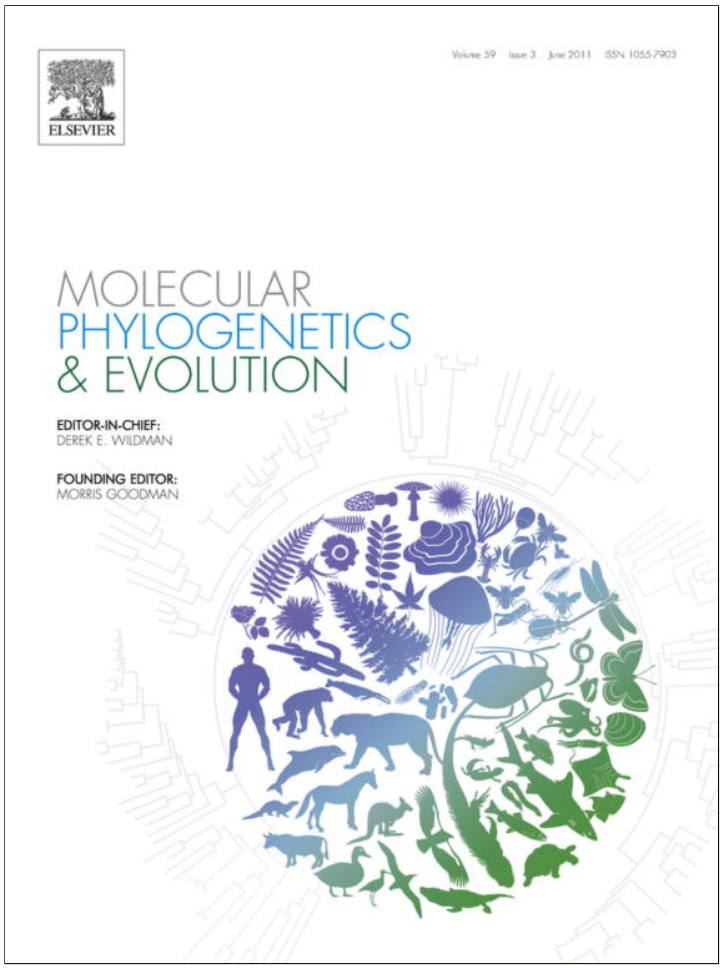

Molecular Phylogenetics and Evolution 59 (2011) 556–566

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

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

First insights into fern *matK* phylogeny

Li-Yaung Kuo [a,1], Fay-Wei Li [b,1,2], Wen-Liang Chiou [c], Chun-Neng Wang [a,b,*]

[a] Institute of Ecology and Evolutionary Biology, National Taiwan University, 1 Roosevelt Road, Sec. 4, Taipei 10617, Taiwan
[b] Department of Life Science, National Taiwan University, 1 Roosevelt Road, Sec. 4, Taipei 10617, Taiwan
[c] Division of Botanical Garden, Taiwan Forestry Research Institute, 53 Nan-Hai Road, Taipei 10066, Taiwan

ARTICLE INFO

Article history:
Received 18 June 2010
Revised 19 February 2011
Accepted 5 March 2011
Available online 12 March 2011

Keywords:
DNA barcoding
Homoplasy
matK
Phylogenetic performance
Substitution evenness

ABSTRACT

MatK, the only maturase gene in the land plant plastid genome, is a very popular phylogenetic marker that has been extensively applied in reconstructing angiosperm phylogeny. However, the use of *matK* in fern phylogeny is largely unknown, due to difficulties with amplification: ferns have lost the flanking *trnK* exons, typically the region used for designing stable priming sites. We developed primers that are either universal or lineage-specific that successfully amplify *matK* across all fern families. To evaluate whether *matK* is as powerful a phylogenetic marker in ferns as in angiosperms, we compared its sequence characteristics and phylogenetic performance to those of *rbcL* and *atpA*. Among these three genes, *matK* has the highest variability and substitution evenness, yet shows the least homoplasy. Most importantly, applying *matK* in fern phylogenetics better resolved relationships among families, especially within eupolypods I and II. Here we demonstrate the power of *matK* for fern phylogenetic reconstruction, as well as provide primers and extensive sequence data that will greatly facilitate future evolutionary studies of ferns.

1. Introduction

Molecular phylogeneticists are always searching for suitable loci with which to reconstruct the tree of life. Compared to nuclear genomes, organellar regions have been more broadly exploited for this purpose because they seldom undergo duplication, are usually uniparentally inherited, and are more easily amplified. The first mitochondrial and chloroplast (cp) loci used were coding regions because they typically evolve more slowly and have few indels, rendering their alignment unambiguous at high taxonomic ranks. However, because coding regions are functionally constrained, they tend to have slower evolutionary rates that usually limit their phylogenetic inferences. The *matK* gene, the only gene to encode for a splicing-associated maturase in the land plant cp genome, has helped overcome this obstacle: it is sufficiently fast evolving to be useful at a wide range of taxonomic depths, while still permitting unambiguous alignment. In angiosperms, for example, *matK* has considerable variation, and is even comparable to non-coding regions, yet with few alignment problems (Müller et al., 2006; Lohne et al., 2007; Hilu and Barthet, 2008; Hilu et al., 2008). Consequently, *matK* is frequently employed in phylogenetic analyses at various taxonomic levels of angiosperms (e.g., Hilu et al., 2003; Döring et al., 2007; Xie et al., 2010), and was even proposed as a DNA barcoding region for land plants in combination with *rbcL* (CBOL Plant Working Group, 2009).

In seed plants, *matK* is usually nested within the *trnK* intron, in a *trnK–matK–trnK* order, in the large single copy region of the chloroplast genome (Sugita et al., 1985; Neuhaus and Link, 1987; Hilu and Barthet, 2008). As a result, *matK* can be amplified by using primers targeted at the conserved *trnK* exons. Most leptosporangiate ferns, however, lack stable priming regions due to the loss of the *trnK* exons (Wolf et al., 2003; Duffy et al., 2009; Gao et al., 2009). The absence of suitable priming sites thus impaired the development of this locus in fern phylogenetics. The frequently used chloroplast loci including *rbcL*, *atpA*, and *atpB*, however, have been unable to unambiguously resolve family-level relationships (Schuettpelz et al., 2006; Schuettpelz and Pryer, 2007; Smith et al., 2006, 2008), especially in eupolypods II (Sano et al., 2000). Applying regions that can be aligned and yet evolve in a rapid manner, such as *matK*, to reconstruct ancient radiations is a more-efficient route than is combining multiple-conservative coding regions (Hilu et al., 2003).

In this study, we designed new primer sets aimed at the *matK* coding region and its 5′ upstream genes, including *trnK*, *chlB*, and *rps16*. We attempted to: (1) ascertain where the *trnK* loss event took place in the evolution of ferns; (2) characterize the sequence properties and quantify the phylogenetic performance of fern *matK*, and further compare it with *rbcL* and *atpA*; (3) evaluate

* Corresponding author at: Institute of Ecology and Evolutionary Biology, Department of Life Science, National Taiwan University, 1 Roosevelt Road, Sec. 4, Taipei 10617, Taiwan. Fax: +886 2 23673374.
E-mail addresses: salvinia@hotmail.com (L.-Y. Kuo), fayweili@gmail.com (F.-W. Li), botwang@ntu.edu.tw (C.-N. Wang).
[1] These authors contributed equally to this work.
[2] Present address: Department of Biology, Duke University, Durham, NC 27708, USA.

1055-7903/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2011.03.010

whether applying *matK* to phylogenetic analyses can bring new insights to the evolutionary history of ferns; and (4) discuss the utility of the fern *matK* gene in future phylogenetic and DNA barcoding studies.

2. Materials and methods

2.1. Taxonomic sampling, DNA extraction, and sequence amplification

We sampled 78 fern species from all 37 families (*sensu* Smith et al., 2006), with four seed plants and two lycophytes used as outgroups. DNA was extracted using a modified CTAB procedure (Wang et al., 2004) or a Plant Genomic DNA Mini Kit (GeneAid, Taipei, Taiwan) following the manufacturer's protocol.

We chose a subset of fern taxa (see Table 1) to screen for the presence of *trnK* exons, using the newly designed *trnK* primers, trnK 0118 and FernK r1 (Table 2; Fig. 1a). When *trnK* exons were present, the *matK* gene was within the *trnK–matK–trnK* amplicon and could be sequenced directly. For taxa without *trnK* exons, *matK* sequences were obtained via two separate amplifications. The first used a primer pair situated near the ends of the *matK* coding region (Fig. 1b). The primer pair, FERmatK fEDR and FERmatK rAGK, worked in most of ferns; however, in certain lineages specific primers had to be used (see Table 2 for details). The second amplification, to acquire the complete 5′ end of *matK*, required lineage-specific primers in the *matK* coding region (indicated by superscript d in Table 2; Fig. 1b) in combination with primers situated upstream in the *chlB* and *rps16* genes (indicated by superscript c in Table 2; Fig. 1b). Sequences from the two amplifications overlapped, and were combined to create a single contig. For species without available *rbcL* or *atpA* sequences on Genbank, these regions were also amplified and sequenced, using the published (Wolf et al., 1994; Schuettpelz et al., 2006) or newly designed primers (Table 2). PCR reactions were performed in a 15-μL volume, including >20 ng genomic DNA, 1× PCR buffer, 200 μM dNTP, 15 pmol of each primer, and 1 U polymerase (ProTaq, PROTECH, Taipei, Taiwan; or Phusion DNA polymerase, FINNZYMES, Espoo, Finland). For amplification of the complete 5′ end *matK* region (Fig. 1b), nested PCR was used by adding 1 μL of the first-round PCR product to second-round PCR with the same conditions described above.

2.2. Phylogenetic analyses

After ambiguous regions in the alignments were removed, maximum likelihood (ML) and Bayesian inference (BI) analyses were respectively carried out using Garli v0.96 (Zwickl, 2006) and MrBayes v3.1 (Ronquist and Huelsenbeck, 2003) for each of the seven datasets, including 3 one-gene datasets (*matK*, *rbcL*, and *atpA*); 3 two-gene datasets (*matK* + *rbcL*, *matK* + *atpA*, and *rbcL* + *atpA*); and a three-gene dataset (*matK* + *rbcL* + *atpA*).

For the ML analyses, a GTR + I + G model was used. The proportion of invariant sites and state frequencies were estimated by the program. The genthreshfortopoterm option was set to 20,000 while others followed the default settings. To calculate ML bootstrap support (BS) values, 1000 replicates were run under the same criteria. For the BI analyses, model selection and substitution parameters values were taken from Modeltest v3.7 (Posada and Crandall, 1998) as selected by the Akaike Information Criterion. Models and substitution parameters were individually assigned and unlinked across the gene partitions. Two simultaneous runs were carried out with four chains (of 10^6 generations each), in which each chain was sampled every 1000 generations. The first 25% of the sample was discarded as burn-in, and the rest used to calculate the majority-rule consensus tree.

2.3. Sequence characterization

For each gene, the nonsynonymous/synonymous substitution ratio (dN/dS) and variability (the proportion of variable sites) were calculated by PAML 4 (Yang, 2007) and Dnasp v4.50.3 (Rozas et al., 2003), respectively. MEGA v4.0 (Tamura et al., 2007) was used to infer the GC content and the transition/transversion ratio (ts/tv) for the three genes (*matK*, *rbcL*, and *atpA*). After excluding autoapomorphic insertions and incomplete sequence data at the ends of the alignments, MacClade v4.06 (Maddison and Maddison, 2003) was used to calculate the consistency index (CI), retention index (RI), and number of substitutions at each nucleotide site for all three genes based on the most likely tree from each one-gene dataset. A total of 1209 *matK*, 1176 *rbcL*, and 1491 *atpA* nucleotide sites were included. The value, $1-(\mathrm{RI}\times\mathrm{CI})$, was calculated for each nucleotide site to represent the degree of homoplasy at different codon positions (only nucleotide sites with more than one substitution event were included). To calculate the nucleotide substitution evenness (SE) of each codon position, the following formula based on the Shannon index (Shannon, 1948) was used, in which *pi* is the proportion of substitutions of a given nucleotide site to the total number of substitutions, and *S* is the total number of nucleotide sites:

$$\mathrm{SE}=\frac{-\sum_{i=1}^{s} pi\ln pi;}{\ln S}$$

3. Results

A total of 69 fern *matK* sequences were newly obtained in this study, spanning all recognized fern families (*sensu* Smith et al., 2006). Each sequence covered at least 4/5 of the gene length, and included a complete 5′ end (Table 1). Our *trnK* primers amplified *trnK-matK-trnK* amplicons from *Equisetum ramosissimum* (Equisetaceae), *Psilotum nudum* (Psilotaceae), *Ophioderma pendula* (Ophioglossaceae), *Botrychium formosanum* (Ophioglossaceae), *Danaea elliptica* (Marattiaceae), and *Osmunda japonica* (Osmundaceae), from which full-length *matK* sequences were obtained. For *matK* without flanking *trnK* exons, sequences were combined from two separate amplifications that gave two overlapping sequences (see Materials and Methods). We inadvertently sequenced the complete 3′ end of *Schizaea dichotoma*, due to the reverse primer, Sch matK rRDS, annealing to the 3′ downstream region of the expected priming site.

The log likelihoods of the maximum likelihood trees for the *matK* (**Supplementary material 1**) and three-gene (Figs. 2a and 2b) datasets were −57011.177 and −114004.4819. Maximum likelihood bootstrap support and Bayesian posterior probabilities for each branch are summarized in **Supplementary material 2**. No well-supported relationships were in conflict among the loci or phylogenetic methods.

Table 3 summarizes the sequence characteristics of *matK*, *rbcL*, and *atpA* genes in ferns, and compares these with angiosperms. The degree of homoplasy and the distribution of nucleotide substitutions for different codon positions are shown in Figs. 3 and 4, respectively.

4. Discussion

4.1. Loss of trnK gene in ferns

The loss of *trnK* exons (but with *matK* remaining intact) has been reported from several plant groups, including ferns (Wolf et al., 2003), *Selaginella* (Tsuji et al., 2007), *Epifagus* (Wolfe et al., 1992), and *Cuscuta* (Funk et al., 2007). Our results show that in

Table 1
Taxa included in this study, their voucher information and GenBank accession numbers. Taxa marked by asterisks were used in the initial screen for *trnK* exons.

Family	Tax on	*matK*		*rbcL*		*atpA*	
		Voucher/ citation	GenBank	Voucher/ citation	GenBank	Voucher/ citation	GenBank
Lycophytes							
Selaginellaceae	*Selaginella uncinata* (Desv. *ex* Poir.) Spring	Tsuji et al. (2007)	AB197035	Tsuji et al. (2007)	AB197035	Tsuji et al. (2007)	AB197035
Lycopodiaceae	*Huperzia lucidula* (Michx.) Trevisan	Wolf et al. (2005)	NC_006861	Wolf et al. (2005)	NC_006861	Wolf et al. (2005)	NC_006861
Spermatophytes							
Cycadaceae	*Cycas taitungensis* C.F. Shen, K.D. Hill, C.H. Tsou and C.J. Chen	Wu et al. (2007)	NC_009618	Wu et al. (2007)	NC_009618	Wu et al. (2007)	NC_009618
Ginkgoaceae	*Ginkgo biloba* L.	Chaw et al. (2005)	AF279806	Leebens-Mack et al. (2005)	DQ069500	Schuettpelz et al. (2006)	DQ390561
Pinaceae	*Pinus thunbergii* Parl.	Wakasugi et al. (1994)	NC_001631	Wakasugi et al. (1994)	NC_001631	Wakasugi et al. (1994)	NC_001631
Magnoliaceae	*Liriodendron tulipifera* L.	Cai et al. (2006)	NC_008326	Cai et al. (2006)	NC_008326	Cai et al. (2006)	NC_008326
Monilophytes							
Equisetaceae	*Equisetum ramosissimum* Desf.	L.Y Kuo 823[a] (Taiwan)	JF303895	Des Marais et al. (2003)	AY226132	L.Y. Kuo 823[a] (Taiwan)	JF303982
	Equisetum arvense L.	Hausner et al. (2006)	AY348551	Des Marais et al. (2003)	AY226140	K.M. Pryer 00-02[b] (Canada)	JF303981
Psilotaceae	*Psilotum nudum* (L.) P. Beauv.*	Wakasugi et al. (1998)	NC_003386	Wakasugi et al. (1998)	NC_003386	Wakasugi et al. (1998)	NC_003386
Ophioglossaceae	*Ophioderma pendula* (L.) C. Presl*	K013186[c] (Taiwan)	JF303896	Hauk et al. (2003)	AY138420	K013186[c] (Taiwan)	JF303979
	Botrychium formosanum Tagawa	K016552[c] (Taiwan)	JF303897	K016552[c] (Taiwan)	JF303963	K016552[c] (Taiwan)	JF303980
Marattiaceae	*Danaea elliptica* Sm.*	C.J. Rothfels 08-182[b] (Costa Rica)	JF303894	Hauk et al. (2003)	AY138398	Schuettpelz and Pryer (2007)	EF463784
	Angiopteris evecta (J.R. Forst.) Hoffmann	Roper et al. (2007)	NC_008829	Roper et al. (2007)	NC_008829	Roper et al. (2007)	NC_008829
Osmundaceae	*Osmundastrum cinnamomeum* (L.) C. Presl	Duffy et al. (2009)	EU223823	Metzgar et al. (2008)	EF588711	Schuettpelz and Pryer (2007)	EF463795
	Osmunda japonica Thunb.*	K019686[c] (Taiwan)	JF303900	Metzgar et al. (2008)	EF588701	Metzgar et al. (2008)	EF588679
Hymenophyllaceae	*Vandenboschia radicans* (Sw.) Copel.*	L.Y. Kuo K7[a] (Taiwan)	JF303901	Ebihara et al. (2005)	AB196367	Schuettpelz and Pryer (2007)	EF463767
	Hymenophyllum polyanthos (Sw.) Sw.	F.W. Li 615[a] (Taiwan)	JF303898	Ebihara et al. (2003)	AB083276	Schuettpelz and Pryer (2007)	EF463759
Gleicheniaceae	*Dicranopteris linearis* (Burm. f.) Underw.*	L.Y. Kuo 824[a] (Taiwan)	JF303902	Perrie et al. (2007)	DQ910495	Schuettpelz and Pryer (2007)	EF463733
Dipteridaceae	*Cheiropleuria integrifolia* (D.C. Eaton *ex* Hook) M, Kato, Y. Yatabe, Sahashi & N. Murak.*	L.Y. Kuo K8[a] (Taiwan)	JF303904	Kato et al. (2001)	AB042569	Schuettpelz and Pryer (2007)	EF463661
Matoniaceae	*Matonia pectinata* R. Br.	E. Schuettpelz 752[b] (Malaysia)	JF303903	Kato and Setoguchi (1998)	AF012267	Schuettpelz and Pryer (2007)	EF463789
Lygodiaceae	*Lygodium japonicum* (Thunb.) Sw.*	Duffy et al. (2009)	EU223821	Korall et al. (2006)	AM177360	Schuettpelz and Pryer (2007)	EF463781
Schizaeaceae	*Schizaea dichotoma* (L.) J. Sm.*	F.W. Li 359[a] (Taiwan)	JF303899	Pryer et al. (2004)	AY612683	Schuettpelz and Pryer (2007)	EF463861
Anemiaceae	*Anemia phyllitidis* (L.) Sw.	A.L. Grusz 08-007[b] (Costa Rica)	JF303905	Wikström et al. (2002)	AJ303391	Schuettpelz and Pryer (2007)	EF463584
Marsileaceae	*Marsilea mutica* Mett.	Duffy et al. (2009)	EU223822	Nagalingum et al. (2007)	DQ643309	Schuettpelz and Pryer (2007)	EF463787
Salviniaceae	*Salvinia molesta* D.S. Mitch.*	F.W. Li 628[a] (cultivated)	JF303906	Nagalingum et al. (2008)	EU269669	F.W. Li 628[a] (cultivated)	JF303983
Loxomataceae	*Loxoma cunninghamii* R. Br.	A.R. Smith 2612[b] (New Zealand)	JF303912	Pryer et al. (2004)	AY612679	Schuettpelz and Pryer (2007)	EF463779
Plagiogyriaceae	*Plagiogyria euphlebia* (Kunze) Mett.*	L.Y. Kuo K14[a] (Taiwan)	JF303911	L.Y. Kuo K14[a] (Taiwan)	JF303965	L.Y. Kuo K14[a] (Taiwan)	JF303986
Culcitaceae	*Culcita macrocarpa* C. Presl	M. Christenhusz 3872[b] (cultivated)	JF303913	Korall et al. (2006)	AM177334	Korall et al. (2006)	AM176436

Table 1 (*continued*)

Family	Tax on	*matK*		*rbcL*		*atpA*	
		Voucher/ citation	GenBank	Voucher/ citation	GenBank	Voucher/ citation	GenBank
Thyrsopteridaceae	*Thyrsoptetis elegans* Kunze	C. Morter 18[b] (cultivated)	JF303910	Korall et al. (2006)	AM177353	Schuettpelz and Pryer (2007)	EF463900
Metaxyaceae	*Metaxya lanosa* A. R. Sm. and Tuomisto	H. Tuomisto 13054[b] (Peru)	JF303909	Smith et al. (2001)	AF317701	H. Tuomisto 13054[b] (Peru)	JF303985
Dicksoniaceae	*Dicksonia antarctica* Labill.	Duffy et al. (2009)	EU223820	Wolf et al. (1994)	U05919	Schuettpelz and Pryer (2007)	EF463659
Cibotiaceae	*Cibotiwn barometz* (L.) J. Sm.*	L.Y. Kuo K 18[a] (Taiwan)	JF303908	Korall et al. (2006)	AM177328	L.Y. Kuo K13[a] (Taiwan)	AM176429
Cyatheaceae	*Alsophila podophylla* Hook.*	F.W. Li 313[a] (Taiwan)	JF303907	F.W. Li 313[a] (Taiwan)	JF303964	F.W. Li 313[a] (Taiwan)	JF303984
Lindsaeaceae	*Sphenomeris chinensis* (L.) Maxon*	F.W. Li 595[a] (Taiwan)	JF303915	Hasebe et al. (1994)	U05651	Schuettpelz and Pryer (2007)	EF463774
Saccolomataceae	*Saccoloma inaequale* (Kunze) Mett.	M. Christenhusz 4233[b] (Puerto Rico)	JF303914	Schuettpelz and Pryer (2007)	EF463265	Schuettpelz and Pryer (2007)	EF463858
Dennstaedtiaceae	*Pteridium aquilinum* (L.) Kuhn	Duffy et al. (2009)	EU223824	Schneider et al. (2004b)	AY300097	L.Y. Kuo 825[a] (Taiwan)	JF303987
	Dennstaedtia scabra (Wall, *ex* Hook.) T. Moore	F.W. Li 608[a] (Taiwan)	JF303916	F.W. Li 608[a] (Taiwan)	JF303966	F.W. Li 608[a] (Taiwan)	JF303988
	Monachosorum henryi Christ*	L.Y. Kuo K18[a] (Taiwan)	JF303917	Wolf et al. (1994)	U05932	Korall et al. (2006)	AM176469
Pteridaceae	*Coniogramme japonica* (Thunb.) Diels*	K017026[c] (Taiwan)	JF303920	Zhang et al. (2007)	DC432658	K017026[c] (Taiwan)	JF303990
	Acrostichum aureum L.	K017715[c] (Taiwan)	JF303921	Said et al. (unpublished)	AB246708	K017715[c] (Taiwan)	JF303991
	Pteris fauriei Hieron.	K017569[c] (Taiwan)	JF303919	Hasebe et al. (1994)	U05647	K017569[c] (Taiwan)	JF303989
	Parahemionitis arifolia (Burm. f.) Panigrahi	K019685[c] (Taiwan)	JF303918	Geiger and Ranker (unpublished)	AY357706	K019685[c] (Taiwan)	JF303992
	Adiantum capillus-veneris L.*	Wolf et al. (2003)	NC_004766	Wolf et al. (2003)	NC_004766	Wolf et al. (2003)	NC_004766
	Haplopieris taeniophylla (Copel.) E. H. Crane	F.W. Li 547[a] (Taiwan)	JF303922	F.W. Li 547[a] (Taiwan)	JF303967	F.W. Li 547[a] (Taiwan)	JF303993
Dryopteridaceae	*Didymochlaena truncatula* (Sw.) J. Sm.	K017011[c] (cultivated)	JF303942	K017011[c] (cultivated)	JF303975	Schuettpelz et al. (2007)	EF452091
	Hypodematium crenatum (Forsk.) Kuhn*	K013155[c] (Taiwan)	JF303944	Schuettpelz and Pryer (2007)	EF463205	Schuettpelz and Pryer (2007)	EF463705
	Leucostegia immersa C. Presl	L.Y. Kuo 170[a] (Taiwan)	JF303943	Tsutsumi and Kato (2006)	AB232388	L.Y. Kuo 170[a] (Taiwan)	JF304009
	Pleocnemia rufinervis (Hayata) Nakai	K016926[c] (Taiwan)	JF303948	K016926[c] (Taiwan)	JF303976	K016926[c] (Taiwan)	JF304012
	Elaphoglossum callifolium (Blume) T. Moore	F.W. Li 399[a] (Taiwan)	JF303949	Tsutsumi and Kato (2006)	AB232400	F.W. Li 399[a] (Taiwan)	JF304013
	Ctenitis eatonii (Baker) Ching*	L.Y. Kuo 441[a] (Taiwan)	JF303947	Tsutsumi and Kato (2006)	AB232391	L.Y. Kuo 441[a] (Taiwan)	JF304011
	Dryopteris sparsa (Buch.-Ham. *ex* D. Don) Kuntze	W.T. Liu 481[a] (Taiwan)	JF303946	Li and Lu (2006)	AY587124	W.T. Liu 481[a] (Taiwan)	JF304010
	Cyrtomium falcatum (L. f.) C. Presl	L.Y. Kuo 828[a] (Taiwan)	JF303945	Schuettpelz and Pryer (2007)	EF463176	Schuettpelz and Pryer (2007)	EF463671
Lomariopsidaceae	*Lomariopsis spectabilis* (Kunze) Mett.	F.W. Li 568[a] (Taiwan)	JF303952	Tsutsumi and Kato (2006)	AB232401	F.W. Li 568[a] (Taiwan)	JF304015
	Cyclopeltis crenata (Fee) C. Chr.	K016940[c] (Thailand)	JF303954	Li and Lu (2006)	DQ054517	K016940[c] (Thailand)	JF304016
	Nephrolepis cordifolia (L.) C. Presl*	F.W. Li 335[a] (Taiwan)	JF303953	Tsutsumi and Kato (2006)	AB232404	Schuettpelz et al. (2007)	EF452103
Tectariaceae	*Tectaria zeylanica* (Houtt.) Sledge	K017295[c] (Taiwan)	JF303951	Schuettpelz and Pryer (2007)	EF463275	Schuettpelz and Pryer (2007)	EF463871
	Arthropteris palisotii (Desv.) Alston*	L.Y. Kuo 829[a] (Taiwan)	JF303950	L.Y. Kuo 829[a] (Taiwan)	JF303977	L.Y. Kuo 829[a] (Taiwan)	JF304014
Oleandraceae	*Oleandra wallichii* (Hook.) C. Presl*	L.Y. Kuo 830[a] (Taiwan)	JF303955	Tsutsumi and Kato (2005)	AB212687	L.Y. Kuo 830[a] (Taiwan)	JF304017
Davalliaceae	*Araiostegiella perdurans* (Christ) M. Kato and Tsutsumi	W.T. Liu 898[d] (Taiwan)	JF303957	Tsutsumi and Kato (2005)	AB212691	W.T. Liu 898[d] (Taiwan)	JF304019
	Humata repens (L. f.) J. Small *ex* Diels*	C.W. Chen 399[a] (Taiwan)	JF303956	Tsutsumi and Kato (2005)	AB212722	C.W. Chen 399[a] (Taiwan)	JF304018
Polypodiaceae	*Colysis wrightii* (Hook.) Ching	F.W. Li 596[a] (Taiwan)	JF303959	Kreier et al. (2008)	EU482954	F.W. Li 596[a] (Taiwan)	JF304021

(*continued on next page*)

Table 1 (*continued*)

Family	Tax on	*matK*		*rbcL*		*atpA*	
		Voucher/ citation	GenBank	Voucher/ citation	GenBank	Voucher/ citation	GenBank
	Aglaomorpha meyeniana Schott	K016952[c] (Taiwan)	JF303958	Janssen and Schneider (2005)	AY529153	K016952[c] (Taiwan)	JF304020
	Ctenopteris curtisii (Baker) Copel.	F.W. Li 593[a] (Taiwan)	JF303960	F.W. Li 593[a] (Taiwan)	JF303978	F.W. Li 593[a] (Taiwan)	JF304022
Woodsiaceae	*Gymnocarpium remote-pinnatum* (Hayata) Ching	L.Y. Kuo 149[a] (Taiwan)	JF303926	L.Y. Kuo 149[a] (Taiwan)	JF303969	L.Y. Kuo 149[a] (Taiwan)	JF303996
	Acystoptetis taiwaniana (Tagawa) A. Love and D. Love*	L.Y. Kuo 175[a] (Taiwan)	JF303925	L.Y. Kuo 175[a] (Taiwan)	JF303968	L.Y. Kuo 175[a] (Taiwan)	JF303995
	Diplaziopsis javanica (Blume) C. Chr.	L.Y. Kuo 138[a] (Taiwan)	JF303928	L.Y. Kuo 138[a] (Taiwan)	JF303970	L.Y. Kuo 138[a] (Taiwan)	JF303997
	Homalosorus pycnocarpos (Spreng.) Small *ex* Pic.Serm.	L.Y. Kuo 122[a] (N.America)	JF303929	Sano et al. (2000)	AB021722	L.Y. Kuo 122[a] (N. America)	JF303999
	Hemidictyum marginatum (L.) C. Presl	M. Christenhusz 2476[b] (French Guiana)	JF303927	Schuettpelz and Pryer (2007)	EF463318	Schuettpelz and Pryer (2007)	EF463922
	Rhachidosorus pulcher (Tagawa) Ching*	H.S. Cheng 441[d] (Taiwan)	JF303962	H.S. Cheng 441[d] (Taiwan)	JF303971	H.S. Cheng 441[d] (Taiwan)	JF303998
	Woodsia polystichoides D.C. Eaton*	F.W. Li 420[a] (Taiwan)	JF303930	Hasebe et al. (1994)	U05657	F.W. Li 420[a] (Taiwan)	JF304000
	Deparia lancea (Thunb.) Fras.-Jenk.*	L.Y. Kuo 112[a] (Taiwan)	JF303940	Schuettpelz and Pryer (2007)	EF463306	Schuettpelz and Pryer (2007)	EF463908
	Athyrium filix-femina (L.) Roth *ex* Mert.	L.Y. Kuo 117[a] (N.America)	JF303941	Skog et al. (2004)	AY818676	Schuettpelz and Pryer (2007)	EF463902
	Diplazium proliferum (Lam.) Thouars	K014215[c] (cultivated)	JF303939	K014215[c] (cultivated)	JF303974	K014215[c] (cultivated)	JF304008
Aspleniaceae	*Hymenasplenium unilaterale* (Lam.) Hayata*	F.W. Li 619[a] (Taiwan)	JF303924	Schuettpelz et al. (2007)	EF452140	Schuettpelz et al. (2007)	EF452078
	Asplenium tripteropus Nakai	L.Y. Kuo 186[a] (Taiwan)	JF303923	Murakami et al. (1999)	AB014699	L.Y. Kuo 186[a] (Taiwan)	JF303994
Thelypteridaceae	*Macrothelypteris torresiana* (Gaud.) Ching	L.Y. Kuo 826[a] (Taiwan)	JF303931	Schuettpelz and Pryer (2007)	EF463277	Schuettpelz and Pryer (2007)	EF463873
	Phegopieris connectilis (Michx.) Watt	L.Y. Kuo 151[a] (Taiwan)	JF303932	Smith and Cranfill (2002)	AF425111	L.Y. Kuo 151[a] (Taiwan)	JF304001
	Thelyptetis uraiensis (Rosenst.) Ching	L.Y. Kuo 139[a] (Taiwan)	JF303933	L.Y. Kuo 139[a] (Taiwan)	JF303972	L.Y. Kuo 139[a] (Taiwan)	JF304002
	Cyclosorus gymnopteridifrons (Hayata) C.M. Kuo	L.Y. Kuo 111[a] (Taiwan)	JF303961	L.Y. Kuo 111[a] (Taiwan)	JF303973	L.Y. Kuo 111[a] (Taiwan)	JF304003
	Cyclosorus pozoi (Lag.) C.M. Kuo	L.Y. Kuo 110[a] (Taiwan)	JF303934	Yatabe et al. (1998)	AB013340	L.Y. Kuo 110[a] (Taiwan)	JF304004
Onocleaceae	*Onoclea sensibilis* L.	L.Y. Kuo 115[a] (N.America)	JF303935	Gastony and Ungerer (1997)	U62036	Schuettpelz and Pryer (2007)	EF463793
	Matteuccia struthiopteris (L.) Tod.	K019687[c] (cultivated)	JF303936	Tsutsumi and Kato (2006)	AB232415	K019687[c] (cultivated)	JF304005
Blechnaceae	*Woodwardia japonica* (L. f.) Sm.	K017010[c] (Taiwan)	JF303937	Nakahira and Kato (unpublished)	AB040600	K017010[c] (Taiwan)	JF304006
	Blechnum orientale L.	L.Y. Kuo 827[a] (Taiwan)	JF303938	Nakahira and Kato (unpublished)	AB040568	L.Y. Kuo 827[a] (Taiwan)	JF304007

[a] Deposited in the Herbarium of the Taiwan Forestry Research Institute (TAIF).
[b] From fern DNA Database (http://www.pryerlab.net/DNA_database.shtml).
[c] Living collections from the Dr. Cecilia Koo Botanic Conservation Center (KBCC).
[d] Deposited in the Herbarium of the Biology Department of National Taiwan Normal University (TNU).

ferns, the *trnK* gene was lost after the earliest extant lineage of leptosporangiates (Osmundaceae) diverged – *Osmunda japonica* has it, but all other assayed leptosporangiates do not. In Hymenophyllaceae, the second earliest-diverging leptosporangiate lineage, no *trnK* exon-like sequence was found between *matK* and its closest gene in the 5′ region, *rps16* (data not shown).

The loss of the *trnK* gene in leptosporangiates has been attributed to an inversion in the cp genome that presumably disrupted the *trnK-matK-trnK* order (Stein et al., 1992; Wolf et al., 2003; Duffy et al., 2009; Gao et al., 2009). However, the inversion occurred after Hymenophyllaceae diverged (Wolf and Roper, 2008), and thus our findings indicate that the loss of *trnK* predates this event. A recent study by Wolf et al. (2010) drew the same inference based on chloroplast genome mappings.

4.2. Comparison of sequence characteristics among matK, rbcL, and atpA

In this study, we provide the first similarity comparison of *matK* between ferns and angiosperms and revealed that it has substantial variability in both plant groups (Table 3). We also contrasted the sequence characteristics of fern *matK* with *rbcL* and *atpA*, including variability, nonsynonymous/synonymous substitution ratio (dN/dS), homoplasy, and substitution evenness. Among these

Table 2
Primers used in this study, including sequences, target regions, target taxon and references. Primers marked by asterisks are applicable to most ferns.

Primer names	Primer sequences (5′–3′)	Target region	Target taxon[e]	References
Lb matK fDAP[a]	GATGCTYCGTTTCYACAT	*matK*	For gleichenoid ferns	This study
SchmatK fEMS[a]	GAGATGTCTACTTCGTTTCCT	*matK*	For *Schizaea*	This study
FERmatK fEDR*[a]	ATTCATTCRATRTTTTTATTTHTGGARGAYAGATT	*matK*	For all ferns	This study
Sch matK rRDS[b]	GGATGCCCAGTACTATCACG	*matK*	For *Schizaea* and *Lygodium*	This study
DeLin matK rNRD[b]	CTACGCAAYSCATCYCGATTT	*matK*	For lindsaeoids ferns	This study
FERmatK rAGK*[b]	CGTRTTGTACTYYTRTGTTTRCVAGC	*matK*	For all ferns	This study
Eul matK rGLR[b]	ATCTCAATCTMCGCAATCCAT	*matK*	For eupolypods I ferns and certain basal leptosporangiates	This study
FERN ch1B fYAA*[c]	GATGTRAYGTATGCRGCYAAAGA	*chlB*	For all ferns	This study
FERN rpsl6 f QCGR*[c]	CRMTRTGGTAGRAAGCAAC	*rpsl6*	For all ferns	This study
FERN rpsl6 fSRQE*[c]	CCCGRMRAGAAGGGARAG	*rpsl6*	For all ferns	This study
Ophio matK rIAI[d]	TRAGGAATYGATCGCAATTG	*matK*	For ophioglossoid ferns	This study
Equi matK rPFR[d]	CTTTGTTGAGAACCTGAAAGG	*matK*	For *Equisetum*	This study
MarOsm matK rWKR[d]	CGGGARAAYCGTTTCCAT	*matK*	For marattioid and osmundaceous ferns	This study
Lb matK rHYG[d]	GATTYYCGTATCTYCCATARTGRAT	*matK*	For filmy and gleichenoid ferns	This study
Lb matK rAPF[d]	GATAGATGKAGAAACGGAGC	*matK*	For gleichenoid ferns	This study
Lb matK rLIR[d]	GTCGRCGAAAYRTTCGARTCAAAG	*matK*	For schizaeoid ferns	This study
Lb matK rYIY[d]	ATYTCGTAGATRTARWRATTCCAA	*matK*	For tree and heterosporous ferns	This study
Tf matK rRLA[d]	GTATTTATGGAGAGCCAACCTTG	*matK*	For tree ferns	This study
Lin matK rRRR[d]	CATCTTGAATTCGTCGGCG	*matK*	For basal polypods ferns	This study
Lb matK rRKE[d]	GTGYTTCYTTCCGRRTAATG	*matK*	For basal polypods ferns	This study
Pt matK rIHY[d]	TTTCTMYATCTTSCRTARTGAAT	*matK*	For pteroid ferns	This study
EuII matK rHLL[d]	GTGARAAACYATCCTTARTAGATG	*matK*	For eupolypods II ferns	This study
EuII matK rTRK[d]	GARACACGTATTTTCTTTCTAGTCA	*matK*	For eupolypods II ferns	This study
EuII matK rHTY[d]	CACRARGTTTTGTACGTGTGA	*matK*	For eupolypods II ferns	This study
EuII matK rSDA[d]	CGATRCTTATYGCGTCTGA	*matK*	For eupolypods II ferns	This study
Eul matK rTYE[d]	TAACAATRTYGAATCRATTTCRTAAGT	*matK*	For eupolypods I ferns	This study
Eul matK rRLF[d]	GTCGRCGAAACARDCGAAC	*matK*	For eupolypods I ferns	This study
trnK 0118	GGGTTGCTAACTCAATGG	*trnK*	For ferns with *trnK* exons	This study
FernK rl	GGAACTAGTCGGATGAAAG	*trnK*	For ferns with *trnK* exons	This study
RbcLFlF	ATGTCACCACAAACAGAAACTAAAGCAAGT	*rbcL*	For all ferns	Wolf et al. (1994)
RbcLF1379R	TCACAAGCAGCAGCTAGTTCAGGACTC	*rbcL*	For all ferns	Wolf et al. (1994)
FERN atpA flSR	GCTCCRGGDATTATYTCRAGACG	*atpA*	For all ferns	This study
FERN atpA rAVA	ATTGTATCTGTAGCTACYGC	*atpA*	For all ferns	This study
ESTRNR46F	GTATAGGTTCRARTCCTATTGGACG	*trnR*	For all ferns	Schuettpelz et al. (2006)
ESATPF412F	GARCARGTTCGACAGCAAG	*atpF*	For all ferns	Schuettpelz et al. (2006)

[a] Forward primers that situated near the 5′ end of *matK* coding region; for amplification and sequencing.
[b] Reverse primers that situated near the 3′ end of *matK* coding region; for amplification and sequencing.
[c] Forward primers that situated on the 5′ upstream genes of *matK*; for amplification.
[d] Reverse lineage-specific primers that situated on the center of *matK* coding region; for amplification and sequencing.
[e] Names follow Schneider et al. (2004a,b) and Pryer et al. (2004).

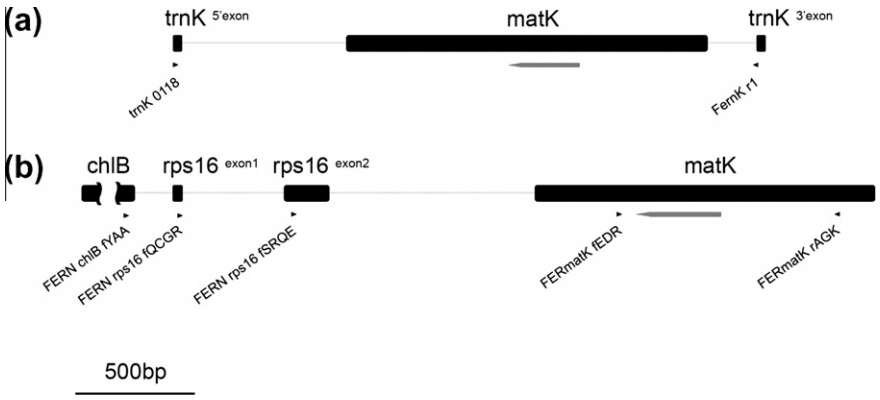

Fig. 1. Primer map and the gene order of fern *matK* and its flanking genes in the plastid genome. (a) The gene order in ferns with *trnK*, such as *Angiopteris* (Roper et al., 2007) and *Psilotum* (Wakasugi et al., 1998). (b) The gene order in ferns where *trnK* has been lost, such as *Alsophila* (Gao et al., 2009) and *Adiantum* (Wolf et al., 2003). Arrowheads indicate the position and direction of the primers that are applicable to most ferns (marked by asterisks in Table 2). Gray areas below *matK* show where lineage-specific reverse primers (marked by 2r in Table 2) were designed for amplifying the *matK* 5′ end and upstream regions.

Table 3
Sequence characteristics of fern *malK*, *rbcL* and *atpA*, and comparisons with their angiosperm counterparts. "–" indicates data not available.

	matK	*rbcL*	*atpA*	Taxon (number of species)	References
Characteristics					
Variability	96%	47%	55%	Ferns (78)	This study
	–	60%	62%	Leptosporangiate ferns (400)	Schuettpelz and Pryer (2007)
	71%	36%	–	Basal angiosperms (42)	Müller et al. (2006)
	70%	–	–	Seed plants (386)	Hilu et al. (2003)
GC content	35.4%	46.8%	43.7%	Ferns (78)	This study
	35.3%	45.2%	–	Basal angiosperms (42)	Müller et al. (2006)
Transition/transversion	2.084	3.943	5.702	Ferns (78)	This study
	1.454	2.108	–	Basal angiosperms (42)	Müller et al. (2006)
dN/dS	0.39	0.04	0.06	Fems (78)	This study
	–	0.17	–	Angiosperms (122)	Kapralov and Filatov (2007)
Substitution evenness					
1st codon position	0.97	0.70	0.78	Ferns (78)	This study
2nd codon position	0.96	0.65	0.77	Ferns (78)	This study
3rdcodon position	0.98	0.97	0.97	Ferns (78)	This study
All codon positions	0.97	0.85	0.88	Ferns (78)	This study

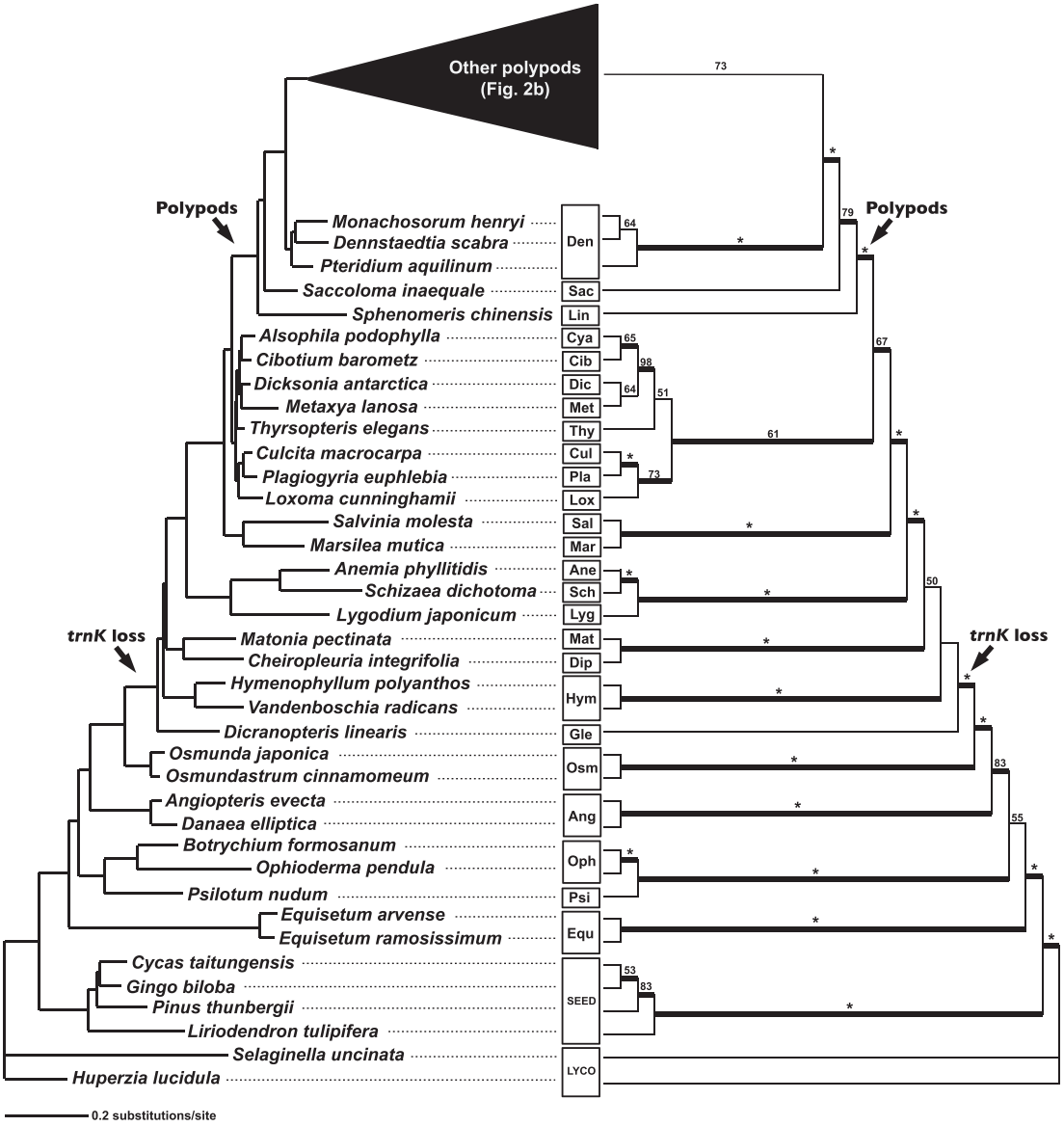

Fig. 2a. ML tree based on the three-gene dataset (*matK* + *atpA* + *rbcL*) from 78 fern taxa. Branches with Bayesian posterior probabilities ⩾0.95 are bolded in the cladogram. Values above the branches are ML bootstrap support (BS), and asterisks indicate BS = 100. Abbreviations in the boxes between the trees are: LYCO, lycophytes; SEED, seed plants; Equ, Equisetaceae; Psi, Psilotaceae; Oph, Ophioglossaceae; Ang, Marattiaceae; Osm, Osmundaceae; Gle, Gleicheniaceae; Hym, Hymenophyllaceae; Dip, Dipteridaceae; Mat, Matoniaceae; Lyg, Lygodiaceae; Sch, Schizaeaceae; Ane, Anemiaceae; Mar, Marsileaceae; Sal, Salviniaceae; Lox, Loxomataceae; Pla, Plagiogyriaceae; Cul, Culcitaceae; Thy, Thyrsopteridaceae; Met, Metaxyaceae; Dic, Dicksoniaceae; Cib, Cibotiaceae; Cya, Cyatheaceae; Lin, Lindsaeaceae; Sac, Saccolomataceae; Den, Dennstaedtiaceae. Family classification follows Smith et al. (2006).

three genes, fern *matK* is both the most variable and the least homoplastic (Fig. 3 and Table 3), a finding similar to what has been reported from angiosperms for these same plastid genes (Müller et al., 2006). The low homoplasy of *matK* is likely to have resulted from a more-neutral evolution pattern: *matK* shows the highest substitution evenness (SE) and the dN/dS ratio closest to 1 (Table 3). In contrast to *atpA* and *rbcL*, nucleotide substitutions in *matK* were uniformly dispersed even in the first and second codon positions (Fig. 4), where selection pressure is strongest.

4.3. New insights into fern phylogeny

Our three-gene tree topology is generally congruent with other multi-locus fern phylogenies (Pryer et al., 2001, 2004; Schneider et al., 2004a; Schuettpelz et al., 2006; Schuettpelz and Pryer, 2007; Rai and Graham, 2010). However, it differs in several noteworthy aspects.

Equisetum was resolved as the earliest diverging lineage of extant ferns, but without strong support (Fig. 2a). This relationship was also inferred by Wolf (1997) based on *atpB*, and by Rai and Graham (2010) based on a large multi-locus dataset. However, in most previous fern phylogenetic studies, *Equisetum* has been grouped with Marattiaceae (Pryer et al., 2001; Wikström and Pryer, 2005; Qiu et al., 2006, 2007), Psilotaceae (Schneider et al., 2009), or sister to Marattiaceae + leptosporangiates (Wikström and Pryer, 2005; Schuettpelz et al., 2006). All of these groupings have been only weakly supported and so the phylogenetic position of *Equisetum* remains ambiguous.

The Gleicheniaceae were shown to be sister to Matoniaceae + Dipteridaceae with high support in previous studies (Pryer

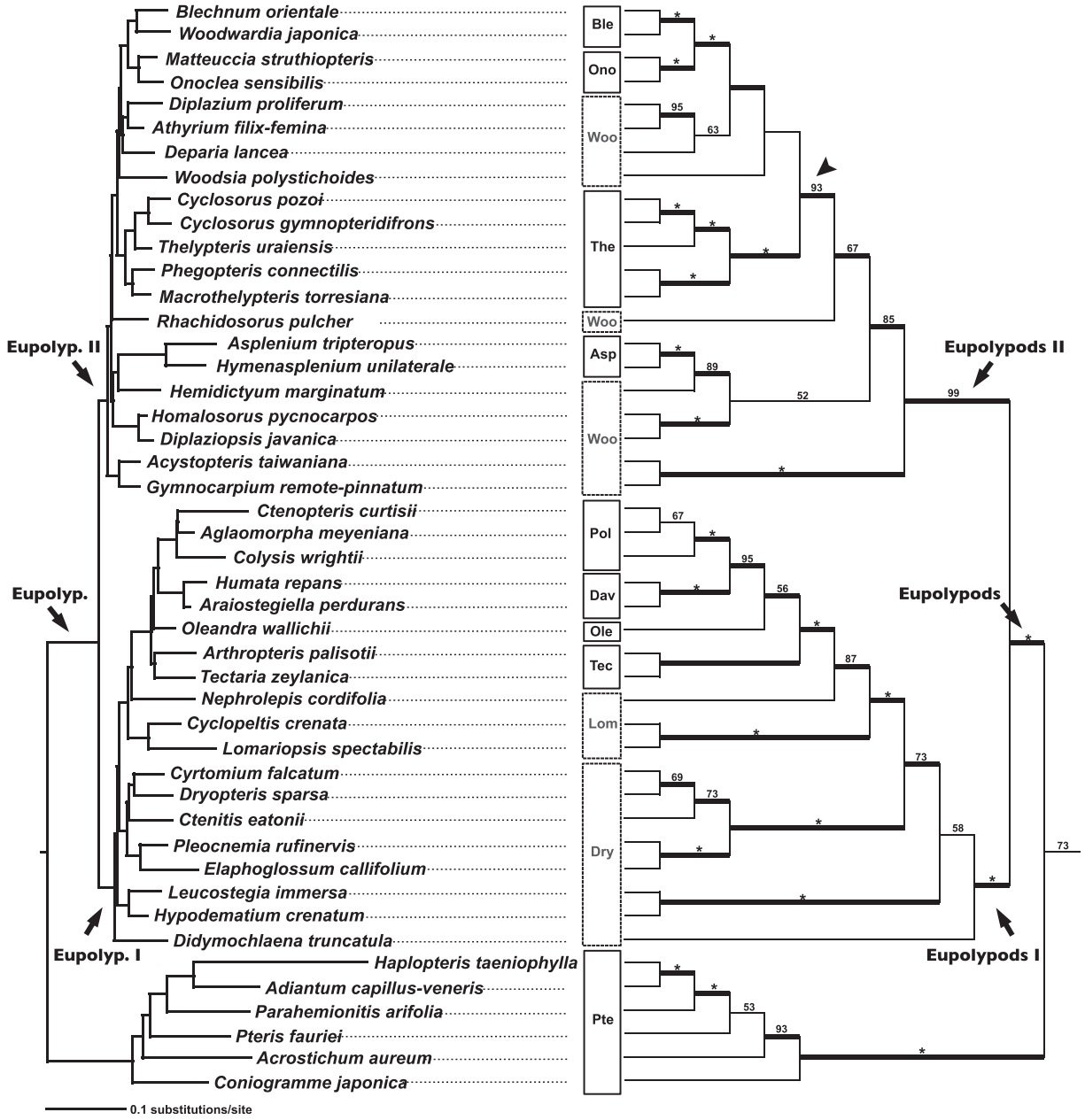

Fig. 2b. Continued from Fig. 2a. Abbreviations in the boxes between the trees are: Pte, Pteridaceae; Dry, Dryopteridaceae; Lom, Lomariopsidaceae; Tec, Tectariaceae; Ole, Oleandraceae; Dav, Davalliaceae; Pol, Polypodiaceae; Woo, Woodsiaceae; Asp, Aspleniaceae; The, Thelypteridaceae; Ono, Onocleaceae; Ble, Blechnaceae. Families in dashed boxes are shown here to be non-monophyletic and are the text also more faint. Arrowhead indicates the monophyly of Thelypteridaceae + Blechnaceae + Onocleaceae + *Woodsia* + *Deparia* + *Athyrium* + *Diplazium*. Arrows indicate Eupolypods (Eupolyp.), Eupolypods I (Eupolyp. I) and Eupolypods II (Eupolyp. II).

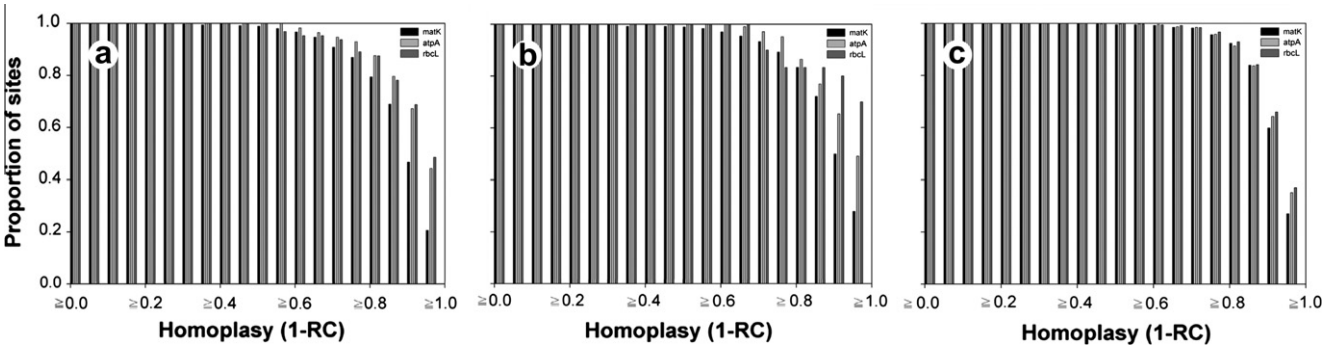

Fig. 3. Proportion of nucleotide sites with different homoplasy degrees (1 − RC) in the (a) first, (b) second, and (c) third codon positions.

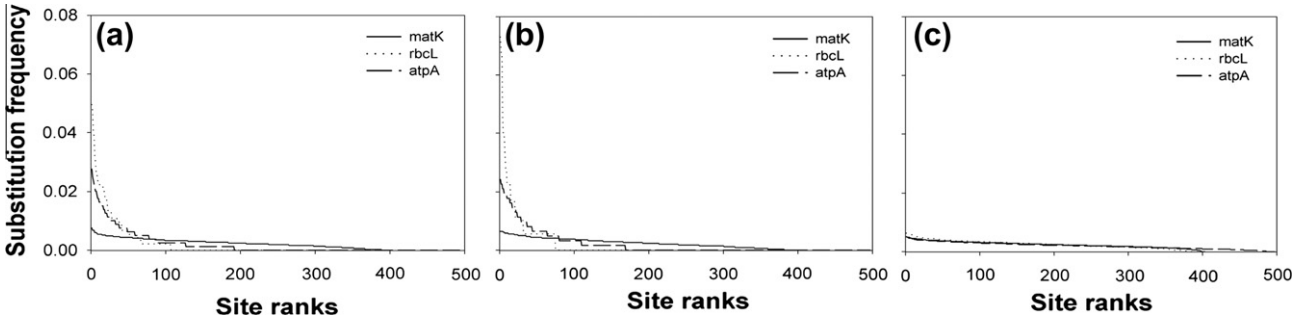

Fig. 4. Distribution of nucleotide substitution frequencies of fern *matK*, *rbcL*, and *atpA* in the (a) first, (b) second, and (c) third codon positions. Nucleotide sites are arranged along the *x*-axis from the highest substitution frequency to the lowest.

et al., 2004; Schuettpelz et al., 2006; Schuettpelz and Pryer, 2007). However, our analyses based on either the *matK* or three-gene datasets did not reveal the same relationship. Rather, our sole Gleicheniaceae accession (*Dicranopteris linearis*) was isolated as sister to all other leptosporangiates excluding Osmundaceae (Fig. 2a), but without support.

The first branch of the polypods was resolved and well-supported as Lindsaeaceae, which is in turn sister to Saccolomataceae + the rest of the polypods (Fig. 2a); however this grouping contradicts Schuettpelz and Pryer (2007) who resolved Lindsaeaceae and Saccolomataceae as a monophyletic clade sister to the remaining polypods. Future studies incorporating more loci and sampling are needed to better elucidate the basal divergence in polypods.

In eupolypods I, Smith et al. (2006) circumscribed the Lomariopsidaceae to comprise *Nephrolepis*, *Lomariopsis*, and *Cyclopeltis*, a relationship that was weakly supported by Schuettpelz and Pryer (2007) and Li et al. (2009). However, in our analyses, *Nephrolepis* was separated as sister to the Tectariaceae + Davalliaceae + Polypodiaceae with strong support (Fig. 2b). This position, relatively distant from *Cyclopeltis* + *Lomariopsis*, renders the Smith et al. (2006) circumscription of the Lomariopsidaceae as non-monophyletic.

The utility of *matK* is particularly evident in eupolypods II. Because of poor phylogenetic markers or a lack of sampling of important taxa in previous phylogenetic studies (Sano et al., 2000; Wang et al., 2003; Schuettpelz and Pryer, 2007), the overall family-level relationships in eupolypods II remains uncertain. Here we specifically focused on the genera in Woodsiaceae (*sensu* Smith et al., 2006). Compared to Schuettpelz and Pryer (2007), our three-gene dataset provided stronger support for: (1) *Gymnocarpium* and *Acystopteris* being the earliest-diverging branch in eupolypods II; (2) the sister relationship of *Hemidictyum* and Aspleniaceae; and (3) the monophyly of Thelypteridaceae + Blechnaceae + Onocleaceae + *Woodsia* + *Deparia* + *Athyrium* + *Diplazium* (indicated by the arrowhead in Fig. 2b). Our study is also the first to map three enigmatic Woodsiaceae genera, *Diplaziopsis*, *Homalosorus*, and *Rhachidosorus* onto a phylogenetic framework that comprises all fern families. *Diplaziopsis* and *Homalosorus* were resolved as a monophyletic group that is in turn sister to *Hemidictyum* + Aspleniaceae. *Rhachidosorus* was resolved, but not supported, as an isolated lineage that sister to the clade marked by the arrowhead in Fig. 2b. The placement of *Woodsia* is still ambiguous in our analyses, and clarification of its placement awaits further investigations.

4.4. Applying matK in future phylogenetic studies and DNA barcoding

matK and *rbcL* were approved as the DNA barcodes for land plants (http://www.barcoding.si.edu/plant_working_group.html) despite the failure of previous assessments to recover *matK* from ferns (Kress and Erickson, 2007; CBOL Plant Working Group, 2009). Here, we filled in the long-missing piece of the *matK* puzzle, and showed that *matK* in ferns is substantially variable, similar to its behavior in angiosperms. Our sequences and primers will serve as the basis for the ongoing endeavors of designing barcoding primers. In addition, we further demonstrated that fern *matK* has outstanding performance in phylogenetic reconstructions. Although non-coding regions may be as good or better in the properties we tested, *matK* sequence alignment is straightforward and unambiguous, making inferences at a broader scale possible. We hope that the introduction of *matK* to fern phylogenetics will facilitate future attempts to resolve uncertain and unexplored relationships, especially for ancient radiations such as those in eupolypods II.

Acknowledgments

We are grateful to Carl Rothfels, Dr. Kathleen Pryer and two anonymous reviewers for their detailed and valuable revision on

the manuscript, Dr. Jer-Ming Hu for his comments in analyses, and Aaron Duffy for sharing his unpublished *matK* sequences. We thank Mr. Cheng-Wei Chen, Mr. Hung-Hsin Chen, and Mr. Wei-Ting Liu for collection assistance, Dr. Kathleen Pryer, as well as the contributors to the Fern DNA Database, for providing silica-dried samples of *Equisetum arvense*, *Danaea elliptica*, *Matonia pectinata*, *Culcita macrocarpa*, *Thyrsopteris elegans*, *Metaxya lanosa*, *Saccoloma inaequale*, and *Hemidictyum marginatum*, and Dr. Cecilia Koo Botanic Conservation Center (KBCC) in Taiwan for providing living materials. KBCC was established in 2007 and dedicated to *ex situ* worldwide tropical plant conservation and protection. This study was partially supported by the National Science Council of Taiwan under an Undergraduate Research Grant (NSC97-2815-C-002-067-B) to F.-W. Li and and Academic Research Projects (NSC 95-2621-B-002-009-MY2; NSC 99-2923-B-002-007-MY2) to C.-N. Wang.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ympev.2011.03.010.

References

Cai, Z.-Q., Penaflor, C., Kuehl, J.V., Leebens-Mack, J., Carlson, J.E., dePamphilis, C.W., Boore, J.L., Jansen, R.K., 2006. Complete plastid genome sequences of *Drimys*, *Liriodendron*, and *Piper*: implications for the phylogenetic relationships of magnoliids. BMC Evol. Biol. 6, 77.

CBOL Plant Working Group, 2009. A DNA barcode for land plants. Proc. Natl. Acad. Sci. USA 106, 12794–12797.

Chaw, S.-M., Walters, T.W., Chang, C.-C., Hu, S.-H., Chen, S.-H., 2005. A phylogeny of cycads (Cycadales) inferred from chloroplast *matK* gene, *trnK* intron, and nuclear rDNA ITS region. Mol. Phylogenet. Evol. 37, 214–234.

Des Marais, D.L., Smith, A.R., Britton, D.M., Pryer, K.M., 2003. Phylogenetic relationships and evolution of extant horsetails, equisetum, based on chloroplast DNA sequence data (*rbcL* and *trnL-F*). Int. J. Plant Sci. 164, 737–751.

Döring, E., Schneider, J., Hilu, K.W., Röser, M., 2007. Phylogenetic relationships in the Aveneae/Poeae complex. Kew Bull. 62, 407–424.

Duffy, A.M., Kelchner, S.A., Wolf, P.G., 2009. Conservation of selection on *matK* following an ancient loss of its flanking intron. Gene 438, 17–25.

Ebihara, A., Iwatsuki, K., Ohsawa, T.A., Ito, M., 2003. *Hymenophyllum paniense* (Hymenophyllaceae), a new species of filmy fern from New Caledonia. Syst. Bot. 28, 228–235.

Ebihara, A., Ishikawa, H., Matsumoto, S., Lin, S.-J., Iwatsuki, K., Takamiya, N., Watano, Y., Ito, M., 2005. Nuclear DNA, chloroplast DNA, and ploidy analysis clarified biological complexity of the *Vandenboschia radicans* complex (Hymenophyllaceae) in Japan and adjacent areas. Am. J. Bot. 92, 1535–1547.

Funk, H.T., Berg, S., Krupinska, K., Maier, U.G., Krause, K., 2007. Complete DNA sequences of the plastid genomes of two parasitic flowering plant species, *Cuscuta reflexa* and *Cuscuta gronovii*. BMC Plant Biol. 7, 45.

Gao, L., Yi, X., Yang, Y.-X., Su, Y.-J., Wang, T., 2009. Complete chloroplast genome sequence of a tree fern *Alsophila spinulosa*: insights into evolutionary changes in fern chloroplast genomes. BMC Evol. Biol. 9, 130.

Gastony, G.J., Ungerer, M.C., 1997. Molecular systematics and a revised taxonomy of the onocleoid ferns (Dryopteridaceae: Onocleeae). Am. J. Bot. 84, 840–849.

Hasebe, M., Omori, T., Nakazawa, M., Sano, T., Kato, M., Iwatsuki, K., 1994. *RbcL* gene sequences provide evidence for the evolutionary lineages of leptosporangiate ferns. Proc. Natl. Acad. Sci. USA 7, 5730–5734.

Hauk, W.D., Parks, C.R., Chase, M.W., 2003. Phylogenetic studies of Ophioglossaceae: evidence from *rbcL* and *trnL-F* plastid DNA sequences and morphology. Mol. Phylogenet. Evol. 28, 131–151.

Hausner, G., Olson, R., Simon, D., Johnson, I., Sanders, E.R., Karol, K.G., McCourt, R.M., Zimmerly, S., 2006. Origin and evolution of the chloroplast *trnK* (*matK*) intron: a model for evolution of group II intron RNA structures. Mol. Biol. Evol. 23, 380–391.

Hilu, K.W., Barthet, M.M., 2008. Mode and tempo of *matK*: gene evolution and phylogenetic implications. In: Pontarotti, P. (Ed.), Evolutionary Biology from Concept to Application. Springer, Berlin and Heidelberg, Germany, pp. 165–179.

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

Hilu, K.W., Black, C., Diouf, D., Burleigh, J.G., 2008. Phylogenetic signal in *matK* vs. *trnK*: a case study in early diverging eudicots (angiosperms). Mol. Phylogenet. Evol. 48, 1120–1130.

Janssen, T., Schneider, H., 2005. Exploring the evolution of humus collecting leaves in drynarioid ferns (Polypodiaceae, Polypodiidae) based on phylogenetic evidence. Plant Syst. Evol. 252, 175–197.

Kapralov, M.V., Filatov, D.A., 2007. Widespread positive selection in the photosynthetic Rubisco enzyme. BMC Evol. Biol. 7, 73.

Kato, M., Setoguchi, H., 1998. An *rbcL*-based phylogeny and heteroblastic leaf morphology of Matoniaceae. Syst. Bot. 23, 391–400.

Kato, M., Yatabe, Y., Sahashi, N., Murakami, N., 2001. Taxonomic studies of *Cheiropleuria* (Dipteridaceae). Blumea 46, 513–525.

Korall, P., Pryer, K.M., Metzgar, J.S., Schneider, H., Conant, D.S., 2006. Tree ferns: monophyletic groups and their relationships as revealed by four protein-coding plastid loci. Mol. Phylogenet. Evol. 39, 830–845.

Kreier, H.-P., Zhang, X.-C., Muth, H., Schneider, H., 2008. The microsoroid ferns: inferring the relationships of a highly diverse lineage of paleotropical epiphytic ferns (Polypodiaceae, Polypodiopsida). Mol. Phylogenet. Evol. 48, 1155–1167.

Kress, W.J., Erickson, D.L., 2007. A two-locus global DNA barcode for land plants: the coding *rbcL* gene complements the non-coding *trnH-psbA* spacer region. PLoS ONE 2, e508.

Leebens-Mack, J., Raubeson, L.A., Cui, L.-Y., Kuehl, J.V., Fourcade, M.H., Chumley, T.W., Boore, J.L., Jansen, R.K., dePamphilis, C.W., 2005. Identifying the basal angiosperm node in chloroplast genome phylogenies: sampling one's way out of the Felsenstein zone. Mol. Biol. Evol. 10, 1948–1963.

Li, C.-X., Lu, S.-G., 2006. Phylogenetic analysis of Dryopteridaceae based on chloroplast *rbcL* sequences. Acta Phytotax. Sin. 44, 503–515.

Li, F.-W., Tan, B.C., Buchbender, V., Moran, R.C., Rouhan, G., Wang, C.-N., Quandt, D., 2009. Identifying a mysterious aquatic fern gametophyte. Plant Syst. Evol. 281, 77–86.

Lohne, C., Borsch, T., Wiersema, J.H., 2007. Phylogenetic analysis of Nymphaeales using fast-evolving and noncoding chloroplast markers. Bot. J. Linn. Soc. 154, 141–163.

Maddison, D.R., Maddison, E.P., 2003. MacClade 4.06. Sinauer Associates, Sunderland, Massachusetts.

Metzgar, J.S., Skog, J.E., Zimmer, E.A., Pryer, K.M., 2008. The paraphyly of *Osmunda* is confirmed by phylogenetic analyses of seven plastid loci. Syst. Bot. 33, 33–36.

Müller, K.F., Borsch, T., Hilu, K.W., 2006. Phylogenetic utility of rapidly evolving DNA at high taxonomical levels: contrasting *matK*, *trnT-F*, and *rbcL* in basal angiosperms. Mol. Phylogenet. Evol. 41, 99–117.

Murakami, N., Nogami, S., Watanabe, M., Iwatsuki, K., 1999. Phylogeny of Aspleniaceae inferred from *rbcL* nucleotide sequences. Am. Fern J. 89, 232–243.

Nagalingum, N.S., Schneider, H., Pryer, K.M., 2007. Molecular phylogenetic relationships and morphological evolution in the heterosporous fern genus *Marsilea*. Syst. Bot. 32, 16–25.

Nagalingum, N.S., Nowak, M.D., Pryer, K.M., 2008. Assessing phylogenetic relationships in extant heterosporous ferns (Salviniales), with a focus on *Pilularia* and *Salvinia*. Bot. J. Linn. Soc. 157, 673–685.

Neuhaus, H., Link, G., 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.

Perrie, L.R., Bayly, M.J., Lehnebach, C.A., Brownseyand, P.J., 2007. Molecular phylogenetic and molecular dating of the New Zealand Gleicheniaceae. Brittonia 59, 129–141.

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

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

Pryer, K.M., Schuettpelz, E., Wolf, P.G., Schneider, H., Smith, A.R., Cranfill, R., 2004. Phylogeny and evolution of ferns (monilophytes) with a focus on the early leptosporangiate divergences. Am. J. Bot. 91, 1582–1598.

Qiu, Y.-L., Li, L., Wang, B., Chen, Z., Knoop, V., Groth-Malonek, M., Dombrovska, O., Lee, J., Kent, L., Rest, J., Estabrook, G.F., Hendry, T.A., Taylor, D.W., Testa, C.M., Ambros, M., Crandall-Stotler, B., Duff, R.J., Stech, M., Frey, W., Quandt, D., Davis, C.C., 2006. The deepest divergences in land plants inferred from phylogenomic evidence. Proc. Natl. Acad. Sci. USA 103, 15511–15516.

Qiu, Y.-L., Li, L., Wang, B., Chen, Z., Dombrovska, O., Lee, J., Kent, L., Li, R., Jobson, R.W., Hendry, T.A., Taylor, D.W., Testa, C.M., Ambros, M., 2007. A nonflowering land plant phylogeny inferred from nucleotide sequences of seven chloroplast, mitochondrial, and nuclear genes. Int. J. Plant Sci. 168, 691–708.

Rai, H.S., Graham, S.W., 2010. Utility of a large, multigene plastid data set in inferring higher-order relationships in ferns and relatives (monilophytes). Am. J. Bot. 97, 1444–1456.

Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574.

Roper, J.M., Hansen, S.K., Wolf, P.G., Karol, K.G., Mandoli, D.F., Everett, K.D.E., Kuehl, J., Boore, J.L., 2007. The complete plastid genome sequence of *Angiopteris evecta* (G. Forst.) Hoffm. (Marattiaceae). Am. Fern J. 97, 95–106.

Rozas, J., Sanchez-DelBarrio, J.C., Messeguer, X., Rozas, R., 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19, 2496–2497.

Sano, R., Takamiya, M., Ito, M., Kurita, S., Hasebe, M., 2000. Phylogeny of the lady fern group, tribe Physematieae (Dryopteridaceae), based on chloroplast *rbcL* gene sequences. Mol. Phylogenet. Evol. 15, 403–413.

Schneider, H., Schuettpelz, E., Pryer, K.M., Cranfill, R.B., Magallón, S., Lupia, R., 2004a. Ferns diversified in the shadow of angiosperms. Nature 428, 553–557.

Schneider, H., Russell, S.J., Cox, C.J., Bakker, F., Henderson, S., Rumsey, F., Barrett, J., Gibby, M., Vogel, J.C., 2004b. Chloroplast phylogeny of asplenioid ferns based on *rbcL* and *trnL-F* spacer sequences (Polypodiidae, Aspleniaceae) and its implications for biogeography. Syst. Bot. 29, 260–274.

Schneider, H., Smith, A.R., Pryer, K.M., 2009. Is morphology really at odds with molecules in estimating fern phylogeny? Syst. Bot. 29, 455–475.

Schuettpelz, E., Pryer, K.M., 2007. Fern phylogeny inferred from 400 leptosporangiate species and three plastid genes. Taxon 56, 1037–1050.
Schuettpelz, E., Korall, P., Pryer, K.M., 2006. Plastid *atpA* data provide improved support for deep relationships among ferns. Taxon 55, 897–906.
Schuettpelz, E., Schneider, H., Huiet, L., Windham, M.D., Pryer, K.M., 2007. A molecular phylogeny of the fern family Pteridaceae: assessing overall relationships and the affinities of previously unsampled genera. Mol. Phylogenet. Evol. 44, 1172–1185.
Shannon, C.E., 1948. A mathematical theory of communication. Bell Syst. Tech. J. 27 (379–423), 623–656.
Skog, J.E., Mickel, J.T., Moran, R.C., Volovsek, M., Zimmer, E.A., 2004. Molecular studies of representative species in the fern genus *Elaphoglossum* (Dryopteridaceae) based on cpDNA sequences *rbcL*, *trnL-F*, and *rps4-trnS*. Int. J. Plant Sci. 165, 1063–1075.
Smith, A.R., Cranfill, R.B., 2002. Intrafamilial relationships of the thelypteroid ferns (Thelypteridaceae). Am. Fern J. 92, 131–149.
Smith, A.R., Tuomisto, H., Pryer, K.M., Hunt, J.S., Wolf, P.G., 2001. *Metaxya lanosa*, a second species in the genus and fern family Metaxyaceae. Syst. Bot. 26, 480–486.
Smith, A.R., Pryer, K.M., Schuettpelz, E., Korall, P., Schneider, H., Wolf, P.G., 2006. A classification for extant ferns. Taxon 55, 705–731.
Smith, A.R., Pryer, K.M., Schuettpelz, E., Korall, P., Schneider, H., Wolf, P.G., 2008. Fern classication. In: Ranker, T.A., Haufler, C.H. (Eds.), Biology and Evolution of Ferns and Lycophytes. Cambridge University Press, New York, pp. 159–174.
Stein, D.B., Conant, D.S., Ahearn, M.E., Jordan, E.T., Kirch, S.A., Hasebe, M., Iwatsuki, K., Tan, M.K., Thomson, J.A., 1992. Structural rearrangements of the chloroplast genome provide an important phylogenetic link in ferns. Proc. Natl. Acad. Sci. USA 89, 1856–1860.
Sugita, M., Shinozaki, K., Sugiura, M., 1985. Tobacco chloroplast $tRNA^{Lys}$(UUU) gene contains a 2.5-kilobase-pair intron: an open reading frame and a conserved boundary sequence in the intron. Proc. Natl. Acad. Sci. USA 82, 3557–3561.
Tamura, K., Dudley, J., Nei, M., Kumar, S., 2007. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol. Biol. Evol. 24, 1596–1599.
Tsuji, S., Ueda, K., Nishiyama, T., Hasebe, M., Yoshikawa, S., Konagaya, A., Nishiuchi, T., Yamaguchi, K., 2007. The chloroplast genome from a lycophyte (microphyllophyte), *Selaginella uncinata*, has a unique inversion, transpositions and many gene losses. J. Plant Res. 120, 281–290.
Tsutsumi, C., Kato, M., 2005. Molecular phylogenetic study on Davalliaceae. Fern Gaz. 17, 147–162.
Tsutsumi, C., Kato, M., 2006. Evolution of epiphytes in Davalliaceae and related ferns. Bot. J. Linn. Soc. 151, 495–510.
Wakasugi, T., Tsudzuki, J., Ito, S., Nakashima, K., Tsudzuki, T., Sugiura, M., 1994. Loss of all *ndh* genes as determined by sequencing the entire chloroplast genome of the black pine *Pinus thunbergii*. Proc. Natl. Acad. Sci. USA 91, 9794–9798.
Wakasugi, T., Nishikawa, A., Yamada, K., Sugiura, M., 1998. A complete nucleotide sequence of the plastid genome from a fern, *Psilotum nudum*. Endocyt. Cell Res. 13(Suppl.), 147.
Wang, M.-L., Chen, Z.-D., Zhang, X.-C., Lu, S.-G., Zhao, G.-F., 2003. Phylogeny of the Athyriaceae: evidence from chloroplast *trnL-F* region sequences. Acta Phytotax. Sin. 4, 416–426.
Wang, C.-N., Moller, M., Cronk, Q.C.B., 2004. Phylogenetic position of *Titanotrichum oldhamii* (Gesneriaceae) inferred from four different gene regions. Syst. Bot. 29, 407–418.
Wikström, N., Pryer, K.M., 2005. Incongruence between primary sequence data and the distribution of a mitochondrial atp 1 group II intron among ferns and horsetails. Mol. Phylogenet. Evol. 36, 484–493.
Wikström, N., Kenrick, P., Vogel, J.C., 2002. Schizaeaceae: a phylogenetic approach. Rev. Palaeobot. Palynol. 119, 5–50.
Wolf, P.G., 1997. Evaluation of *atpB* nucleotide sequences for phylogenetic studies of ferns and other pteridophytes. Am. J. Bot. 84, 1429–1440.
Wolf, P.G., Roper, J.M., 2008. Structure and evolution of fern plastid genomes. In: Ranker, T.A., Haufler, C.H. (Eds.), Biology and Evolution of Ferns and Lycophytes. Cambridge University Press, New York, pp. 159–174.
Wolf, P.G., Soltis, P.S., Soltis, D.E., 1994. Phylogenetic relationships of dennstaedtioid ferns: evidence from *rbcL* sequences. Mol. Phylogenet. Evol. 3, 383–392.
Wolf, P.G., Rowe, C.A., Sinclair, R.B., Hasebe, M., 2003. Complete nucleotide sequence of the chloroplast genome from a leptosporangiate fern, *Adiantum capillus-veneris* L. DNA Res. 10, 59–65.
Wolf, P.G., Karol, K.G., Mandoli, D.F., Kuehl, J., Arumuganathan, K., Ellis, M.W., Mishler, B.D., Kelch, D.G., Olmstead, R.G., Boore, J.L., 2005. The first complete chloroplast genome sequence of a lycophyte, *Huperzia lucidula* (Lycopodiaceae). Gene 350, 117–128.
Wolf, P.G., Roper, J.M., Duffy, A.M., 2010. The evolution of chloroplast genome structure in ferns. Genome 53, 731–738.
Wolfe, A.D., Morden, C.W., Palmer, J.D., 1992. Function and evolution of a minimal plastid genome from a non-photosynthetic plant. Proc. Natl. Acad. Sci. USA 89, 10648–10652.
Wu, C.-S., Wang, Y.-N., Liu, S.-M., Chaw, S.-M., 2007. Chloroplast genome (cpDNA) of *Cycas taitungensis* and 56 cp protein-coding genes of *Gnetum parvifolium*: insights into cpDNA evolution and phylogeny of extant seed plants. Mol. Biol. Evol. 24, 1366–1379.
Xie, L., Yi, T.-S., Li, R., Li, D.-Z., Wen, J., 2010. Evolution and biogeographic diversification of the witch-hazel genus (Hamamelis L., Hamamelidaceae) in the Northern Hemisphere. Mol. Phylogenet. Evol. 56, 675–689.
Yang, Z.-H., 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24, 1586–1591.
Yatabe, Y., Takamiya, M., Murakami, N., 1998. Variation in the *rbcL* sequence of *Stegnogramma pozoi* subsp. *mollissima* (Thelypteridaceae) in Japan. J. Plant Res. 111, 557–564.
Zhang, G.-M., Zhang, X.-C., Chen, Z.-D., Liu, H.-M., Yang, W.-L., 2007. First insights in the phylogeny of Asian cheilanthoid ferns based on sequences of two chloroplast markers. Taxon 56, 69–378.
Zwickl, D.J., 2006. Genetic Algorithm Approaches for the Phylogenetic Analysis of Large Biological Sequence Datasets under the Maximum Likelihood Criterion. PhD Dissertation, University of Texas at Austin, Austin, TX.