American Journal of Botany 94(8): 1402–1412. 2007.

EXPRESSION OF *MATK*: FUNCTIONAL AND EVOLUTIONARY IMPLICATIONS[1]

MICHELLE M. BARTHET[2] AND KHIDIR W. HILU

Virginia Tech, Blacksburg, Virginia 24061 USA

Strong phylogenetic signal from *matK* has rendered it an invaluable gene in plant systematic and evolutionary studies at various evolutionary depths. Further, *matK* is proposed as the only chloroplast-encoded group II intron maturase, thus implicating MATK in chloroplast posttranscriptional processing. For a protein-coding gene, *matK* has an unusual evolutionary mode and tempo, including relatively high substitution rates at both the nucleotide and amino acids levels. These evolutionary features have raised questions about *matK* function. In the current study, we examined *matK* RNA and protein from representative land plant species to provide insight into functional aspects of this unusual gene. We report the first evidence of a transcript for *matK* separate from the *trnK* precursor and demonstrate that a full-length MATK protein exists in five angiosperm species. We also show that *matK* RNA and protein levels are regulated by light and developmental stage, suggesting functional roles for this putative maturase. Specifically, *matK* expression increased after etiolation and decreased at 4 weeks after germination. This work provides evidence for the expression of the only putative chloroplast-encoded group II intron maturase and insight into regulation mechanisms relating to plant development and, indirectly, to photosynthesis.

Key words: developmental stage; etiolation; expression; *matK*; maturase; systematics.

In plant systematics, *matK* has recently emerged as an invaluable gene because of its high phylogenetic signal compared with other genes used in this field (Müller et al., 2006). The 1500 bp *matK* gene is nested in the group II intron between the 5′ and 3′ exons of *trnK* in the large single copy region of the chloroplast genome of most green plants (e.g., Sugita et al., 1985; Steane, 2005; Daniell et al., 2006; Turmel et al., 2006). Phylogenetic analysis of a data set composed of *matK*, *rbcL*, and *trnT-F* sequences from basal angiosperms demonstrated that *matK* contributes more parsimony informative characters and significantly more phylogenetic structure on average per parsimony-informative site than the highly conserved chloroplast gene *rbcL* (Müller et al., 2006). Sequence information from *matK* alone has generated phylogenies as robust as those constructed from data sets comprised of 2–11other genes combined (see Hilu et al., 2003). Further, the molecular information generated from *matK* has been used to resolve phylogenetic relationships from shallow to deep taxonomic levels (Johnson and Soltis, 1994; Hayashi and Kawano, 2000; Hilu et al., 2003; Cameron, 2005).

The *matK* gene stands out among plastid genes used in plant systematics in its distinct mode and tempo of evolution. The rate of substitution in *matK* is three times higher at the nucleotide level and is six times higher at the amino acid level than that of *rbcL* (Johnson and Soltis, 1994; Olmstead and Palmer, 1994), denoting it as a fast or rapidly evolving gene (Soltis and Soltis, 2004). The accelerated rate of amino acid substitution in *matK* is due to almost even distribution of substitution rates among the three codon positions compared with most protein-coding genes where the rates are skewed toward the third codon position. For example, substitution rates for the three codon positions of *matK* were 62%, 57%, and 66%, respectively, when comparing angiosperms at the ordinal level (Hilu et al., 2003), and 32%, 28%, and 39%, respectively, in Orchidaceae (Whitten et al., 2000). Insertions and deletions (indels) are frequent in *matK*, though these indels primarily occur in multiples of three, maintaining the reading frame. Further, the transition/transversion ratio in the gene approaches unity in sequences examined at both deep (order) and shallow (subfamily) taxonomic levels (Liang and Hilu, 1996; Hilu et al., 2003). The rapid rate of substitution, along with the rare presence of frameshift indels and a few cases of premature stop codons, prompted some researchers to suggest that *matK* may not be functional in some plants (Kores et al., 2000; Whitten et al., 2000; Kugita et al., 2003; Hidalgo et al., 2004; Jankowiak et al., 2004).

Sequence features alone, however, provide inconclusive evidence for lack of function. RNA editing mechanisms previously reported in *matK* (Vogel et al., 1997; Tillich et al., 2005) may correct the reading frame in species with frameshift indels and premature stop codons and restore the codon identities needed to form the proper amino acids for function. Further, genome studies of the holoparasite *Epifagus virginiana* (Ems et al., 1995) and *Adiantum capillus-veneris* (Wolf et al., 2003) support that *matK* has a function in the plant. *Epifagus virginiana,* which has lost more than 60% of its chloroplast genome including *trnK*, retained *matK* (Ems et al., 1995). Similarly, a large rearrangement in the chloroplast genome of *Adiantum* resulted in the loss of *trnK* but retention of *matK* (Wolf et al., 2003). The maintenance of the *matK* gene in both species after loss of several other genes suggests that this gene is expressed and serves an essential and irreplaceable function in the plant.

In addition to the importance of *matK* in plant phylogenetics, it is also the only putative group II intron maturase encoded in the chloroplast genome (Neuhaus and Link, 1987). Maturases

[1] Manuscript received 27 August 2006; revision accepted 15 June 2007.

The authors appreciate the support of this research by NSF grant EF-0431057, NSF Deep Time, Sigma Xi, Virginia Academy of Sciences, and Virginia Tech. They also thank D. Quandt for help collecting plant material and primer design, C. Neinhaus and D. Jarrell for providing plant material, and D. Wiley-Vawter and T. Wieboldt for help with voucher specimens. Special thanks to D. Bevan for assistance with peptide antigen design, S. Drysdale and S. Majumder for assistance in the laboratory, and S. Parker and the anonymous reviewers for comments on the manuscript.

[2] Author for correspondence (e-mail: michelle.barthet@usyd.edu.au); current address: School of Biological Sciences, University of Sydney, NSW 2006, Australia

TABLE 1. Species for which an RT-PCR product for *matK* was observed, their respective taxonomic group, and size of RT-PCR product.

Major taxonomic group	Lineage	Family	Species	Product size (bp)
Bryophytes		Bartramiaceae	*Bartramia pomiformis*	690
		Polytrichaceae	*Atrichum altecristatum*	690
		Anthocerotaceae	*Phaeoceros laevis*	665
Monilophytes		Adiantaceae	*Adiantum hispidulum*	450
Angiosperms	Basal lineage	Nymphaeaceae	*Nymphea oderata*	522
	Monocots	Alismataceae	*Sagitaria latifolia*	503
		Poaceae	*Oryza sativa*	876
		Poaceae	*Zea mays*	876
		Orchidaceae	*Spathoglottis plicata*	449
	Core eudicots	Solanaceae	*Solanum tuberosum*	333
		Brassicaceae	*Arabidopsis thaliana*	621

Note: RT = reverse-transcriptase.

are enzymes that catalyze nonautocatalytic intron removal from premature RNAs. Maturases generally contain three domains: a reverse-transcriptase (RT) domain, domain X (the proposed maturase functional domain), and a zinc-finger-like domain (Mohr et al., 1993). The 3′ region of *matK* has homology to the domain X of mitochondrial group II intron maturases (Neuhaus and Link, 1987). This region of *matK* also lacks indels (Hilu and Liang, 1997), indicating evolutionary constraint and conservation of function. Among higher plants, *matK* is the only plastid gene containing this putative maturase domain (Neuhaus and Link, 1987). The gene, however contains only remnants of the RT domain and no evidence of the zinc-finger-like domain found in other maturases (Mohr et al., 1993). Although at present there is no direct evidence for the maturase function of MATK, indirect evidence from the white barley ribosomal mutant *albostrians* supports this proposed function. Vogel et al. (1997, 1999) demonstrated that the *albostrians* mutant, which lacks all chloroplast ribosome activity and, subsequently, all chloroplast proteins including MATK, lacks group II intron excision from certain premature RNAs. Based on this and other studies, researchers have proposed that the following RNA transcripts require MATK for intron excision: *trnK, trnA, trnI*, *rps12*, *rpl2*, and *atpF* (Ems et al., 1995; Jenkins et al., 1997; Vogel et al., 1999). The tRNA or protein products from these genes are required for normal chloroplast function including photosynthesis. The above considerations suggest that MATK has an essential function in the chloroplast as a posttranscription splicing factor.

Previous studies examining *matK* RNA and protein have yielded conflicting results concerning whether this gene is expressed. While several studies have observed a *matK*/*trnK* transcript (Kanno and Hirai, 1993; Vogel et al., 1997; Kugita et al., 2003; Nakamura et al., 2003; Wolf et al., 2004), the majority of these studies did not determine whether this transcript proceeded to translation of MATK protein or was merely a premature unspliced *trnK* tRNA that happened to include the *matK* open reading frame (ORF). Only one study, which used *Hordeum vulgare* L. (barley, Poaceae) investigated whether a monocistronic *matK* transcript existed separate from *trnK* (Vogel et al., 1997). The results from this study indicated that a *matK* transcript may not exist separately from a *trnK* unspliced precursor. This finding suggests that *matK* is not translated into protein, or is translated from a transcript present in very low levels (Vogel et al., 1997), or cannot be easily distinguished from the *trnK* precursor. Thus, the question of whether *matK* is an expressed gene with its own transcript separate from premature *trnK* remains largely unanswered.

Three studies have presented evidence for the existence of a MATK protein. Two of these studies, du Jardin et al. (1994) using *Solanum tuberosum* L. (potato, Solanaceae) and Liere and Link (1995) using *Sinapis alba* L. (Brassicaceae), identified a MATK protein of much lower molecular mass than expected for the full-length ORF. These results indicate that MATK is truncated. In contrast, a later study using barley identified a protein product close to the expected molecular mass for full-length MATK (Vogel et al., 1999), confirming that *matK* is expressed and that the protein is not truncated in this plant species. Consequently, further studies at the RNA and protein level are needed to examine expression and function of this gene.

The objectives of the current study were two fold: (1) to determine the presence of transcription and translation products for *matK* and assess the relationship of *matK* to the expression of *trnK*, and (2) to examine possible functions for this putative maturase. To address these objectives, we first determined the number and size of transcripts generated from the *matK*/*trnK* gene region in two model species *Oryza sativa* L. (rice, Poaceae) and potato. Second, we used several methods to examine whether *matK* is transcribed independently from *trnK*. Third, we investigated whether a protein product of the expected size for full-length MATK is produced in several angiosperm species. Fourth, we investigated possible functions for MATK in rice by assessing RNA and protein levels in response to etiolation and during postgermination development. This work provides the first detailed examination of the expression of this potentially essential chloroplast maturase.

MATERIALS AND METHODS

Plant material for RNA analyses—Two angiosperms models, one monocot (rice) and one eudicot (potato), were used in this study to avoid possible bias to a single angiosperm species. The choice of using potato vs. a more common eudicot model such as *Arabidopsis* for our expression studies was based on a previous investigation of *matK* expression and function, which used potato as the model (du Jardin et al., 1994). In addition to these two models, 14 species representing the four major land plant lineages (bryophytes, monilophytes, gymnosperms, and angiosperms) were used to survey *matK* RNA (Appendix 1). Species chosen within angiosperms represented basal and late-diverging lineages (Table 1). Taxa sampling in monocots was extended to represent sections of their phylogeny (Hilu et al., 2003) from early diverging Alismatales (*Sagitaria latifolia* Willd., Alismataceae), to later diverging Poales (rice and *Zea mays* L., Poaceae) and Asparagales (*Spathoglottis plicata* Blume, Orchidaceae). The orchid *Spathoglottis* was specifically chosen to investigate comments that *matK* is a pseudogene in the chloroplast genome of this plant (Freudenstein et al., 2004). By including sample taxa of basal-, mid-, and late-diverging lineages, our survey accounts for differences in angiosperms due to evolutionary distance and presents a broad representation across this plant group. Plants used in the current work were obtained from three sources: seed stock at the Virginia Tech Biology greenhouse, field collections, or potted plants purchased from private companies (for details of plant collections see Appendix 1). Both young and adult leaf material were collected for each plant, placed in zip-seal freezer bags, and stored at −80°C. Herbarium vouchers are deposited in the herbarium at Virginia Tech (VPI), Technische Universität Dresden (DR), or as part of the personal collection of Dietmar Quandt (Quandt), Technische Universität Dresden.

Probe synthesis—Nonradioactive digoxygenin (Dig)-labeled probes for nucleic acid hybridization were generated from rice or potato genomic DNA that was isolated using the CTAB method (Doyle and Doyle, 1990). The rice *matK* ORF probe consisted of 876 base pairs (bp) that included 250 bp of the conserved maturase domain X (Mohr et al., 1993) and was PCR-amplified

using primers W and 9R (for information on primers, see Appendix S1 in Supplemental Data accompanying the online version of this article) (Fig. 1A). The potato *matK* ORF probe consisted of a 333-bp product of the *matK* N-terminal region and was amplified using the primers corematK1 and corematK2. Specificity of these two probes to *matK* was tested by Southern blot (see following section on Southern blot). Two rice *trnK* probes were designed to exclusively hybridize to the 5′ or 3′ exon of *trnK* (Fig. 1A). The primers 5extrnKF and trnK5exR were used to amplify a sequence segment starting from 132 bp upstream of the beginning of the 5′ *trnK* exon to its end to generate the 5′ *trnK* exon probe. Primers trnK3exF and trnK3exR were used to amplify a sequence section starting from 71 bp upstream of the beginning of the *trnK* 3′ exon to the end of that exon to generate the 3′ *trnK* exon probe. Dig-labeled probe synthesis was performed according to supplier instructions (Roche, Indianapolis, Indiana, USA).

Southern hybridization—The 876-bp rice *matK* ORF probe, which included part of domain X, and 333 bp Dig-labeled potato *matK* ORF probe were tested for target specificity against Southern blot membranes cross-linked by ultraviolet light with PCR product for *matK*, *rbcL*, and *mat-r*. The *mat-r* gene codes for a mitochondrial maturase containing domain X (Farré and Araya, 1999). The chloroplast genes *matK* and *rbcL* and the mitochondrial gene *mat-r* were amplified from rice genomic DNA using primer combinations W/9R (*matK*), rbcLOsF/rbcLOsR (*rbc*L), and matRF/matRrevdX (*mat-r*). Amplification product of *mat-r* included 186 bp of domain X. Amplified products were separated on a 1.5% agarose gel, transferred to a nylon membrane by capillary blotting in Church buffer (Church and Gilbert, 1984), and hybridized with the *matK* ORF probes from rice and potato at low (50°C) and high (65°C) stringency. Probe signal was detected with anti-Dig and the chemiluminescent substrate CDP-Star (Roche, Indianapolis, Indiana, USA) followed by exposure on film.

RNA isolation and Northern hybridization analyses—Total RNA was isolated by grinding tissue under liquid nitrogen followed by phenol/chloroform/LiCl extraction and ethanol precipitation (Altenbach and Howell, 1981). RNA was separated on a 1% formaldehyde gel according to Gerard and Miller (1986). RNA was transferred to a nylon membrane via capillary blotting in Church buffer and hybridized with the Dig-labeled *matK* ORF probes (described in the previous sections) at both low and high stringency or with the Dig-labeled *trnK* 5′ and 3′ exon probes at high stringency. Because RNA editing of *matK* has been observed previously (Vogel et al., 1997; Wolf et al., 2004), a low stringency hybridization was performed to ensure binding of *matK* ORF DNA probes to possible RNA-edited transcripts of this gene, which may vary slightly in sequence. Transcripts on membranes were visualized by incubation with CDP-Star followed by exposure on film.

3′ RACE—Total RNA was isolated as described previously from rice tissue and stored at −80°C. RNA was DNase-treated using RQ1 RNase-free DNase (Promega, Madison, Wisconsin, USA) with the following modifications to the manufacturer's protocol: 1 μL of RNase inhibitor (40 u/μL, Invitrogen, Carlsbad, California, USA) was added along with 6.7 μL of 5× reverse transcriptase buffer (Invitrogen), 4.3 μL of DNase (1 u/μL), and up to 10 μg of total RNA for a 20-μL reaction. 3′ RACE was performed using the TaKaRa 3′-Full RACE Core Set (TAKARA BIO, Otsu, Shiga, Japan) according to the manufacturer's directions with the exception of using trnK3exR as the reverse primer instead of oligo dT in the RT reaction for mature *trnK* cDNA. The primer combinations W/trnK3exR, W/9R, and 3914/trnK3exR were used in a PCR reaction with cDNA generated from 3′ RACE to amplify transcripts of the *matK* ORF to the end of the 3′ *trnK* exon, the *matK* ORF alone, and the mature transcript of *trnK*, respectively. A negative control lacking reverse transcriptase (no-RT control) was included in the synthesis of cDNA using 3′ RACE.

Reverse transcription (RT)-PCR—Prior to RT-PCR, all RNA samples were amplified in a PCR reaction that included RNase to check for DNA contamination. RNA samples contaminated with DNA were treated with DNAse as described previously. First strand synthesis was performed according to Shirley and Hwang (1995) with the following modifications: 1 μL Superscript II reverse transcriptase (200 u/μL, Invitrogen) was used instead of Moloney Murine Leukemia Virus reverse-transcriptase and an oligo $dT_{(15)}$ primer was used for the reverse primer. The reaction was incubated at 37°C for 30 min for first strand synthesis.

The cDNA was amplified using gene-specific primers designed to anneal to the 5′ region of *matK* distant from the conserved domain X. Some primers were species- or family-specific; however, degenerate primers that could span large groups of plants and remain specific to *matK* were made where possible (for information on primers, see Appendix S1 in Supplemental Data accompanying the online version of this article). A negative control containing RNA not reverse transcribed (no-RT) and RNase was co-amplified with cDNA, confirming that all PCR products from RT-PCR were the result of amplifying cDNA and not genomic DNA. All PCR products were separated on 1.5% agarose gels.

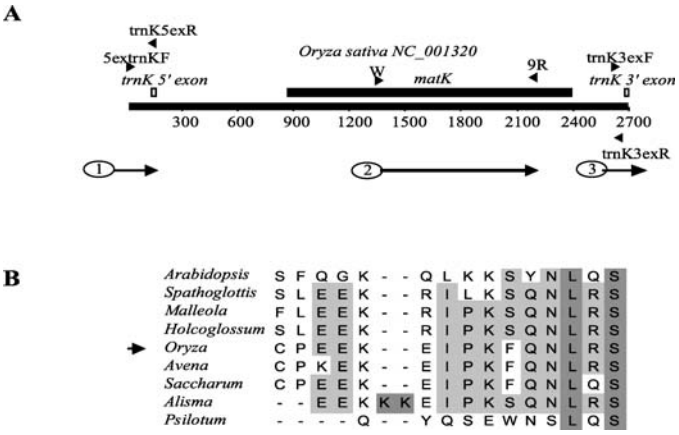

Fig. 1. Schematic representation of the *matK*/*trnK* gene region and peptide antigen alignment. (A) Annotation of the *matK*/*trnK* DNA region in rice with primers used to construct the *trnK* 5′ and 3′ exon probes and *matK* open reading frame (ORF) probe noted. Probes are (1) *trnK* 5′ exon probe, (2) *matK* ORF probe, and (3) *trnK* 3′ exon probe. (B) Alignment of the peptide region used to generate the rice MATK antibody. Arrow indicates the rice sequence used to generate the MATK antibody. Dark gray shading, 100% identity; light gray shading, consensus match; white, mismatch. GenBank accessions used: *Arabidopsis thaliana* BAA84366, *Spathoglottis plicata* DQ52860, *Malleola baliensis* AB217737, *Holcoglossum tsii* AB217732, *Oryza sativa* NP_039361, *Avena sativa* Q9MUZ6, *Saccharum officinarum* YP_054609, *Alisma canaliculatum* Q9GHE4, and *Psilotum nudum* NP_569609.

Sequencing—Amplified products from all PCR reactions, including generated probes and 3′ RACE products, were gel-extracted using the Qiagen Qiaquick gel extraction kit (Qiagen, Valancia, California, USA) and sequenced to confirm their gene identity. Direct sequencing was performed using the BigDye Terminator Sequencing Kit (ABI, Framingham, Massachusetts, USA) with the same primers used in the initial reaction. Annealing temperature of the sequencing reaction varied depending on the primer used. Products of the sequencing reaction were separated through capillary gel electrophoresis at the Virginia Bioinformatics Core Laboratory Facility (Blacksburg, Virginia, USA).

Plant material for protein analyses—Nine plant species (eight angiosperms and one monilophyte) were used to determine whether *matK* protein was expressed in various plant species. Because the antigen used to generate the MATK antibody utilized in this study was from rice, sample taxa were predominantly limited to monocots in order to increase the likelihood of antibody binding. Sample taxa included three species of the Orchidaceae to address previous assertions that *matK* might not be functional in some members of this plant family (Kores et al., 2000; Whitten et al., 2000; Goldman et al., 2001). The eudicot *Arabdiopsis thaliana* L. (Heynh.) (Brassicaceae) was used to test the threshold of homology required for antibody binding to protein from more evolutionary diverged angiosperms. Protein extract from the monilophyte *Psilotum nudum* L. (Psilotaceae) was used as a negative control. Although *P. nudum* may express MATK protein, the corresponding region in this fern to the peptide sequence used to produce our rice MATK antibody has a four-amino acid deletion (Fig. 1B) and low identity (13%) to the rice antigenic peptide sequence, rendering the cross-immunological reaction very unlikely. Plant material was obtained as described for RNA analyses (see Appendix 1). Leaf material was collected for each plant, placed in zip-seal freezer bags, and stored

at −80°C. Herbarium vouchers are located at VPI. Vouchers are not available for specimens obtained from the University of Mary Washington.

MATK antibody design and testing—An antibody was produced against the 15-amino-acid rice MATK peptide sequence CPEEEKEIPKFQNLRS and synthesized by Cocalico Biologics (Reamstown, Pennsylvania, USA). The choice of peptide region for generating the MATK antibody was based on two criteria: (1) specificity to MATK and (2) applicability to a broad range of angiosperm species. To increase specificity of this antibody against MATK and not other maturases that may contain domain X, we chose the peptide sequence from the N-terminal region of MATK. This antigen sequence was demonstrated to be specific to MATK by blastp search in GenBank (www.ncbi.nlm.nih.gov/BLAST/) using default settings, Viridiplantae as the organism group, and an inclusion threshold of 0.005. The 15-amino-acid peptide CPEEE KEIPKFQNLRS was also found to be fairly well-conserved across angiosperms by amino acid alignment (Fig. 1B). Comparison of this peptide sequence with the same region in other plant genera revealed almost 100% homology in the closely related grasses *Avena sativa* L. and *Saccharum officinarum* L. (Poaceae), 80% in *Alisma canaliculatum* A. Braun & Bouché (Alismataceae), 67% in *Spathoglottis plicata* Blume (Orchidaceae), 73% in *Malleola baliensis* J.J. Sm. and *Holcoglossum tsii* T. Yukawa (Orchidaceae), 40% in the eudicot *A. thaliana*, and 13% in the fern *P. nudum* (Fig. 1B). The specificity of this region to MATK, as determined by blastp and the conserved nature of this peptide across angiosperms, made this peptide region a strong candidate for the generation of a highly specific MATK antibody that could be used against protein extract from a variety of angiosperm taxa.

Antibody production in rabbits was carried out by Cocalico Biologicals. MATK antibody from the first bleed was purified by nitrocellulose absorption (A. Esen, Virginia Tech, personal communication) because of low antibody titer compared to background signal. Purification was not necessary for subsequent bleeds because of a much higher antibody titer than in the first bleed. Binding specificity of this antibody to MATK protein was tested by competition assay utilizing synthesized peptide antigen as the competing agent. Competition of antibody binding was demonstrated using 50 μg rice crude protein extract and 30 μg synthesized antigen resolved on a 7.5% SDS-PAGE gel. Resolved protein was then transferred to nitrocellulose and incubated with either unpurified rabbit MATK antibody (third bleed diluted 1 : 3000 in phosphate-buffered saline plus 0.1% Tween-20 [PBS-T] plus 3% bovine serum albumin [BSA]) or a mixture of this antibody plus synthesized antigen as the primary antibody. The MATK antibody and synthesized antigen mixture was incubated at room temperature for 1 h followed by 4°C overnight prior to being added to the membrane for the competition assay.

Protein extraction and Western blotting—Leaf or whole plant tissue was ground under liquid nitrogen and total protein extracted by the addition of Laemmli SDS sample buffer (62.5 mM Tris, pH 6.8, 2% SDS, 10% glycerol, and 5% β-mercaptoethanol) plus 1 mM of the protease inhibitor phenylmethanesulfonyl fluoride (PMSF) followed by boiling at 95°C for 15 min. Denatured protein was centrifuged twice at 15 000 × g for 10 min each time, and the supernatant was retained after each spin. The final supernatant was stored at −20°C as crude protein extract. Protein concentration was determined by Bradford assay according to Jones et al. (1989) using BSA as the standard. Fifty or 75 μg of crude protein extract from each plant species was fractionated by 7.5% SDS-PAGE and transferred onto nitrocellulose membrane. Equal loading of protein and transfer efficiency were confirmed by Ponceau S staining. For Western blotting, membranes were incubated with either nitrocellulose-absorption purified rabbit MATK antibody or pre-immune rabbit sera (diluted 1 : 50 or 1 : 300 in PBS-T) for the primary antibody. Horse radish peroxidase (HRP)-conjugated anti-rabbit IgG (Cell Signaling Technology, Danvers, Massachusetts, USA, diluted 1 : 2000 in 5% nonfat dry milk/PBST) was used as the secondary antibody for detection of immunoreactive protein. Bound antibody was detected using the ECL peroxidase/luminol system (Amersham Biosciences, Piscataway, New Jersey, USA) or West Pico chemiluminescent detection system (Pierce Biotechnology, Rockford, Illinois, USA).

Etiolation—Rice seed stock stored at 4°C was planted in vermiculite at the Virginia Tech Biology greenhouse and placed in two principal treatments: (1) under uniform light with a 10/14 day/night photoperiod in the greenhouse (the control) or (2) in the dark for 2 wk, hereafter referred to as either dark treatment or 0 h, reflecting the lack of light exposure. Both control and dark-treatment plants were uniformly watered. Dark-treatment plants remained in the dark while water was applied. Control plants from the greenhouse were harvested after 2 wk of growth. At 2 wk postgermination, a subset of plants from the dark treatment was harvested in a dark room. Remaining rice plants from the dark treatment were placed in a Percival growth chamber (Percival Scientific, Perry, Iowa, USA) and exposed to light for up to 24 h. Whole plant tissue was collected in two sets from these light-exposed plants: (1) after 4 h of light exposure and (2) after 24 h of light exposure. All tissue was placed in zip-seal freezer bags, frozen immediately in liquid nitrogen, and stored at −80°C. RNA and protein were isolated as described earlier. Transcript levels for *matK*/*trnK* and protein from each treatment were evaluated by Northern hybridization or Western blot (MATK antibody diluted 1 : 50 in PBS-T), respectively, as described earlier.

Developmental stage—Rice and potato plants were grown from seed stock and tubers, respectively, in the Virginia Tech Biology greenhouse under uniform light and water conditions. Plant leaves were collected at four times (2, 4, 6, and 8 wk postgermination), placed in zip-seal freezer bags, and stored at −80°C. RNA was isolated from both rice and potato, resolved on formaldehyde gels, and transferred to nylon membranes by the Northern blot method as described earlier. Membranes were hybridized with rice and potato *matK* ORF-Dig-labeled probes and detected by chemiluminescence using CDP-Star. Crude protein extract was isolated from rice leaves collected 2, 4, 6, and 8 wk postgermination, and analyzed by Western blot (unpurified MATK antibody diluted 1 : 300 in PBS-T) as described earlier.

RESULTS

matK RNA and protein—Hybridization of probes specific to the *matK* ORF to total RNA from rice and potato identified two predominant transcripts of 2.6 and 2.9 kilobases (kb) (Fig. 2A). Three smaller, less prominent, transcripts of approximately 1.8, 1.5, and 0.9 kb were also observed in both species (Fig. 2A). It is likely that these smaller transcripts represent splicing intermediates of the *trnK*/*matK* transcript. Transcripts of these same sizes were also identified in total rice RNA when using probes specific to the 5′ and 3′ *trnK* exons (Fig. 2B). In addition, the 5′ *trnK* exon probe hybridized to a transcript of less than 0.5 kb, while the 3′ *trnK* exon probe hybridized to a transcript of approximately 5.7 kb (Fig. 2B). A monocistronic *matK* transcript lacking the *trnK* exons was not found by Northern hybridization.

The specificity of the *matK* ORF probes to *matK* RNA was supported by Southern hybridization results. Hybridization of probes specific for the rice and potato *matK* ORF at high stringency (65°C) to PCR-amplified products of *matK*, *rbcL*, and *mat-r* from rice genomic DNA resulted in a strong band in the *matK* lane but no cross-reactivity to *rbcL* or *mat-r* (Fig. 2C). Only very weak cross-reactivity to *mat-r* was observed when the same hybridization was repeated at low stringency (data not shown). Amplified products for *matK, rbcL,* and *mat-r* were sequenced to confirm their gene identity.

RT-PCR using 3′ RACE amplified an 876-bp product for *matK,* a 1377-bp product for part of the *matK* ORF to the end of the 3′ *trnK* exon, and a 120-bp product for mature *trnK* (Fig. 3). The 3′ RACE products were sequenced and confirmed as *matK*, *matK* with the 3′ *trnK* exon, and mature *trnK* transcripts, respectively. The 120-bp size of the mature *trnK* transcript from 3′ RACE corresponded well to the low molecular mass (<0.5 kb) transcript identified using the Dig-labeled 5′ *trnK* exon probe in Northern hybridization. A no-RT control for the 3′ RACE reaction using primers specific to the *matK* ORF (Fig. 3A) did not result in detectable product. These 3′ RACE results confirm Northern blot data that a transcript containing both *matK* and the 3′ exon of *trnK* does exist and that the mature

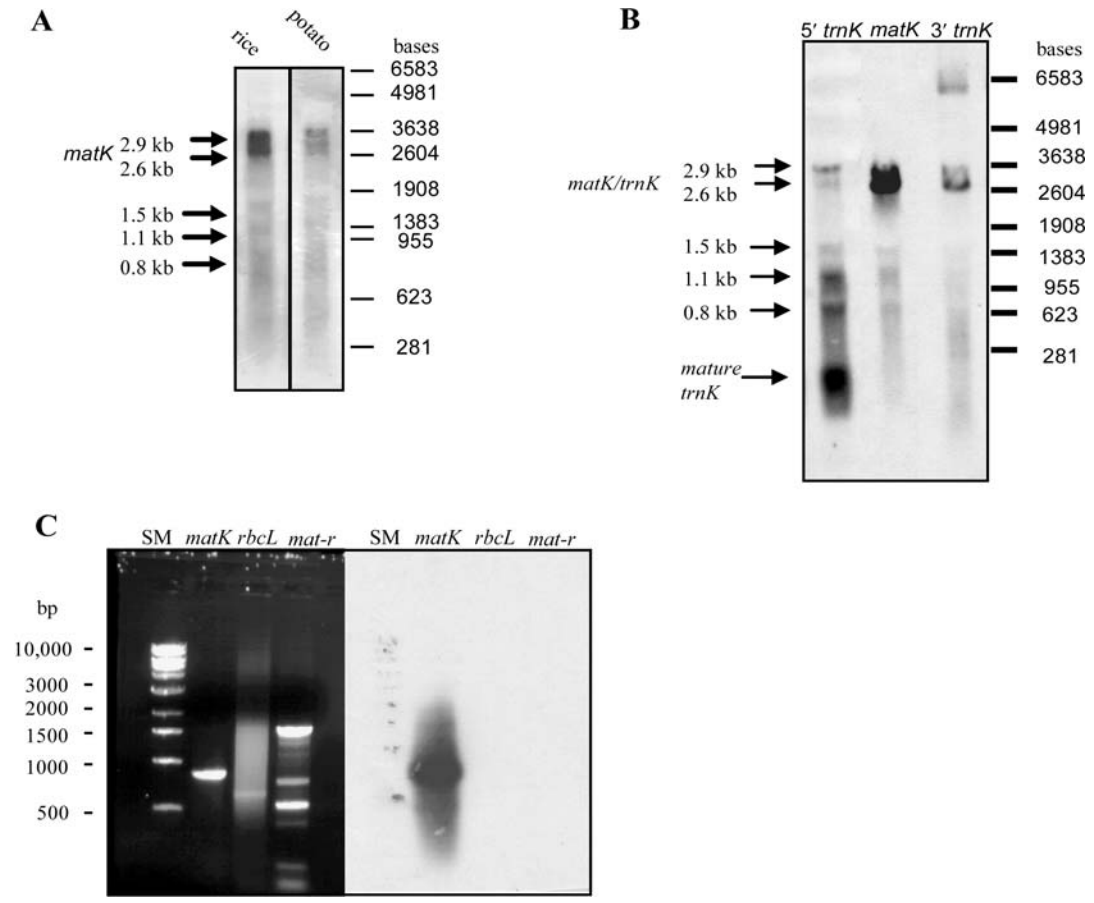

Fig. 2. Transcript sizes and probe specificity of *matK*/*trnK*. (A) Size of *matK* transcripts from rice and potato determined using the *matK* open reading frame (ORF) probe. RNA size markers (Promega) are noted on the right. Arrows indicate transcripts identified with the *matK* ORF probe. Size of transcript is listed beside the arrow. Results represent replicates of four independent biological experiments, each repeated technically 1–4 times. (B) Northern blot comparing the size of transcripts found using the *trnK* 5′ exon probe, *matK* ORF probe, or *trnK* 3′ exon probe. Transcripts for the *matK*/*trnK* gene unit are noted along with corresponding sizes. RNA size markers are noted on the right. Results represent three independent biological experiments with one or two technical replicates within each experiment. (C) Test of the specificity of the *matK* ORF probe to *matK* verses *rbcL* and *mat-r*. Shown are the 1.5% agarose gel (left) and the nylon membrane (right) of the Southern blot using PCR products for these genes and the *matK* ORF probe. SM, 1-kb ladder size marker (Promega). The top band in the *mat-r* lane was sequenced as *mat-r*.

trnK tRNA is approximately 120 bp in size and corresponds to a transcript band less than 0.5 kb on Northern blots.

An immunoreactive ~55-kDa protein was identified from rice protein extracts in Western blot experiments using an antibody targeted against rice MATK (Fig. 4A). This immunoreactive protein is close to the expected size (~61 kDa) of the rice full-length MATK protein. Four additional protein bands of approximately 40, 30, 25, and 20 kDa were also identified (Fig. 4A). The 55-, 30-, and 20-kDa bands were not found when pre-immune serum was used on these same membranes (Fig. 4A). The 20-kDa band was not consistently observed with immunoblotting (data not shown). Competition experiments using an excess of synthesized antigen demonstrated successful competition of the MATK antibody to binding with the three protein bands of 55, 30, and 20 kDa (Fig. 4B), although less so for the 30-kDa band. Therefore, these immunoreactive protein signals are not the result of contaminating antibodies in the antipeptide serum but are specific to the MATK antibody. The 30- and 20-kDa immunoreactive bands are most likely products of MATK protein proteolysis. The 40- and 25-kDa bands were observed when blots were incubated with both MATK antibody and pre-immune serum (Fig. 4A) and are, therefore, considered the result of background pre-immune contaminating antibodies. These same background bands were not found on Western blots when using the third bleed of antibody serum, which contains a much higher titer of the MATK antibody (Fig. 4B).

matK/trnK RNA and protein levels with etiolation—The 2.6- and 2.9-kb transcripts identified using the *trnK* and *matK* ORF probes were present at high levels in rice plants grown under light–dark cycles (control) but at very low levels in plants from the dark treatment (Fig. 5A). Exposure of dark-treatment plants to light increased RNA levels of the 2.6- and 2.9-kb transcripts (Fig. 5B). In contrast, the mature *trnK* tRNA level remained consistently high in all treatments (control, dark treatment, and exposure to light after dark treatment; Fig. 5B). Levels of the ~55-kDa immunoreactive protein decreased by about half in dark-treatment rice plants compared to the control. Levels of this protein increased slightly when etiolated plants were exposed to light (Fig. 5B). These results of protein

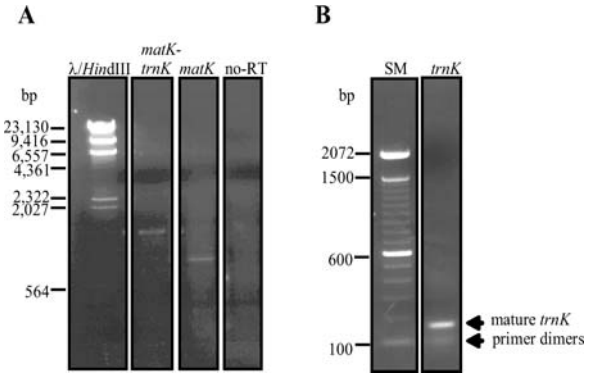

Fig. 3. PCR products from 3′ RACE performed on rice RNA. (A) 1% Agarose gel, lanes from left to right: *Lambda*/*Hin*dIII size markers; 3′ RACE product using primers W and trnK3exR; 3′ RACE product using primers W and 9R; and no-RT control. (B) Agarose gel (1.7%), lanes from left to right: 100-bp ladder (Gibco BRL, Gaithersburg, MD, USA); and 3′ RACE products using primers 3914 and trnK3exR. The mature transcript of *trnK* tRNA and primer dimers are indicated with arrows to the right. The results were reproduced once experimentally.

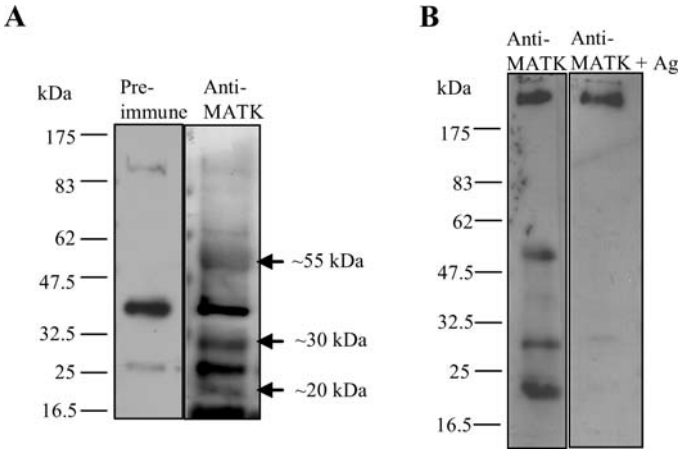

Fig. 4. Immunodetection of MATK in rice and competition assay. (A) Immunoreactive protein detected with either pre-immune serum (left) or MATK antibody (Anti-MATK, right). Protein molecular mass standards (New England BioLabs, Beverly, MA, USA) are noted. Arrows indicate protein bands observed only when blots were incubated with MATK antibody. Results replicated in three independent biological experiments. (B) Competition assay using an excess of synthesized MATK antigen. Protein detected on Western blot after incubation with MATK antibody (left) or after MATK antibody incubated with an excess of synthesized antigen (Anti-MATK + Ag, right). The results were reproduced once experimentally.

levels for the ~55-kDa immunoreactive protein parallel those of the 2.6- and 2.9-kb *matK* RNA transcripts and demonstrate light-induced expression or regulation for these RNA transcripts and the ~55-kDa immunoreactive protein.

Effect of developmental stage on matK RNA and protein—The 2.6-kb and 2.9-kb predominant *matK* transcripts were present at 2, 4, 6, and 8 wk postgermination in both rice and potato. However, these transcripts decreased substantially at 4 wk postgermination in both species (Fig. 5C). The highest relative level of the 2.6- and 2.9-kb RNA transcripts occurred 8 wk postgermination in rice (Fig. 5C) and 6 wk postgermination in potato (data not shown). A large decrease in protein expression was evident from Western blot analyses for the ~55-kDa immunoreactive protein at 4 wk postgermination compared to the other time points (2, 6, and 8 wk postgermination) in rice (Fig. 5C). The decrease in the ~55-kDa protein level corresponded to the decrease in RNA levels observed for the 2.6- and 2.9-kb *matK*/*trnK* RNA transcripts at 4 wk postgermination (Fig. 5C). These results indicate a relationship between plant developmental stage and expression of these RNA transcripts and the ~55-kDa immunoreactive protein.

Survey of matK RNA and protein across land plants—RT-PCR on isolated DNA-free RNA produced a single predominant *matK* PCR product from 11 plant species representing three major plant lineages: bryophytes, monilophytes, and angiosperms (Table 1). The PCR products ranged from 333 bp to 876 bp, depending on the particular primer pair utilized for amplification (Table 1). All RT-PCR products resulting from this study were sequenced and confirmed as *matK,* with the exception of *Adiantum hispidulum* Swartz (Pteridaceae), in which low amplification prevented adequate sequencing. Amplification of no-RT controls did not result in PCR product, confirming that the RNA used for RT-PCR was devoid of genomic DNA contamination and that RT-PCR products were the result of cDNA amplification (Fig. 6A).

In spite of several attempts using different protocols, an adequate amount of DNA-free RNA for RT-PCR could not be isolated from any of the five gymnosperm species used in this study. The high concentration of polyphenolics in gymnosperms may have been responsible for our inability to isolate DNA-free RNA from these plants. A high polyphenolic concentration has been previously noted to hinder RNA extraction from gymnosperms (Kiefer et al., 2000).

A survey for expressed MATK protein in nine land plants resulted in the identification of MATK antibody-immunoreactive protein bands in six angiosperms (*O. sativa, A. sativa, S. officinarum, Sagitarria latifolia* Willd. [Alismataceae]*, Spathoglottis gracilis* Rolfe ex Hook.f. [Orchidaceae], and *A. thaliana*) (Fig. 6B). The immunoreactive protein from five of these angiosperms was similar to the predicted size of the full-length MATK from each of these species (predicted/observed in kDa: 60/~55 in *O. sativa*, 61/~55 in *A. sativa*, 60/~55 in *S. latifolia*, 62/~60 in *S. officinarum*, and 60/~55 in *A. thaliana*, Fig. 6). Protein bands of these same molecular masses were absent when immunoblots were incubated with pre-immune serum (Fig. 6). One additional anti-MATK immunospecific protein band of ~20 kDa was seen on immunoblots with *A. thaliana* protein extract (Fig. 6). This additional band is most likely a product of MATK proteolysis. A unique MATK antibody-immunoreactive protein of 23 kDa was observed in *S. gracilis* (Fig. 6). No bands specific to the MATK antibody were found in extracts from the orchids *M. ligulata* and *H. kimballianum* (data not shown), or the fern *P. nudum* (Fig. 6B).

DISCUSSION

On the basis of the proposed substrates of MATK maturase activity, this putative enzyme critically impacts all chloroplast function including photosynthesis. In this study, we investigated whether *matK* is expressed and functional. Our results demonstrate that *matK* RNA and protein exist in several plant

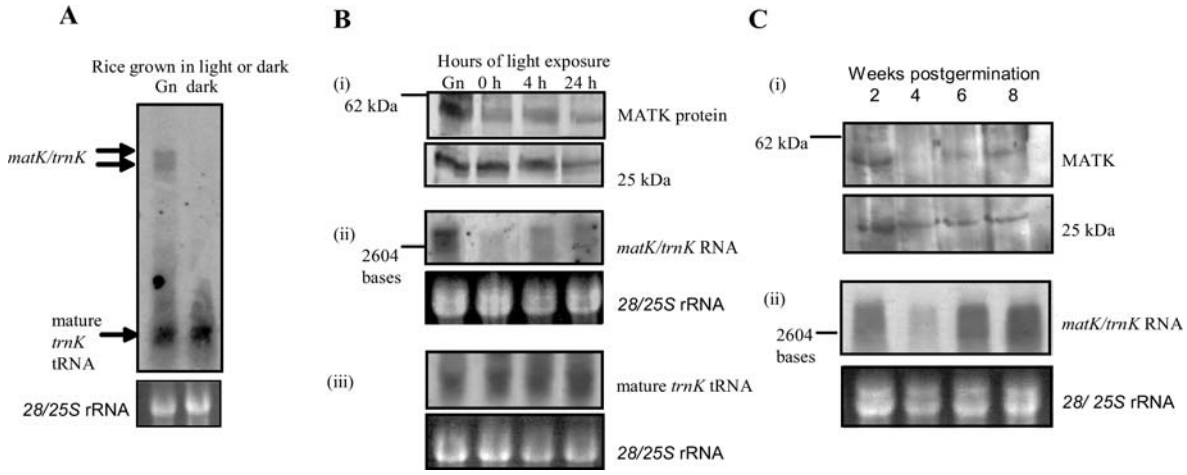

Fig. 5. The effect of light and developmental stage on *matK* RNA and protein levels in rice. (A) Levels of *matK* RNA and the mature transcript of *trnK* from light and dark treatments. Gn, greenhouse-grown control plants. (B) Levels in response to etiolation and exposure to light for 4 and 24 h for (i) the ~55-kDa immunoreactive protein; bottom, 25-kDa background protein band used as loading control; (ii) the 2.6- and 2.9-kb *matK*/*trnK* RNA transcripts; bottom, ethidium bromide stained rRNA loading control; and (iii) the mature *trnK* tRNA; bottom, rRNA loading control. Molecular mass bands noted on left. Hours of light exposure after etiolation are noted. Results represent two independent biological experiments. (C) Levels in response to developmental stage in rice for (i) the ~55-kDa immunoreactive protein; bottom, 25-kDa protein loading control; and (ii) the 2.6- and 2.9-kb *matK*/*trnK* RNA transcripts; bottom, rRNA loading control. Molecular mass bands noted on left. Weeks postgermination of seeds are noted above frames. RNA results represent two independent biological experiments for rice. Northern blots for each biological replicate were repeated one or two times.

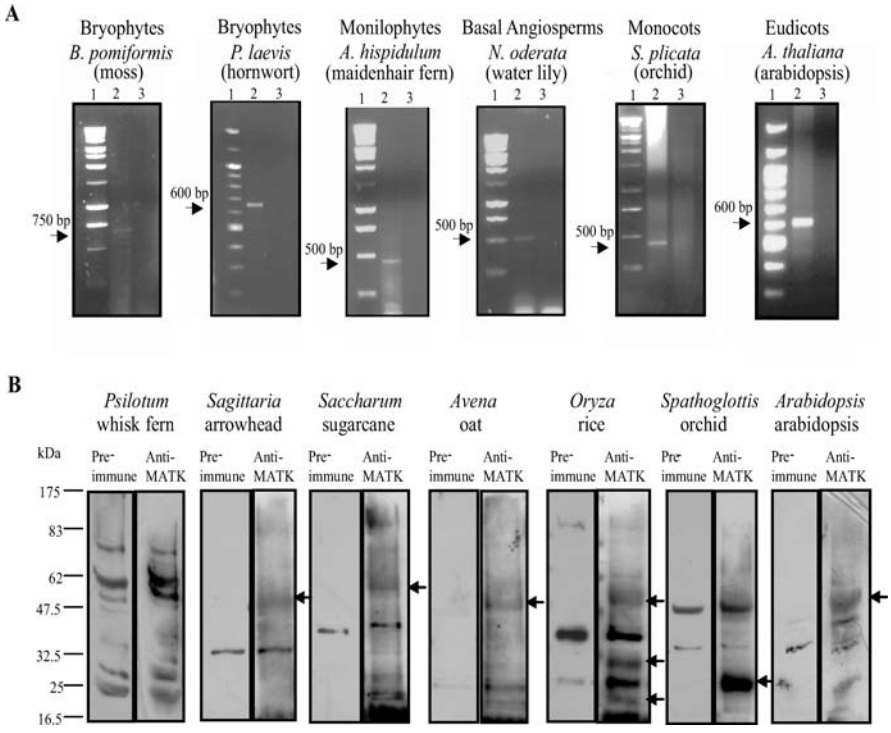

Fig. 6. Detection of *matK* ribonucleic acid (RNA) and protein across land plants. (A) Reverse-transcriptase (RT)-PCR products of *matK* cDNA across land plants. Lanes: (1) DNA size marker (a 250- or 100-bp ladder); (2) RT-PCR product; and (3) no-RT control. Key ladder sizes for PCR products are indicated. All amplifications were repeated at least once experimentally. (B) Western blot analysis of protein extract from seven land plants using pre-immune serum (left) or MATK antibody (right). Protein molecular weight standards (New England BioLabs, Beverly, Massachusetts, USA) noted to the far left. Arrows indicate protein bands observed only when blots were incubated with the MATK antibody. Western blot results represent three experimental replicates.

species. Further, we demonstrated that expression of *matK* is influenced by light and development stage, suggesting possible functions for this important maturase.

Independent transcripts for matK from the trnK precursor—Identification of the same 2.6- and 2.9-kb predominant transcripts by probes specific to the *matK* ORF and the 5′ and 3′ *trnK* exons (Fig. 2B) suggests that *matK* and *trnK* are transcribed as one dicistronic transcript. Vogel et al. (1997) found similar results using RNA from barley and a probe specific to the 5′ *trnK* exon. They identified two predominant transcripts of approximately 2.6 and 2.9 kb for the *matK*/*trnK* gene region and no evidence for a monocistronic *matK* transcript. Because it is unlikely that the same RNA transcript would form the mature *trnK* tRNA and at the same time proceed to translation for MATK protein, we hypothesize that the 2.6- and 2.9-kb transcripts represent mRNA transcripts dedicated to the formation of the MATK protein and are separate entities from the unspliced *trnK* precursor. While the *trnK* exons are present in this mRNA transcript, these exons act as the 5′ and 3′ untranslated regions (UTRs) of the mRNA and will therefore not form the mature *trnK* tRNA. We believe that the unspliced *trnK* precursor is very rapidly processed to produce the mature *trnK* tRNA-lysine and is not observable on Northern blots.

Several lines of evidence support our hypothesis that the 2.6- and 2.9-kb transcripts are dedicated to forming the MATK protein and are not unspliced *trnK* precursors. First, RNA levels for the 2.6-kb and 2.9-kb transcripts were light induced, while RNA levels of mature *trnK* tRNA remained constant (Fig. 5A, B). Thus, light is impacting expression of the high molecular mass transcripts differently than expression of the mature *trnK* tRNA, suggesting independent regulation of these transcripts. Second, levels of the 2.6- and 2.9-kb transcripts paralleled the expression of a ~55-kDa protein immunoreactive to the MATK antibody in both etiolation and developmental stage experiments (Fig. 5). It appears, therefore, that these high molecular mass transcripts, which we have already determined by Northern hybridization to contain the *matK* ORF, are translated into a protein product that corresponds to the expected molecular mass of full-length MATK. Finally, RT-PCR experiments revealed that a *matK* transcript is present in the leptosporangiate fern *Adiantum hispidulum* (Fig. 6A). Similar results were also observed for *A. capillus-veneris* (Wolf et al., 2004). *Adiantum*, unlike other plants used in this study, has a large rearrangement in its chloroplast genome in the *trnK* intron, resulting in the loss of *trnK* but retention of *matK* (Wolf et al., 2003). The *matK* transcript identified in *Adiantum* by our study and by Wolf et al. (2004), therefore, could not possibly be the result of an unspliced *trnK* precursor transcript. Collectively, these results strongly support our hypothesis that the 2.6- and 2.9-kb predominant transcripts are dedicated *matK* mRNA transcripts and not the unspliced *trnK* precursor.

Rice MATK protein—We conclude that the ~55-kDa immunoreactive band observed on Western blots containing rice crude protein extract is a full-length MATK protein (Fig. 4A). The rice MATK protein sequence (GenBank accession: P12175) is 511 amino acids long, which corresponds to a 61.4-kDa protein. The ~55-kDa protein band we observed on Western blots of rice protein extract corresponds closely to this predicted size for MATK from GenBank. Further, competition experiments verified the specificity of our antibody to the MATK protein by successful competition of binding by the 15-amino-acid MATK antigen peptide sequence used to generate this antibody (Fig. 4B). The identity of this immunoreactive ~55-kDa protein as MATK is also supported by the correlation between RNA levels for the 2.6- and 2.9-kb *matK* transcripts and the pattern of expression for this protein in etiolation and developmental stage experiments (Fig. 5). We propose that the additional 30- and 20-kDa protein bands immunoreactive to MATK antibody observed in rice immunoblots are the products of MATK protein proteolysis. The absence of the 30- and 20-kDa molecular mass bands in pre-immune controls and lack of these bands after addition of antigen in competition assays indicates that these bands are the MATK protein (Fig. 4). The low molecular mass of these bands, however, suggests that these are not the full-length MATK protein but most likely degraded products. Multiple bands have been observed on SDS-PAGE gels and/or immunoblots as a result of protein proteolysis even when using highly purified antibodies and protease inhibitors (Bialek et al., 1989; Cramer et al., 1998). We conclude, therefore, that the ~55-, 30-, and 20-kDa immunoreactive bands on Western blots are the MATK protein, with the ~55-kDa protein being the full-length polypeptide. Although three studies have previously reported finding a MATK protein from plant extracts (du Jardin et al., 1994; Liere and Link, 1995; Vogel et al., 1999), only Vogel et al. (1999), using barley, identified a MATK protein close to the expected molecular mass (predicted 56 kDa, observed 60 kDa). Our findings concur with those of Vogel et al. (1999) and demonstrate that expressed full-length MATK protein in rice is close to the predicted size.

Etiolation of matK/trnK—Protein levels for the ~55-kDa MATK protein increased less than the RNA after light exposure (Fig. 5A, B). This may reflect differential regulation of expression at the RNA and protein level for *matK* or delayed translation of protein from RNA transcripts. Nonetheless, these results indicate that the expression of *matK* is influenced by light. Similarly, a five-fold increase in *matK* transcription was observed in a *Nicotiana tabacum* L. (tobacco, Solanaceae) chloroplast microarray study when tissues were grown in light (Nakamura et al., 2003). Thus, the two studies provide independent evidence in support of light-induced *matK* expression and a function for MATK related to light.

Several chloroplast genes have light-induced expression (Klein and Mullet, 1990; Klein, 1991; Baumgartner et al., 1993). These genes are involved in two major activities of the chloroplast: photosynthesis and chloroplast development (Mullet, 1988; Klein and Mullet, 1990; Klein, 1991). Potential substrates for MATK maturase activity include RNAs from one photosynthesis-related gene, *atpF* (Herrmann et al., 1993; Kostrzewa and Zetsche, 1993; Jenkins et al., 1997; Vogel et al., 1999), and several other genes needed to form the chloroplast translation machinery (e.g., *trnK, trnA, trnI*, *rpl2*, and *rps12* (Hess et al., 1994; Ems et al., 1995; Jenkins et al., 1997; Vogel et al., 1999). Chloroplast development requires turning on protein translation in this organelle and increases the expression of all RNAs and proteins related to the translation machinery (Baumgartner et al., 1993). Levels of the maturase (MATK?) needed for processing introns in these transcripts should, therefore, increase with light exposure after etiolation in order to generate the needed proteins and tRNAs for photosynthesis and/or the chloroplast translation machinery.

matK RNA and protein levels during plant development—We detected *matK* RNA and protein at all times examined; however, the levels dropped significantly at 4 wk postgermination (Fig. 5C). This downregulation in *matK* expression may be associated with reduced expression of one of its proposed RNA substrates. While several genes are known to be involved directly or indirectly in the regulation of plant development (Jan et al., 2006; Wang et al., 2006), including the chloroplast posttranscription factor PNPase (Sauret-Güeto et al., 2006), substrates for the proposed maturase activity of MATK mainly include tRNAs and mRNAs needed for chloroplast protein translation (Hess et al., 1994; Ems et al., 1995). These proposed substrates are not directly related to developmental stage. Further investigation is required to identify potential substrates related to plant development that reflect this observed downregulation for MATK at 4 wk postgermination.

matK transcription and protein expression across land plants—A *matK* transcript has been previously identified in the grasses *O. sativa* and *H. vulgare* (Kanno and Hirai, 1993; Vogel et al., 1997), tobacco (Nakamura et al., 2003), *Anthoceros formosae* Steph. (Anthocerotaceae) (Kugita et al., 2003), and *A. capillus-veneris* L. (Pteridaceae) (Wolf et al., 2004). These studies, however, primarily focused on the *matK*/*trnK* gene unit as a whole and did not separate the expression of these two genes or correlate transcription data with protein expression. Our study has provided strong evidence that *trnK* and *matK* are transcribed independently of each other in rice and that the most abundant transcripts for this gene region correlate with the expression of *matK,* not a *trnK* precursor. Under this assumption, we expanded the survey for a *matK* transcript to 13 land plant families (14 species). A *matK* transcript was found in 11 plant species spanning the land plant phylogenetic tree (Table 1). These species included the orchid, *S. plicata* (Fig. 6A), noted previously to contain *matK* as a possible pseudogene (Freudenstein et al., 2004). Interestingly, the translated amino acid sequence from the *matK* cDNA of this orchid did not contain any premature stop codons (GenBank accession DQ525860).

Three of the angiosperm species identified to contain *matK* RNA (*O. sativa*, *S. latifolia,* and *A. thaliana*) also contained full-length MATK protein (Fig. 6). Although these results do not demonstrate a direct relationship between *matK* RNA and protein expression, they strongly support that *matK* RNA is indicative of expressed MATK protein. While the homology of our antigen to nonmonocot angiosperms is substantially less than within the monocots (nearly 100% to closely related monocots vs. 40% to *Arabidopsis*), it appears that the homology between our target antigen in rice and *Arabidopsis* occurs at key amino acid positions sufficient for antibody binding. At present, there is no way to reliably predict the minimum amount of homology required between sequences for antibody binding. For example, only four contiguous amino acids appear to be required for cross-reactivity of the GroEL antibody from the bacteria *Actinobacillus actinomycetemcomitans* to human fibronectin (Yoshioka et al., 2004). The amount of homology needed for the MATK antibody we developed in this study appears to be greater than 13% as determined by the lack of specific binding to protein from the fern *P. nudum* (Fig. 6B).

A full-length MATK protein was not found in extracts of the three orchids examined by Western blot. This could be due to protein proteolysis as indicated by the ~23-kDa immunoreactive protein band recognized with our MATK antibody against extract of *S. gracilis* (Fig. 6B). On the basis of the cDNA sequence of *S. plicata* and detection of full-length MATK protein in several other angiosperms (Fig. 6B), this protein is likely to be translated in the orchids.

Conclusion—Establishing a functional role for MATK has important implications for plant phylogenetics, gene evolution, and plant molecular biology. The results of this study support independent transcription for the *matK* gene from the *trnK* unspliced precursor and demonstrate MATK protein expression in several plant species. Furthermore, we provide evidence of a function for MATK in plant physiology by demonstrating an influence of light and developmental stage on its expression. Maintenance of expression and functionality for *matK* despite its high nucleotide and amino acid substitution rate suggests that genetic buffers are in place, which constrain the evolution of this gene. Assessment of protein evolution in relation to function and gene structure is required to better understand the nature of evolutionary constraints inherent to *matK* and other functional genomic regions. Such insights about genes used in molecular systematics are valuable in gene choices for systematic studies and interpretations of patterns of species and gene evolution.

In addition to the importance of *matK* in phylogenetic and evolutionary studies, this gene also encodes a putative group II intron maturase of the chloroplast. The relationship we found between light and developmental stage and amount of *matK* expressed suggests important functions for this protein in plant physiology. Review of potential substrates for MATK activity supports a role for this maturase at the posttranscriptional level in essential plant activities influenced by light, such as photosynthesis. Only one study (Sauret-Güeto et al., 2006) has previously reported a link between chloroplast posttranscription processing factors and regulation of plant developmental stage. MATK may be a second posttranscription regulatory connection between the chloroplast and plant development. To our knowledge, ours is the first study to show these functional aspects of MATK activity. Further investigations are needed to determine how MATK functions in these plant processes and to provide a more comprehensive understanding of the evolution of this unusual gene.

LITERATURE CITED

ALTENBACH, S. B., AND S. H. HOWELL. 1981. Identification of satellite RNA associated with turnip crinkle virus. *Virology* 112: 25–33.

BAUMGARTNER, B. J., J. C. RAPP, AND J. E. MULLET. 1993. Plastid genes encoding the transcription/translation apparatus are differentially transcribed early in barley (*Hordeum vulgare*) chloroplast development. *Plant Physiology* 101: 781–791.

BIALEK, G., H.-P. NASHEUER, H. GOETZ, AND F. GROSSE. 1989. Exonucleolytic proofreading increases the accuracy of DNA synthesis by human lymphocyte DNA polymerase α-DNA primase. *EMBO Journal* 8: 1833–1839.

CAMERON, K. M. 2005. Leave it to the leaves: a molecular phylogenetic study of Malaxideae (Orchidaceae). *American Journal of Botany* 92: 1025–1032.

CHURCH, G. M., AND W. GILBERT. 1984. Genome sequencing. *Proceedings of the National Academy of Sciences, USA* 81: 1991–1995.

CRAMER, H., M. MEINHARD, M. SCHULZ, AND H. SCHNABL. 1998. Immunological detection of inositol 1,4,5-trisphosphate receptor in microsomes from *Vicia faba* (L.) and *Zea mays* (L.) and broadbean vacuoles. *International Journal of Plant Sciences* 159: 750–752.

DANIELL, H., S.-B. LEE, J. GREVICH, C. SASKI, T. QUESADA-VARGAS, C. B.

GUDA, J. TOMKINS, AND R. K. JANSEN. 2006. Complete chloroplast genome sequences of *Solanum bulbocastanum*, *Solanum lycopersicum* and comparative analyses with other Solanaceae genomes. *Theoretical and Applied Genetics* 112: 1503–1518.

DOYLE, J. J., AND J. L. DOYLE. 1990. Isolation of plant DNA from fresh tissue. *Focus* 12: 13–25.

DU JARDIN, P., D. PORTETELLE, L. HARVENGT, M. DUMONT, AND B. WATHELET. 1994. Expression of intron-encoded maturase-like polypeptides in potato chloroplasts. *Current Genetics* 25: 158–163.

EMS, S. C., C. W. MORDEN, C. K. DIXON, K. H. WOLFE, C. W. DEPAMPHILIS, AND J. D. PALMER. 1995. Transcription, splicing and editing of plastid RNAs in the nonphotosynthetic plant *Epifagus virginiana*. *Plant Molecular Biology* 29: 721–733.

FARRÉ, J., AND A. ARAYA. 1999. The *mat-r* open reading frame is transcribed from a non-canonical promoter and contains an internal promoter to co-transcribe exons *nad1e* and *nad5III* in wheat mitochondria. *Plant Molecular Biology* 40: 959–967.

FREUDENSTEIN, J. V., C. V. D. BERG, D. H. GOLDMAN, P. J. KORES, M. MOLVRAY, AND M. W. CHASE. 2004. An expanded plastid DNA phylogeny of Orchidaceae and analysis of jackknife branch support strategy. *American Journal of Botany* 91: 149–157.

GERARD, G. F., AND K. MILLER. 1986. Comparison of glyoxal and formaldehyde gels for sizing rRNAs. *Focus* 8: 5–6.

GOLDMAN, D. H., J. V. FREUDENSTEIN, P. J. KORES, M. MOLVRAY, D. C. JARRELL, W. M. WHITTEN, K. M. CAMERON, R. K. JANSEN, AND M. W. CHASE. 2001. Phylogenetics of Arethuseae (Orchidaceae) based on plastid *matK* and *rbcL* sequences. *Systematic Botany* 26: 670–695.

HAYASHI, K., AND S. KAWANO. 2000. Molecular systematics of *Lilium* and allied genera (Liliaceae): phylogenetic relationships among *Lilium* and related genera based on the *rbcL* and *matK* gene sequence data. *Plant Species Biology* 15: 73–93.

HERRMANN, R. G., J. STEPPUHN, G. S. HERRMANN, AND N. NELSON. 1993. The nuclear-encoded polypeptide Cfo-II from spinach is a real, ninth subunit of chloroplast ATP synthase. *FEBS Letters* 326: 192–198.

HESS, W. R., B. HOCH, P. ZELTZ, T. HÜBSCHMANN, H. KÖSSEL, AND T. BÖRNER. 1994. Inefficient *rpl2* splicing in barley mutants with ribosome-deficient plastids. *Plant Cell* 6: 1455–1465.

HIDALGO, O., T. GARNATJE, A. SUSANNA, AND J. MATHEZ. 2004. Phylogeny of Valerianaceae based on *matK* and ITS markers, with reference to *matK* individual polymorphism. *Annals of Botany* 93: 283–293.

HILU, K. W., T. BORSCH, K. MÜLLER, 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, G. R. JENS, C. S. CAMPBELL, AND L. W. CHATROU. 2003. Angiosperm phylogeny based on *matK* sequence information. *American Journal of Botany* 90: 1758–1776.

HILU, K. W., AND H. LIANG. 1997. The *mat*K gene: sequence variation and application in plant systematics. *American Journal of Botany* 84: 830–839.

JAN, A., H. NAKAMURA, H. HANDA, H. ICHIKAWA, H. MATSUMOTO, AND S. KOMATSU. 2006. Gibberellin regulates mitochondrial pyruvate dehydrogenase activity in rice. *Plant and Cell Physiology* 47: 244–253.

JANKOWIAK, K., J. LESICKA, A. PACAK, A. RYBARCZYK, AND Z. SZWEYKOWSKA-KULIŃSKA. 2004. A comparison of group II introns of plastid tRNALysUUU genes encoding maturase protein. *Cellular & Molecular Biology Letters* 9: 239–251.

JENKINS, B. D., D. J. KULHANEK, AND A. BARKAN. 1997. Nuclear mutations that block group II RNA splicing in maize chloroplasts reveal several intron classes with distinct requirements for splicing factors. *Plant Cell* 9: 283–296.

JOHNSON, L. A., AND D. E. SOLTIS. 1994. *matK* DNA sequences and phylogenetic reconstruction in Saxifragaceae s. str. *Systematic Botany* 19: 143–156.

JONES, C. G., J. D. HARE, AND S. J. COMPTON. 1989. Measuring plant protein with the Bradford assay. *Journal of Chemical Ecology* 15: 979–992.

KANNO, A., AND A. HIRAI. 1993. A transcription map of the chloroplast genome from rice (*Oryza sativa*). *Current Genetics* 23: 166–174.

KIEFER, E., W. HELLER, AND D. ERNST. 2000. A simple and efficient protocol for isolation of functional RNA form plant tissue rich in secondary metabolites. *Plant Molecular Biology Reporter* 18: 33–39.

KLEIN, R. R. 1991. Regulation of light-induced chloroplast transcription and translation in eight-day-old dark-grown barley seedlings. *Plant Physiology* 97: 335–342.

KLEIN, R. R., AND J. E. MULLET. 1990. Light-induced transcription of chloroplast genes. *psbA* transcription is differentially enhanced in illuminated barley. *Journal of Biological Chemistry* 265: 1895–1902.

KORES, P. J., P. H. WESTON, M. MOLVRAY, AND M. W. CHASE. 2000. Phylogenetic relationships within the Diurideae (Orchidaceae): inferences from plastid *mat*K DNA sequences. *In* K. L. Wilson and D. A. Morrison [eds.], Monocots: systematics and evolution, 449–456. CSIRO Publishing, Collingwood, Victoria, Australia.

KOSTRZEWA, M., AND K. ZETSCHE. 1993. Organization of plastid-encoded ATPase genes and flanking regions including homologues of *infB* and *tsf* in the thermophilic red alga *Galdieria sulphuraria*. *Plant Molecular Biology* 23: 67–76.

KUGITA, M., A. KANEKO, Y. YAMAMOTO, Y. TAKEYA, T. MATSUMOTO, AND K. YOSHINAGA. 2003. The complete nucleotide sequence of the hornwort (*Anthoceros formosae*) chloroplast genome: insight into the land plants. *Nucleic Acids Research* 31: 716–721.

LIANG, H., AND K. W. HILU. 1996. Application of the *matK* gene sequences to grass systematics. *Canadian Journal of Botany* 74: 125–134.

LIERE, K., AND G. LINK. 1995. RNA-binding activity of the *matK* protein encoded by the chloroplast *trnK* intron from mustard (*Sinapis alba* L.). *Nucleic Acids Research* 23: 917–921.

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. *Nucleic Acids Research* 21: 4991–4997.

MÜLLER, K. F., T. BORSCH, AND K. W. HILU. 2006. Phylogenetic utility of rapidly evolving DNA at high taxonomical levels: contrasting *matK, trnT-F* and *rbcL* in basal angiosperms. *Molecular Phylogenetics and Evolution* 41: 99–117.

MULLET, J. E. 1988. Chloroplast development and gene expression. *Annual Review of Plant Physiology and Plant Molecular Biology* 39: 475–502.

NAKAMURA, T., Y. FURUHASHI, K. HASEGAWA, H. HASHIMOTO, K. WATANABE, J. OBOKATA, M. SUGITA, AND M. SUGIURA. 2003. Array-based analysis on tobacco plastid transcripts: preparation of a genomic microarray containing all genes and all intergenic regions. *Plant and Cell Physiology* 44: 861–867.

NEUHAUS, H., AND G. LINK. 1987. The chloroplast $tRNA^{Lys}$ (UUU) gene from mustard (*Sinapsis alba*) contains a class II intron potentially coding for a maturase-related polypeptide. *Current Genetics* 11: 251–257.

OLMSTEAD, R. G., AND J. D. PALMER. 1994. Chloroplast DNA systematics: a review of methods and data analysis. *American Journal of Botany* 81: 1205–1224.

SAURET-GÜETO, S., P. BOTELLA-PAVÍA, Ú. FLORES-PÉREZ, J. F. MARTÍNEZ-GARCÍA, C. S. OMÁN, P. LEÓN, A. BORONAT, AND M. RODRÍGUEZ-CONCEPCIÓN. 2006. Plastid cues posttranscriptionally regulate the accumulation of key enzymes of the methylerythritol phosphate pathway in *Arabidopsis*. *Plant Physiology* 141: 75–84.

SHIRLEY, B. W., AND I. HWANG. 1995. The interaction trap: in vivo analysis of protein–protein associations. *Methods in Cell Biology* 49: 401–416.

SOLTIS, D. E., AND P. S. SOLTIS. 2004. *Amborella* not a "basal angiosperm"? Not so fast. *American Journal of Botany* 91: 997–1001.

STEANE, D. A. 2005. Complete nucleotide sequence of the chloroplast genome from the Tasmanian blue gum, *Eucalyptus globulus* (Myrtaceae). *DNA Research* 12: 215–220.

SUGITA, M., K. SHINOZAKI, AND M. SUGIURA. 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. *Proceedings of the National Academy of Sciences, USA* 82: 3557–3561.

TILLICH, M., H. T. FUNK, C. SCHMITZ-LINNEWEBER, P. POLTNIGG, B. SABATER, M. MARTIN, AND R. M. MAIER. 2005. Intra editing of plastid RNA in *Arabidopsis thaliana* ecotypes. *Plant Journal* 43: 708–715.

TURMEL, M., C. OTIS, AND C. LEMIEUX. 2006. The chloroplast genome

sequence of *Chara vulgaris* sheds new light into the closest green algal relatives of land plants. *Molecular Biology and Evolution* 23: 1324–1338.

VOGEL, J., T. BÖRNER, AND W. HESS. 1999. Comparative analysis of splicing of the complete set of chloroplast group II introns in three higher plant mutants. *Nucleic Acids Research* 27: 3866–3874.

VOGEL, J., T. HUBSCHMANN, T. BÖRNER, AND W. R. HESS. 1997. Splicing and intron-internal RNA editing of *trnK-matK* transcripts in barley plastids: support for MatK as an essential splicing factor. *Journal of Molecular Biology* 270: 179–187.

WANG, X., Y. XU, Y. HAN, S. BAO, J. DU, M. YUAN, Z. XU, AND K. CHONG. 2006. Overexpression of *RAN1* in rice and *Arabidopsis* alters primordial meristem, mitotic progress, and sensitivity to auxin. *Plant Physiology* 140: 91–101.

WHITTEN, W. M., N. H. WILLIAMS, AND M. W. CHASE. 2000. Subtribal and generic relationships of Maxillarieae (Orchidaceae) with emphasis on Stanhopeinae: combined molecular evidence. *American Journal of Botany* 87: 1842–1856.

WOLF, P. G., C. A. ROWE, AND M. HASEBE. 2004. High levels of RNA editing in a vascular plant chloroplast genome: analysis of transcripts from the fern *Adiantum capillus-veneris*. *Gene* 339: 89–97.

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 research* 10: 59–65.

YOSHIOKA, M., D. GRENIER, D. HINODE, M. FUKUI, AND D. MAYRAND. 2004. Antigenic cross-reactivity and sequence homology between *Actinobacillus actinomycetemcomitans* GroEL protein and human fibronectin. *Oral Microbiology and Immunology* 19: 124–128.

APPENDIX 1. Plant species used in this study along with their respective lineage, specimen origin, and voucher information. Vouchers for specimens are deposited in the following herbaria: DR = Technische Universität Dresden; Quandt = Herbarium Quandt (personal collection of Dietmar Quandt); and VPI = Virginia Polytechnic Institute and State University. Voucher specimens are not available for *Malleola ligulata* and *Holcoglossum kimballianum*.

Plant species/Major group	Family	Location/Source	Voucher
Bryophytes			
Bartramia pomiformis Hedw.[a]	Bartramiaceae	Pembroke, VA	D. Quandt (Quandt-A10012)
Atrichum altecristatum (Ren. & Card.) Smyth & Smyth[a]	Polytrichaceae	Pembroke, VA	D. Quandt (Quandt-A10015)
Phaeoceros laevis L. (Prosk).[a]	Anthocerotaceae	Mandanici I, Sicily	C. Neinhuis (DR-026538)
Monilophytes			
Adiantum hispidulum Swartz[a]	Pteridaceae	Biology greenhouse, Virginia Tech., Blacksburg, VA	M. Barthet (VPI-102885)
Psilotum nudum L. Beauv.[b]	Psilotaceae	Biology greenhouse, Virginia Tech., Blacksburg, VA	D. Wiley-Vawter (VPI-102875)
Gymnosperms			
Pinus strobes L.[a]	Pinaceae	Blowing Rock, NC	M. Barthet (VPI-95326)
Ginkgo biloba L.[a]	Ginkgoaceae	Virginia Tech, Blacksburg, VA	M. Barthet (VPI-93330)
Ephedra viridis Coville[a]	Gnetaceae	Richters Herbs, Ontario, Canada	D. Wiley-Vawter (VPI-102876)
Ephedra nevadensis S. Wats.[a]	Gnetaceae	Richters Herbs, Ontario, Canada	D. Wiley-Vawter (VPI-102877)
Gnetum gnemon L.[a]	Gnetaceae	University of Connecticut	D. Wiley-Vawter (VPI-102878)
Angiosperms			
Nymphea oderata Ait.[a]	Nympheaceae	Burleson County, TX	Kristi Niehaus (VPI-102874)
Sagittaria latifolia Willd.[ab]	Alismataceae	Claytor Lake, VA	M. Barthet (VPI-102883)
Oryza sativa L.[ab]	Poaceae	Valley Seed Service, CA	M. Barthet (VPI-102887)
Avena sativa L. 'Norle'[b]	Poaceae	Southern States, VA	K. W. Hilu (VPI-102888)
Zea mays L.[a]	Poaceae	Wetsel Seed Co., VA	D. Wiley-Vawter (VPI-102880)
Saccharum officinarum L.[b]	Poaceae	Glasshouse Works, OH	D. Wiley-Vawter (VPI-102879)
Spathoglottis plicata Blume[a]	Orchidaceae	Gardino Nursery Corp., FL	M. Barthet (VPI-102882)
Spathoglottis gracilis Rolfe ex Hook.f.[b]	Orchidaceae	Gardino Nursery Corp., FL	M. Barthet (VPI-102881)
Malleola ligulata (J.J. Sm.) J.J. Sm.[b]	Orchidaceae	David Jarrell, Ph.D., University of Mary Washington	Not available
Holcoglossum kimballianum (Rchb.f.) Garay[b]	Orchidaceae	David Jarrell, Ph.D., University of Mary Washington	Not available
Solanum tuberosum L.[a]	Solanaceae	Biology greenhouse, Virginia Tech., Blacksburg, VA	M. Barthet (VPI-102884)
Arabidopsis thaliana L. (Heynh.)[ab]	Brassicaceae	Eric Beers, Ph.D., Virginia Tech., Blacksburg, VA	M. Barthet (VPI-102886)

Note: State abbreviations (USA) are VA = Virginia; NC = North Carolina; TX = Texas; CA = California; OH = Ohio; and FL = Florida.

[a] Plants used for RNA analyses.

[b] Plants used for protein analyses.