted for this character.

1.3 Optimization of ancestral distributions

The estimation of ancestral distributions was performed using the same phylogenetic trees as for character ancestral state reconstructions. Three methods, the parsimony-based dispersal-vicariance method (DIVA) (Ronquist, 1997), the ML-based methods of Ree et al. (2005) and Ree & Smith (2008), and the Bayesian method using a character model (Pagel et al., 2004) were employed. Seven geographic areas were

defined, eastern Asia (EAS), eastern North America (ENA), western North America (WNA), Central and South America (CSAM), Europe (EUR), Australia (AUS), and Africa (AFR) to cover all the endemic distributional areas of *Cornus* and its outgroups. The method of DIVA finds the best biogeographic pathways given the tree topology and distributions of taxa by minimizing dispersal and extinction events. In the analysis, each taxon was scored for presence (1) or absence (0) for each of these areas. The distributions of *Cornus racemosa*, *C. oligophlebia* Merr., and *C. walteri* Wangerin, each representing a subclade in the BW group, were scored according to the occurrence of all species in the represented subclade. To explore influences of maximum number of areas constrained for each node, analyses with constraints of maximum areas of 4 and 2, respectively, were performed for comparison. Furthermore, analyses using two different models of area coding for the outgroup *Alangium* were also conducted under these two different maximum area constraints. One model coded *Alangium* for all areas of its occurrence including fossils (all of the seven areas except Central and South America). The other model coded *Alangium* for only the root area of the genus (eastern Asia) that was inferred from biogeographic analysis of the genus based on the phylogeny using DIVA (Feng et al., in review). For the other outgroup *Diplopanax*, the distribution was scored for occurrence of all taxa of the clade it represents in the Cornales (nyssoids, mastixioids, *Grubbia* P. Bergius and *Curtisia* (Burm.) G. A. Sm.; Fan & Xiang, 2003; Xiang et al., 2002; Xiang et al., 2007). Optimizations were conducted using the same tree topology as for character ancestral state reconstructions described above.

ML inferences of geographic range evolution were conducted for the same distribution matrix under the constraints of maximum areas 4 and 2, respectively, on a chronogram from the Bayesian analysis of six gene regions, the same chronogram as the one for ancestral character state reconstruction. The program LAGRANGE (Ree & Smith, 2008) was employed to run the analysis with a simple model of one rate of dispersal and extinction constant over time and among lineages. This program not only finds the most likely ancestral areas at a node and the split of the areas in the two descendant lineages, it also calculates the probabilities of these most-likely areas (Ree & Smith, 2008) at each node, a feature lacking in the program AReA that was based on Ree et al. (2005). The values of rate of dispersal and extinction implemented in the analysis with LAGRANGE were estimated based on the tree using a maximum likelihood method by the program. To reduce computation time, the following areas were disallowed at all nodes due to their wide disjunction, which requires prior extinction in their intervening areas or long distance dispersal (EAS-CSAM, EAS-AFR, ENA-AUS, ENA-AFR, WNA-AUS, WNA-AFR, CSAM-EUR, EUR-AUS, EAS-ENA-FR, EAS-WNA-FR, ENA-WNA-AUS, ENA-WNA-FR, AM-EUR-AUS, EAS-ENA-WNA-FR, ENA-NA-US-AFR). To explore the impacts of excluding these areas as a prior and influences of rates of extinction and dispersal, changes of physical connection between geographic areas through time, analyses were ran using AReA v. 2.1 (Smith, 2006; based on Ree et al., 2005) without excluding areas as a prior using a range of rates of V (dispersal) and L (extinction) including those estimated from LAGRANGE as well as the following: V=L=0.005; V= 0.09, L=0.01; V=0.01, L=0.09; V=0.09, L=0.01; V= 0.02, L=0.18; V=0.18, L=0.02; V=0.0047, L=0.0033. The analyses with AReA and LAGRANGE were both conducted for the chronogram with and without fossils. For analysis with fossils, the branch length in time of the fossil lineages in the chronogram was arbitrarily estimated as the length of time between the earliest and latest appearances of the fossil taxa. If there was only one report for a fossil taxon, the branch length in time of that fossil was arbitrarily estimated as the time range of the geological period of its occurrence. For example, the fossil species of BB was found in the mid Oligocene bed. The branch length in time of the BB fossil lineage was arbitrarily given as 5.1 million years (my), approximately the length of time of the mid Oligocene. The fossil lineages were connected to the stem lineage of the represented crown group at the time point corresponding to the age of the fossils. This analysis was done mainly to see the impact of new geographic areas represented by fossils, but not occupied by extant species of a lineage, on optimization of ancestral areas of the perspective lineages. Information of age, occurrence, and phylogenetic placements of fossils is from Xiang et al. (2006). The analyses were run for 1,000,000 replicates as suggested in the manual. For analyses with AReA, the Mesophytic model for probabilities of dispersal between geographic areas at different geological time periods was implemented (Smith, 2006). The different models have not been coded into LAGRANGE (Ree, pers. comm.). Optimization of ancestral areas using the Bayesian method in BAYESTRAITS v.1.0 was done as described above for character mapping, but using "rjhp exp 0.0 30 and "ratedev 15".

1.4 Estimation of divergence time using BEAST

Phylogenetic dating using five-gene data (*rbc*L, *mat*K, *ndh*F, *atp*B, and 26S rDNA) was conducted using r8s (Sanderson, 2002) and Multidivtime (Thorne et al., 1998; Kishino et al., 2001) in a separate study (Xiang et al., in review). These methods relied on only a single phylogeny. Here we estimated the divergence time using the same data set again using BEAST 1.4.3 (Drummond & Rambaut, 2006) for a comparison. BEAST applies Bayesian methods and searches for the optimal phylogeny and estimates divergence time simultaneously. The divergence time estimated from BEAST was a summary of estimations based on all the optimal Bayesian trees found in the analysis, thus taking the phylogenetic uncertainty into account. The analyses with BEAST were performed using the model suggested by Modeltest 3.7 (Posada & Crandall, 1998) with an uncorrelated exponential relaxed molecular clock model for 30 million generations. Multiple runs were done to determine the number of generations required to allow all node time estimates to reach a stationary distribution. Four nodes were constrained by fossils, as done for Multidivtime (Xiang et al., 2005; Xiang et al., in review).

2 Results

2.1 Optimization of ancestral states

Results from ancestral character state optimizations from MP, ML, and SCM without temporal constraints are presented in Figs. 2 and 3 for all nodes. Results from BAYESTRAITS are provided in Table 2 as well as Figs. 2 and 3. These results showed that for binary characters without homoplasy, e.g., character 5, the results at all nodes are identical and certain among MP, ML, SCM, and Bayesian methods (Fig. 2). For characters with homoplasy, there are substantial differences in the relative certainty of optimal states at some nodes (e.g., nodes m & l for character 1, node r for character 2, nodes g & r for character 3, nodes f, g, m, & r for character 4 among the four methods, as well as nodes l, p, r, & s for character 6 between MP and Bayesian inference). There are only a few nodes among all characters showing striking difference in the optimal states inferred from different methods. These are nodes g and r for character 3 (inflorescence type) and nodes a, b, c, f, g, r for character 4 (winter condition of inflorescence bud) (Fig. 2B). For character 3, nodes g and r were convincingly resolved as state 0 by MP, ML, and SCM, but BAYESTRAITS resulted in extremely low posterior probabilities for state 0 at these two nodes and suggested the most probable state to be 2 for node g (pp=0.38) and 1 for node r (pp=0.5), which implies some degree of uncertainty. For character 4, the MP method did not resolve the ancestral states at g and r nodes (equivocal). However, the ML and Stochastic character mapping analyses show that the ancestral states of these two nodes are state 0 (ML, most likely; SCM, 1.0 pp), while BAYESTRAITS shows that the ancestral state at this node is most probable state 2 (Fig. 2B; pp: 0.65). For nodes a & b of character 4, ML and SCM suggested alternative states (2 vs. 1), while BAYESTRAITS suggested the two states are nearly equal probable, similar to MP (Fig. 2B). For nodes c and f of Character 4, the ancestral states are equivocal in MP and ML analyses, but SCM and BAYESTRAITS resolved it as state 1 (Fig. 2B). In general, the optimal states at the lower nodes of the phylogeny estimated from BAYESTRAITS using 1,000 trees are less certain than estimations from MP, ML, and SCM based on a single, "best" phylogeny for all characters analyzed.

When temporal information was taken into account by using a chronogram for estimation with MP, ML, and SCM, the results showed a few differences for Characters 1, 2 and 4 under the same method. For Character 1, node m (Fig. 2A) was reconstructed as state 1 (green dots) by SCM and as equivocal by ML based on the tree topology; whereas in the reconstructions based on a chronogram, this node was inferred as state *2* (black dots) by both SCM and ML methods (Fig. 3). For Character 2, node r (Fig. 2A) was inferred to be equivocal by ML method in the topology-based analysis, but in the chronogram-based analysis, this node changed to be most likely state 0 (open circle) (Fig. 3). For character 4, the topology-based analysis with the ML method optimized nodes a and b to be mostly likely state 2 (green dots) (Fig. 2), but in the chronogram-based ML analysis, state 2 and state 1 (blue dots) at this node were equally probable (Fig. 3). In the topology-bsed SCM analysis, nodes a and b were both certain to be state 1 (blue dots) (Fig. 2) while in the chronogram-based SCM analysis, both nodes were reconstructed as state 2 (green dots) (Fig. 3). In general, discrepancies between ML and SCM in topology-based analyses disappeared in the chronogram-based analyses (compare node m for character 1 and 4 and node r for character 2 in Figs. 2 & 3).

2.2 Ancestral distributions

2.2.1 Between coding models for *Alangium* The results from analyses using AReA and DIVA with *Alangium* coded for its root (indicated by a "*" in Table 3) or for all areas of its occurrence (column

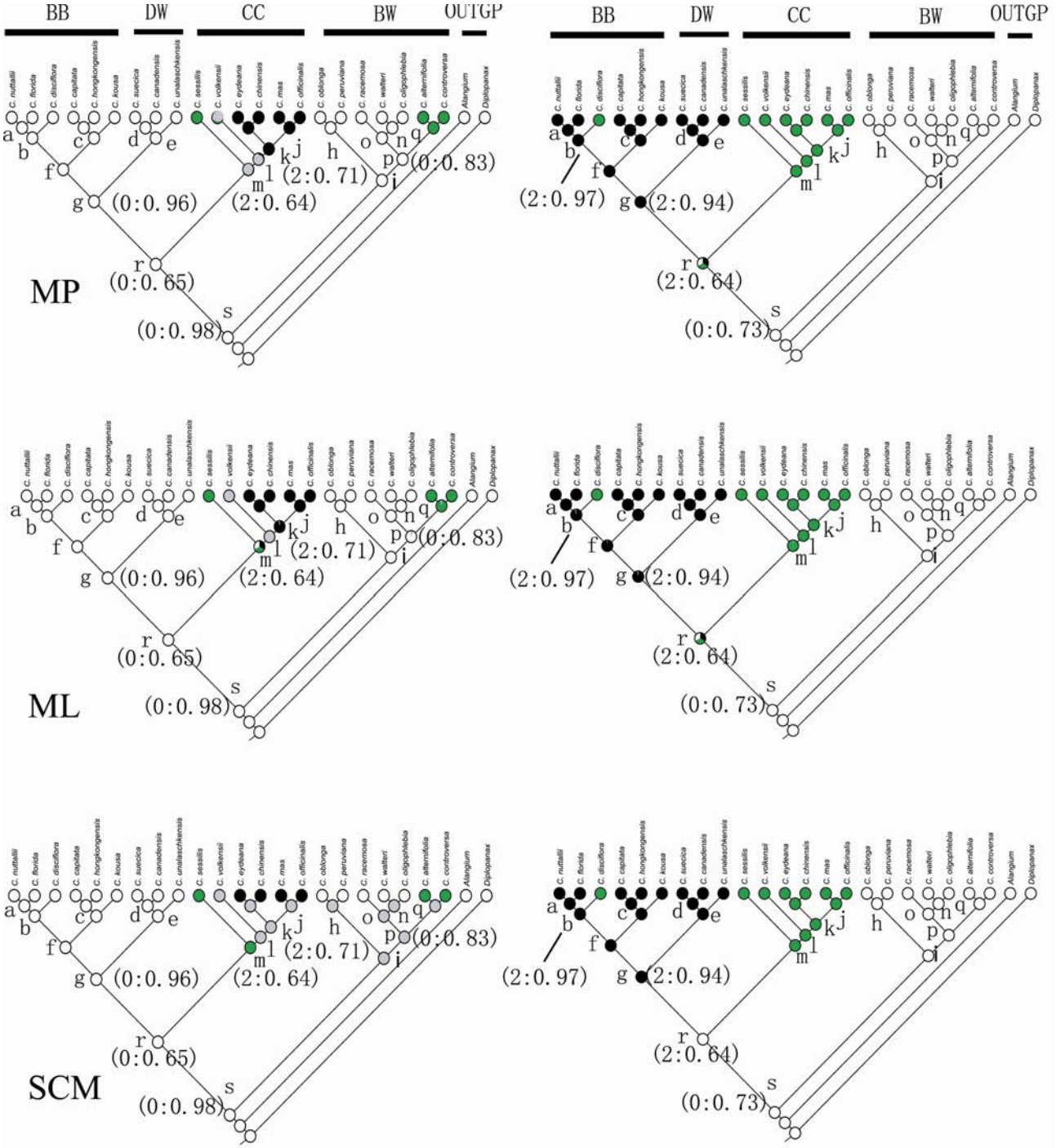
BB
DW
CC
BW
OUTGP
MP
ML
SCM
(0:0.96)
(2:0.64)
(2:0.71)
(0:0.83)
(0:0.65)
(0:0.98)
(2:0.97)
(2:0.94)
(2:0.64)
(0:0.73)
Character 1
Chromosome number:
x = 11 (0)
x = 10 (1)
x = 9 (2)
equivocal
Character 2
Inflorescence bracts:
Rudimentary, triangular, early disciduous (0)
Small, bowl-shaped scales, 4, at the base (1)
Large petaloid, 4 or 6, at the base (2)

Fig. 2A

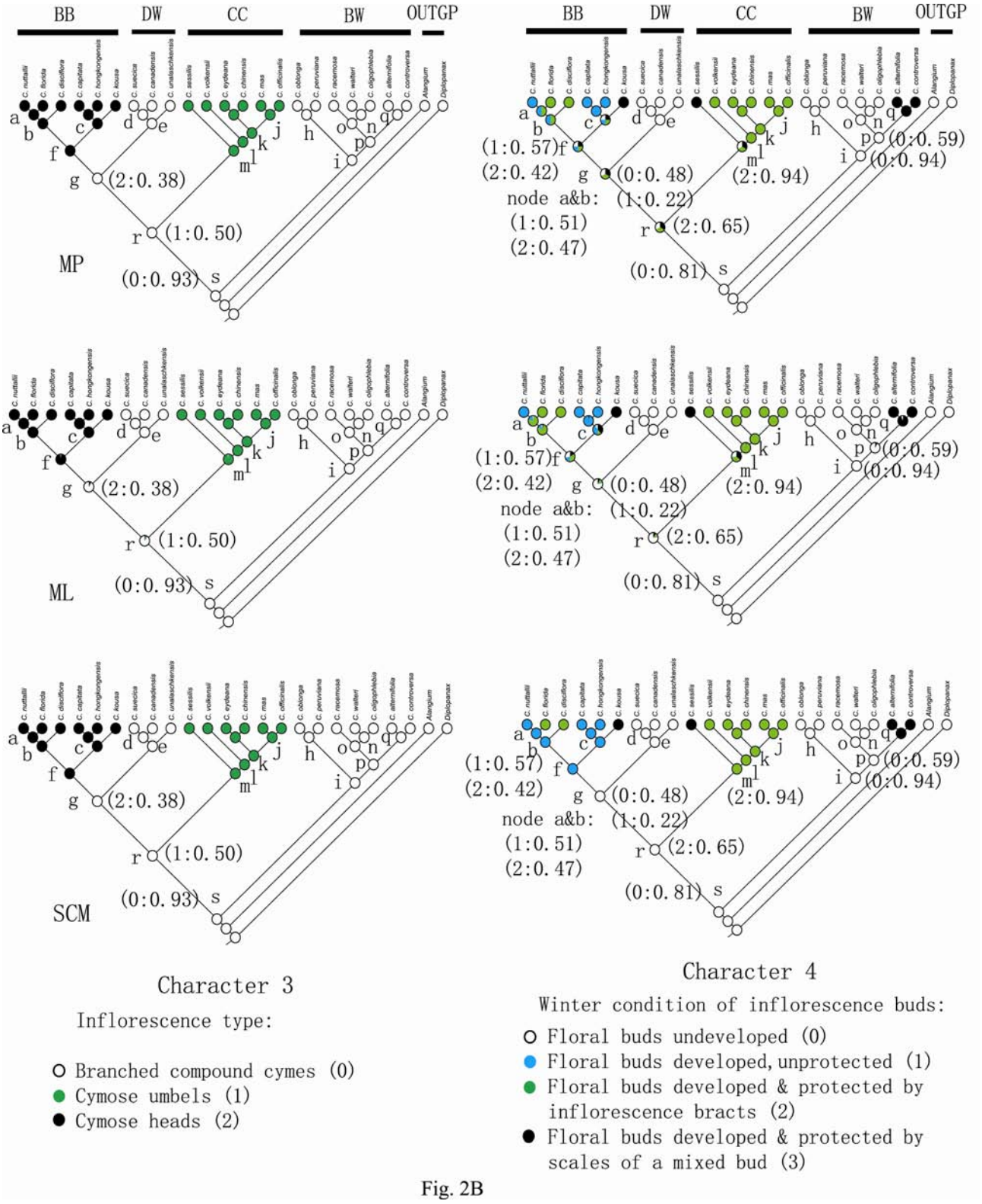
BB
DW
CC
BW
OUTGP
MP
ML
SCM
(2:0.38)
(1:0.50)
(0:0.93)
(1:0.57)
(2:0.42)
(0:0.48)
(2:0.94)
node a&b:
(1:0.51)
(2:0.47)
(1:0.22)
(2:0.65)
(0:0.81)
(0:0.59)
(0:0.94)
Character 3
Inflorescence type:
Branched compound cymes (0)
Cymose umbels (1)
Cymose heads (2)
Character 4
Winter condition of inflorescence buds:
Floral buds undeveloped (0)
Floral buds developed, unprotected (1)
Floral buds developed & protected by inflorescence bracts (2)
Floral buds developed & protected by scales of a mixed bud (3)

Fig. 2B

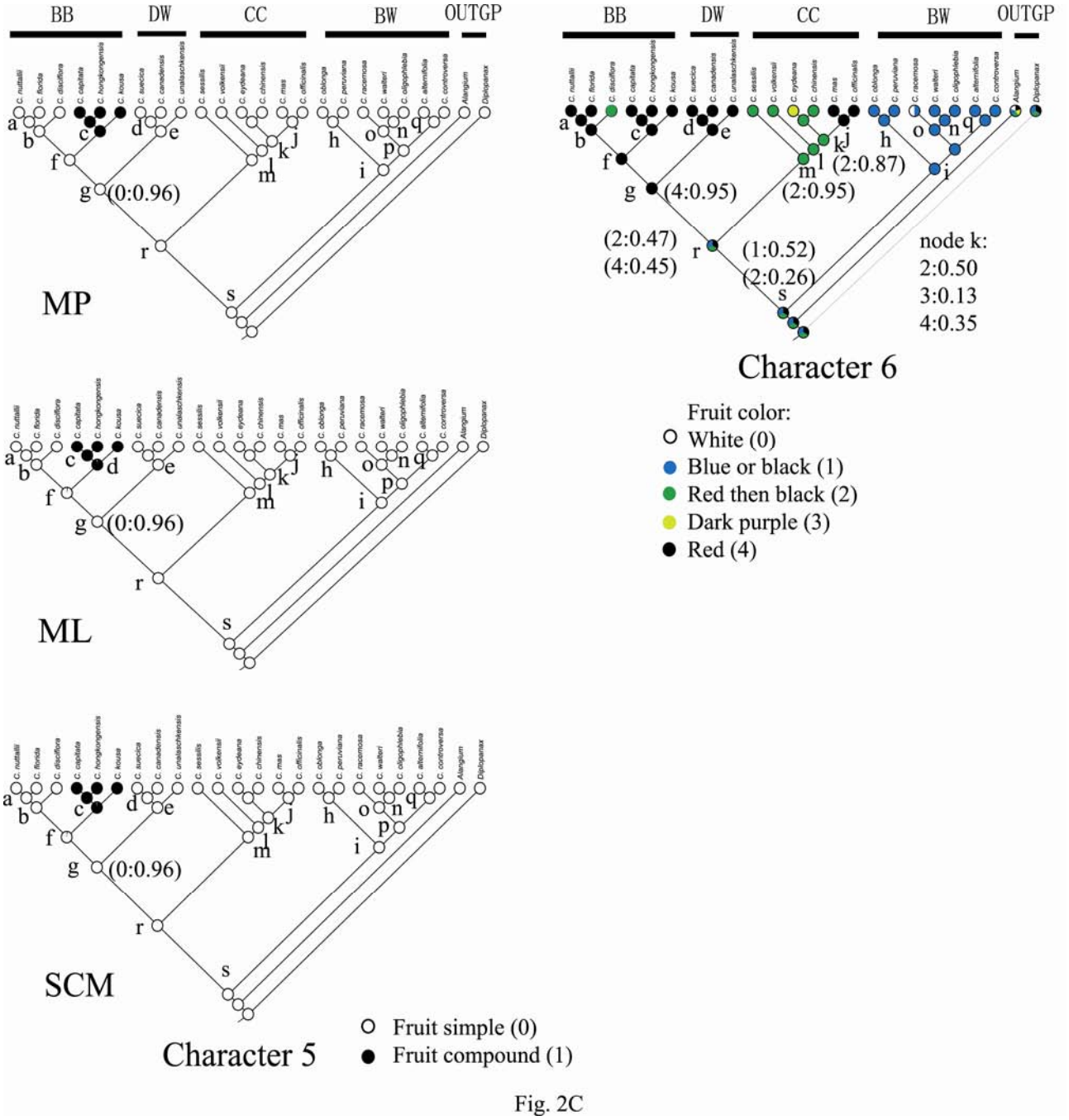

Fig. 2C

Fig. 2. Evolutionary histories of six key morphological characters on inflorescence and fruit morphology reconstructed using four different methods (MP, maximum parsimony; ML, maximum likelihood; SCM, stochastic character mapping in MESQUITE 2.01, and BAYESTRAITS 1.0). Results for the same character from different methods are presented in the same column. Posterior probabilities of <100% from BAYESTRAITS are presented as numbers. Character state and posterior probability at each node is presented in Table 2. Characters are referred to in Table 1. Species from left to right on each tree: *Cornus nuttallii, C. florida, C. disciflora, C. capitata, C. hongkongensis, C. kousa, C. suecica, C. canadensis, C. unalaschkensis, C. sessilis, C. volkensii, C. eydeana, C. chinensis, C. mas, C. officinalis, C. oblonga, C. peruviana, C. racemosa, C. walteri, C. oligophlebia, C. alternifolia, C. controversa, Alangium*, and *Diplopanax*. **A,** For characters 1 & 2. **B,** For characters 3 & 4. **C,** For characters 5 & 6.

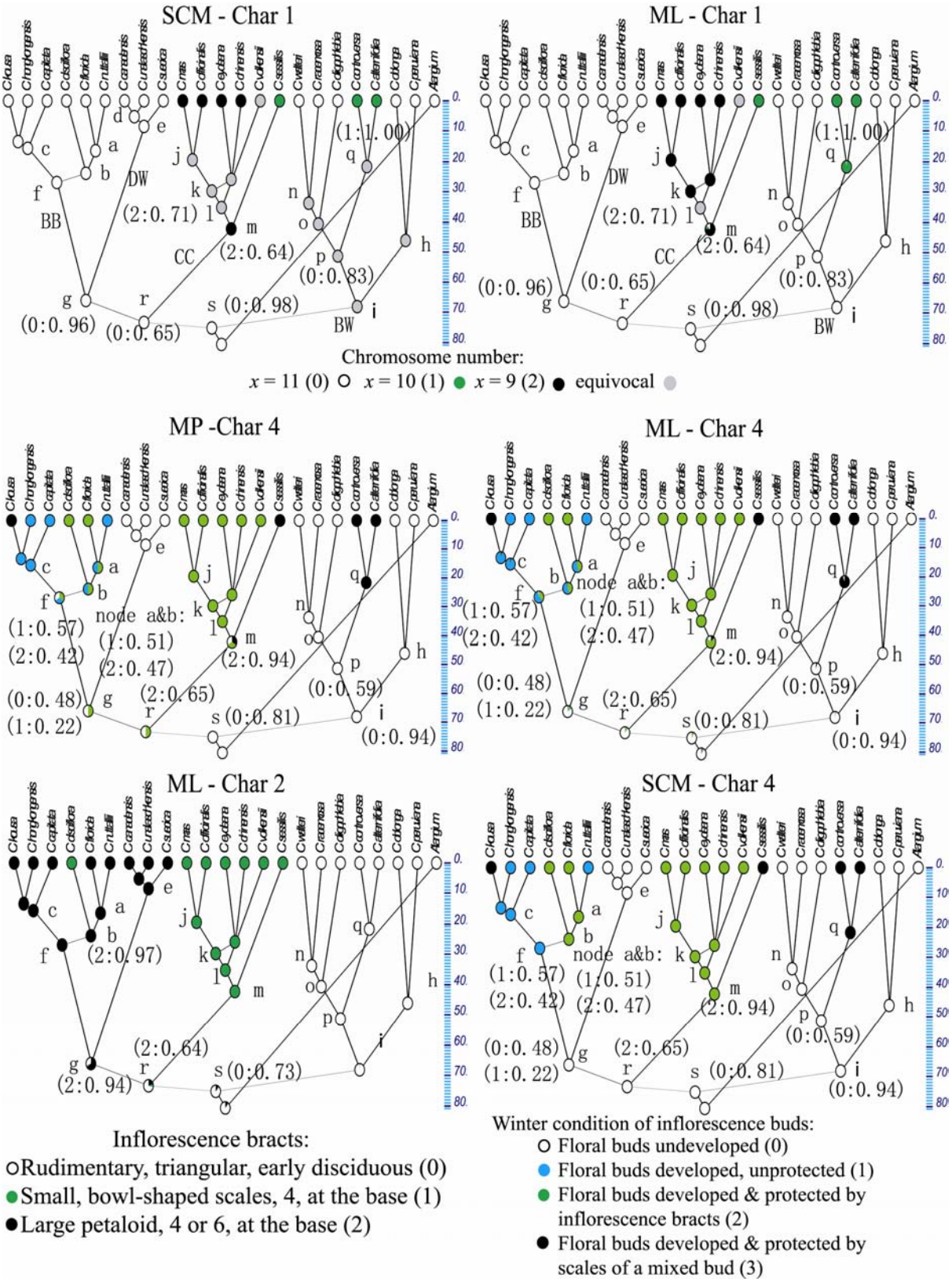

Fig. 3. Evolutionary histories of characters inferred from the chronogram of the "best" tree. Only those different from inferences based on the "best" tree topology are presented.

Table 2 Results of reconstruction of ancestral states for six characters (explained in Table 1) from BAYESTRAITS 1.0 based on 1,000 Bayesian trees

Character node	1	2	3	4	5	6
a	1.00+0.00 (0)	1.00+0.00 (0)	1.00+0.00 (0)	0.50+0.27 (1) 0.47+0.27 (2)	1.00+0.00 (0)	1.00+0.00 (4)
b	1.00+0.00 (0)	0.97+0.10 (2)	1.00+0.00 (2)	0.52+0.35 (1) 0.48+0.35 (2)	0.99+0.02 (0)	1.00+0.00 (4)
c	1.00+0.00 (0)	1.00+0.00 (2)	1.00+0.00 (0)	0.96+0.09 (1)	1.00+0.00 (1)	1.00+0.00 (4)
d	1.00+0.00 (0)	1.00+0.00 (2)	1.00+0.00 (0)	1.00+0.00 (0)	1.00+0.00 (0)	1.00+0.00 (4)
e	1.00+0.00 (0)	1.00+0.00 (2)	1.00+0.00 (0)	1.00+0.00 (0)	1.00+0.00 (0)	1.00+0.00 (4)
f	1.00+0.00 (0)	1.00+0.00 (2)	1.00 +0.00 (2)	0.42+0.33 (2) 0.57+0.34 (1)	0.98+0.05 (0)	1.00+0.00 (4)
g	0.96+0.04 (0)	.094+0.07 (2)	0.38+0.19 (2)	0.22+0.48 (1) 0.48+0.19 (0)	0.96+0.04 (0)	0.95+0.05 (4)
h	0.99+0.01 (0)	0.99+0.01 (0)	0.99+0.01 (0)	0.99+0.01 (0)	0.99+0.01 (0)	0.99+0.01 (1)
i	0.99+0.02 (0)	1.00+0.00 (0)	1.00+0.00 (0)	0.94+0.12 (0)	1.00+0.00 (0)	0.99+0.01 (1)
j	1.00+0.00 (2)	1.00+0.00 (1)	1.00+0.00 (1)	1.00+0.00 (2)	1.00+0.00 (0)	0.99+0.01 (4)
k	1.00+0.00 (2)	1.00+0.00 (1)	1.00+0.00 (1)	1.00+0.00 (2)	1.00+0.00 (0)	0.50+ 0.01 (2) 0.35+0.29 (4) 0.13+0.18 (3)
l	0.70+0.26 (2)	1.00+0.00 (1)	1.00+0.00 (1)	0.99+0.03 (2)	1.00+0.00 (0)	0.87+0.20 (2)
m	0.64+0.30 (2)	1.00+0.00 (1)	1.00+0.00 (1)	0.94+0.10 (2)	1.00+0.00 (0)	0.95+0.14 (2)
n	1.00+0.00 (0)	1.00+0.00 (0)	1.00+0.00 (0)	1.00+0.00 (0)	1.00+0.00 (0)	0.99+0.01 (1)
o	1.00+0.00 (0)	1.00+0.00 (0)	1.00+0.00 (0)	1.00+0.00 (0)	1.00+0.00 (0)	0.99+0.01 (1)
p	0.83+0.13 (0)	1.00+0.00 (0)	1.00+0.00 (0)	0.59+0.28 (0) 0.33+0.28 (3)	1.00+0.00 (0)	1.00+0.00 (1)
q	0.99+0.01 (1)	1.00+0.00 (0)	1.00+0.00 (0)	1.00+0.00 (3)	1.00+0.00 (0)	0.99+0.01 (1)
r	0.65+0.21 (0)	0.94+0.07 (2)	0.38+0.19 (2)	0.48+0.19 (0) 0.22+0.18 (1)	0.96+0.05 (0)	0.95+0.05 (4)
s	0.98+0.05 (0)	0.73+0.12 (0)	0.93+0.10 (0)	0.81+0.21 (0)	1.00+0.00 (0)	0.53+0.30 (1) 0.26+0.25 (2)

First number in each cell is the posterior probability and the second number is the standard deviation. Numbers in parenthesis indicate the character state (explained in Table 1). Nodes a–s are marked on the phylogeny in Figs. 2 and 3.

without indicated by a star in Table 3) showed little difference for possible optimal ranges at each node, although there were more alternative solutions at some nodes from one of the coding models in DIVA and AReA (Table 3). There was also little difference between optimizations with and without the outgroups in a rooted tree.

2.2.2 Constraints of maximum areas at a node For DIVA, optimal solutions at each node was usually more numerous when maximum areas were constrained as four and contained the solutions from maximum area constrained as 2 (Table 3). An exception was found in the analysis without fossils and coded *Alangium* for its root area only. In this analysis, solutions at most of the eight compared nodes were either identical between the two constraints or more inclusive from analyses with constraint of maximum areas equal to 2. For ML analyses from LAGRANGE, the most probable areas with the highest likelihood at each node were highly consistent among the two different constraints (Table 3). These results were also quite consistent, especially at the four lowest nodes, with the analyses from AReA without restriction on maximum areas if fossils were included.

2.2.3 With and without fossils The reconstruction of ancestral areas at some nodes were strikingly different between including fossils and excluding fossils (Table 3). With fossils, the reconstructions at deepest nodes (nodes 30, 41, 28, 42) favored Europe. This area was not included in any of the solutions at these nodes from analyses without fossils except at node 42 from the ML analysis with LAGRANGE with a constraint of maximum area of 4.

2.2.4 Among DIVA, ML, and Bayesian methods When all conditions were identical in the analyses, results from DIVA and ML showed differences at most nodes, and most of the time, DIVA suggested more alternatives while ML was able to narrow down to one or two most probable areas with the highest likelihood (Table 3). ML and DIVA remarkably

Table 3 Comparisons of ancestral distributions optimized from LAGRANGE (ML), AReA, and DIVA

Node	With fossils						Without fossils								
	Maxareas=4			No restriction	Maxareas=2		Maxareas=2			No restriction		Maxareas=4			
	A: ML[a] (LAGRANGE) V=0.0057, L=0.0052 Ln= –77.12	B: DIVA[b] (DIVA[a])	C: DIVA* [b]	D: ML[b] AReA (AReA*) V=0.09 L=0.01; Ln= –343.43 *Ln= –342.92	E: DIVA[b] (DIVA[a])	F: DIVA* [b]	G: ML[a] (LAGRANGE) V=0.0054 L=0.0078 Ln= –70.52	H: DIVA[b] (DIVA[a])	I: DIVA*[b]	J: ML[b] AReA (AReA*) V=0.09 L=0.01 Ln= –291.31 *Ln= –290.73	K: Bayes Traits[b]	L: ML[a] (LAGRANGE) V=0.0047 L=0.0033 Ln= –63.43	M: DIVA[b] (DIVA[a])	N: DIVA*[b]
34	1–234 (0.23)	135 (35/135)	15/35/135	**1**/3 (**1**/13)	15/35 (35)	15/35	1–3 (0.53)	13 (13)	13	**1**/3 (**1**/3)	3 (0.54)	1–234 (0.60)	13 (13)	13
30	5–5 (0.36)/ 3–3(0.25)	3 (3)	1/3/13	1/**5** (**1**/5)	1/3 (3)	1/3	3–3 (0.53)	1 (1/13)	1/3/13	**1**/3 (**1**/3)	3 (0.40)	3–3 (0.45)	1/3 (3)	1
27	157–3 (0.23)	3 (3)	3/13/15/ 135/1357	1/**5** (**1**/13/15)	3/13/ 15 (13/35)	3/13/15	5–3 (0.31)	13 (13/35)	13	**1**/3/15 (**1**/3/15)	1 (0.42) 3 (0.31)	157–3 (0.45)	13/135 (13/5/ 135)	13/135
37	2–5 (0.53)	2 (2)	2/125	1/2/**5** (1/2/**5**)	2/ 25 (2/12/25)	2	2–2 (0.67)	2 (2/12)	2	1/2/**3**/23 (1/2/**3**/ 13/23)	2 (0.65)	2–2 (0.05)	2 (2)	2
40	1–2 (0.07)/ 2–2(0.06)/ 2–4(0.05)	14 (14)	14	1/**5**/13… (1/**5**…)	14 (14)	14	2–2 (0.22)	14 (14)	14	1/**3**/… (1/3/…)	1(0.49); 4(0.35)	1–2 (0.14) 2–4 (0.12)	14 (14)	14
41	5–5 (0.40)	25/125/245/ 1245 (125/ 245/1245)	15/25/125/ 245/1245	1/**5** (1/**5**/25)	5/15/ 25/24 (5/15/25)	15	2–2 (0.67)	1 (1/12/24)	1	1/**3** (1/2/**3**)	2(0.34); 1(0.29); 3/4(0.14)	2–2 (0.66)	1/24/124 (124)	1
28	5–5 (0.41)/3–3(0.21)	3(3)	1/3/13	1/**5** (**1**/5)	1/3/ 13 (3)	1/3/13	3–3 (0.53)	1 (1/3)	1/3/13	1/3 (1/3)	3(0.40); 1(0.29)	3–3 (0.24)	1/3 (3)	1
42	5–5 (0.13)	23/123/234/ 1234/235/ 1235/2345 (1234/1235/ 1345/2345)	1/13/23/ 123/34/ 134/125/ 1235/1245/ 1345	1/**5**/15 (1/5/15)	1/13/23/34/ 35 (13/23/ 34/35)	1/13	3–2 (0.28)	1 (1/12/13/ 23/14/34)	1/13	1/3 (1/3)	1(0.36)	357–2 (0.28)	1/23/123/ 124/234/ 1234 (1234)	1

Distributions: eastern Asia, 1; eastern North America, 2; western North America, 3; Central and South America, 4; Europe, 5; Australia 6; Africa, 7. Inferred ancestral areas without physical connections were excluded from DIVA results except at node 5 for which there is only one solution. A star indicates results from analyses coding *Alangium* for its root area (eastern Asia) estimated in Feng et al. (in review). Columns without a star show results from analyses with the wide coding model for *Alangium* (all areas of its occurrenc). Superscript "a" indicates analyses with a rooted tree without outgroups. Superscript "b" indicates analyses with rooted tree(s) including outgroups. The optimal ancestral areas at each node presented under LAGRANGE are the ones with the highest likelihood scores and highest probabilities among the alternatives. The probablility of other solutions with likelihood scores nearly equal to the highest (within a value of 2) are much lower (two magnitudes smaller), thus are not presented. A dash in the result under LAGRANGE indicates the split of areas in the two daughter lineages. Alternative most likely solutions (with score difference in log likelihood units within 2) from AReA at each node are presented and separated by slashes. The one with the highest likelihood score is in bold face. Dots "…" indicate there are many alternative solutions with slightly lower likelihood score (within a value of 2). Results under AReA are from analyses with the V and L rates resulting in the highest overall likelihood score. Results from analyses with other combinations of V and L rates are not presented. For BAYESTRAITS, the areas with posterior probability greater than 10%s are presented.

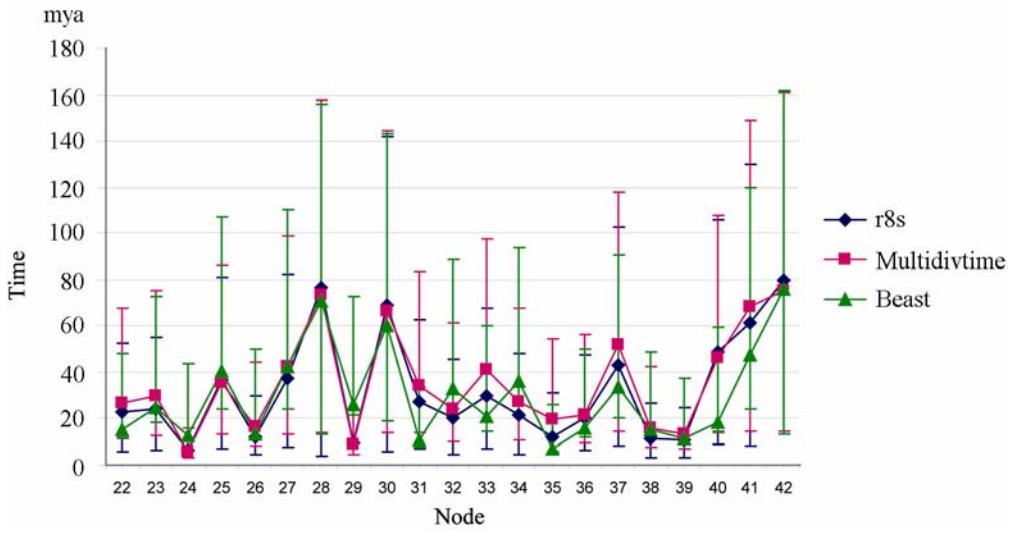

Fig. 4. Comparison of divergence times estimated from BEAST with those from Multidivtime and r8s in Xiang et al. (in review). Node numbers correspond to those marked in Fig. 5 (22–42). Estimations from BEAST at nodes 22, 29, 31, 40, and 41 (shaded in green) are significantly younger than estimates from Multidivtime and r8s.

differed in the estimations at nodes 30, 40, and 41 in analyses without fossils and with maximum areas constrained as 2 (Table 3, comparing ML[a] in Column G and DIVA[a] in Column H) and differed at the nodes 34, 27, 40, 41, 42 in analyses without fossils and with maximum areas constrained as 4 (Table 3, comparing ML[a] in Column L and DIVA[a] in Column M). The differences between ML and BayesTraits were much less, however, only at nodes 37 and 40 (Table 3, comparing ML[b] in column J and BAYESTRAITS[b] in column K). In analyses with fossils, ML and DIVA differed at all of the eight nodes (Table 3, comparing ML[a] in column A and DIVA[a] in column B). At nodes 30 and 28, DIVA favored western North America (area 3) while ML favored Europe (area 5). At node 40, DIVA consistently optimized a disjunct ancestral distribution in eastern Asia and Central and South America (areas 14), but ML suggested distributions involving eastern Asia and eastern North America (areas 12) and a number of alternatives all with low posterior probabilities (Table 3, column A). In general, reconstruction of ancestral distributions at the lower nodes of the *Cornus* phylogeny using different methods showed substantial differences and uncertainty. Probabilities of the most likely ancestral area estimated from ML methods at each node were mostly below 50% (Table 3).

2.3 Divergence time

Estimations of divergence time using BEAST were congruent at most nodes with those from previous studies using Multidivtime and r8s (Xiang et al., in review) (Fig. 4). Exceptions were found for five nodes (shaded in green in Fig. 5). The estimations from BEAST were much younger for nodes 22, 31, and 40, 41, but older for node 29 from BEAST. Divergence time estimations indicate that *Cornus* split from its sister *Alangium* in the late Cretaceous and early radiation of the genus into several clades occurred also during the late Cretaceous and the earliest Tertiary (Fig. 5).

3 Discussion

3.1 Does choice of methods matter?

The comparison of ancestral state reconstruction for six characters shows that the impact of method choice depends on the nature of characters. For characters with no homoplasy, no polymorphism, and no missing data (e.g., character 5), the reconstruction of the ancestral state was consistent among all methods compared and between topology-based and chronogram-based analyses (see results & Fig. 2C). This suggests that for perfect and clean data meeting these conditions, choice of methods does not affect the inference of evolutionary trend, although analysis using a chronogram is better in that it provides temporal information on the events of character state transitions (see Fig. 3). However, for characters with homoplasy, but no missing data and no polymorphism (characters 2, 3, 4), character state reconstructed at nodes including the taxa carrying the homoplasy (e.g.,

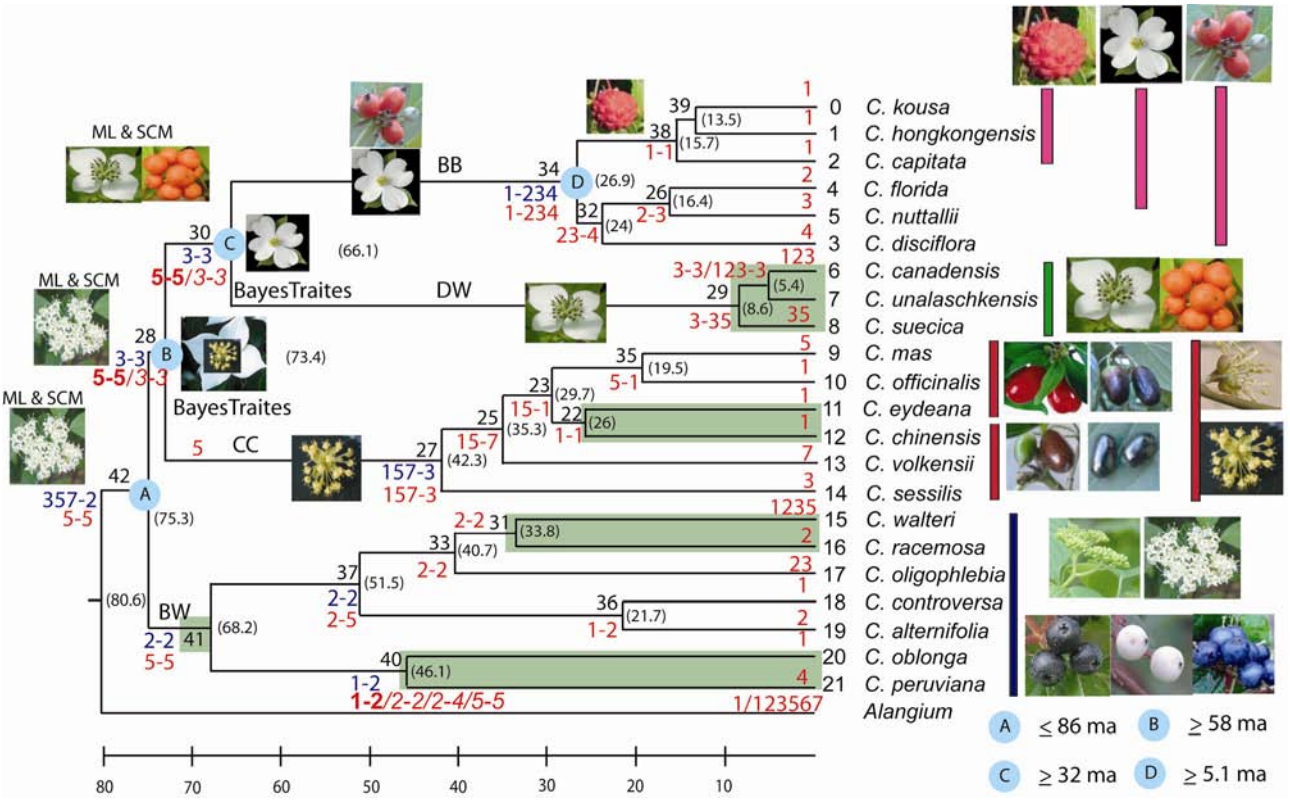

Fig. 5. Chronogram of *Cornus* from showing major variation and evolutionary trends of inflorescence and fruit characters, as well as ancestral areas estimated from the maximum likelihood method. The chronogram was derived from Bayesian inference of the phylogeny of the six gene regions *rbc*L,*mat*K, *ndh*F, *atp*B, ITS, and 26S rDNA and divergence time estimation based on *rbc*L,*mat*K, *ndh*F, *atp*B, and 26S rDNA using Multidivtime in Xiang et al. (in review). This topology and chronogram was used for analyses of character evolution and biogeographic history. BB, big-bracted dogwoods; BW, blue- or white-fruited dogwoods; CC, cornelian cherries; DW, dwarf dogwoods. All nodes are strongly supported with 1.00 posterior probability and bootstrap values >90%, except a few terminal nodes, 32, 39, 22, 31 and one deep node, 28. Nodes 42, 34, 30, and 28 were constrained by fossils in all divergence time analyses. Numbers connected by a dash at each node indicate geographic area splitting in the two descendant lineages; the number before the dash represents areas inherited down to the upper daughter lineage while the number after the dash represents areas inherited by the lower daughter. Alternative area combinations at a node are separated by forward slash. Areas in italics have lower likelihood and lower probability (see Table 3). Areas in blue at eight deep nodes are inferred from ML analyses without fossils by LAGRANGE and areas in red are inferred from ML analyses including fossils using LAGRANGE and AReA. 1=eastern Asia, 2=eastern North America, 3=western North America, 4=Central and South America, 5=Europe, 6=Australia, and 7=Africa. The mean divergence time estimation from Multidivtime for each node is indicated by number in parentheses in million years ago (mya). The horizontal axis shows the time scale in million years ago. Nodes numbered from 22–42 correspond to those in Fig. 4. Nodes with divergence times estimated from BEAST that are significantly different from estimations from Multidivtime and r8s (nodes 22, 29, 31, 40, and 41) are shaded in green.

nodes g and r for characters 2, 3, nodes c, f, g, m, and r for character 4) is method-dependent (Figs. 1, 2). In such cases, the certainty on the reconstructed state increases from parsimony, to ML, and then to SCM methods, while the reconstructions by BAYESTRAITS often suggested a different character state as being most probable. Furthermore, analyses based on the best tree topology only or chronogram can result in different reconstructions at some of these nodes (see Figs. 2 & 3 for characters 2, 3, 4). For characters with missing data and homoplasy, but no polymorphism (character 1), reconstruction of the ancestral state based on a chronogram at the node including the lineage with missing data was strikingly different from that based on the topology only (Figs. 2A, 3), suggesting that reconstruction of the ancestral state with missing data is sensitive to branch length (or time available for evolution). Finally, for characters with polymorphism and homoplasy, but no missing data, the inference from parsimony and BAYESTRAITS was quite congruent (character 6; Figs. 2C, 3). Comparisons of results from ML and SCM for this type of character cannot be made because the ML and SCM in MESQUITE cannot perform estimation for data with polymorphism within a taxon at the present. Development of algorithms that can cope with this type of

character is desirable since polymorphism is not uncommon. Nonetheless, these results indicate that the incongruence among analyses mostly involves homoplasious characters, implying difficulty and uncertainty in reconstructing the ancestral state of characters with homoplasy. This finding, although may not be surprising, is particularly important to bear in mind in studies of comparative biology, e.g., when attempting to demonstrate adaptive evolution of species or genes with repeated gain or loss of a certain ecological trait or gene function in different species by reconstructing the ancestral state on a phylogeny. It is recommended that when reconstructing the ancestral state for homoplasious characters, sensitivity to methods and models of character evolution should be explored. For example, one can use BAYESTRAITS to explore the influence of different rates of character transition and prior distributions on the reconstruction. Careful analyses can help to avoid over confident or incorrect conclusions on important biological problems derived by simple estimations using maximum parsimony on a single tree topology.

The results from this study further indicate that, for the nodes with phylogenetic uncertainty (e.g., relatively lower bootstrap support), the ML and SCM analyses with the simple Mk1 model and a single best tree (no matter whether including or excluding temporal information) can result in confident reconstructions that are not favored by BAYESTRAITS (e.g., node r for characters 2, 3, 4; corresponding to node 28 in Fig. 5). Furthermore, the reconstructions with ML and SCM at nodes with strong support (e.g., node g, for characters 3, 4; Figs. 2B, 3) can also be overconfident compared to BAYESTRAITS when some level of uncertainty exists at other nodes of the phylogeny. The discrepancy between ML, SCM, and BAYESTRAITS found in this study reflects a combination of uncertainties in both ancestral state and phylogenetic reconstructions. If the Mk1 model implemented in the ML and SCM analyses is not the correct model for the characters under investigation, the use of this model in the ML and SCM analyses could have been responsible for the incongruence. Unfortunately, there are few alternative models available for comparison in MESQUITE 2.01. At present, the Asymmetrical 2-parameter Markov-K model is the only alternative model in MESQUITE under the ML and SCM methods. Yet this model is limited to binary data for defining differential rates for forward and backward changes. There is no option in MESQUITE for applying differential rates of changes between states of multistate morphological characters, which are probably most common and interesting. User-defined step matrix for cost of change for multistate characters can only be applied to parsimony analysis. An experiment with the ML analysis using the Asymmetrical 2-parameter Markov-k model for the only binary character in this study (character 5) with a cost of 0.5 for state 0 (clustered fruits) to state 1 (fused fruits) and cost of 1 for state 1 to state 0 did not change the result. Experiments with MP analyses using step matrices for multistate characters also showed little difference. Only at node m (common ancestor of CC) of character 1, the reconstruction changed from equivocal to state 1 (x=10), a state inconsistent with the inferences from BAYESTRAITS, ML, and SCM (Figs. 2A, 3).

Compared to the ML and SCM methods in MESQUITE, the full Bayesian influences approach implemented in BAYESTRAITS estimates the parameters of a model of discrete trait evolution and infer ancestral state by combining information from uncertainty of phylogeny and uncertainty in the estimate of ancestral state (Pagel et al., 2004). It uses the continuous-time Markov model (allowing more than one change along a branch) to estimate the maximum likelihood values and their posterior probability distribution of rate parameters based on the phylogeny and observations on the value of the traits in each species. The method uses the most recent common ancestor approach to overcome the limitation posed from variation in tree topologies (e.g., not all of the trees necessarily contain the internal node or nodes of interest) (Pagel et al., 2004). In addition, the method takes into account branch length (or temporal information) in the estimation of ancestral state. These features probably make BAYESTRAITS a better choice for studying homoplasious character evolution. The program is quite user-friendly and allows authors to explore various parameters for prior distributions and rate of character transitions.

3.2 Character evolution in *Cornus*

The following discussion on character evolution in *Cornus* is based on results from BAYESTRAITS and chronogram-based ML and SCM analyses.

3.2.1 Chromosome numbers In *Cornus*, most species have x=11, three species have x=10 and three have x=9 (see Table 1). In species with x=9 there are two large metacentric chromosomes while in species with x=10 one pair of the large metacentric chromosomes are replaced by two telocentric pairs and in species with x=11, the two pairs of large metacentric chromosomes are replaced by four pairs of telocentric chromosomes (Dermen, 1932). Based on this evidence, Dermen (1932) proposed chromosomal

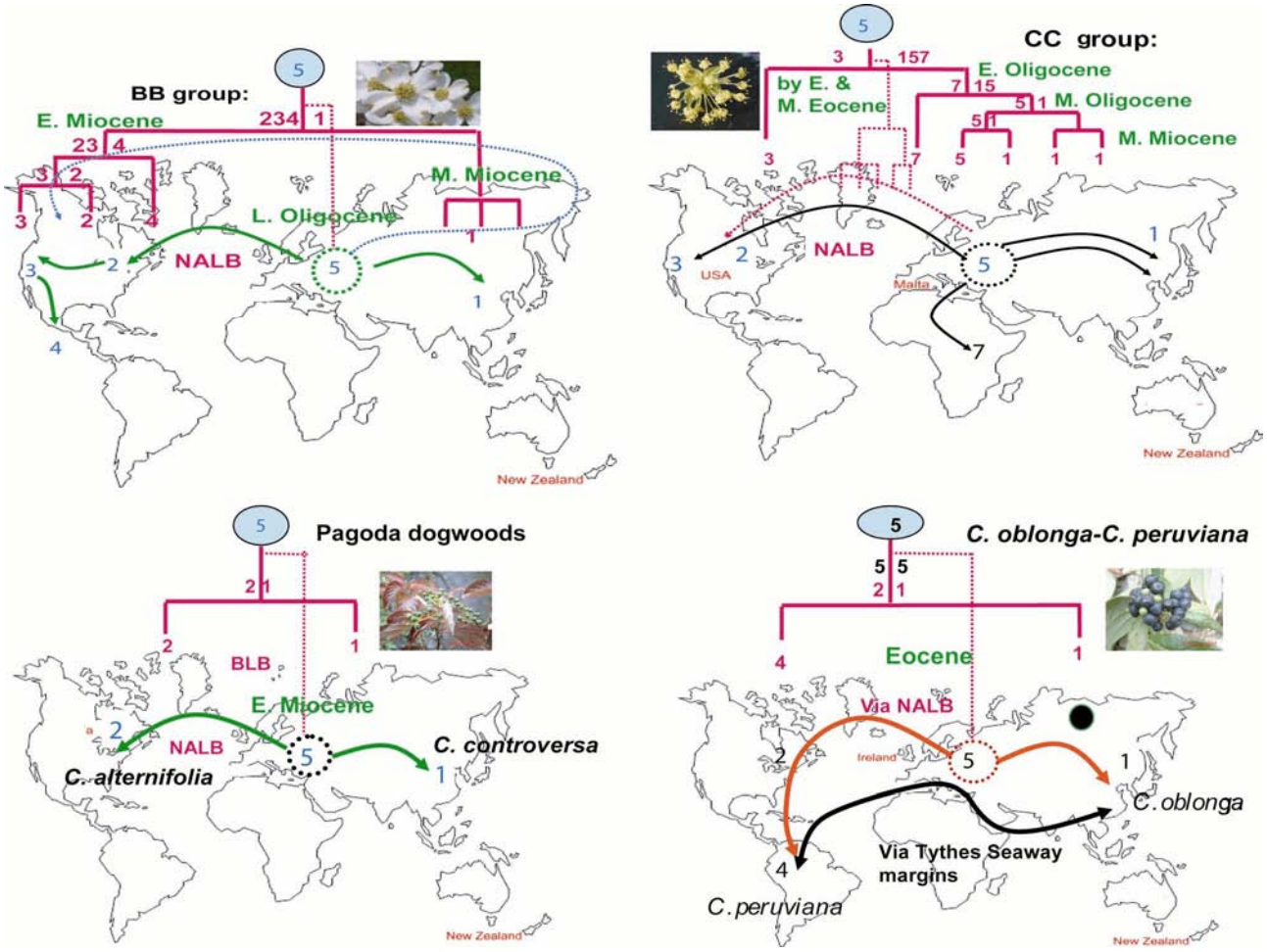

Fig. 6. Biogeographic histories of the four intercontinental disjunct lineages of *Cornus* favored by the ML analysis including fruit fossils. BB, big-bracted dogwoods; CC, cornelian cherries. Numerical numbers represent geographic areas corresponding to those defined in Table 1, Table 3, and Fig. 5.

Robinsonian fission as the mechanism for chromosome number evolution in *Cornus*. This scenario was not supported by the chloroplast DNA phylogeny (Xiang et al., 1996). Based on the parsimony principle and the chloroplast DNA phylogeny, Xiang et al. (1996) and Xiang & Eyde (1995) suggested that the chromosome number in *Cornus* evolved from x=11 to 10 and 9; with two independent reductions to x=10, once in the common ancestor of the CC lineage (node m) and once in the common ancestor of pagoda dogwoods (node q), and a further reduction from x=10 to x=9 later within the CC lineage. This implies three independent chromosomal fusion events. The hypothesis partially agrees with the reconstruction of ancestral state in the present study. The ML, SCM, and BAYESTRAITS all suggested x=11 being ancestral, but the evolutionary pathway reconstructed using these model-based methods is more dynamic. These methods inferred a change from x=11 to x=9 (rather than 10 as in parsimony) in the common ancestor of CC (node m) and x=9 to x=10 in *C. sessilis* Torr. ex Durand (Fig. 3). This implies two chromosomal fusions (at node m) and one fission (in *C. sessilis*) in the CC lineage and one chromosomal fusion in the pagoda dogwoods (node q) occurred during the evolution of *Cornus*. Furthermore, divergence time dating and the reconstructions from ML, SCM and BAYESTRAITS suggest that the fusion events in the cornelian cherries occurred back in the Eocene, while the fusion event in the pagoda dogwoods occurred not earlier than the Miocene (Figs. 3, 4).

Chromosome fission has been considered a major mechanism leading to speciation and increases chromosome numbers in closely related species that are

derived rapidly (see Godfrey & Masters, 2000; Kolnicki, 2000), while chromosome fusion is considered the major mechanism causing a decrease in chromosome numbers (Murnane, 2006). Both, especially fusion, have been reported from plants and animals, e.g., fission in flies (Rousselet et al., 2000) and beans (Schubert & Rieger, 1990), and fusion in human (Ijdo et al., 1991), moths (Traut & Clarke, 1997), *Drosophila* (Yu et al., 1999), and *Arabidopsis* (Heacock et al., 2004). Loss of telomeres, occurring spontaneously or via exogenous DNA damage, (Murnane, 2006) has been considered the major cause for genome instability which provide "sticky ends" for chromosome fusion (Murnane, 2006). In *Cornus*, it is possible that spontaneous loss of telomeres in acrocentric chromosomes had led to chromosome fusion and resulted in a decrease of chromosome numbers in three lineages at different geological times.

3.2.2 Inflorescence architecture and morphology The different major lineages of *Cornus* exhibit strikingly different inflorescence architectures (Fig. 5). All species of the big-bracted (BB) lineage bear cymose heads (glomerule) subtended by four petaloid bracts except in two species. The pacific flowering dogwood *C. nuttallii* Audubon usually has six bracts and the Mexican flowering dogwood *C. disciflora* Moc. & Sesse ex DC. has four, small, broadly ovate, scale-like bracts that are deciduous prior to anthesis and before they expand to become petaloid. The conditions of inflorescence buds among species of the BB lineage also vary. In some species the bud is preformed in the summer or fall and are covered completely by the scale-like bracts (e.g., in *C. florida* L. and *C. disciflora*). These bracts will expand and develop into the large petaloid structure during the next growing season in *C. florida*, but will fall off before their expansion in *C. disciflora*. In two species (e.g., in *C. kousa* and *C. multinervosa* (Pojark.) Q. Y. Xiang), the inflorescence bud is covered by the expanded leaf bud scales beneath the inflorescence bud (thus the whole bud is mix bud; Xiang, 1987). In a few species (e.g., *C. nuttallii*, *C. hongkongensis* Hemsley, *C. capitata*, & *C. elliptica*), the inflorescence bud is not protected. The young bracts are non-scaly and only partially cover the bud. In the cornelian cherry group (CC), species bear cymose umbels subtended by four broadly ovate, scale-like bracts (like those in *C. disciflora*) (see Fig. 5), which are persistent through the anthesis and young fruit stage. The inflorescence buds are also preformed in the fall and covered by the scale-like bracts, but the bracts never expand to become petaloid during the growing season of the following year. The blue- or white-fruited lineage (BW) produce large and branched compound cymes with one rudimentary, early deciduous (before anthesis) and minute bract at the branching point of the lower part of the inflorescence (see Fig. 5). The flower buds are not preformed in this group, except probably in the pagoda dogwoods (*C. alternifolia* L.f. and *C. controversa* Hemsley). In these two species, the inflorescence is preformed in the fall and covered by several alternatively arranged and imbricate scales that also protect the leaf buds. In the fourth lineage, the dwarf dogwood group (DW), the inflorescence architecture is an intermediate between the BB and BW (see Fig. 5). These species have a highly reduced branched compound cyme that is not preformed, but subtended by four large petaloid bracts.

The ancestral form of inflorescence architecture in *Cornus* was inferred to be branched compound cymes with rudimentary bracts, like those found in the BW lineage (Fig. 5). Two alternative pathways of subsequent evolution of inflorescences were suggested, one revealed by BAYESTRAITS and the other by the ML and SCM analyses. The two pathways mainly differ in the reconstruction of character state at the two deep nodes (g and r in Figs. 2, 3, corresponding to nodes 28 and 30 in Fig. 5). The BAYESTRAITS reconstructions at these two nodes are illustrated with images at the right side of the nodes in Fig. 5, while those by ML and SCM are illustrated with images at the left of the nodes in Fig. 5. The Bayesian pathway implies a series of dramatic changes of inflorescence architecture during the early diversification of the genus in the late Cretaceous or early Tertiary. Suppression of inflorescence branching, preformation of inflorescence bud, development of four base bracts into broadly ovate scales, and enlargement and petaloidy of the scales all occurred together during the intial divergence of the genus, leading to umbels subtended by four large petaloid bracts for the common ancestor of BB , CC, and DW (image at right of node 28 in Fig. 5; node r in Figs. 2, 3). During the divergence between CC and BB+DW lineages, bract expansion and petaloidy were lost in the CC stem lineage (node 27 in Fig. 5; node m in Figs. 2, 3), while umbels were replaced by a head via suppression of pedicel development along with a reversal to a non-preformed inflorescence bud in the BB+DW stem lineage, resulting in glomerules subtended by four large petaloid bracts in the common ancestor of BB and DW (image at right of node 30 in Fig. 5; node f in Figs. 2, 3). When DW diverged from BB, a reversal to the branched compound cyme occurred in the DW

stem lineage through slight elongation of inflorescence branches, leading to the mini-form of branched compound cymes with 4, large petaloid bracts observed in the modern dwarf dogwood species (*C. canadensis* L. f., *C. suecica* L., and *C. unalaschkensis* Ledeb.) (image on the branch of DW in Fig. 5). In the BB group, later losses of bract enlargement and petaloidy occurred in *C. disciflora* in the mid Tertiary, associated with a southward migration in North America into moister mountains of Mexico and further to Central America.

This scenario, although favored (has the highest posterior probability) by BAYESTRAITS among the alternatives, has much uncertainty as seen from the variable posterior probability values at the two deep nodes (all pp < 0.95) (see characters 2, 3, 4 in Figs. 2, 3). The uncertainty does not seem too closely correlated with the phylogenetic uncertainty. First, the posterior probabilities at nodes with uncertain reconstructions of ancestral states are not similar to the posterior probabilities in the phylogenetic trees (Figs. 2, 3, 5). Nodes with high pp values in the phylogenetic tree (e.g., nodes f and g) can have low pp values for the favored state or two alternative favored states of the inflorescence characters (Figs. 2, 3). Thus, the uncertainty must come from a combination of phylogenetic uncertainty and ambiguity in ancestral state reconstruction. However, it is noteworthy that this scenario suggested by BAYESTRAITS on inflorescence evolution is novel. An ancestral inflorescence of umbels subtended by petaloid bracts has never been proposed and is not found in extant species or fossils. If it is true, an umbel subtended by petaloid bracts must have become extinct (due to unknown disadvantages). The early Tertiary cornelian cherry fossils (all represented by fruits; see Eyde, 1988; Crane et al., 1990; Xiang et al., 2003, 2005) might have been the lineage bearing this type of inflorescence that has become extinct. It would be interesting to find out if there are any fossils with such morphology from the sites where the fossil fruit stones were discovered. If this evolutionary pathway is true, the umbel with four scale-like bracts in the modern cornelian cherries (CC) represents a retention of the ancestral inflorescence from the common ancestor of CC and BB+DW, while the compound cyme in the dwarf dogwoods (DW) represents a reversal and the four showy bracts in the DW species are plesiomorphic.

In the alternative pathway suggested by ML-SCM ("best"-phylogeny based) the early ancestors of the four major lineages of *Cornus* (nodes g, r, s in Figs. 2, 3) all retained the ancestral inflorescence form (compound cymes) (images at left of nodes 42, 28, 30 in Fig. 5). Major changes of inflorescence architecture occurred more recently, after the diversification into the four major clades. The suppression of inflorescence branching occurred in CC and BB lineages separately. The development of 4-fold bracts and their petaloidy occurred before the condensation of the inflorescence branches in the BB lineage. This pathway implies that the umbels with four scale-like bracts in the cornelian cherries (CC) are an apomorphy and evolved from the ancestral inflorescence type (compound cymes), while the minicompound cymes with four showy bracts in the dwarf dogwoods (DW) represent a plesiomorphy and a retention of the ancestral inflorescence (Fig. 5), in contrast to the inferences from BAYESTRAITS. It is difficult to distinguish the two hypotheses without additional evidence, e.g., developmental and genetic data, to determine which pathway is more likely. Given that the ML and SCM analysis used the "best" phylogeny based on six gene regions (ITS, 26S rDNA, *mat*K, *rbc*L, *atp*B, and *ndh*F) (Xiang et al., in review), which is also supported by four other low copy nuclear genes (PI & AP3 genes: Zhang, 2006; Zhang et al., 2008. *antR*-Cor: Fan et al., 2004, 2007. *waxy*: Xiang, unpublished), this "best" phylogeny may well represent the species phylogeny, adding some support to the ML and SCM pathway. In other words, the ancestor of BB-DW-CC was more likely bearing branched compound cymes with early deciduous rudimentary bracts and the inflorescence was not preformed, and the ancestor of BB-DW was more likely to bear a condensed compound cyme with four petaloid bracts, like the DW species (Fig. 5). Nonetheless, the ancestral state of inflorescence architecture at the two deep nodes above the root of the genus remains uncertain to some extent. Histological studies on the morphogenesis of different inflorescence types will be helpful to evaluate which ancestral state at these nodes may be more likely from a developmental perspective. Coupling with a molecular genetic approach, such studies would further unravel the molecular basis underlying the evolution of inflorescence architectures, a project now is being on-going in Xiang's lab.

3.3 Evolution of fruits and potential ecological causes

Fruits of the dogwoods are drupaceous developing from a 2-locular inferior ovary, with a fleshy layer of pulp and a stony endocarp. In most species, the fruits are simple, but in the Asian species of the big-bracted (BB) lineage, fruits are compound (fused, appearing like a strawberry with stones) (see Fig. 5).

Character ancestral state reconstruction and divergence time dating indicate that this dramatic change in fruit structure of *Cornus*, fusion of clustered fruits, occurred in the mid or late Miocene, associated with an isolation of the BB lineage in eastern Asia, and was proposed to be a result of selections by monkeys (Fig. 5; Eyde, 1985).

The fleshy fruits exhibit various colors at the mature stage among species, including blue, black, white, red, red-black and dark purple (Fig. 5). The blue and black fruits are found in most species of the BW lineage and a few species of the group have white fruits. Red fruits characterize the BB and DW lineages, except for *C. disciflora* which has red fruits maturing into black color. The fruit colors in the cornelian cherries (CC lineage) vary among species from red, red-black, to dark purple. Ancestral state reconstructions using parsimony and BayesTraits (analyses with ML and SCM methods in MESQUITE were not permitted due to polymorphism) suggested that the fruit color at the root of *Cornus* is more likely to be blue or black, and several evolutionary changes in fruit color occurred during the genus radiation in different geological time periods. The early changes appeared to be accompanied with the major alterations in inflorescence architecture (Fig. 2C, character 6; Fig. 5). A red color stage in fruits evolved from the blue and black fruits in the early Tertiary (node r; Fig. 2C). Red fruits evolved in lineages with petaloid bracts (node g) in the early Tertiary and also in the ancestor of *C. mas* L. and *C. officinalis* Seib. & Zucc. of the cornelian cherry group (CC) in the mid to late Tertiary (node j in Figs. 2C, 5). Red maturing into black fruits also evolved twice, once in *C. disciflora* in the mid Tertiary and once in ancestor of CC lineage (node m) or in the ancestor of CC+BB+DW clade (node r) in the very early Tertiary (Fig. 2C). Furthermore, dark purple fruit evolved in *C. eydeana* QY Xiang & YM Shui from red-black fruits in the CC lineage, and white fruits evolved from blue-black fruits in the BW clade (Fig. 2C).

Evolution of fruit color in fleshy fruited plants is an old topic in evolutionary ecology, but our understanding of the underlying ecological causes has been very limited (Willson et al., 1989; Willson & Whelan, 1990). Most fleshy fruits consumed by vertebrates in the temperate zones are red and/or black which are also common in tropical and subtropical fruits (Willson et al., 1989; Willson & Whelan, 1990). Willson and Whelan (1990) invoked a wide array of ecological and evolutionary causes to explain the globally most frequent occurrence of red and black fruits along with the occurrence of other colored fruits at lower frequency. A partial list includes effects on avian foraging and against natural enemies, physiological adaptations, results of selection acting on correlated characters, constraint by metabolic costs, competition for dispersal agents, as well as constraint by phylogenetic history (Willson & Whelan, 1990). However, as indicated by the authors, there appears to be no known data to strongly support any of these hypotheses.

In *Cornus*, given that the differentiation in fruit color seems to correlate with phylogenetic lineages and with changes in inflorescence architecture and morphology, it is possible that phylogenetic constraint coupled with selection acting on the inflorescences could have contributed to the fruit color divergence among the major lineages of *Cornus*. However, other factors, such as avian foraging, defense against pathogens, and differences in taste or nutrients might have also played important roles in the fruit color evolution in the genus as discussed below.

The fruits of dogwoods are eaten by mostly birds, but also by rodents and other mammals (Eyde, 1988). There is no clear evidence for color preference (e.g., white fruits of *C. alba* Linn. Mant. and black fruits of *C. sanguinea* L. are both eaten by European redstarts), but there seems to be a preference in fruit size and display positions (on branches vs. on ground). For example, small birds do not swallow and disperse the relatively large cornelian cherry fruits (CC) and some species have fruits persistent on the branch (e.g., *C. sanguinea* with black fruit) while other species have fruits fallen to the ground (e.g., *C. florida* with red fruit) (Baird, 1980; Eyde, 1988). In the later case, color may make a difference. Red color is easier for birds to find on the ground than blue or black fruits. The red fruits of the rhizomatous little dwarf dogwoods (DW) are dispersed mostly by ground-feeding birds and mammals (see Eyde, 1988). The fused red fruits in the Asian big-bracted dogwoods (like the fruits in *C. kousa*) are favored by monkeys that take the sweet pulp and discard the seeds, and were considered as the mechanism for the evolution of the fused large fruits (Eyde, 1985). Furthermore, phytochemistry studies showed that the different colors in fruits are a result of production of different kinds of anthocyanins (Vareed et al., 2006; Bjorøy et al., 2007) and their relative quantities. These studies reported that fresh fruits of *C. alternifolia* and *C. controversa* (BW lineages, black fruits) contain mostly delphinidin 3-*O*-glucoside and delphinidin 3-*O*-rutinoside, as well as a very small amount of cyanidin 3-*O*-glucoside (Vareed et al., 2006). Fresh fruits of *C. florida* and *C.*

kousa (BB lineage, red fruits) contain mostly *cyanidin* 3-*O*-galactoside and *cyanidin* 3-*O*-glucoside, with a negligible amount of delphinidin 3-*O*-glucoside, while major anthocyanins in *C. mas* and *C. officinalis* fruits (CC lineage, red fruits) are delphinidin 3-*O*-galactoside, cyanidin 3-*O*-galactoside and pelargonidin 3-*O*-galactoside. In *Cornus alba* fruits (BW, white), five anthocyanins were detected (delphinidin 3-*O*-β-galactopyranoside-3′,5′-di-*O*-β-glucopyranoside, delphinidin 3-*O*-β-galactopyranoside-3′-*O*-β-glucopyranoside cyanidin 3-*O*-β-galactopyranoside-3′-*O*-β-glucopyranoside, the 3-*O*-β-galactopyranosides of delphinidin and cyanidin), with delphinidin 3-*O*-β-galactopyranoside-3′,5′-di-*O*-β-glucopyranoside being the most abundant (Bjorøy et al., 2007). The relative amounts of the major anthocyanins within each species also vary. Therefore, the alternations in fruit color of *Cornus* must be related to changes in regulation of the anthocyanin pathway. These data also indicate that the red fruits in cornelian cherries and big-bracted dogwoods do not contain the same kinds of anthocyanins, suggesting the origins of red fruits in the two lineages were mostly likely via different developmental pathways.

Although functions of anthocyanins in plants are still debated (Willson & Whelan, 1990; Gould & Lister, 2006), masking of chlorophylls by anthocyanins was found to reduce risk of photo-oxidative damage to leaf cells in some plants as they senesce, including the white-fruited red osier dogwood (*Cornus sericea* L.) and black-fruited Pagoda dogwood (*Cornus alternifolia*) (Feild et al., 2001; Hoch et al., 2003; Lee et al., 2003). There also have been reports on anti-cancer, anti-inflammatory and antioxidant effects of *Cornus* anthocyanins (Seeram et al., 2002; Vareed et al., 2006) as well as their uses in treatment of diabetes mellitus-related disorders (Jayaprakasam et al., 2006; Nair et al., 2006). The pigments delphinidin 3-*O*-β-galactopyranoside-3′,5′-di-*O*-β-glucopyranoside and delphinidin 3-*O*-β-galactopyranoside-3′-*O*-β-glucopyranoside rich in the black fruits of *C. controversa* and *C. alternifolia* were confirmed to show growth inhibitory activity towards various human cancer cell lines (Vareed et al., 2006). The evidence suggests that differences in avian foraging, alteration in nutrient and taste, and defense against pathogens could have played a role in fruit color evolution in *Cornus*.

3.4 Biogeographic history

Ancestral area estimation using parsimony and maximum likelihood methods under different conditions show substantial differences at some nodes and indicate uncertainty in ancestral area reconstructions at lower phylogenetic nodes (Table 3). Inclusion and exclusion of fossils have a striking effect on the uncertainty (see Table 3 and Fig. 5). Other factors, such as coding models for outgroup distributions, maximum area constraints, and including or excluding outgroups in the rooted tree influenced the ancestral area estimation to a lesser extent (Table 3). The differences among the analyses clearly suggest challenge and limitation in estimating lineage biogeographic histories. At one hand, adding fossils in the analyses could have biased the results if the fossils were not correctly placed on the phylogeny and the branch length leading to the fossil taxa were erroneously constrained in the ML analyses. Thus caution must be taken to evaluate the fossil data (e.g., taxonomic and age identifications) and to determine the phylogenetic affinities of the fossil taxa to lessen this problem. On the other hand, one might take an extreme approach by excluding fossils in the analysis to avoid the problems associated with fossils. The shortcoming of this approach is obvious. The biogeographic history inferred with incomplete information (e.g., without considering the fossils) could also be heavily biased. A cautious and feasible approach would be to compare results from analyses including and excluding fossils.

Results from this study and those from Xiang et al. (2006) have also shown that DIVA is sensitive to approaches of geographic area coding for terminal taxa of the ingroup. Coding only the ancestral area of a terminal taxon (recommended by DIVA) vs. coding distributional areas of all of the constitutent species of the terminal taxon could lead to different conclusions. For example, when the fossil lineage of cornelian cherries was coded for occurrence only in Europe (the ancestral area of the lineage estimated from Xiang et al., 2005), the root area of *Cornus* was optimized as Europe under the maximum area constraint of 2 using DIVA (Xiang et al., 2006). However, in the present study, the fossil lineage of cornelian cherries was coded for all distribution areas of its occurrrence (Europe and western North America), a case probably most common since the root areas of most lineages included in an analysis are unknown, the same analysis with DIVA under maxiareas of 2 optimized the root area of *Cornus* to be quite uncertain, most including western North America (Table 3, Column E). The ancestral areas optimized in Xiang et al. (2006) for these lower phylogenetic nodes with the coding approach favored by DIVA (coding for ancestral areas for higher hierarchical terminal taxa) are largely

congruent with the results inferred from ML analyses in the present study (see Table 3, column A and D), which mostly include Europe. However, without divergence time information in Xiang et al. (2006), evaluation of alternative intercontinental dispersal routes for the genus was limited.

The evolutionary pathway of *Cornus* in space and time favored by ML analyses using LAGRANGE including fossils is illustrated in Figures 5 and 6. The origin and early diversification of *Cornus* was suggested to be in the Eurasian continent in the late Cretaceous and early Tertiary (Fig. 5). Three intercontinental dispersals out of Europe into eastern North America occurred in the BW lineage in different geological times, twice in the early Tertiary at ~52 mya and ~46 mya and once in the mid ~22 mya (Fig. 5). These migrations must have crossed the North Atlantic Land Bridge (NALB). Although NALB was believed to have been broken by the early Eocene, migration of temperate plants up to the early Miocene by hopping across the island chains was considered possible (Tiffney, 1985). This result is congruent with post-Eocene Trans-NALB migration of plants (see Xiang et al., 2005). The migrations of the BB lineage from Europe into eastern and western North America, as well as Central America at ~27 mya could have, however, occurred via either NALB and the Bering Land Bridge (BLB) which was believed to be available throughout most of the Tertiary (Tiffney & Manchester, 2001). The early dispersal of the CC lineage from Europe into western North America at ~42 mya was more likely to have taken the NALB route because in the Eocene, Tran-BLB dispersal through Asia was blocked by the Turgai Strait (Tiffney & Manchester, 2001). However, the recent discovery of dogwood leaf fossils from the Paleocene of western North America and northeastern Asia where fruit fossils of the CC lineage of the same age were also found (Manchester et al., in press) may lend support to early Tertiary Trans-BLB migration of the CC lineage at high latitudes. The leaf fossils were not used in the biogeographic analysis due to their phylogenetic uncertainty within the genus.

It must be noted that the biogeographic history inferred here and shown in Fig. 5 and Fig. 6 is contingent upon the fruit fossil evidence available at the present and only the ancestral ranges with the highest likelihood scores and greatest probabilities. Given that there are alternative ranges with slightly lower likelihood scores and lower probabilities at many nodes, uncertainty clearly remains with this biogeographic history. Furthermore, long distance dispersal can never be ruled out. This again demonstrates the limitation and difficulty in biogeographic analyses. Considering all of the fossil evidence and results from different analyses for *Cornus*, it is likely that the genus diversified into major lineages and spread around the globe via NALB and BLB quickly in the early Tertiary, soon after its origin in the late Cretaceous. Taken at face value, the results support post-Eocene migration of plants across the NALB in the BW lineage (from node 37 to 36) (Xiang et al., 2005), suggesting that the NALB may have been more important for plant migration during the later Tertiary time than previously thought (Tiffney and Manchester, 2001; see Xiang et al., 2005). The results are in agreement with the role of the NALB in connecting the floras in Tropical Asia, America and Africa (Fig. 6, CC lineage, *C. oblonga*-*C. peruviana* J. F. Macbr.), as suggested by Xiang et al. (2005, 2006), Lavin et al. (2000), and Davis et al. (2002).

The divergence time between *Cornus oblonga* Wall. and *C. peruviana* J. F. Macbr. was estimated to be in the Eocene by Multidivtime and r8s and in the Miocene by BEAST (node 40 in Fig. 4; Figs. 5, 6). Based on the ancestral areas reconstructed by LAGRANGE and AReA (Table 3), both of these divergence times agree with migration from Europe eastward to Asia and westward to eastern North America via NALB and further southward via long distance dispersal into South America by birds or by water (see discussion in Xiang et al., 2006 and references therein). However, a divergence time in the Eocene also agrees with migration along the Tethys seaway margins, which was considered to be a viable migration route for the thermophilic boreotropical flora in the early Eocene (Tiffney, 1985) and was proposed to be used for spread of the tropical family Alangiaceae (Feng et al., in review).

In summary, this study showed that reconstructions of the ancestral state of characters and ancestral area distributions are affected by both analytical methods and phylogenetic uncertainty. Researchers are encouraged to apply the most sophisticated method possible available and conduct vigorous analyses to explore alternative hypotheses before reaching a conclusion for evolution of characters that are homoplasious. The substantial differences between DIVA and ML analyses that included and excluded fossils suggest the need of a reevaluation of the global biogeographic patterns previously assembled using DIVA and raise special attentions to the use of fossil data in biogeographic analyses. Researchers reconstructing ancestral areas with a phylogeny

containing widely distributed higher hierarchical terminal taxa (e.g., genus, family, etc.) should be particularly careful with the choice of methods.

Acknowledgements The authors thank R. Ree, S. Smith, and B. Moore for assistance and guidance in biogeographic analyses using LAGRANGE and AReA. We also thank the Deep Time Research Coordination Network supported by the National Science Foundation (NSF) funded to D. E. Soltis (DEB-0090283), the Phytogeography of the Northern Hemisphere working group supported by NESCent funded to M. Donoghue and P. Manos, and the ClockWork group supported by NESCent funded to J. Clark and B. Wiegmann for the workshops, symposia, and discussion with the participants of each group. We also thank the two reviewers and the journal editors Y-L Qiu and Z-D Chen for valuable comments to improve this manuscript. This research was supported by the National Science Foundation (DEB-0444125) funded to Xiang and benefited from a Multidisciplinary grant for faculty research and professional development grant at North Carolina State University funded to Xiang, Franks and Xie.

References

APG (Angiosperm Phylogeny Group). 2003. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG II. Botanical Journal of Linnean Society 141: 399–436.

Baird JW. 1980. The selection and use of fruit by birds in an eastern USA forest. Wilson Bulletin 92: 63–73.

Bjorøy Ø, Fossen T, Andersen ØM. 2007. Anthocyanin 3-galactosides from Cornus alba "Sibirica" with glucosidation of the B-ring. Phytochemistry 68: 640–645.

Blanchette M, Green ED, Miller W, Haussler D. 2004. Reconstructing large regions of an ancestral mammalian genome in silico. Genome Research 14: 2412–2423.

Bollback JP. 2006. SIMMAP: Stochastic character mapping of discrete traits on phylogenies. BMC Bioinformatics 7: 88.

Boufford DE, Spongberg SA. 1983. Eastern Asian-eastern North American phytogeographical relationships—a history from the time of Linnaeus to the twentieth century. Annals of the Missouri Botanical Garden 70: 423–439.

Bullock AA. 1959. *Nomina familiarum conservanda proposita*. Taxon 8: 154–181.

Chase MC, Soltis DE, Olmstead RG, Morgan D, Les DH, Mishler BD, Duvall MR, Price RA, Hills HG, Qiu Y-L, Kron KA, Rettig JH, Conti E, Palmer JD, Manhart JR, Sytsma KJ, Michaels HJ, Kress WJ, Karol KG, Clark WD, Hedren M, Gaut BS, Jansen RK, Kim K-J, Wimpee CF, Smith JF, Furnier GR, Strauss SH, Xiang Q-Y, Plunkett GM, Soltis PS, Swensen S, Williams SE, Gadek PA, Quinn CJ, Eguiarte LE, Golenberg E, Learn GH Jr, Graham SW, Barrett SCH, Dayanandan S, Albert VA. 1993. Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbc*L. Annals of the Missouri Botanical Garden 80: 528–580.

Crane PR, Manchester SR, Dilcher DL. 1990. A preliminary survey of fossil leaves and well-preserved reproductive structures from the Sentinel Butte Formation (Paleocene) near Almont, North Dakota. Fieldiana Geol. 1418: 1–63.

Cunningham C. 1999. Some limitations of ancestral character-state reconstruction when testing evolutionary hypotheses. Systematic Biology 48: 665–674.

Davis CC, Bell CD, Mathews S, Donoghue MJ. 2002. Laurasian migration explains Gondwanan disjuncts: evidence from Malpighiaceae. Proceedings of the National Academy Sciences USA 99: 6833–6837.

de Candolle AP. 1828. Prodromus Systematis Naturalis Regni Vegetabilis 3: 203.

Dermen H. 1932. Cytological studies of *Cornus*. Journal of Arnold Arboretum 13: 401–417.

Donoghue MJ, Smith SA. 2004. Patterns in the assembly of temperate forests around the Northern Hemisphere. Philosophical Transactions of the Royal Society, London B 359: 1633–1644.

Drummond AJ, Rambaut A. 2006. BEAST v1.4 [online]. Available from http://beast.bio.ed.ac.uk/.

Ekman S, Andersen LH, Wedin M. 2008. The limitations of ancestral state reconstruction and the evolution of the Ascus in the Lecanorales (Lichenized Ascomycota). Systematic Biology 57: 141–156.

Eyde RH. 1985. The case for monkey-mediated evolution in big-bracted dogwoods. Arnoldia (Jamaica Plain) 45: 2–9.

Eyde RH. 1988. Comprehending Cornus: Puzzles and progress in the systematics of the dogwoods. Botanical Review (Lancaster) 54: 233–351.

Fan CZ, Purugganan MD, Thomas DT, Wiegmann BM, Xiang QY. 2004. Heterogeneous evolution of the Myc-like Anthocyanin regulatory gene and its phylogenetic utility in *Cornus* L. (Cornaceae). Molecular Phylogenetics and Evolution 33: 580–594.

Fan CZ, Xiang QY. 2001. Phylogenetic relationships within Cornus L. (Cornaceae) based on 26S rDNA sequences. American Journal of Botany 88: 1131–1138.

Fan C, Xiang QY. 2003. Phylogenetic analyses of Cornales based on 26S rRNA and combined 26S rDNA-matK-rbcL sequence data. American Journal of Botany 90: 1357–1372.

Fan C, Xiang QY, Remington DL, Purugganan MD, Wiegmann BM. 2007. Evolutionary patterns in the *ant*R-*Cor* gene in the dwarf dogwood complex (*Cornus*, Cornaceae). Genetica 130: 19–34.

Feild TS, Lee DW, Holbrook NM. 2001. Why leaves turn red in autumn. The role of anthocyanins I senescing leaves of red-osier dogwood. Plant Physiology 127: 566–574.

Feng CM, Manchester SR, Xiang QY(J). Phylogeny and biogeography of Alangiaceae (Cornales) inferred from DNA sequences, morphology, and fossils. Molecular Phylogenetics and Evolution. (in review, available from the first author upon request)

Godfrey LR, Masters JC. 2000. Kinetochore reproduction theory may explain rapid chromosome evolution Proceedings of National Academy of Sciences USA 97: 9821–9823.

Gould KS, Lister C. 2006. Flavonoid functions in plants. In: Andersen OM, Markham KR eds. Flavonoids: Chemistry, Biochemistry and Applications. Boca Raton: CRC Press. 397–441.

Harvey PH, Pagel MD. 1991. The Comparative Method in Evolutionary Biology. Oxford: Oxford University Press.

Heacock M, Spangler E, Riha K, Puizina J, Shippen DE. 2004. Molecular analysis of telomere fusions in Arabidopsis: multiple pathways for chromosome end-joining. The EMBO Journal 23: 2304–2313.

Hines HM. 2008. Historical biogeography, divergence times, and diversification patterns of Bumble Bees (Hymenoptera: Apidae: *Bombus*). Systematic Biology 57: 58–75.

Hoch WA, Singsaas EL, McCown BH. 2003. Resorption protection. Anthocyanins facilitate nutrient recovery in autumn by shielding leaves from potentially damaging light levels. Plant Physiology 133: 1296–1305.

Huelsenbeck JP, Nielsen R, Bollback JP. 2003. Stochastic mapping of morphological characters. Systematic Biology 52: 131–158.

Huelsenbeck JP, Rannala B, Masly JP. 2000. Accommodating phylogenetic uncertainty in evolutionary studies. Science 288: 2349–2350.

Huelsenbeck JP, Ronquist F, Nielsen R, Bollback JP. 2001. Bayesian inference of phylogeny and its impact on evolutionary biology. Science 294: 2310–2314.

Ijdo JW, Baldini A, Ward DC, Reeders ST, Wells RA. 1991. Origin of human chromosome 2: an ancestral telomere-telomere fusion. Proceedings of the National Academy of Sciences USA 88: 9051–9055.

Jayaprakasam B, Olson LK, Schutzki RE, Tai MH, Nair MG. 2006. Amelioration of obesity and glucose intolerance in high-fat-fed C57BL/6 mice by anthocyanins and ursolic acid in Cornelian cherry (*Cornus mas*). Journal of Agriculture and Food Chemistry 54: 243–248.

Kishino H, Thorne JL, Bruno WJ. 2001. Performance of a divergence time estimation method under a probabilistic model of rate evolution. Molecular Biology and Evolution 18: 352–361.

Kolnicki RL. 2000. Kinetochore reproduction in animal evolution: Cell biological explanation of karyotypic fission theory. Proceedings of the National Academy of Sciences USA 97: 9493–9497.

Kondo B, Omland KE. 2007. Ancestral state reconstruction of migration: multistate analysis reveals rapid changes in new world orioles (*Icterus* spp.) The Auk 124: 410–419.

Lavin M, Thulin M, Labat JN, Pennington RT. 2000. Africa, the odd man out: molecular biogeography of *Dalbergioid legumes* (Fabaceae) suggests otherwise. Systematic Botany 25: 449–467.

Lee DW, O'Keefe J, Hobrook NM, Field TS. 2003. Pigment dynamics and autumn leaf senescence in a New England deciduous forest, eastern USA. Ecological Research 18: 677–694.

Lewis PO. 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. Systematic Biology 50: 913–925.

Maddison DR, Maddison WP. 1992. MacClade: Analysis of phylogeny and character evolution. Version 3.0. Sunderland, MA: Sinauer Associates.

Maddison DR, Maddison WP. 2001. MacClade4: Analysis of phylogeny and character evolution. Version 4.03. Sunderland, MA: Sinauer Associates.

Maddison WP, Maddison DR. 2007. Mesquite: a modular system for evolutionary analysis [online]. Version 2.01. Available from http://mesquiteproject.org.

Manchester SR. 1994. Fruits and seeds of the Middle Eocene Nut Beds flora, Clarno Formation, North Central Oregon. Palaeontography of America 58: 1–205, 70 pls.

Manchester SR, Xiang QY, Kodrul TM, Akhmetiev M. Leaves of *Cornus* (Cornaceae) from the paleocene of North American and Asia confirmed by trichome characters. International Journal of Plant Sciences. (in press)

Milne RI, Abbott RJ. 2002. The origin and evolution of Tertiary relict floras. Advances in Botanical Research 38: 282–314, pl. 5–7.

Moore BR, Smith SA, Donoghue MJ. 2006. Increasing data transparency and estimating phylogenetic uncertainty in supertrees: Approaches using nonparametric bootstrapping. Systematic Biology 55: 662–676.

Murnane JP. 2006. Telomeres and chromosome instability DNA Repair. Vol. 5. No. 9-10. (8 September 2006). 1082–1092.

Murrell ZE. 1993. Phylogenetic Relationships in *Cornus* (Cornaceae). Systematic Botany 18: 469–495.

Murrell ZE. 1996. A new section of *Cornus* in South and Central America. Systematic Botany 21: 273–288.

Nair MG, Jayaprakasam B, Olson LK, Vareed SK. 2006. Insulin secretion by anthocyanins and anthocyanidins. US Patent and Trademark Office, United States Patent Application No:20060025353.

Nepokroeff M, Sytsma KJ, Wagner WL, Zimmer EA. 2003. Reconstructing ancestral patterns of colonization and dispersal in the Hawaiian understory tree genus *Psychotria* (Rubiaceae): a comparison of parsimony and likelihood approaches. Systematic Biology 52: 820–838.

Nielsen R. 2002. Mapping mutations on phylogenies. Systematic Biology 51: 729–739.

Ouzounis CA. 2005. Ancestral state reconstructions for genomes. Current Opinion in Genetics & Development 15: 595–600.

Pagel M. 1999. The maximum likelihood approach to reconstructing ancestral character states of discrete characters on phylogenies. Systematic Biology 48: 612–622.

Pagel M, Meade A. 2006. Bayesian analysis of correlated evolution of discrete characters by reversible-jump Markov chain Monte Carlo. American Naturalist 167: 808–825.

Pagel M, Meade A, Barker D. 2004. Bayesian estimation of ancestral character states on phylogenies. Systematic Biology 53: 673–684.

Pedersen N, Holyoak DT, Newton AE. 2007. Systematics and morphological evolution within the moss family Bryaceae: A comparison between parsimony and Bayesian methods for reconstruction of ancestral character states. Molecular Phylogenetics and Evolution 43: 891–907.

Posada D, Crandall KA. 1998. Modeltest: testing the model of DNA substitution. Bioinformatics 14: 817–818.

Ree RH, Moore BR, Webb CO, Donoghue MJ. 2005. A likelihood framework for inferring the evolution of

geographic range on phylogenetic trees. Evolution 59: 2299–2311.

Ree RH, Smith SA. 2008. Maximum-likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. Systematic Biology 57: 4–414.

Renner SS, Beenken L, Grimm GW, Kocyan A, Ricklefs RE. 2007. The evolution of dioecy, heterodichogamy, and labile sex expression in *Acer*. Evolution 61: 2701–2719.

Ronquist F. 1996. DIVA version 1.1 [online]. Computer program and manual available by anonymous FTP from Uppsala University (ftp.uu.se or ftp.systbot.uu.se).

Ronquist F. 1997. Dispersal-vicariance analysis: a new approach to the quantification of historical biogeography. Systematic Biology 46: 195–203.

Ronquist F. 2004. Bayesian inference of character evolution. Trends in Ecology and Evolution 19: 475–481.

Rossnes R, Eidhammer I, Liberles DA. 2005. Phylogenetic reconstruction of ancestral character states for gene expression and mRNA splicing data. BMC Bioinformatics, 6: 127.

Rousselet J, Monti L, Auger-Rozenberg MA, Parker JS, Lemeunier F. 2000. Chromosome fission associated with growth of ribosomal DNA in *Neodiprion abietis* (Hymenoptera: Diprionidae). Proceedings of the Royal Society: Biological Sciences 267: 1819–1823.

Sanderson MJ. 2002. Estimating absolute rates of molecular evolution and divergence times: A penalized likelihood approach. Molecular Biology and Evolution 19: 101–109.

Sanmartín I, Enghof H, Ronquist F. 2001. Patterns of animal dispersal, vicariance and diversification in the Holarctic. Biological Journal of Linnean Society 73: 345–390.

Sanmartín I, Ronquist F. 2004. Southern Hemisphere biogeography analyzed with event-based models: Plant versus animal patterns. Systematic Biology 53: 216–243.

Schluter D, Price T, Mooers AO, Ludwig D. 1997. Likelihood of ancestor states in adaptive radiation. Evolution 51: 1699–1711.

Schubert I, Rieger R. 1990. Alteration by centric fission of the diploid chromosome number in *Vicia faba* L. Genetica 81: 67–69.

Seeram NP, Schutzki R, Chandra A, Nair MG. 2002. Characterization, quantification, and bioactivities of anthocyanins in *Cornus* species. Journal of Agricultural and Food Chemistry 50: 2519–2523.

Smith SA. 2006. AReA: Ancestral Range Analysis v.2.1 [online]. Available from http://blackrim.org/programs/area.html.

Soltis DE, Soltis PS, Endress PK and Chase MW. 2005. Phylogeny and Evolution of Angiosperms. Sunderland, MA: Sinauer Associates.

Takhtajan AL. 1987. Sistema Magnoliofitov. Izdatelístvo ìNauka,î Leningrad.

Thorne JL, Kishino H, Painter S. 1998. Estimating the rate of evolution of the rate of molecular evolution. Molecular Biology and Evolution 15: 1647–1657.

Thorne RF. 1972. Major disjunctions in the geographic ranges of seed plants. Quaternary Review in Biology 47: 365–411.

Tiffney BH. 1985. The Eocene North Atlantic land bridge: its importance in Tertiary and modern phytogeography of the Northern Hemisphere. Journal of Arnold Arboretum 66: 243–273.

Tiffney BH, Manchester SR. 2001. The use of geological and paleontological evidence in evaluating plant phylogeographic hypotheses in the Northern Hemisphere tertiary. International Journal of Plant Sciences 162: S3–S17.

Traut W, Clarke CA. 1997. Karyotype evolution by chromosome fusion in the moth genus *Orgyia*. Hereditas 126: 77–84.

Vareed SK, Reddy MK, Schutzki RE, Muraleedharan G. 2006. Anthocyanins in *Cornus alternifolia, Cornus controversa, Cornus kousa*. Life Sciences 78: 777–784.

Willson MF, Irvine AK, Walsh NG. 1989. Vertebrate dispersal syndromes in some Australian and New Zealand plant communities, with geographic comparisons. Biotropica 21: 133–147.

Willson MF, Whelan CJ. 1990. The evolution of fruit color in fleshy-fruited plants. The American Naturalist 136: 790–809.

Wu ZY. 1983. On the significance of Pacific intercontinental discontinuity. Annals of the Missouri Botanical Garden 70: 577–590.

Xiang QY(J). 1987. System and synopsis of *Cornus* subgen. *Syncarpea* (Nakai) Q. Y. Xiang (Cornaceae). Bulletin of Botanical Research (Harbin) 7 (2): 33–52.

Xiang QY(J), Boufford DE. 2005. Cornaceae, Mastixiaceae, Toricelliaceae, Helwingiaceae, Aucubaceae. In: Wu ZY, Raven PH eds. Flora of China. Beijing: Science Press; St. Louis: Missouri Botanical Garden Press. 14: 206–234.

Xiang QY(J), Brunsfeld SJ, Soltis DE, Soltis PS. 1996. Phylogenetic relationships in *Cornus* based on chloroplast DNA restriction sites: implications for biogeography and character evolution. Systematic Botany 21: 515–534.

Xiang QY(J), Eyde RH. 1995. Chromosome number of *Cornus sessilis* (Cornaceae): phylogenetic affinity and evolution of chromosome numbers in *Cornus*. Sida 16: 765–768.

Xiang QY(J), Manchester SR, Thomas D, Zhang WH, Fan CZ. 2005. Phylogeny, biogeography, and molecular dating of cornelian cherries (*Cornus*, Cornaceae)—tracking Tertiary plant migration. Evolution 59: 1685–1700.

Xiang QY(J), Moody ML, Soltis DE, Fan CZ, Soltis PS. 2002. Relationships within Cornales and circumscription of Cornaceae—*matK* and *rbcL* sequence data and effects of outgroups and long branches. Molecular Phylogenetics and Evolution 24: 35–57.

Xiang QY(J), Shui Y, Murrell Z. 2003. *Cornus eydeana* (Cornaceae), A new cornelian cherry from China—Notes on Systematics and Evolution. Systematic Botany 28: 757–764.

Xiang QY(J), Soltis DE. 2001. Dispersal-Vicariance analyses of intercontinental disjuncts: historical biogeographical implications for angiosperms in the Northern Hemisphere. International Journal of Plant Sciences 162: S29–39.

Xiang QY(J), Soltis DE, Morgan DR Soltis PS. 1993. Phylogenetic relationships of *Cornus* L. *sensu lato* and putative relatives inferred from *rbcL* sequence data. Annals of the Missouri Botanical Garden 80: 723–734.

Xiang QY(J), Soltis DE, Soltis PS. 1998. Phylogenetic relationships of Cornaceae and close relatives inferred from *matK* and *rbcL* sequences. American Journal of Botany 85: 285–297.

Xiang QY(J), Thomas DT, Zhang W, Manchester SR, Murrell Z. 2006. Species level phylogeny of the Dogwood genus *Cornus* (Cornaceae) based on molecular and morphological evidence—implication in taxonomy and Tertiary intercontinental migration. Taxon 55: 9–30.

Xiang QY(J), Thomas DT, Manchester MR. 2007. Phylogeny and Biogeography of Cornales—Increased Sampling in Taxa and Characters, Placement of *Hydrostachys*, BEAST vs. Multidivtime, and DIVA vs. AreA. Abstract presented at meeting of Botany & Plant Biology. July, 2007, Chicago.

Xiang QY(J), Thone JL, Seo TK, Zhang WH, Thomas DT, Ricklefs RE. Rate of nucleotide substitution in *Cornus* L. (Cornaceae)—A remarkable increase in the circumboreal lineage. Molecular Phylogenetics and Evolution. (in review, available from the first author upon request)

Yu Y-C, Linà F-J, Chang H-Y. 1999. Stepwise chromosome evolution in *Drosophila albomicans*. Heredity 83: 39–45.

Zhang WH. 2006. Molecular evolution of floral homeotic B-class genes in the dogwood genus *Cornus* (Cornaceae)—gene duplication, selection, and coevolution. Ph.D. Dissertation. North Carolina State University. Available from http://www.lib.ncsu.edu/theses/available/etd-10142006-234225.

Zhang WH, Xiang QY, Thomas DT, Wiegmann BM, Frohlich MW, Soltis DE. 2008. Molecular evolution of *PISTILLATA*-like genes in the dogwood genus *Cornus* (Cornaceae). Molecular Phylogeny and Evolution 47: 175–195.