Molecular Phylogenetics and Evolution 59 (2011) 123–138

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

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

# Resolving and dating the phylogeny of Cornales – Effects of taxon sampling, data partitions, and fossil calibrations

Qiu-Yun (Jenny) Xiang [a,*], David T. Thomas [a,1], Qiao Ping Xiang [b]

[a] Department of Plant Biology, North Carolina State University, Raleigh, NC 27695, USA
[b] State Key Laboratory in Systematic Botany and Evolution, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China

ARTICLE INFO

*Article history:*
Received 29 September 2010
Revised 18 January 2011
Accepted 26 January 2011

*Keywords:*
Cornales
Divergence time
Effects of data partitions and taxa sampling
Fossil calibration
Hydrostachyaceae
Phylogeny relationships
26S rDNA
rbcL
matK
ndhF
atpB
trnL-F
trnH-K

ABSTRACT

The order Cornales descends from the earliest split in the Asterid clade of flowering plants. Despite a few phylogenetic studies, relationships among families within Cornales remain unclear. In the present study, we increased taxon and character sampling to further resolve the relationships and to date the early diversification events of the order. We conducted phylogenetic analyses of sequence data from 26S rDNA and six chloroplast DNA (cpDNA) regions using parsimony (MP), maximum likelihood (ML), and Bayesian inference (BI) methods with different partition models and different data sets. We employed relaxed, uncorrelated molecular clocks on BEAST to date the phylogeny and examined the effects of different taxon sampling, fossil calibration, and data partitions. Our results from ML and BI analyses of the combined cpDNA sequences and combined cpDNA and 26S rDNA data suggested the monophyly of each family and the following familial relationships ((Cornaceae–Alangiaceae)–(Curtisiaceae–Grubbiaceae))–(((Nyssaceae–Davidiaceae)–Mastixiaceae)–((Hydrostachyaceae–(Hydrangeaceae–Loasaceae))). These relationships were strongly supported by posterior probability and bootstrap values, except for the sister relationship between the N–D–M and H–H–L clades. The 26S rDNA data and some MP trees from cpDNA and total evidence suggested some alternative alignments for Hydrostachyaceae within Cornales, but results of SH tests indicated that these trees were significantly worse explanations of the total data. Phylogenetic dating with simultaneous calibration of multiple nodes suggested that the crown group of Cornales originated around the middle Cretaceous and rapidly radiated into several major clades. The origins of most families dated back to the late Cretaceous except for Curtisiaceae and Grubbiaceae which may have diverged in the very early Tertiary. We found that reducing sampling density within families and analyzing partitioned data sets from coding and noncoding cpDNA, 26S rDNA, and combined data sets produced congruent estimation of divergence times, but reducing the number and changing positions of calibration points resulted in very different estimations.



## 1. Introduction

Cornales represents the earliest diverging lineage among the living members of the largest clade, Asterids, on the angiosperm phylogeny (APG, 2003, 2009; Soltis et al., 2000). The order, as currently defined, is a morphologically heterogeneous assemblage containing the dogwood genus (*Cornus L. s.l.*), the hydrangea family (Hydrangeaceae), the blazing star family (Loasaceae) and several small genera with generally isolated geographic distributions (Xiang et al., 1998, 2002; Fan and Xiang, 2003). These genera have been classified into several families, including Alangiaceae (23 spp., *Alangium*, Old World Tropics), Nyssaceae (~8 spp., *Nyssa*, *Camptotheca*, eastern Asia, eastern North America), Davidiaceae (1 sp., *Davidia*, China), Mastixiaceae (15 spp., *Mastixia*, *Diplopanax*, southern China, Southeastern Asia), Curtisiaceae (1 or 2 spp., *Curtisia*, southeastern Africa), Grubbiaceae (*Grubbia*, 1–3 spp., southern Africa), and Hydrostachyaceae (~22 spp., *Hydrostachys*, southeastern Africa) (see Fan and Xiang, 2003; Xiang et al., 2002, 2005). The order is exceptional in having excellent fossil records from the late Cretaceous throughout the Tertiary with many from areas beyond the restricted modern distribution of most genera, (Crane et al., 1990; Eyde, 1963, 1988, 1997; Eyde et al., 1969; Eyde and Xiang, 1990; Tiffney and Haggard, 1996; Mai, 1993; Manchester, 1994, 1999; Manchester et al., 1999, 2007; Takahashi et al., 2002). These fossils provide excellent information for phylogenetic dating analyses.

The Cornales clade also represents challenges in a phylogenetic study due to a possibly rapid radiation and presence of long branches in some taxa, causing difficulty in resolving relationships at deep nodes. Previous molecular phylogenetic studies using data from two or three genes resolved the cornalean families into eight

* Corresponding author. Address: Department of Plant Biology, North Carolina State University, Campus Box 7612, Raleigh, NC 27695, USA. Fax: +1 919 515 3436.
*E-mail address:* jenny_xiang@ncsu.edu (Q.-Y. (Jenny) Xiang).
[1] Present address: Johnston Community College, Smithfield, NC 27577, USA.

1055-7903/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2011.01.016

distinct lineages, but the relationships among them remained unconvincingly resolved (supported by <75% bootstrap values) (Fan and Xiang, 2003; Xiang et al., 1998, 2002). The eight clades are: (1) Cornaceae, (2) Alangiaceae, (3) Nyssaceae–Davidiaceae, (4) Mastixiaceae, (5) Grubbiaceae–Curitisiaceae, (6) Hydrangeaceae, (7) Loasaceae, and (8) Hydrostachyaceae. Several hypotheses were suggested from these studies, including the sister relationships between clades 1 and 2, between clades 3 and 4, between clades 6 and 7. A monophyletic group consisting of all the small families (lineages 1–5), a basal placement of Grubbiaceae and Curtisiaceae, and various placements of Hydrostachyaceae were also suggested in previous studies, but all were weakly supported (Xiang et al., 1998, 2002; Fan and Xiang, 2003). Furthermore, inclusion of the aquatic monogeneric Hydrostachyaceae in Cornales has also been debated for several reasons: (1) a long branch connecting the family, (2) unstable positions that were sometimes outside of Cornales (e.g., in Lamiales) in broad phylogenetic analyses (mostly with parsimony method; but see Burleigh et al., 2009), and (3) generally very weak support for all placements resolved in various studies (Albach et al., 2001; Hempel et al., 1995; Xiang, 1999; Xiang et al., 2002). Although the Hydrostachyaceae was more frequently resolved within Cornales when using a method that can better correct for long branch attraction in phylogenetic analyses (e.g., maximum likelihood [ML] method vs. parsimony [MP]), its phylogenetic placement within Cornales has been inconsistent, mostly within or near Hydrangeaceae and Loasaceae, and sometimes at the base of Cornales (Albach et al., 2001; Fan and Xiang, 2003; Xiang, 1999; Xiang et al., 2002; Schenk and Hufford, 2010). Furthermore, in the most recent phylogenetic analyses of angiosperms using the ML method, including one species of Hydrostachyaceae, the family was placed in Hydrangeaceae in the strongly supported Cornales clade when using a 3-gene data set (18S rDNA, *atpB*, and *rbcL*; 100% boostrap value), but in a strongly-supported Lamiales clade when using a 5-gene data set with missing data (18S rDNA, *atpB*, *rbcL*, *matK*, and 26S rDNA; 97% boostrap value). In both cases, the family was connected by a long branch (Burleigh et al., 2009).

The lack of clear relationships among the families hampers the ability to decipher evolutionary events and the time and place of their occurrence. In a recent study of using *matK* sequence data of Cornales to address the effects of substitution models on divergence time estimation using the Penalized Likelihood (PL) method (Sanderson, 2002), the origin of the Cornales crown group (including Hydrostachyaceae) and the divergence of the major clades were suggested to have occurred in the mid-Cretaceous (~110 million years ago). This hypothesis needs to be evaluated with more data using approaches accounting for phylogenetic and calibration uncertainties (e.g., BEAST of Drummond and Rambaut, 2007), given what was reviewed regarding the weak support of phylogenetic relationships among clades and certain limitations of the PL method (e.g., assuming autocorrelation of divergence rates among lineages and fixation of a nodal time for calibration). In the present study, we use DNA sequence data from seven regions (*rbcL*, *matK*, *ndhF*, *atpB*, *trnH-K*, *trnL-F*, and 26S rDNA, ~12,506 bp) with extensive sampling of outgroups from the Asterids clade and increased sampling from Hydrostachyaceae and other families to (1) further resolve relationships among the five family groups in Cornales, and (2) date the early diversification events and ages of major families using data from all seven regions and BEAST.

## 2. Material and methods

### 2.1. Taxon sampling

The general sampling strategy was to include representative taxa from all Cornelean genera and families with increased sampling from lineages exhibiting relatively long branches in previous phylogenetic analyses, e.g., Hydrostachyaceae, Alangiaceae, Nyssaceae, and Mastixiaceae. A total of 98 accessions (representing 79 species from Cornales, 17 from other Asterid lineages, and two from Rosids) were included in the analyses. Among the Cornales samples, there are 22 species of Cornaceae representing all the major clades of the family, four species of Alangiaceae, four species of Mastixiaceae from both genera, four species of Nyssaceae from two genera, the monotypic Davidiaceae and Curtisiaceae, one of the two species of Grubbiaceae, six of the 22 species of Hydrostachyaceae, all 17 genera of Hydrangeaceae with multiple species from the large genera (e.g., *Deutzia*, *Philadelphus*, and the polyphyletic *Hydrangea* - Soltis et al., 1995), and 10 species of five genera of Loasaceae, representing the three monophyletic subfamilies (Table 1). In most published phylogenetic analyses including Hydrostachyaceae, only a single species was included (Albach et al., 2001; Xiang, 1999; Xiang et al., 1998; Burleigh et al. 2009). We increased both the species and character sampling in the present study in an effort to break up particularly long branches and place the taxa more reliably. We also included extensive sampling of outgroup taxa (a total of 19 genera, at least one from each order of the Asterid clade, and two genera from the Rosid clade, *Peonia* and *Vitis*), to assess the phylogenetic affinity of Hydrostachyaceae to Cornales (Table 1). Outgroup species were selected based on the availability of DNA sequences from GenBank.

### 2.2. DNA sequence data

DNA sequences for some taxa and regions were available from previously published studies. In the present study, DNA sequences of the chloroplast *ndhF*, *atpB*, *trnL-F*, and *trnH-K* were generated for a majority of the Cornales taxa, and sequences of chloroplast *matK* and *rbcL* genes and 26S rDNA were generated for a portion of the Cornalean taxa and outgroups (see Table 1). Sequencing of *matK*, *rbcL*, and 26S rDNA was conducted following previous studies on *Cornus* and Cornales (Xiang et al., 1998; Fan and Xiang, 2001, 2003). Sequences of *ndhF* and *atpB* were generated with primers described in Olmstead et al. (2000) and Hoot et al. (1995). For the *trnL–trnF* region, primer sequences and PCR procedures followed Sang et al. (1997), and PCR primers used to amplify the *trnH–trnK* region were obtained from Demesure et al. (1995). Additional internal primers used to amplify and sequence the *trnH–trnK* region were either obtained from Modliszewski et al. (2006) or designed for this study (see Supplementary materials). Sequences of 26S rDNA were generated following Fan and Xiang (2003).

### 2.3. Phylogenetic analyses

The sequences of 26S rDNA, *rbcL*, *matK*, *ndhF*, and *atpB* were all easily aligned manually with MacClade 4.06 (Maddison and Maddison 2003) using the previous *rbcL–matK* and combined *rbcL–matK*–26S rDNA data matrices of Fan and Xiang (2003) as the reference. Sequences of *trnL–trnF* and *trnH–trnK* were initially aligned using Clustal X 1.81 (Thompson et al., 1997) and subsequently adjusted in MacClade. For some of the outgroup genera, DNA sequences of different regions were available for different species. For these genera, hybrid or synthetic sequences were assembled, i.e., combining the sequences from different species to have all or most of the DNA regions covered for that genus (Table 1).

Phylogenetic analyses were initially conducted for individual regions of the cpDNA data using maximum parsimony (MP). Because no strongly supported conflicting nodes were found among the trees from different regions, detailed analyses were conducted using MP, Bayesian Inference (BI), and Maximum Likelihood (ML) methods for the combined sequences of *rbcL–matK–ndhF–atpB–trnL-F–trnH-K* from the cpDNA genome (referred to as cpDNA data