Taiwania, **49**(4): 273-287, 2004

The Evolution of Chloroplast *matK* Genes, Including Identification of New Homologues from *Ophioglossum petiolatum* and Two Lycophytes

Shu-Lan Chuang[1] and Jer-Ming Hu[1, 2]

(Manuscript received 27 October, 2004; accepted 24 November, 2004)

ABSTRACT: The introns of chloroplast *trnK*^{UUU} belong to Group II introns and contain an open reading frame denoted as *matK*. The *trnK*$^{5'}$-*matK*-*trnK*$^{3'}$ structure is consistent in almost all examined higher land plants and in Characeae, but not in other green algae examined. The putative gene product MatK is the only maturase in chloroplasts. Functional chloroplast *matK* genes are retained even in the nonphotosynthetic parasite, *Epifagus virginiana* and the fern, *Adiantum capillus-veneris*, in which chloroplast genome rearrangement has left *matK* free-standing, apart from *trnK* exons. Among lower land plants, the chloroplasts of *Psilotum*, mosses and liverworts all have *trnK*$^{5'}$-*matK*-*trnK*$^{3'}$ structure, but *matK* is a pseudogene in hornwort *Anthoceros formosae*. In this study we found a clear *trnK*$^{5'}$-*matK*-*trnK*$^{3'}$ structure in *Ophioglossum petiolatum*, *Lycopodiella cernua* and *Selaginella doederleinii*, but PCR with degenerate primers failed to amplify any *trnK* or *matK* fragments from other ferns and fern allies. However, dot blot hybridization showed distinct signals in these plants that failed to amplify *matK* fragments by PCR, indicating that the *matK* sequences in those taxa may be too divergent to amplify by an ordinary PCR approach. RT-PCR results showed *matK* genes are expressed in *Ophioglossum petiolatum* and *Lycopodiella cernua*, but no signal was detected in *Selaginella doederleinii*. Overall, the expression patterns of *matK* are not consistent in lower land plants. Phylogenetic analysis of *matK* sequence showed that *Pinus*, *Ginkgo*, and *Cycas* form a monophyletic group, which is sister to angiosperms. Together, they form a clade that is sister to Gnetales. This *ad hoc* reconstruction is likely due to the high evolutionary rate in *matK*.

KEY WORDS: Chloroplast *matK*, *Lycopodiella cernua*, *Selaginella doederleinii*, *Ophioglossum petiolatum*, Evolution.

INTRODUCTION

Over forty plastid genomes, including more than twenty land plants, have been completely sequenced and are available in GenBank, thus providing fruitful information on gene structures of plastid genomes. Although introns are not common in organelle genomes of land plants, at least 18 plastid genes have been found to harbor introns (Odintsova and Yurina, 2003). All but one of the chloroplast introns belong to Group II or III subclass introns with specific RNA secondary structure (Michel *et al.*, 1989; Sugiura, 1992). The only exception is the intron of tRNA^{Leu} gene, a group I intron that seems to have an ancient origin dated back to cyanobacteria (Kuhsel *et al.*, 1990; Besendahl *et al.*, 2002). Chloroplast *matK*, which is encoded by the *trnK* intron, is commonly present in land plants and is the only maturase-like gene in plant plastids. The gene was first characterized in tobacco (Sugita *et al.,* 1985), and further designated as *matK* in mustard (*Sinapis alba*) based on its similar structure and composition to mitochondrial Group II intron-encoded maturases in yeast (Neuhaus and Link, 1987).

1. Institute of Ecology and Evolutionary Biology, National Taiwan University, No.1, Sec. 4, Roosevelt Road, Taipei 106, Taiwan.
2. Corresponding author. Email: jmhu@ntu.edu.tw

Group II introns are known for their self-splicing ability under certain conditions (Henke *et al.*, 1995), however, the splicing is usually facilitated by other nuclear-encoded proteins (Schmelzer *et al.*, 1983) or by their own intron-encoded ORF (Guo *et al.*, 1997), which shows a reverse transcriptase activity. The putative protein encoded by this ORF contains a highly conserved maturase domain (X-domain) near the C-terminal region. This domain is about 500 amino acids in length, and exists in all the Group II intron genes with ORF responsible for the maturase activity of chloroplasts and mitochondria, or in a freestanding form (Mohr *et al.*, 1993). This X-domain has a strongly conserved sequence SX_{3}-$_{6}TLAXKXK$, and most of the sequences have a large excess of basic over acidic amino acids (Mohr *et al.*, 1993) and are mostly hydrophilic (Sugita *et al.*, 1985).

The putative maturase function of chloroplast MatK protein is mostly based on sequence comparisons and sparse data for the presence of a spliced form RNA in rice (Chiba *et al.*, 1996), the detection of proteins in *Solanum* (du Jardin *et al.*, 1994), and RNA-binding activity in *Sinapis* (Liere and Link, 1995). Nonetheless, *matK* gene is indispensable since it is intact and free-standing in the highly reduced plastid genome of the parasitic *Epifagus virginiana* (Wolfe *et al.*, 1992; Mohr *et al.*, 1993; Ems *et al.*, 1995). Interestingly, six Group II introns are left in the remaining 21 likely functional genes in *Epifagus*, and these might be the substrates for *matK* gene (Ems *et al.*, 1995; Wolfe *et al.*, 1992).

These X-domain containing genes in eukaryotes, however, vary in gene structure. There is a reverse transcriptase (RT)-like domain present at N-terminal end in most of the X-domain genes of mitochondria (e.g. *cox1* in *Marchantia*), and sometimes a Zn^{2+}-finger-like (Zn) region at the C-terminus end (e.g. *cox*1aI2 in yeast) (Mohr *et al.*, 1993). Both the Group II intron-encoded ORFs in cyanobacterium *Calothrix* (Mohr *et al.*, 1993) and *Lactococcus lactis* (Matsuura *et al.*, 1997) have the RT-X-Zn motifs. It is reasonable to suggest that this might be the most complex and complete structure of the X-domain genes. Mohr *et al.* (1993) demonstrated that the region upstream of the X-domain in MatK protein has some similarity to conserved blocks V, VI, and VII of the RT-domains of other Group II intron ORFs. The sequenced plastid *matK* genes from GenBank all lack the necessary RT domain I-IV and Zn domain found in other X-domain containing genes, indicating that plastid *matK* may have lost the RT function.

To date, the chloroplast *trnK* ^{Lys}(UUU) intron-encoded *matK* has only been found in Charophytes (except for *Mesostigma* and *Chlorokybus*) and land plants (Ems *et al.*, 1995; Wakasugi *et al.*, 1997; Turmel *et al.*, 2002; Sanders *et al.*, 2003). The intron and *matK* are absent in all other green algae, *Euglena*, and the plastid precursor cyanobacteria (Kotani and Tabata, 1998; Sanders *et al.*, 2003). *Mesostigma* and *Chlorokybus*, two algae that are usually placed in Charophytes, do not have intron in their chloroplast *trnK* genes. However, they have dubious phylogenetics positions among green algae based on recent analyses, therefore may not belong to the Charophyte clade (Bhattacharya *et al.*, 1998; Turmel *et al.*, 2002). These results suggest that chloroplast *trnK* intron+*matK* might have been recruited by the ancestors of Charophytes *s. l.* and the land plants, from a mobile Group II intron, though the actual source is unknown.

Interestingly, the chloroplast *matK* has been found to be a pseudogene in hornwort *Anthoceros formosae* (Kugita *et al.*, 2003), and in some bryophytes like *Porella* and *Plagiomnium* (Jankowiak *et al.*, 2004). Other bryophytes such as *Marchantia polymorpha* (Shimada and Sugiura, 1991), *Physcomitrella patens* (Sugiura *et al.,* 2003), and *Sphagnum* (Jankowiak *et al.*, 2004), all harbor an intact *matK* ORF. In comparison, *matK* is free standing without franking *trnK* exons in *Adiantum capillus-veneris* (Wolf *et al.*, 2003). The universal

presence and/or functional studies of MatK among other ferns and fern allies are largely unexplored.

Despite the uncertain function of *matK*, the conserved DNA sequences provide a useful marker for phylogenetic analyses at familial to generic levels as demonstrated in many studies (Steele and Vilgalys, 1994; Hilu and Liang, 1997; Hu *et al.,* 2000; Hilu *et al.*, 2003). In this study, *matK*-like sequences were identified and characterized from selected land plants, focusing on lycophytes, eusporangiophytes, and true ferns.

MATERIALS AND METHODS

Plant materials and DNA isolation

Plant materials were all collected from Taiwan; for voucher information see Table 1. Genomic DNAs were extracted based on a modified CTAB method of Porebski *et al.* (1997), to reduce the effects of polysaccharide and polyphenols.

Table 1. Voucher information of plant materials used in PCR-screening and dot blot analysis. The right end of the table shows the PCR results. Check marks indicate the presence of PCR products of the appropriate size.

Family	Species	Voucher information	18S	*rbcL*	*matK*
Bryophytes					
Anthoceroceae	*Anthoceros formosae* Steph.	CSL009, Taipei	√	√	
Marchantiaceae	*Marchantia polymorpha* L.	CSL021, Taipei (cult.)	√	√	√
Lycophytina					
Lycopodiaceae	*Lycopodiella cernua* (L.) Pic.Serm.	CSL013, Taipei Co.	√	√	√
	Lycopodium pseudoclavatum Ching	CSL028, Taichung Co.		√	
Selaginellaceae	*Selaginella doederleinii* Hieron.	CSL012, Taipei Co.	√	√	√
	Selaginella delicatula (Desv. *ex* Poir) Alston	CSL011, Taipei Co.		√	
	Selaginella tamariscina (P. Beauv.) Spring	CSL002, Nantou Co.	√	√	
	Selaginella involuens (Sw.) Spring	CSL030, Taipei Co.	√	√	
	Selaginella stauntoniana Spring	CSL029, Hualien Co.			
Isoetaceae	*Isoetes taiwanensis* DeVol	CSL003, Taipei (cult.)	√	√	
Euphyllophytina					
Equisetaceae	*Equisetum ramosissimum* Desf. subsp. *debile* (Roxb. *ex* Vaucher) Hauke	CSL001, Pingtung Co.	√	√	
Ophioglossaceae	*Ophioglossum petiolatum* Hook.	CSL005, Taipei (cult.)	√	√	√
Marattiaceae	*Angiopteris palmiformis* (Cav.) C. Chr.	CSL004, Taipei (cult.)	√	√	
Osmundaceae	*Osmunda banksiifolia* (Presl) Kuhn	CSL024, Taipei Co.	√	√	
Schizaeaceae	*Lygodium japonicum* (Thunb.) Sw.	CSL028, Taipei	√	√	
Gleicheniaceae	*Dicranopteris linearis* (Burm. f.) Underw.	CSL032, Taipei	√		
Petridaceae	*Adiantum capillus-veneris* L.	CSL035, Taipei	√		
Lindsaeaceae	*Sphenomeris biflora* (Kaulf.) Tagawa	CSL039, Taipei (cult.)	√		

Genomic PCR amplification and sequence analysis

Primers used to amplify chloroplast *trnK/matK* region and *rbcL* genes are listed in Table 2. These include some species-specific primers used for RT-PCR reaction. PCR was performed with a T-Gradient thermocycler (Biometra, Goettingen, UK). PCR reactions contained 0.5 µL Advantage™2 Taq polymerase (BD Biosciences Clontech, Palo Alto, CA, USA), 5 µL buffer, 4 µL dNTPs (2.5 mM each), 2 µL each primer (10 mM), 20-200 ng genomic DNA, and

distilled water to 50 μL. The PCR program started with 5 min of 94°C incubation, followed by 35 cycles of 30 sec at 94°C denaturing, 90 sec at 54-60°C annealing, and 120 sec at 72°C extension. The reaction was finished with 5 min of 72°C incubation and stopped at 4°C. Amplified products were purified by QIAquick PCR purification kit (QIAGEN, GmbH, Germany) and cloned into pGEM-T Easy vector (Promega, Madison, WI, USA). Nucleotide sequences were determined by automated sequencer ABI PRISM 337 (Applied Biosystems), and assembled by Sequencher 4.0 (Gene Codes Corp., Ann Arbor, MI, USA).

Table 2. Primers used for genomic sequence amplification and RT-PCR.

Primer name	Sequence (5' to 3')	Taxon	Direction
trnK1L	CTCAATGGTAGAGTACTCG	Universal primer	Forward
trnK2R	AACTAGTCGGATGGAGTAG	Universal primer	Reverse
ophio_matK317F	CAGAGATTCTTATTCGACTTCTCG	*Ophioglossum petiolatum*	Forward
ophio_matK872R	GCTTTAATTCCAACCATTTCAGATA	*Ophioglossum petiolatum*	Reverse
Lyco_matK348F	CGTAATCCTAGTTCTGCAAATTGTT	*Lycopodium cernua*	Forward
Lyco_matK935R	ATTAAAAATTTAGTCCCTCCCACAG	*Lycopodium cernua*	Reverse
Sela_d_matK456F	ACCCCAATCTCTTCATCCAG	*Selaginella doederleinii*	Forward
Sela_d_matK1049R	TCCATCTTGGCTTGAACCTT	*Selaginella doederleinii*	Reverse
rbcL35F	GATTCAAGGCTGGCGTTAAAGAT	Universal	Forward
rbcL700R	GCGAATTCTGCCCTTTTCATCAT	Universal	Reverse
rbcL50R	AACACCAGCTTTRAATCCAA	Universal	Reverse
atpBR	ACATCKARTACKGGACCAATAA	Universal	Forward
trnLc	CGAAATCGGTAGACGCTACG	Universal	Forward
trnLd	GGGGATAGAGGGACTTGAAC	Universal	Reverse

Genomic dot blot hybridization

For each sample, 0.1 micrograms of genomic DNA (1 μg/μL) was denatured by adding half volume of 2 M NaCl and half volume of 1 M NaOH for 10 min at room temperature. The DNAs were then dotted on a pre-soaked nylon membrane (NEN™, Boston, USA). After washing with 50 ml 1 N NaCl and 50 ml 1 M Tris-HCl for 5 mins, the membrane was crosslinked with 120,000 μjoule using Spectrolinker SL-1000 (Spectronics, Westbury, New York, USA). Hybridization procedure basically followed Sambrook *et al.* (2001). Hybridization was carried out in a buffer containing 5X SSC, 0.1% n-Lauroylsarcosine, 0.1% SDS, and 1% blocking reagent (Roche, Indianapolis, IN, USA) at 40-45°C. The PCR fragments amplified by species-specific primers from five obtained clones were used as probes labeled by DIG-11dUTP (Roche, Indianapolis, IN, USA). Four of the probes were from *matK* homologues: *Lycopodiella matK*, *Ophioglossum matK*, *Selaginella matK*, and *Adiantum matK*. One *rbcL* probe (*Ophioglossum rbcL*) was used as a positive control. Other hybridization procedures were performed as suggested by manufacture (Roche, Indianapolis, IN, USA).

RNA isolation and reverse-transcription PCR assay

Total RNA from young vegetative tissues were isolated using the Pine Tree Method (Chang *et al.,* 1993). RNAs were treated with RNase-free DNase (2 units/μg RNA) (Promega, Madison, WI, USA) at 37°C for 30 min. First strand cDNA was synthesized by Superscript™ II RNaseH$^-$ reverse transcriptase system (Invitrogen, life technologies, Carlsbad, CA, USA). The reverse primers used to synthesize first strand DNA are listed in Table 2, and the forward primers were used in secondary PCR reaction to amplify specific products. PCR conditions were the same as previously described.

Phylogenetic analysis of *matK*

Forty nucleotide sequences of *matK* coding regions, including three from this study, were compiled into a data matrix aligned by ClustalX (Thompson *et al.*, 1997). The plant information and their accession numbers are listed in Table 3. Neighbor-joining (NJ) and maximum parsimony (MP) analyses were performed with PAUP* 4.0b10 (Swofford, 2002). For all analyses, gaps were treated as missing data, and no sites containing insertion/deletions were excluded. Neighbor-joining (NJ) analysis was conducted employing an HKY85 model (Hasegawa *et al.*, 1985) to estimate the distances between sequences. Parsimony search options invoked 100 random addition sequences, tree bisection-reconnection branch-

Table 3. Sequence information used in this study. Accession numbers refer to the NCBI GenBank database. Asterisk marks indicate new sequences obtained in this study.

	Family	Genus and species	Accession
Green algae	Chaetosphaeridiaceae	*Chaetosphaeridium globosum*	NC_004115
	Characeae	*Chara connivens*	AY170442
	Characeae	*Tolypella prolifera*	AY170451
	Characeae	*Lychnothamnus barbatus*	AY170448
	Characeae	*Nitellopsis obtusa*	AY170447
Bryophytes	Funariaceae	*Physcomitrella patens* (Hedw.) Bruch & Schimp.	NC_005087
	Marchantiaceae	*Marchantia polymorpha* L.	NC_001319
	Mniaceae	*Plagiomnium insigne* (Mitt.) T. J. Kop.	AY522574
	Porellaceae	*Porella platyphylla* (L.) Pfeiff.	AY168655
	Sphagnaceae	*Sphagnum inudatum* Russow	AY342156
Lycopodiophytes	Lycopodiaceae	*Lycopodiella cernua* (L.) Pic.Serm.*	AY826399
	Selaginellaceae	*Selaginella doederleinii* Hieron.*	AY826400
Ferns	Adiantaceae	*Adiantum capillus-veneris* L.	NC_004766
	Ophioglossaceae	*Ophioglossum petiolatum* Hook.*	AY826401
	Psilotaceae	*Psilotum nudum* (L.) P. Beauv.	NC_003386
Cycads	Cycadaceae	*Cycas taitungensis* C. F. Shen *et al.*	AF279795
	Cycadaceae	*Zamia floridana* A. DC.	AF279804
Ginkgo	Ginkgoaceae	*Ginkgo biloba* L.	AF543736
Gnetophytes	Ephedraceae	*Ephedra sinica* Stapf	AF279805
	Gnetaceae	*Gnetum africanum* Welw.	AY449631
	Welwitschiaceae	*Welwitschia mirabilis* Hook. f.	AF280996
Conifers	Araucariaceae	*Agathis borneensis* Warb.	AB023975
	Pinaceae	*Pinus thunbergii* Parl.	D17510
	Taxaceae	*Amentotaxus argotaenia* (Hance) Pilg.	AF152219
	Taxodiaceae	*Taxodium distichum* (L.) Rich.	AF152212
Basal angiosperms	Amborellaceae	*Amborella trichopoda* Baill.	AF543721
	Annonaceae	*Anaxagorea acuminata* (Dunal) A. DC.	AY220436
	Cabombaceae	*Brasenia schreberi* J. F. Gmelin	AF092973
	Calycanthaceae	*Calycanthus fertilis* var. *ferax* (Michx.) Rehder	AJ428413
	Magnoliaceae	*Magnolia henryi* Dunn	AF209199
	Nymphaeaceae	*Nymphaea odorata* Aiton	AF092988
	Piperaceae	*Piper crocatum* Ruiz & Pav.	AF543745
Monocots	Alismataceae	*Alisma canaliculatum* A. Braun & Bouche	AB040179
	Arecaceae	*Nypa fruticans* Wurmb	AF543743
	Asparagaceae	*Asparagus cochinchinensis* (Lour.) Merr.	AB029804
Eudicots	Berberidaceae	*Mahonia japonica* (Thunb. *ex* Murr.) DC.	AB038184
	Cactaceae	*Lepismium cruciforme* (Vell.) Miq.	AY015344
	Ranunculaceae	*Hepatica nobilis* var. *japonica* Nakai	AB110532
	Saxifragaceae	*Saxifraga integrifolia* Hook.	L20131
	Trochodendraceae	*Trochodendron aralioides* Siebold & Zucc.	AF543751

swapping, and retention of multiple parsimonious trees. The internal support was evaluated by bootstrap analyses (Felsenstein, 1985) and decay indices (Bremer, 1988, 1994). In the parsimony analysis, each of 1,000 bootstrap replicates was analyzed with the heuristic search option invoking one random addition replicate each, and not invoking the retention of multiple parsimonious trees. Decay indices (Bremer support) were calculated by incorporating AutoDecay (Eriksson, 1998) and PAUP* 4.0b10 (Swofford, 2002), which quantify the extra length needed to collapse a branch in the consensus of near-most-parsimonious trees (Bremer, 1988, 1994). Five green algal sequences were used as outgroups in the analyses.

RESULTS

PCR amplification of nuclear 18S, chloroplast *rbcL* and *matK*

All but five taxa (*Lycopodium pseudoclavatum*, *Selaginella stauntoniana*, and *Selaginella delicantula*, and the two bryophytes) examined yielded PCR products of nuclear 18S fragments. We have sequenced six of them to confirm the identity of these products, including *Lycopodiella cernua*, *Selaginella doederleinii*, *Ophioglossum petiolatum*, *Isoetes taiwanensis*, *Equisetum ramosissimum* ssp. *debile*, and *Selaginella tamariscina*. Partial *rbcL* PCR products were successfully amplified from all examined taxa except for *S. stauntoniana*, and the two bryophytes. Two of them (*S. doederleinii* and *O. petiolatum*) were sequenced to confirm their identity. Only five of the eighteen taxa examined yielded PCR products amplified by the trnK1L/trnK2R pair. After cloning and sequencing, three of them show high similarities to chloroplast *matK* sequences in GenBank. Sequences of *trnK*(5')-*matK*-*trnK*(3') were then identified from *L. cernua*, *S. doederleinii*, and *O. petiolatum* (information see Table 4). A summary of the PCR result is shown in the right end of Table 1. Amino acid alignment of the X domain of *matK* from *L. cernua*, *S. doederleinii*, *O. petiolatum*, and selected taxa is shown in Fig. 1.

Table 4. Sequence information of chloroplast *trnK/matK* obtained in this study.

Taxa	*trnK* 5' intron	*matK*	*trnK* 3' intron	Total
Lycopodiella cernua	727 bp	1554 bp	197 bp	2478 bp
Selaginella doederleinii	697 bp	1521 bp	189 bp	2407 bp
Ophioglossum petiolatum	853 bp	1344 bp	125 bp	2322 bp

Genomic dot blot hybridization

The result of dot blot hybridization is shown in Fig. 2. Similar results are found among different *matK* and *rbcL* probes. All samples have hybridization signals except for *Anthoceros formosae* and *Dicranopteris linearis*, where signals are very weak or invisible using all four probes. All other samples show distinct *matK* hybridization signals.

Reverse-transcription PCR assay

For the three taxa with new *trnK/matK* sequences, RT-PCR was performed by taxon-specific primers; results are shown in Fig. 3. In *Lycopodiella cernua* and *Ophioglossum petiolatum*, both *rbcL* and *matK* yielded PCR products from RT-PCR (Fig. 3A). Since no intron fragment was amplified on RT-PCR, this suggests that these two genes are indeed expressed in both species. In comparison, only *rbcL*, but not *matK*, is expressed in *Selaginella doederleinii* (Fig. 3B). We used two different intron regions as control and repeated the reactions; all of them show the same pattern.

```
Lychnothamnus         RISTKMPVSNLIHSLSLVDLCNIQGYPIHKATWSVLNDEKIINIFYKLWKNILLYYSGCSNRRDLGKIQYILECSCMKTLAFKHKSSITSTWKKYNKYLSF
Nitellopsis           RISTKMPVFNLIHSLSLVHLCNIEGYPIHKAKWSVLNDEKIINIFSQLWKNILLYYSGCSNRRDLGKIQYILEFSCMKTLAFKHKSSIRSTWKQYNKYLSF
Chara                 RISTKMPVFNLIHRLSVMHLCNLEGYPIHKAAWSVFNDKQIMNIFSNLLRNILLYYSGCSNRSDLGKIQYILEFSCMKTLAFKHKSSIRSTWTQYKKHVSF
Tolypella             KISTEIPVHSLISNLTTIKLCNRKGYPIHKASWSTFSDKDIINIYHKFWNELSLYYCGSSNRFDLSQIQYIFEFSCIKTLAFKHKSNIRLTWEQYKEYVSF
Chaetosphaeridium     KICINVPIKLLIIFLSKNGFCDISGNSKSKLSWSVLQDIEIIEKFRRLWLTISGYYSGSSNKYCLKIVLYILRYSCAKTLACKHKMSLKKIWKKYTLNLSV
Sphagnum              ELCSITPILSLIGLLAREGFCDALGHPISKLAWSTLTDEAIFNRFDQIWRNLFCYYSGCQNRKNLYQVQYILRFSCAKTLACKHKNTIRSVWKKYDLKFLT
Porella               EFCGIIPIVPLIILLARERFCDTSGRPICKLSWTTLADNEIFKQFDQITKNIFRYYSGCIKKKGLYQLQYILRFSCAKTLACKHKSTIRTVWKRYGSNFVT
Marchantia            EFCSIIPVIPLIRLLAKEKFCDVLGRPLCKLSWTTLSDNEIFERFDQIIKHIFSYYSGCINKKGLYQLQYIFRFSCAKTLACKHKSTIRTVWKKYGSNLLT
Plagiomnium           ELYSITPISSLIELLAKEKFCDILGHPISKLAWSTLTDDEIFNRFDQIWKNFFYYYSGCKSKKNLYQVQYILRFSCAKTLACKHKSTIRYVWKKHGSNFFA
Physcomitrella        EIYSITPISSLIELLAKENFCDTLGHPISKLAWTTLTDDEIFNRFDQIWRNFFYYYSGCKNKKNLYQVQYILRFSCAKTLACKHKSTIRYVWKKYGSNFFA
Adiantum              KFYPKIPNSIIITTLAKQRFCDFTGRPIGKSAWVTSTDDKIIDGYVQLWQVFSLYYGASMNQYRLRRLIFLLQMSCDSTLAGKHRSTIRLLRCKSNVEALN
Ophioglossum          VLCPKVPTSLSIRSLAREGFCNGLGFPISRSAWATSTDTDTTNRFNRLWKNLFIYYSGSSGLGGLYRIRYILRFSCAKTLACKHKSTIRAVWKRFGSRFNL
Psilotum              EFCASIPTSSLIESLTREGFCDSSGRPVGRSTWTILKDDDILNKYHQIWGDLSCYYSGSFSRDGLWRAKYILQLSCAKTLAQKHKSTTRVVRNHFGLKFIT
Selaginella           ELCPIIPFLLLVNSLARGGFCTNLGRPVSKLSWTTLTDDDILKKFDQIWRSVYYYYSGSINNHGLFRLRYIFRFSCAKTLACKHKSTTRIVWKRFSLNSFL
Lycopodiella          ELCVIIPVFRLIQLLTKEKFCNTSGRPISKSAWTTFQDDDILNQFNHIWKNLFYYYSGCLNRSDLYQIQYILRFSCAKTLACKHKSTIRVVWKKYGSRLFP
Gnetum                ELSPQIQVISMIEFFSIEGFCDITGKPISKLSWIRFTDDSIFDRYDRSWKFLYYYYSGVINKGSLDRVKYILLFSCFKTLALKHKSTIRVVRKEFDVKLFN
Welwitschia           EFHPKFGIISIMKFLSIEGFCDIMGRPISKLSWTCFTDDDIFDKCDRFWKILYYYYCGAKNKAYLDRIKYILLLSCFKTIAFKHKSTIRVVRKEFDFELRK
Ephedra               ELNSKLSAVFVIQFLSKEGLCDIMGNPKSKLAWLSFTDNSILDKYDHFCRNVDSFYSEAINKRFLDRVKDILFLSCIKTLACKHKSTIRIVRKELGFELRK
Amentotaxus           ELNPIAPIRSILFFLAKERFCDISGQTISKLSWTSLSDDDILDRFDRICRNLFHYYSGSINPDGLYYIKYILLLPCAKTLACKHKSTIRVVREESGSELFT
Taxodium              ELNPIAPIRSILFFLAKEKFCDISGWPISKLSWASLSDDDILDRFDRIWINLFHYYSGSINQDGLYHIKYILLLSCAKTLACKHKSTIRVVREQLGSELFT
Agathis               ELDPIAPIRSIIGLLAKERFCDISGRPICKLAWTSLSDDDILDRFDRICRNLFHYYSGSFNQDGLYSIKYILLLSCAKTLACKHKSTIRVVREELGSELFT
Pinus                 EMDPIVPIVPIIGLLATEKFCDISGRPISKLSWTSLTDDDILDRFDQIWRNLFHYYSGSFDRDGLYRIKYILLLSCAKTLACKHKSTIRVVRKELGPELFK
Zamia                 EFDPIAPTTLLIGSLAKEKFCDISGHPISRLAWTGLTDDDILDRFDRIWRNIFHYYSGSSKKDGLYRMKYILRLPCAKTLACKHKSAIRVVRERFGSELFT
Cycas                 EFDPIAPTTLLIGYLAKERFCDISGRPTGRLAWTGLTDDNILHRFDRIWRNILHYYSGSSKKDGLYRMKYILRLPCAKTLACKHKSAIRVVRERFGSELFT
Ginkgo                EFDSIAPIIPLIGLLAKERFCDISGRPISKLAWTGLKDDDILDRFDRICRNIIDYYSGSFNKDGSYRMKYILRLPCAKTLACRHKSTIRVVWEEFGSELFT
Nymphaea              RFDTIVPIFPLIGSLVKAKFCNVSGYPISKSVWADSSDSDIIARFGWICRNLSHYHSGSSKKHSLCRIKYILRLSCARTLARKHKSTVRAICKRLGSKLLE
Brasenia              RFDTIVPIFPLIGSLVKAKFCNVSGHPTSKSVWADLSDSDIIARFGWICRNLSHYHSGSSKKHSLCRIKYILRLSCARTLARKHKSTVRAICKRLGSKLLE
Amborella             RFDTVVPTIFLIGSLAKVKLCNVSGHPISKSVWADSSDSDILDQFGRICRNLSHYHSGSYKKHSLCRIKYILRLSCARTLARKHKSTVRAILKRLGSEFLD
Anaxagorea            KFETLVPIIPLIGSLAKAKFCDVSG?PISKSARADSSNSDIINRFGRIYRNISHYHSGSSKKQTLYRIKYILRLSCARTLARKHKSTVRAFLKRLGSKFLE
Magnolia              KFETLVPIIPLIGSVAKAKFCNVSGHPISKSVRADSSDSDIINRFGRIYRNLSHYHSGSSKKQTLYRIKYILRLSCARTLARKHKSTVRAFLKRLGSEFLE
Calycanthus           KFETIVPIIPLIGSLAKAKFCNGSGHPISKPFRTDLSDSEIINRFGRICKNLSHYHSGSSKKQSLYRIKFILRLSCARTLSRKHKSTVRAFLKRLGSELLE
Piper                 KFETIVPIISLIDSLSKEKFCNLSGHPTSKAIWSDLSDSDIMERFGRVCRNLSHYYSGCSKKQILYRIKYILRLSCARTLARKHKSTVRTFLKKLGSGFLK
Hepatica              KFDTIVPIIPLIGSLSKAKFCNFSGHPISKPAWADSSDSDIIDRFGRICRNLSHYYSGSSKKKSLYRIKYILRLSCARTLARKHKSTVRSFLKRLGSEFLE
Mahonia               KFYTIVPIITLIGSLANSKFCNASGHPISKSARTDSSDSDIIDRFGRICRNLSHYYSGSSKKKSLYRVKYILRLSCARTLSRKHKSTVRSFFKRLGSELLE
Trochodendron         KFDTIVPISPLIGSLA?AKFCTYQGIPISKPVRADSSDSDIIDRFGRICRNLSHYHSGSSKKKSLYRIKYILRLSCARTLASKHKSTVRAFLKRFGSELLE
Saxifraga             KFDIIVPIIPLIRSLAKAKFCNLVGDPISKPAWADSSDSYIIDRFVRICRNIYHYHSGSSKKNCLYRVKYILRLSCARTLARKHKSTVRAFLKRLGSGLLE
Lepismium             KFDTIVRIIPLVGSLAKAKFCNVLGHPISKSVWTDLLDSDIIDRFGRICRNLSHYYSGSSRKKSLYRIKYILRLSCARTLARKHKSTVRAFLKRLGSEFLE
Nypa                  ?FDTRVPVISFIGSLAKAKFCTVSGHPISKPIWTDLSDCDIIDRFGRICRNLSNYLSGSSKKQSLYRIKYILRFSCARTLARKHKSMVRAFLQRLGSGLLE
Asparagus             KFDTVVPVILLIRSLAKAKFCTVSGHPISKPIWADLSDSDILDRFGRICRNLSHYHSGSSRKRGLYRIKYILRLSCARTLARKHKSTVRTFLRRLGSGLLE
Alisma                KFDTIVPIIPLIGSLSKAKFCNVSGHPISKPIWTDLSDSDIIDRFVRICRNLSHYHSGSSKKQSLYRMKYILRLSCARTLARKHKTTVRAFFQRLGSGFLE
                                                                                                *
```

Fig. 1. Alignment of MatK X-domain comprised of 101 amino acids, for 40 taxa. Asterisk shows the putative RNA binding sites suggested by Mohr *et al.* (1993).

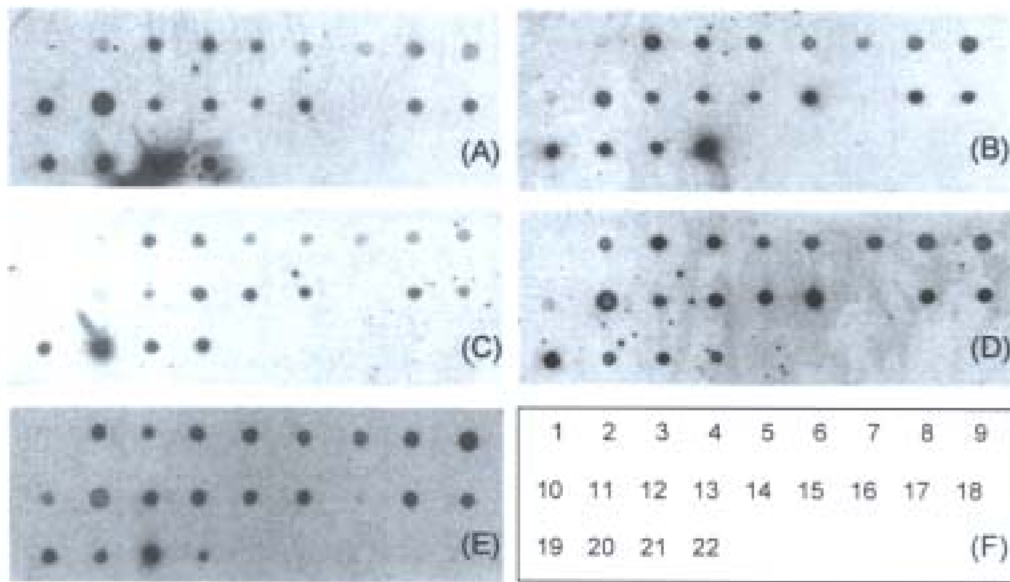

Fig. 2. Results of dot blot hybridization. A: Probed with *Lycopodiella matK*. B: Probed with *Ophioglossum matK*. C: Probed with *Selaginella matK*. D: Probed with *Adiantum matK*. E: Probed with *Ophioglossum rbcL*. F: The number and position of each dot. The corresponding taxa are as follows: 1. *Anthoceros formosae*, 2. *Marchantia polymorpha*, 3. *Equisetum ramosissimum*, 4. *Isoetes taiwanensis*, 5. *Selaginella doederleinii*, 6. *Selaginella delicatula*, 7. *Selaginella stauntoniana*, 8. *Selaginella involuens*, 9. *Selaginella tamariscina*, 10. *Lycopodiella cernua*, 11. *Lycopodium pseudoclavatum*, 12. *Ophioglossum petiolatum*, 13. *Angiopteris palmiformis*, 14. *Osmunda banksiifolia*, 15. *Adiantum capillus-veneris*, 16. *Dicranopteris linearis*, 17. *Lygodium japonicum*, 18. *Sphenomeris biflora*, 19. *Nicotiana sylvestris*, 20. plasmid harboring *Selaginella doederleinii matK*, 21. plasmid harboring *Lycopodiella cernua matK*, 22. plasmid harboring *Ophioglossum petiolatum matK*.

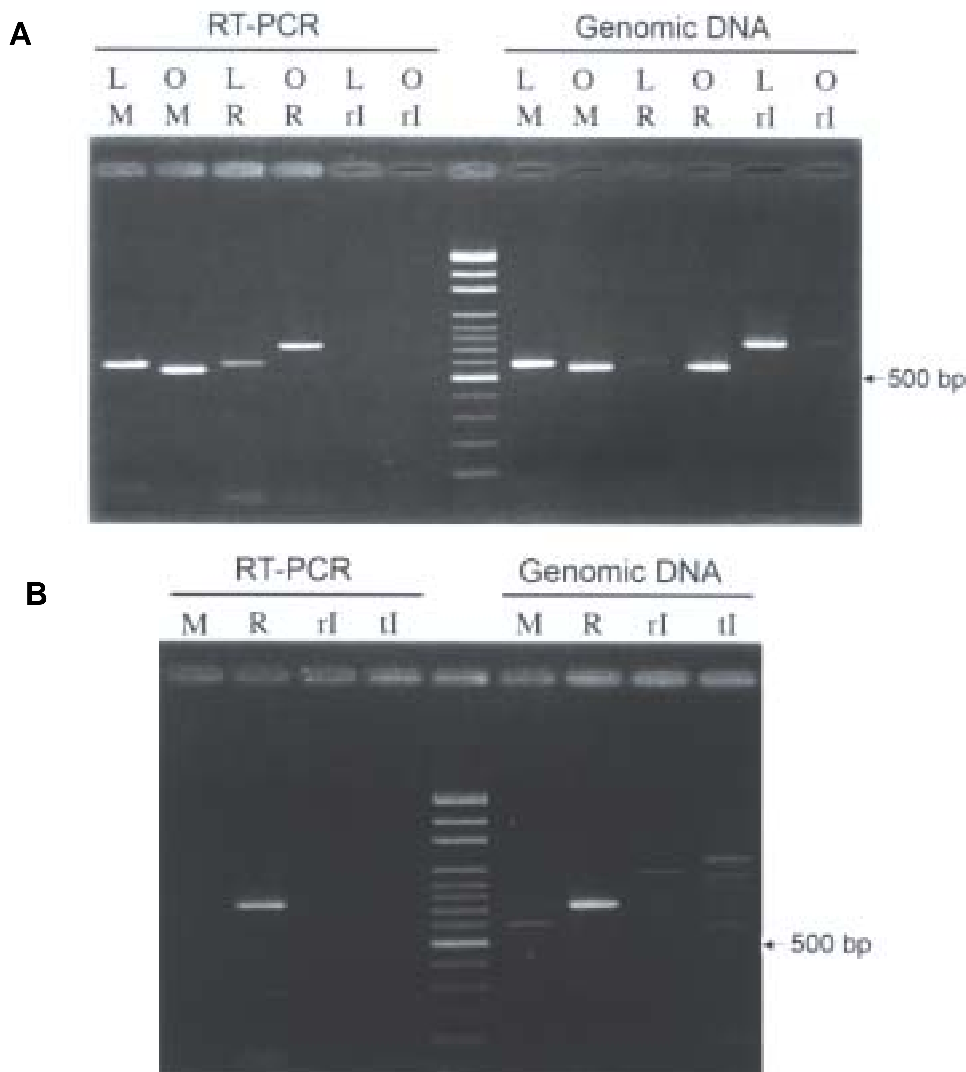

Fig. 3. Result of RT-PCR. A: Results from *Lycopodiella cernua* (denoted as "L") and *Ophioglossum petiolatum* (denoted as "O"). B: Results from *Selaginella doederleinii*. RNA templates were used in the RT-PCR reactions on the left, and genomic DNA templates were used on the right as control in both (A) and (B). M: PCR of *matK* region; R: PCR of *rbcL* region; rI: PCR of *rbcL-atpB* intron; tI: PCR of *trnL* intron.

Phylogenetic analysis of *matK*

The data matrix based on nucleotide sequences of *matK* coding regions contains 1867 characters with 1665 variable sites, 1498 of which are parsimony-informative characters. All characters were included in the phylogenetic analyses. Phylogenies obtained by neighbor-joining and parsimony methods are similar, and the results are shown in Fig. 4. Eighteen equally most parsimonious trees were found with length of 10059, consistency index = 0.37 and retention index = 0.56. Bryophytes, ferns, and fern allies are unresolved at the base of embryophyte clade. The sequences are very difficult to align except for the X domain in these taxa as well. Gnetophytes, conifers+cycads+Ginkgo, and angiosperms form three very well supported clades, all receiving 100% of bootstrap values. The resolution is relatively low within angiosperms, except for well supported eudicots (93-95% bootstrap support), represented by *Hepatica, Mahonia, Trochodendron, Saxifraga,* and *Lepismium*.

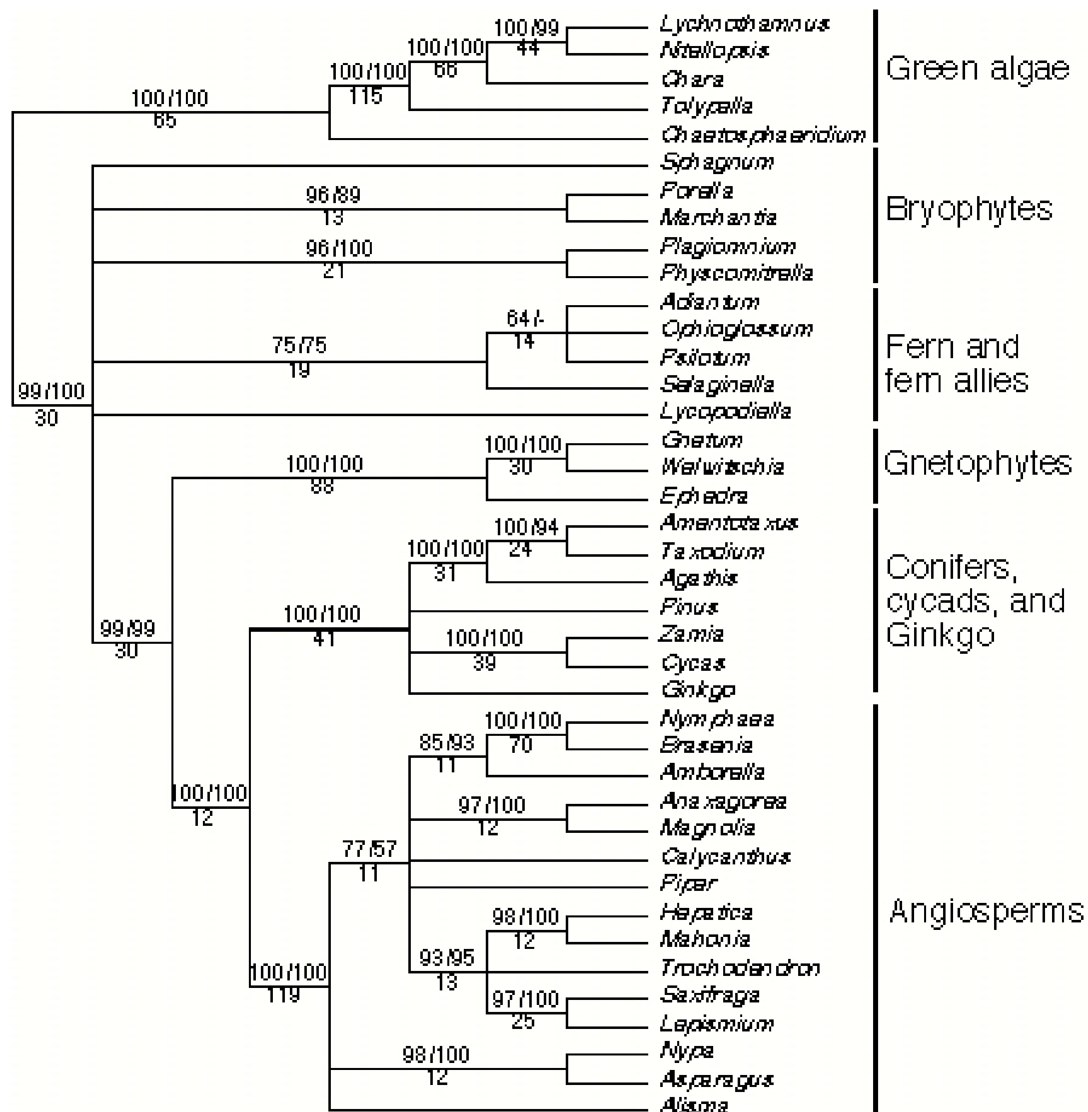

Fig. 4. Strict consensus tree of eighteen equally most parsimonious trees based on chloroplast *matK* nucleotide sequences. Numbers above the branches are bootstrap values using parsimony (before slash) and neighbor-joining (after slash) criteria. Numbers below the branches are decay indices.

DISCUSSION

In this study, we have identified plastid *trnK/matK*, *rbcL* and nuclear 18S rDNA sequences from three taxa: *Lycopodiella cernua*, *Selaginella doederleinii*, and *Ophioglossum petiolatum*. We failed to obtain *matK* sequences from other ferns and fern allies using degenerate primers or employing different PCR conditions (data not shown). However, the presence of *matK* is confirmed by dot blot hybridization, suggesting their *matK* might be too divergent to be obtained using ordinary PCR approach. RT-PCR results show that, even if *matK* gene is present with intact open reading frame, it may not be expressed as demonstrated in *Selaginella doederleinii*. This poses the question of whether or not *matK* is essential for Group II intron splicing in lower land plants. Since the *matK* homologue in *Anthoceros formosae* has been suggested to be a pseudogene (Kugita *et al.*, 2003), we speculate that

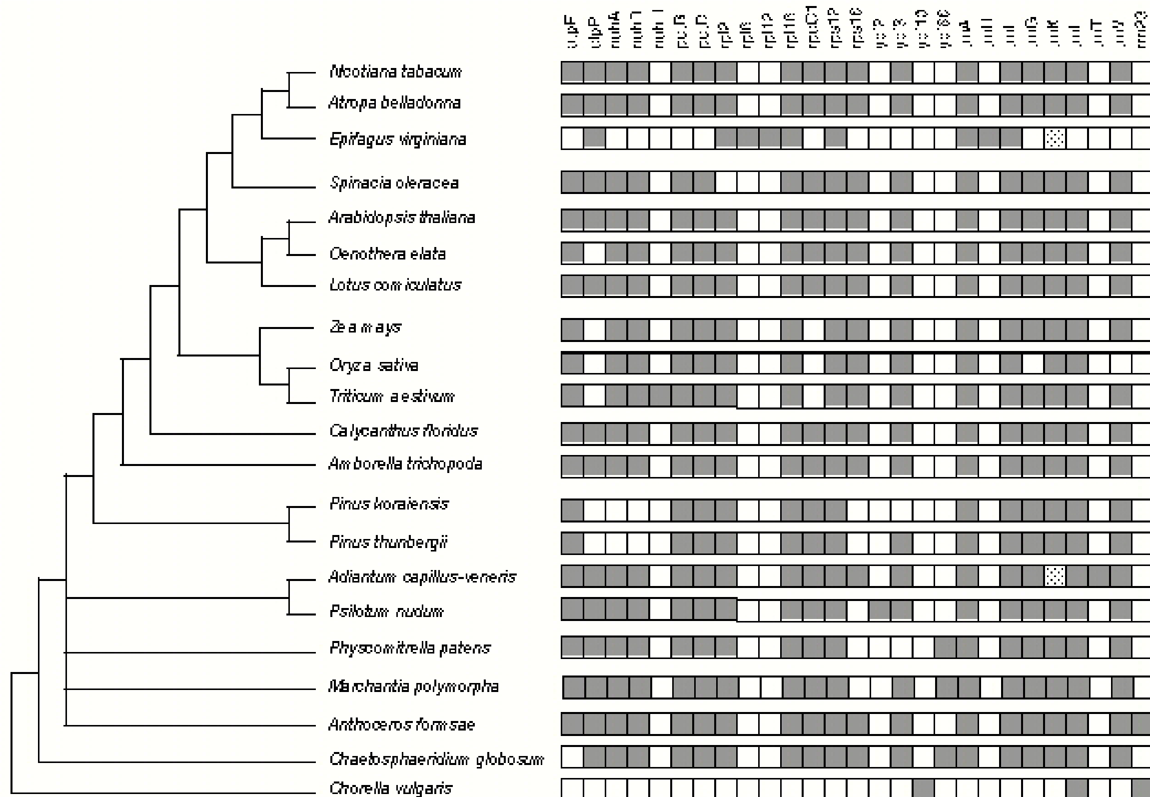

Fig. 5. A simplified diagram of green plant phylogeny and the distribution of introns in their chloroplast genome. The phylogeny is according to Pryer *et al.* (2001), APG II (Bremer *et al*., 2003), and Palmer *et al.* (2004). Gray boxes indicate genes with introns and white boxes indicate genes without intron disruption. The dotted boxes in *Epifagus* and *Adiantum* represent *matK* gene in free-standing form (*trnK* exons are missing).

chloroplast *matK* may indeed not be essential in primitive lineages of land plants. Nontheless, the presence of introns in chloroplast genomes is largely correlated with the presence of *matK*. Fig. 5 shows a simplified diagram of green plant phylogeny and the presence/absence of introns in their chloroplast genome. *Chorella vulgaris* and all green algae other than Characeae, show an intact *trnK* without intron interruption. Characeae and all of the land plants, in comparison, show the *trnK*$^{5'}$-*matK*-*trnK*$^{3'}$ structure. The figure clearly shows that the presence of introns in many genes coincides with the presence of *matK* residing in *trnK*. Although there are several intron gains and losses along the lineages of land plants, most introns persist throughout evolutionary history.

Given the fact that these Group II introns are capable of self-splicing, *matK* probably played a minor role when chloroplasts first harbored introns in their genomes. We speculate that, as evolution proceeded, *matK* gained some more important function in Group II intron splicing, thus becoming indispensible.

The sequences of *matK* are very divergent even in the usually highly conserved X domain, especially among Charophytes, ferns and fern allies. However, the serine (S) residue in the X domain, marked in Fig. 1, is highly conserved among all sequences, consistent with the universal presence among Group II introns (Mohr *et al.*, 1993). Only four taxa show a replacement of proline at this position: *Amentotaxus, Zamia, Cycas,* and *Ginkgo*. Whether or not this has any functional correlation, or evolutionary meaning, is subject to further tests.

Phylogenetic analysis using *matK* sequences poses certain difficulties since they are likely too divergent to provide confident relationships. Although the relationships among ferns and fern allies are unresolved, the peculiar placement of angiosperms, gnetophytes, and the rest of seed plants is quite interesting. The position of Gnetales is one of the most enigmatic questions in seed plant phylogeny (see review by Burleigh and Mathews, 2004). They have been placed with various groups of seed plants, but most recent molecular studies support that Gnetales are related to conifers (Chaw *et al.*, 2000; Rydin *et al.*, 2002; Soltis *et al.*, 2002). The placement of Gnetales sister to the rest of the seed plants ("Gnetales-sister" tree), has only been found in a few studies (Hamby and Zimmer, 1992; Sanderson *et al.*, 2000; Rydin *et al.*, 2002). Although the internal support of Gnetales-sister tree is quite high in the *matK* tree (Fig. 4), we speculate that this might be due to methodological problems imposed by the high evolutionary rate in the *matK* phylogeny, as demonstrated by Burleigh and Mathews (2004). Phylogenetic analysis using nucleotide sequences of the X domain region only showed that all gymnosperms formed a monophyletic group using parsimony criteria (data not shown). However, relationships within angiosperms are peculiar in that some eudicots becoming the basal group of angiosperms in this analysis using X-domain data set, although the support of this tree topology is low (<50% bootstrap value). This suggests that it may not be easy to extract correct phylogenetic information from *matK* genes, at least at the higher taxonomic level, and a thorough analysis is needed to elucidate this problem.

It is clear that *matK* is quite divergent in ferns and fern allies, and the failure to amplify the *matK* region using ordinary PCR indicates that we have to use other approaches in order to obtain these sequences. Such data are much needed especially in true ferns, where there is only one very divergent sequence available (*Adiantum capillus-veneris*).

ACKNOWLEDGEMENTS

We thank Nicole Maturen for comments on the manuscript; members from Hu lab for helping on collecting materials and valuable discussion.

LITERATURE CITED

Bhattacharya, D., K. Weber, S. S. An and W. Berning-Koch. 1998. Actin phylogeny identifies *Mesostigma viride* as a flagellate ancestor of the land plants. J. Mol. Evol. **47**: 544-550.

Besendahl., A., Y.-L. Qiu, J. H. Lee, J. D. Palmer and D. Bhattacharya. 2002. The cyanobacterial origin and vertical transmission of the plastid tRNA(Leu) group-I intron. Curr. Genet. **37**: 12-23.

Bremer, K. 1988. The limits of amino acid sequence data in angiosperm phylogenetic reconstruction. Evolution **42**: 795-803.

Bremer, K. 1994. Branch support and tree stability. Cladistics **10**: 295-304.

Bremer, B., K. Bremer, M. W. Chase, J. L. Reveal, D. E. Soltis, P. S. Soltis, P. F. Stevens, A. A. Anderberg, M. F. Fay, P. Goldblatt, W. S. Judd, M. Kallersjo, J. Karehed, K. A. Kron, J. Lundberg, D. L. Nickrent, R. G. Olmstead, B. Oxelman, J. C. Pires, J. E. Rodman, P. J. Rudall, V. Savolainen, K. J. Sytsma, M. van der Bank, K. Wurdack, J. Q. Y. Xiang and S. Zmarzty. 2003. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG II. Bot. J. Linn. Soc. **141**: 399-436.

Burleigh, J. G. and S. Matthews. 2004. Phylogenetic signal in nucleotide data from seed plants: implications for resolving the seed plant tree of life. Am. J. Bot. **91**: 1599-1613.

Chang, S., J. Puryear and J. Chirney. 1993. A simple and efficient method for isolating RNA from pine trees. Pl. Mol. Biol. Rep. **11**: 113-116.

Chaw, S. M., C. L. Parkinson, Y. C. Cheng, T. M. Vincent and J. D. Palmer. 2000. Seed plant phylogeny inferred from all three plant genomes: Monophyly of extant gymnosperms and origin of Gnetales from conifers. Proceed. Nat. Acad. Sci., USA **97**: 4086-4091.

Chiba, T., T. Harada, S. Goto, R. Ishikawa and M. Niizeki. 1996. Transcription of tRNA genes from large-scale plastid DNA deletion clearly reveals the action of nuclear-encoded RNA polymerase in the plastid. J. Pl. Physiol. **148**: 652-656.

du Jardin, P., D. Portetelle, L. Harvengt, M. Dumont and B. Wathelet. 1994. Expression of intron-encoded maturase-like polypeptides in potato chloroplasts. Curr. Genet. **25**: 158-163.

Ems, S. C., C. W. Morden, C. K. Dixon, K. H. Wolfe, C. W. de Pamphilis and J. D. Palmer. 1995. Transcription, splicing and editing of plastid RNAs in the nonphotosynthetic plant *Epifagus virginiana*. Pl. Mol. Biol. **29**: 721-733.

Eriksson, T. 1998. AutoDecay ver. 4.0 (program distributed by the author). Department of Botany, Stockholm University. Stockholm, Sweden.

Felsenstein, J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. Evolution **39**: 783-791.

Guo, H., S. Zimmerly, P. S. Perlman and A. M. Lambowitz. 1997. Group II intron endonucleases use both RNA and protein subunits for recognition of specific sequences in double-stranded DNA. EMBO J. **16**: 6835-6848.

Hamby, R. K. and E. A. Zimmer. 1992. Ribosomal RNA as a phylogenetic tool in plant systematics. In: Soltis, P. S. *et al.* (eds.). Molecular systematics of plants, pp. 50-91. Chapman and Hall, New York, USA.

Hasegawa, M., H. Kishino and T. Yano. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J. Mol. Evol. **22**: 160-174.

Henke, R. M., R. A. Butow and P. S. Perlman. 1995. Mautrase and endonuclease functions depend on separate conserved domains of the bifunctional protein encoded by the group I intron al4α of yeast mitochondrial DNA. EMBO J. **14**: 5094-5099.

Hilu, K. W., T. Borsch, K. Muller, D. E. Soltis, P. S. Soltis, V. Savolainen, M. W. Chase, M. P. Powell, L. A. Alice, R. Evans, H. Sauquet, C. Neinhuis, T. A. B. Slotta, J. G. Rohwer, C. S. Campbell and L. W. Chatrou. 2003. Angiosperm phylogeny based on *matK* sequence information. Am. J. Bot. **90**: 1758-1776.

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

Hu, J.-M., M. Lavin, M. F. Wojciechowski and M. J. Sanderson. 2000. Phylogenetic systematics of the tribe Millettieae (Leguminosae) based on chloroplast *trnK*/*matK* sequences and its implications for evolutionary patterns in Papilionoideae. Am. J. Bot. **87**: 418-430.

Jankowiak, K., J. Lesicka, A. Pacak, A. Rybarczyk and Z. Szweykowska-Kulinska. 2004. A comparison of Group II introns of plastid $tRNA^{Lys}{}_{UUU}$ genes encoding maturase protein. Cell. Mol. Biol. Lett. **9**: 239-251.

Kotani, H. and S. Tabata. 1998. Lessons from sequencing of the genome of a unicellular cyanobacterium, *Synechocystis* sp. PCC6803. In: Jones, R. L. C. R. S. and V. Walbot (eds.). Annual Review of Plant Physiology and Plant Molecular Biology, Vol. 49, pp.

151-171. Annual Reviews Inc., Palo Alto, California, USA.

Kugita, M., A. Kaneko, Y. Yamamoto, Y. Takeya, T. Matsumoto and K. Yoshinaga. 2003. The complete nucleotide sequence of the hornwort (*Anthoceros formosae*) chloropast genome: insight into the earliest land plants. Nucl. Acid. Res. **31**: 716-721.

Kuhsel, M. G., R. Strickland and J. D. Palmer. 1990. An Ancient Group-I Intron Shared By Eubacteria and Chloroplasts. Science **250**: 1570-1573.

Liere, K. and G. Link. 1995. RNA-binding activity of the MatK protein encoded by the chloroplast *trn*K intron from mustard (*Sinapis alba* L.). Nucl. Acid. Res. **23**: 917-921.

Matsuura, M., R. Saldanha, H. Ma, H. Wank, J. Yang, G. Mohr, S. Cavanagh, G. M. Dunny, M. Belfort and A. M. Lambowitz. 1997. A bacterial group II intron encoding reverse transcriptase, maturase, and DNA endonuclease activities: biochemical demonstration of maturase activity and insertion of new genetic information within the intron. Gen. Devel. **11**: 2910-2924.

Michel, F., K. Umesono and H. Ozeki. 1989. Comparative and functional anatomy of group II catalytic introns--a review. Gene **82**: 5-30.

Mohr, G., P. S. Perlman and A. M. Lambowitz. 1993. Evolutionary relationships among group II intron-encoded proteins and identification of a conserved domain that may be related to maturase function. Nucl. Acid. Res. **21**: 4991-4997.

Neuhaus, H. and G. Link. 1987. The chloroplast tRNA-Lys(UUU) gene from mustard (*Sinapis alba*) contains a class II intron potentially coding for a maturase-related polypeptide. Curr. Genet. **11**: 251-257.

Odintsova, M. S. and N. P. Yurina. 2000. RNA editing in plant chloroplasts and mitochondria [Review]. Russ. J. Pl. Physiol. **47**: 274-284.

Palmer, J. D., D. E. Soltis and M. W. Chase. 2004. The plant tree of life: an overview and some points of view. Am. J. Bot. **91**: 1437-1445.

Porebski, S., L. G. Bailey and B. R. Baum. 1997. Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. Pl. Mol. Biol. Rep. **15**: 8-15.

Pryer, K. M., H. Schneider, A. R. Smith, R. Cranfill, P. G. Wolf, J. S. Hunt and S. D. Sipes. 2001. Horsetails and ferns are a monophyletic group and the closest living relatives to seed plants. Nature **409**: 618-622.

Rydin, C., M. Kallersjo and E. M. Friist. 2002. Seed plant relationships and the systematic position of Gnetales based on nuclear and chloroplast DNA: Conflicting data, rooting problems, and the monophyly of conifers. Int. J. Pl. Sci. **163**: 197-214.

Sambrook, J., D. W. Russell and J. Sambrook. 2001. Molecular cloning, 3rd ed. Cold Spring Harbor Laboratory.

Sanders, E. R., K. G. Karol and R. M. McCourt. 2003. Occurrence of *matK* in a *trnK* group II intron in charophyte green algae and phylogeny of the Characeae. Am. J. Bot. **90**: 628-633.

Sanderson, M. J., M. F. Wojciechowski, J.-M. Hu, T. Sher Khan and S. G. Brady. 2000. Error, bias, and long-branch attraction in data for two chloroplast photosystem genes in seed plants. Mol. Biol. Evol. **17**: 782-797.

Schmelzer, C., C. Schmidt, K. May and R. J. Schweyen. 1983. Determination of functional domains in intron bI1 of yeast mitochondrial RNA by studies of mitochondrial mutations and a nuclear suppressor. EMBO J. **2**: 2047-2052.

Shimada, H. and M. Sugiura. 1991. Fine structural features of the chloroplast genome: comparison of the sequenced chloroplast genomes. Nucl. Acid. Res. **19**: 983.

Soltis, D. E., P. S. Soltis and M. J. Zanis. 2002. Phylogeny of seed plants based on evidence from eight genes. Am. J. Bot. **89**: 1670-1681.

Steele, K. P. and R. Vilgalys. 1994. Phylogenetic analysis of Polemoniaceae using nucleotide sequences of the plastid gene *mat*K. Syst. Bot. **19**: 126-142.

Sugita, M., K. Shinozaki and M. Sugiura. 1985. Tobacco chloroplast tRNALys (UUU) gene contains a 2.5-kilobase-pair intron: An open reading frame and a conserved boundary sequence in the intron. Proc. Nat. Acad. Sci., USA **82**: 3557-3561.

Sugiura, M. 1992. The chloroplast genome. Pl. Mol. Biol. **19**: 149-168.

Sugiura, C., Y. Kobayashi, S. Aoki, C. Sugita and M. Sugita. 2003. Complete chloroplast DNA sequence of the moss *Physcomitrella patens*: evidence for the loss and relocation of *rpoA* from the chloroplast to the nucleus. Nucl. Acid. Res. **31**: 5324-5331.

Swofford, D. L. 2002. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods), version 4.0 beta10. Sinauer Associates, Inc., Sunderland, MA, USA.

Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin and D. G. Higgins. 1997. The CLUSTAL-X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucl. Acid. Res. **25**: 4876-4882.

Turmel, M., C. Otis and C. Lemieux. 2002. The chloroplast and mitochondrial genome sequences of the charophyte *Chaetosphaeridium globosum*: Insights into the timing of the events that restructured organelle DNAs within the green algal lineage that led to land plants. Proc. Nat. Acad. Sci., USA **99**: 11275-11280.

Wakasugi, T., T. Nagai, M. Kapoor, M. Sugita, M. Ito, S. Ito, J. Tsudzuki, K. Nakashima, T. Tsudzuki, Y. Suzuki, A. Hamada, T. Ohta, A. Inamura, K. Yoshinaga and M. Sugiura. 1997. Complete nucleotide sequence of the chloroplast genome from the green alga *Chlorella vulgaris*: The existence of genes possibly involved in chloroplast division. Proc. Nat. Acad. Sci., USA **94**: 5967-5972.

Wolf, P. G., C. A. Rowe, R. B. Sinclair and M. Hasebe. 2003. Complete nucleotide sequence of the chloroplast genome from a leptosporangiate fern, *Adiantum capillus-veneris* L. DNA Res. **10**: 59-65.

Wolfe, K. H., C. W. Morden and J. D. Palmer. 1992. Function and evolution of a minimal plastid genome from a nonphotosynthetic parasitic plant. Proc. Nat. Acad. Sci., USA **89**: 10648-10652.

石松類和瓶爾小草葉綠體 *matK* 基因之鑑定與 *matK* 基因的演化

莊樹嵐[1]、胡哲明[1,2]

(收稿日期：2004 年 10 月 27 日；接受日期：2004 年 11 月 24 日)

摘　　要

高等植物葉綠體 *trnK*UUU 基因的內插子屬於第二類內插子，其中包含一段開讀框，因其序列及結構和酵母菌粒線體的 maturase 相似，故稱之為 *matK*，*matK* 是葉綠體中唯一具有去除內插子功能的基因，只有較高等的輪藻和陸生植物的葉綠體 *trnK* 基因才有內插子和 *trnK*5'-*matK*-*trnK*3'的結構，其他綠藻葉綠體的 *trnK* 並沒有 *matK* 開讀框、也沒有第二類內插子。葉綠體 *trnK*5'-*matK*-*trnK*3'的結構在高等植物相當一致，但先前研究發現兩個 *matK* 脫離 *trnK* 而存在的例子，像是葉綠體嚴重退化的寄生植物 *Epifagus virginiana* 及因葉綠體基因重組而導致 *trnK* 外顯子丟失的鐵線蕨，它們的 *matK* 仍然具有完整的開讀框，顯示了 *matK* 的確十分重要而不可缺失，蘚苔類和松葉蕨也被證實具有 *trnK*5'-*matK*-*trnK*3'的結構，但是在角蘚中卻發現其 *matK* 是偽基因。本論文鑑定原始陸生植物其他未被檢視的重要類群中葉綠體 *matK* 基因的存在，結果發現瓶爾小草、過山龍、生根卷柏都有 *trnK*5'-*matK*-*trnK*3'的結構，以 RT-PCR 偵測 *matK* 表現時，瓶爾小草和過山龍都有偵測到訊號，但生根卷柏卻沒有，顯示 *matK* 的表現在原始陸生植物似乎不是那麼一致。利用 PCR 的偵測並未能在水韭、木賊、觀音座蓮、紫萁、海金沙、芒萁等物種中放大出 *trnK* 或 *matK* 的片段。但是利用點式雜合反應則顯示這些物種可能有 *matK* 序列的存在，顯示這些序列可能因變異太大而無法以 PCR 方式放大。故有關蕨類葉綠體 *matK* 的存在與否仍有待進一步研究。以葉綠體 *matK* 核甘酸序列進行譜系分析，結果顯示部分裸子植物(松柏、銀杏、蘇鐵)和被子植物形成姊妹群，兩者再和麻黃類植物形成姊妹群。作者認為這個結果主要是由於 *matK* 基因的演化速率太快，而造成譜系分析的不可靠性。

關鍵詞：葉綠體 *matK*、過山龍、生根卷柏、瓶爾小草、演化。

1. 國立台灣大學生態學與演化生物學研究所，台北市 106 羅斯福路 4 段 1 號，台灣。
2. 通訊作者。Email: jmhu@ntu.edu.tw