Theor Appl Genet (2000) 101:925–930

S. Shi · H. Jin · Y. Zhong · X. He · Y. Huang · F. Tan
D.E. Boufford

Phylogenetic relationships of the Magnoliaceae inferred from cpDNA *mat*K sequences

Received: 2 January 2000 / Accepted: 12 February 2000

Abstract The coding region of the *mat*K gene was sequenced to infer the phylogeny of the family Magnoliaceae. Phylogenetic analyses of 21 *mat*K sequences representing ten genera of Magnoliaceae and three outgroups suggest relationships among both subfamilies and genera. Monophyly of the subfamily Liriodendroideae (the genus *Liriodendron*) and the subfamily Magnolioideae is strongly supported, respectively. Within the subfamily Magnolioideae, three clades are formed: (1) the genus *Magnlietia*, (2) the subgenus *Magnolia*, and (3) the subgenus *Yulania*, with the genera *Michelia*, *Paramichelia, Tsoongiodendron, Alcimandra*, *Kmeria*, *Parakmeria* and *Manglietiastrum*. However, the genus *Magnolia* is shown to be a polyphyletic group, and the genus *Michelia* a paraphyletic group. Relatively low sequence divergences are detected among genera of the the subfamily Magnolioideae, ranging from 0.14% to 1.70%, especially in the tribe Micheliinae (0.14–0.98%). Molecular evidence from *mat*K sequence data suggests that the phylogenetic positions and the delimitation of the eight genera *Magnolia, Michelia, Tsoongiodendron, Paramichelia, Alcimandra*, *Kmeria*, *Parakmeria* and *Manglietiastrum* need to be reconsidered.

Key words Magnoliaceae · Molecular phylogeny · Chloroplast DNA · *mat*K gene

Communicated by P.M.A. Tigerstedt

S. Shi (✉) · H. Jin · X. He · Y. Huang · F. Tan
School of Life Sciences, Zhongshan University,
Guangzhou 510275, Guangdong, China
e-mail: lssssh@zsu.edu.cn
Tel.: 86-20-84113677, Fax: 86-20-84036551

Y. Zhong
Institute of Biodiversity Science, School of Life Sciences,
Fudan University, Shanghai 200433,
and Wuhan Institute of Botany, Academia Sinica,
Wuhan 430074, China

D.E. Boufford
Harvard University Herbaria, 22 Divinity Avenue, Cambridge,
MA 02138-2020, USA

Introduction

The family Magnoliaceae *sensu stricto* comprises approximately 240 species of evergreen and deciduous trees and shrubs which are distributed widely in the tropical, subtropical and temperate zones of southern and eastern Asia, southeastern North America, and northern South America. In China most taxa of the family occur in tropical and subtropical forests. They have also been introduced and cultivated as ornamental plants (e.g. *Magnolia grandiflora*), timbers with relatively high quality (e.g. *Alcimandra cathcartii*), medicinal plants (e.g. a famous Chinese traditional medicinal material *Magnolia officinalis*), and natural resources of stacte, a sweet spice used in making incense, and flavor (e.g. *Magnolia cylindrica* and *Michelia hedyosperma*), etc.

In the latest classifications of the angiosperms, except for Engler's system, the family Magnoliaceae was placed in a primitive position (Takhtajan 1980; Cronquist 1981; Endress 1990). But there are uncertainties and controversies about the number of genera in the family, the delimitation of the genera, and the relationships among the genera because of a considerable overlap in characters (Dandy 1964; Law 1984; Chen and Nooteboom 1993). Recently, molecular phylogenetis has been challenging the classification of the Magnoliaceae based on morphological characters (Kim et al. 1998; Azuma et al. 1999; Jin et al. 1999).

Here we attempt to reconstruct the phylogeny of the Magnoliaceae with emphasis on intergeneric relationships by using the sequences of the *mat*K gene of chloroplast DNA (cpDNA). The single-copy *mat*K gene, consisting of approximately 1500 base pairs (bp) in length, is located within the intron of the transfer RNA gene for lysine (*trn*K) in the chloroplast; and 102 amino-acid loci at the carboxyl terminus are structurally related to portions of a maturase-like polypeptide and might be involved in splicing Group-II introns (Hilu and Liang 1997). The evolutionary rate of the *mat*K gene has been suggested to be intermediate between the cpDNA *rbc*L gene and the internal transcribed spacer (ITS) regions of nuclear ribosomal DNA (nrDNA); thus, the *mat*K sequences may be appropriate for phylogenetic studies at both inter- and intra-familial levels (Johnson and Soltis 1994). In addition, we employ two different tree-making methods, i.e. maximum parsimony (MP) and neighbor-joining (NJ), to analyze the same sequence data in order to obtain a topologically congruent *mat*K phylogeny of the Magnoliaceae.

Materials and methods

Plant material

Twenty one samples representing 10 of 15 genera of the Magnoliaceae and three species as outgroups were used in this study (Table 1). Leaf samples for DNA extraction were collected in the field or from cultivated plants in China.

DNA sequencing

Total genomic DNA was extracted from 2 g of fresh or 1 g of silica-gel dried leaves using the CTAB method (Doyle and Doyle 1987), then purified with glass powder. Two pairs of primers, *trn*K-3914F and *trn*K-2R (Johnson and Soltis 1994), and MG1 and MG15 (Hilu and Liang 1997), were used for the *mat*K gene amplification. A better effect of amplification was detected in the latter primer pair, MG1 and MG15. The PCR reactions include the following cycles: preheat at 94°C for 4 min followed by 25 cycles of a denaturation step at 94°C for 1 min, an annealing step at 50°C for 1 min, and an extension step at 72°C for 3 min, and then a final extension step of 72°C for 8 min. The amplified products were purified with the Millipore PCR Purification Kit (R7HW30603, MILLIPORE) to remove the redundant small molecular fragments of primers and dNTPs. Purified double-stranded DNAs were sequenced following the dideoxy chain-termination method (Sanger et al. 1977) using the Sequenase Version 2.0 DNA Sequencing Kit (27-1682-01, Amersham Pharmacia Biotech) and alpha ^{35}S-dATP as a radioactive tracer. Five primers, MS3R, MS4R, MS5R, *mat*K-1470R (Johnson and Soltis 1994), MS2F and MatK5, were used as sequencing primers (Table 2 and Fig. 1). MS5R displayed a better result of sequencing than MS3R, even with its similar location and sequence. Products of the sequencing reactions were separated on a 6% polyacrylamide gel at 50 W. The gel was dried in an oven at 50°C for 5–6 h after being transferred to 3 MM filter paper (3030-392. Whatman), and exposed to X-ray film for 3–5 days.

Table 2 Base composition of amplification and sequencing primers used in this study

Primer	Sequence
PCR:	
Trank-3914F	5′TGGGTTGCTAACTCAATGG3′
Trank-2R	5′AACTAG TCGGATGGAGTAG3′
MG15	5′ATCTGGGTTGCTAACTCAATG3′
MG1	5′CTACTGCAGAACTAGTCGGATGGAGTAGAT3′
Sequencing:	
MS3R	5′TA(T/C)TGAATGAATAGATCGTA3′
*mat*K-1470R	5′AAGATGTTGAT(T/C)GTAAATGA3′
MS2F	5′CTATATAATTCTCATGTAT3′

Data analysis

The *mat*K obtained sequences have been assigned GenBank accession numbers (Table 1). The assembled sequences were aligned using Clustal X (Thompson et al. 1997). Sequence divergences were estimated using a Kimura two-parameter distance (Kimura 1980).

Phylogenetic analyses were performed with PAUP* 4.0 (Swofford 1999) using the maximum parsimony (Swofford et al. 1996) and the neighbor-joining (Saitou and Nei 1987) methods. Parsimony analysis was performed using a branch-and-bound search.

Table 1 Accessions of the Magnoliaceae and the outgroups used in this study. SBG = South China Botanical Garden, Guangzhou, Guangdong, China; SYS = Campus of Zhongshan (Sunyatsen) University, Guangzhou, Guangdong, China; ZBG = Zhongshan Botanical Garden, Nanjin, Jiangsu, China. HNR = Heishiding Natural Reserve, Fengkai, Guangdong, China. The classification of the Mangoliaceae follows Law (1984)

Taxon	Voucher	Geographical origin	GenBank accession no.
Illicium henryi	J.P. 274(SBG)	Cultivated in SBG	AF123463
Mytilaria laosensis	S.Shi123(SYS)	Cultivated in SBG	AF128828
Altingia gracilipes	Hao 155(SYS)	Cultivated in ZBG	AF133223
Magnolia alboserícea	T.C. 01(SBG)	Cultivated in SBG	AF123464
Magnolia henryi	T.C. 04(SBG)	Cultivated in SBG	AF209199
Magnolia denudata	S.Shi 306(SYS)	Cultivated in SYS	AF123465
Michelia alba	T.C. 82(SYS)	Cultivated in SYS	AF123466
Michelia figo	T.C. 83(SBG)	Cultivated in SBG	AF123467
Michelia foveolata	S.Shi 147 (SYS)	HNR	AF123468
Paramichelia baillonii	J.P. 260 (SBG)	Cultivated in SBG	AF123469
Tsoongiodendron odorum	J.P. 259 (SBG)	Cultivated in SBG	AF123470
Alcimandra cathcartii	J.P. 271 (SBG)	Cultivated in SBG	AF123473
Kmeria septentrionalis ♀	J.P. 261(SBG)	Cultivated in SBG	AF123471
Kmeria septentrionalis ♂	J.P. 262(SBG)	Cultivated in SBG	AF123472
Parakmeria omeiensis	T.C. C02 (SBG)	Cultivated in SBG	AF123476
Parakmeria lotungensis	J.P. 263(SBG)	Cultivated in SBG	AF123475
Manglietiastrum sinicum	J.P. 279(SBG)	Cultivated in SBG	AF123479
Manglietia hainanensis	T.C. 74(SBG)	Cultivated in SBG	AF123478
Manglietia moto	S.Shi144(SYS)	HNR	AF123477
Liriodendron tulipifera	J.P. 272(SBG)	Cultivated in SBG	AF123480
Liriodendron chinense	J.P. 275(SBG)	Cultivated in SBG	AF123481

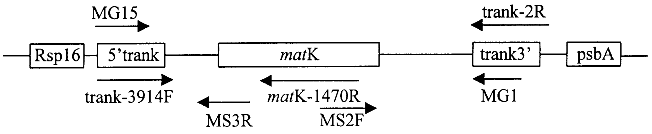

Fig. 1 Diagram showing locations of amplification by sequencing primers of *mat*K used in this study. *Arrowheads* indicate the direction of strand synthesis. *Boxes* represent coding regions

Bootstrap analyses were carried out with 1000 replicates using TBR branch-swapping of the heuristic search (Felsenstein 1985). A neighbor-joining tree was constructed using the Kimura two-parameter distance (Kimura 1980). All phylogenetic trees were rooted using *Illicium henryi* (Illiciaceae), *Mytilaria laosensis* and *Altingia gracilipes* (Hamamelidaceae) as outgroups.

Results

The *mat*K gene yielded a final sequence alignment of 1497 bases of the Magnoliaceae and outgroups. The matrix is available at *http://www.msu.edu/user/zhangxi3/program/magnolia.html.* The Kimura two-parameter distances are presented in Table 3. The *mat*K sequence divergences between all taxa of the currently recognized subfamily Magnolioideae are low (0.10–1.70%) in comparison to other angiosperm families. The sequence divergences between genera in the subtribe Micheliinae were estimated in an even-lower range (0.10–0.98%). In another subfamily Liriodendroideae, however, the sequence divergences between *Liriodendron* and other genera are relatively higher (2.19–2.65%) than that among those genera (the highest is 1.6%).

A strict consensus MP tree, resulting from 108 most-parsimonious trees (MPTs), and the NJ tree are presented in Fig. 2. Both in the MP tree and NJ tree, the ingroup taxa are divided into two main clades with strong bootstrap support. One clade is formed by two *Liriodendron* species, and another clade encompasses all the remaining taxa. Within the latter clade, three clades are formed in the MP tree. One of them can be further divided into three clades. In the NJ tree, these clades are formed with slight difference to the MP tree. The topologically congruent and well-supported clades between the two trees can provide phylogenetic information among the ten genera.

Discussion

Two currently recognized subfamilies of Magnoliaecae, i.e. the Liriodendroideae and the Magnolioideae, form two monophyletic groups with 100% bootstrap support. In particular, the Liriodendroideae, consisting of a single genus *Liriodendron* that is disjunctly distributed in Eastern Asia and eastern North America, is basal in the two trees.

Two tribes within the subfamily Magnolioideae, i.e. the tribe Michelieae with axillary flowers and the tribe Magnolieae with terminal flowers, were established by Law (1984). In his classification of the Michelieae, this tribe was further divided into two subtribes: the Elmerrillinae and the Micheliinae. The Elmerrillinae contains only the genus *Elmerrillia* while the Micheliinae contains three genera: *Michelia, Paramichelia*, and *Tsoongiodendron*. However, Figlar (1998) put *Michelia* in *Magnolia* (at the subgenus level) and *Elmerrillia* in the subgenus *Michelia*, because these taxa give a proleptic branch initiation. The *mat*K data do not support the monophyletic relationship between *Magnolia* and *Michelia* or the monophyly of *Michelia* species. On the other hand, the close relationships among *Michelia, Paramichelia, Alcimandra* and *Tsoongidendron* suggest that the phylogenetic positions of *Alcimandra* and *Tsoongidendron* should be reconsidered.

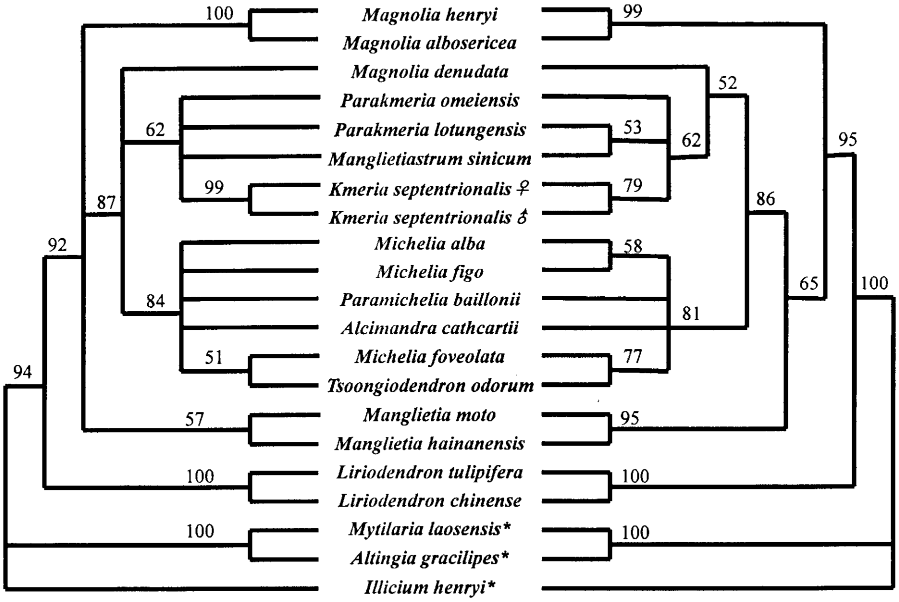

Fig. 2 The strict consensus phylogenetic trees of the Magnoliaceae generated using maximum-parsimony (MP) and neighbor-joining (NJ) methods. On the right is the NJ tree, and on the left the MP tree. The length of the MP tree is 99 steps, CI is 0.626, RI is 0.422, RC is 0.264, and HI is 0.374. Gaps were treated as missing data. The *numbers* represent the bootstrap supports (%) for the clades. An *asterisk* (*) indicates the outgroup used to root both trees

Table 3 Kimura two-parameter distances of taxa in the Magnoliaceae and outgroups. Gaps were eliminated from the comparison. 1 *Illicium henryi*, 2 *Mytilaria laosensis*, 3 *Altingia gracilipes*, 4 *Magnolia henryi*, 5 *Magnolia albosericea*, 6 *Magnolia denudata*, 7 *Michelia alba*, 8 *Michealia figo*, 9 *Michelia foveolata*, 10 *Paramichelia baillonii*, 11 *Tsoongiodendron odorum*, 12 *Kmeria septentrionalis* ♀, 13 *Kmeria septentrionalis* ♂, 14 *Alcimandra cathcartii*, 15 *Parakmeria lotungensis*, 16 *Parakmeria omeiensis*, 17 *Manglietia moto*, 18 *Manglietia hainanensis*, 19 *Manglietiastrum sinicum*, 20 *Liriodendron tulipifera*, 21 *Liriodendron chinense*

No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1	–																				
2	0.16508	–																			
3	0.16425	0.05504	–																		
4	0.08696	0.13310	0.14235	–																	
5	0.8867	0.13347	0.14267	0.00474	–																
6	0.08320	0.13203	0.13919	0.01154	0.01143	–															
7	0.08315	0.13401	0.14080	0.01223	0.01143	0.00538	–														
8	0.08206	0.13316	0.14034	0.01151	0.01072	0.00470	0.00067	–													
9	0.08399	0.13649	0.14372	0.01358	0.01287	0.00677	0.00272	0.00203	–												
10	0.08186	0.13246	0.14066	0.01335	0.01251	0.00486	0.00139	0.00139	0.00352	–											
11	0.08341	0.13573	0.14227	0.01498	0.01415	0.00673	0.00406	0.00337	0.00272	0.00350	–										
12	0.08867	0.13638	0.14490	0.01535	0.01373	0.00648	0.00650	0.00574	0.00798	0.00596	0.00792	–									
13	0.08295	0.l13389	0.14102	0.01489	0.01409	0.00671	0.00672	0.00603	0.00812	0.00625	0.00806	0.00000	–								
14	0.08392	0.13534	0.14189	0.01491	0.01409	0.00671	0.00404	0.00335	0.00541	0.00348	0.00539	0.00791	0.00804	–							
15	0.09897	0.13447	0.14613	0.01699	0.01526	0.00644	0.00728	0.00725	0.00977	0.00751	0.00963	0.00430	0.00800	0.00962	–						
16	0.08512	0.13278	0.14000	0.01221	0.01211	0.00605	0.00471	0.00403	0.00610	0.00558	0.00742	0.00575	0.00605	0.00739	0.00490	–					
17	0.08483	0.13214	0.14130	0.01355	0.01278	0.01075	0.01076	0.01006	0.01218	0.01044	0.01213	0.01226	0.01208	0.01208	0.01215	0.00875	–				
18	0.08426	0.13203	0.14130	0.01167	0.01088	0.01022	0.00886	0.00816	0.01029	0.00989	0.01163	0.01171	0.01156	0.001157	0.01313	0.00952	0.00203	–			
19	0.10023	0.13245	0.14146	0.01198	0.01365	0.00269	0.00457	0.00456	0.00650	0.00475	0.00634	0.00389	0.00360	0.00727	0.00305	0.00186	0.00923	0.01032	–		
20	0.07600	0.13700	0.14640	0.02194	0.02174	0.02311	0.02387	0.02308	0.02536	0.02322	0.02523	0.02421	0.02442	0.02511	0.02674	0.02451	0.02313	0.02268	0.02584	–	
21	0.07615	0.14249	0.15125	0.02222	0.02217	0.02440	0.02518	0.02434	0.02648	0.02457	0.02664	0.02486	0.02502	0.02579	0.02845	0.02580	0.02427	0.02369	0.02776	0.00072	–

Our study also presented three main differences between the *mat*K phylogeny and the traditional classification of the tribe Magnolieae. First, the genus *Magnolia* is not shown to be monophyletic. Law (1984) partitioned the genus into two subgenera, *Magnolia* and *Yulania*, according to their evergreen versus deciduous habits. Azuma et al. (1999) indicated the polyphyly of the subgenus *Magnolia*, based on the sequence data of the *trn*K intron, the *psb*A-*trn*H, and the *atp*B-*rbc*L intergenic spacer. However, Nooteboom (1998) merged *Michelia*, *Kmeria* and *Manglietia* into *Magnolia*. The *mat*K data indicate that *Magnolia denudata* (representing the subgenus *Yulania*) is closer to *Parakmeria*, *Manglietiastrum* and *Kmeria* than to *Magnolia henryi* and *M. albosericea* (representing the subgenus *Magnolia*). Second, Law (1984, 1997) not only insisted that *Kmeria* should be an independent genus, but also separated *K. septentrionalis* from this genus and founded a new genus, *Woonyoungia*. On the basis of its unisexual flowers with only 6–7 tepals and the fact that tripetala sometimes developed in male flowers, *Kmeria* was suggested to be derived from the genus *Magnolia* but not to be worthy of generic rank (Nooteboom 1998). Our study therefore sampled *K. septentrionalis* with male and female flowers, respectively, and detected very low sequence divergences among *Kmeria*, *Parakmeria* and *Manglietiastrum* (0.19–0.39%). The results do not support that one of the three should be recognized as a distinct genus. Third, Canright (1955) suggested that the genus *Manglietia* should be merged into the genus *Magnolia* based on many overlapping characters between them, while Chen et al. (1993) indicated that *Manglietia* could be distinguished from *Magnolia* clearly because it had more than four ovules in each carpel. Nooteboom (1998) merged *Manglietia* into *Magnolia* again, but put it in a separate section, based partially on molecular data. In our analyses of the *mat*K sequences, the two species of *Manglietia* form a monophyletic group but with relatively low bootstrap support (57%). This group and the subgenus *Magnolia*, as well as the clade consisting of the genus *Michelia*, the genus *Alcimandra*, the subgenus *Yulania*, the genera *Kmeria*, *Parakmeria* and *Manglietiastrum*, exhibit a parallel sister-group relationship. The results do not support that *Manglietia* is a distinct genus either, but suggest that *Manglietia* might be treated as a subgenus or a section within the genus *Magnolia*.

Since the Magnoliaceae is a taxonomically and phylogenetically complex group, this work on the chloroplast *mat*K DNA sequences has attempted to provide a new molecular phylogeny and taxonomic interpretation of the family, and found that further studies are needed, especially in estimating the delimitation of the small genera and the relationships among them. Two main tendencies have been shown recently in the study of the phylogeny of the Magnoliaceae. One is in classifying in finer detail at both generic and species levels (Law 1984, 1997). Another is a classification employing broader generic and specific concepts (Chen and Nooteboom 1993), where many genera and species are merged. Both tendencies are based on the above-mentioned similarities and overlaps in characters states and structures. It is necessary to obtain additional molecular data, including sequencing and analyzing more genes, such as the genes and gene families that evolve faster, to resolve the controversies about the phylogeny and relationships in the Magnoliaceae. At the same time, it is also very important to combine and synthesize the molecular and non-molecular data for a phylogenetic reconstruction of the Magnoliaceae.

Acknowledgements We thank Law Yuh-Wu, Wu Qigen, Xu Fengxia, Zeng Qiwen, Liu Yinzhi and Gong Xun for providing leaf material, and Miles Tracy for helpful comments on the manuscript, as well as Zhang Qun for technical assistance. This study was supported by the National Science Funds for Distinguished Young Scholars (39825104), the National Natural Science Foundation of China (39825104, 39970057), and the Natural Science Foundation of Guangdong Province (970190) to S.S.

References

Azuma H, Thien LB, KAwano S (1999) Molecular phylogeny of *Magnolia* (Magnoliaceae) inferred from cpDNA sequence and evolutionary divergence of the floral scents. J Plant Res 112: 291–306

Canright JE (1955) The comparative morphology and relationships of the Magnoliaceae. IV. Wood and nodal anatomy. J Arnold Arbor 36:119–140

Chen BL, Baas P, Wheeler EA, Wu S (1993) Wood anatomy of trees and shrubs from China. VI. Magnoliaceae. Iawa J 14: 391–412

Chen BL, Nooteboom HP (1993) Notes on Mangoliaceae III: The Magnoliaceae of China. Ann Missouri Bot Gard 80:999–1104

Cronquist A (1981) An integrated system of classification of flowering plants. Columbia University Press, New York

Dandy JE (1964) Magnoliaceae. In: Hutchinson J (ed) The genera of flowering plants, vol I. Clarendon Press, Oxford, pp 50–57

Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull 19:11–15

Endress PK (1990) Evolution of reproductive structures and functions in primitive angiosperms (Magnoliidae). Mem New York Bot Gard 55:5–34

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791

Figlar RB (1998) Proleptic branches initiation in *Michelia* and *Magnolia* subgenus *Yulania* provides a basis for combinations in subfamily Magnolioideae. International Symposium on the Family Magnoliaceae (abstracts)

Hilu KW, Liang H (1997) The *mat*K gene: sequence variation and application in plant systematics. Am J Bot 84:830–839

Jin H, Shi S, Pan H, Huang Y, Zhang H (1999) Phylogenetic relationships between *Michelia* (Magnoliaceae) and its related genera based on the *mat*K gene sequences. Acta Sci Nat Univ Sunyatseni 38:93–97

Johnson LA, Soltis DE (1994) *mat*K DNA sequences and phylogenetic reconstruction in the Saxifragaceae *sensu stricto*. Systematic Bot 19:143–156

Kim S, Park C-W, Suh Y (1998) Molecular phylogeny of the Magnoliaceae based on *ndh*F sequences. J Plant Biol 17:135–142

Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions. J Mol Evol 16:111–120

Law Y-W (1984) A preliminary study on the taxonomy of the family Magnoliaceae. Acta Phytotaxon Sinica 22:89–109

Law Y-W (1997) *Woonyoungia*, a new genus of Magnoliaceae from China. Bull Bot Res 17:353–356

Nooteboom HP (1998) Different looks at the classification of the Magnoliaceae. Int Symp on the Family Magnoliaceae (abstracts)

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74:5463–5467

Swofford DL (1999) PAUP* 4.0. Phylogenetic analysis using parsimony (and other methods). Sinauer Associates, Sunderland, Massachusetts

Swofford DL, Olsen GJ, Waddell PJ, Hillis DM (1996) Phylogenetic inference. In: Hillis D, Moritz C, Mable BK (eds) Molecular systematics (2nd edn). Sinauer Associates, Sunderland, Massachusetts, pp 407–514

Takhtajan A (1980) Outline of the classification of flowering plants (Magnoliophyta). Bot Rev 46:225–359

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The Clustal_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 25:4876–4882