Origin and Evolution of the Chloroplast *trnK* (*matK*) Intron: A Model for Evolution of Group II Intron RNA Structures

Georg Hausner, Robert Olson,†[1] Dawn Simon,† Ian Johnson,† Erin R. Sanders,‡ Kenneth G. Karol,§ Richard M. McCourt,‖ and Steven Zimmerly†*

*Department of Microbiology, Buller Building, University of Manitoba, Winnipeg, Manitoba R3T 2N2, Canada; †Department of Biological Sciences, University of Calgary, Alberta T2N 1N4, Canada; ‡Department of Biological Chemistry, David Geffen School of Medicine, University of California, Los Angeles; §Department of Biology, University of Washington, Seattle; ‖Department of Botany, Academy of Natural Sciences, 1900 Benjamin Franklin Parkway, Philadelphia, Pennsylvania

The *trnK* intron of plants encodes the *matK* open reading frame (ORF), which has been used extensively as a phylogenetic marker for classification of plants. Here we examined the evolution of the *trnK* intron itself as a model for group II intron evolution in plants. Representative *trnK* intron sequences were compiled from species spanning algae to angiosperms, and four introns were newly sequenced. Phylogenetic analyses showed that the *matK* ORFs belong to the ML (mitochondrial-like) subclass of group II intron ORFs, indicating that they were derived from a mobile group II intron of the class. RNA structures of the introns were folded and analyzed, which revealed progressive RNA structural deviations and degenerations throughout plant evolution. The data support a model in which plant organellar group II introns were derived from bacterial-like introns that had "standard" RNA structures and were competent for self-splicing and mobility and that subsequently the ribozyme structures degenerated to ultimately become dependent upon host-splicing factors. We propose that the patterns of RNA structure evolution seen for the *trnK* intron will apply to the other group II introns in plants.

Introduction

In plant chloroplasts, the $tRNA^{Lys}$(UUU) gene (*trnK*) contains a group II intron (*trnK*I1), which encodes the *matK* open reading frame (ORF). The *trnK* intron and its encoded *matK* ORF have generated substantial interest in the fields of plant evolution and molecular biology. In evolutionary studies, the *matK* ORF has been used as a marker to construct plant phylogenies because the ORF evolves rapidly yet is ubiquitous in plants (Hilu and Liang 1997; Kelchner 2002). From the perspective of molecular biology, the *trnK* intron is of interest because it represents an unusual form of a group II intron, and the MatK protein has been suggested to have novel properties as a maturase.

Group II introns are self-splicing RNAs and mobile elements found in eubacteria, archaea, and the organelles of fungi, plants, and algae (Bonen and Vogel 2001; Lambowitz and Zimmerly 2004). Although some group II introns consist of an RNA structure alone, many encode a reverse transcriptase (RT) protein within the intron RNA structure, which gives the intron the ability to invade new sites in a genome. The protein also has a second important function in facilitating intron splicing, which it does by binding to the intron RNA structure and stimulating its innate self-splicing properties. Because such a splicing (maturation) function is generally specific to its host intron, such proteins are called maturases.

The *trnK* intron differs from typical group II introns because its encoded protein, MatK, is a degenerate version of an RT (fig. 1*A*). Canonical group II intron ORFs contain three conserved domains: an RT domain containing subdomains 0–7; an X domain associated with maturase activity; and an optional En (endonuclease) domain. A fourth DNA-binding domain (D) is located between X and En and has been characterized functionally but is not conserved in sequence (San Filippo and Lambowitz 2002). In contrast, the *matK* ORF aligns only for RT subdomains 5–7 (poorly conserved) and X (highly conserved), although it is roughly the same size as other group II intron ORFs (Mohr, Perlman, and Lambowitz 1993). The retention of domain X argues that MatK proteins retain maturase activity; however, mobility functions appear to have been lost because the RT catalytic site residues and other RT motifs are not present (Mohr, Perlman, and Lambowitz 1993). Despite these predictions, the complete biochemical functions of MatK have not been demonstrated definitively (below).

In contrast to the ORF structure, the RNA structures of *trnK* introns are fairly typical of group II introns and fall into the IIA1 class of introns (Michel, Umesono, and Ozeki 1989). Group II intron RNA structures consist of six domains arranged around a central wheel (fig. 1*B*) (Qin and Pyle 1998). Domain I is largest and constitutes about half of the intron's size, while domain V is the most highly conserved in sequence and is considered the active site of the ribozyme. The *matK* ORF is encoded within an extended loop of about 2 kb within domain IV, as is typical of other group II introns. Although there are no gross defects in the RNA structure of *trnK*I1, this intron along with other plant organellar introns has not been reported to self-splice in vitro (Michel and Ferat 1995; Barkan 2004). Presumably, splicing in vivo relies on protein cofactors, of which the MatK protein is the most obvious candidate.

There has been considerable speculation about a function of MatK beyond its presumed host intron–specific maturase activity. Over 10 years ago, it was noticed that *Epifagus virginiana*, a nonphotosynthetic plant, possesses a reduced chloroplast genome that has shed large portions of DNA, including the *trnK* gene and its intron, yet *matK* is retained as a free-standing ORF (Wolfe et al. 1992). MatK was therefore suggested to have an important function in chloroplasts beyond splicing of its resident intron. This function was suggested to be splicing of other group II introns still present in the reduced chloroplast genome. If true,

[1] Present address: Department of Radiation Oncology, BC Cancer Agency, 600 West 10th Avenue, Vancouver BC V5Z 4E6, Canada.

Key words: chloroplast, group II intron, matK, maturase, trnK.

E-mail: zimmerly@ucalgary.ca.

Mol. Biol. Evol. 23(2):380–391. 2006
doi:10.1093/molbev/msj047
Advance Access publication November 2, 2005

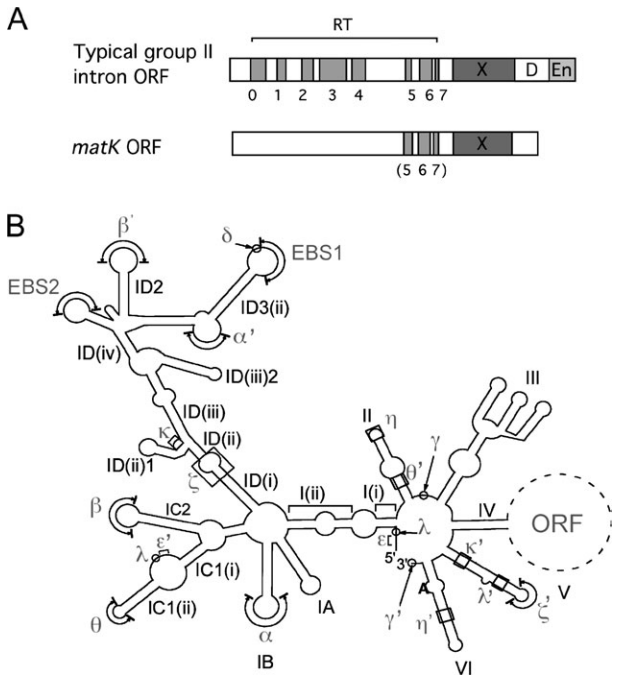

FIG. 1.—Structures of the *matK* ORF and typical group II intron ORFs. (*A*) A typical group II intron ORF contains four domains, RT, X, D, and En, while *matK* ORFs have only remnants of domains 5–7 and a well-conserved domain X. The schematics are derived from the *Lactococcus lactis* Ll.ltrB intron (GenBank accession number U50902) and *Arabidopsis trnK*I1 (GenBank accession number NC_000932), and are drawn to scale. (*B*) A typical secondary structure of the A1 class, to which *trnK*I1 belongs. The structure consists of six domains surrounding a central wheel. Notation of subdomains is shown in black, and long-range tertiary contacts are in gray (e.g., α forms six Watson-Crick base pairs with α′; θ is a non–Watson-Crick interaction with θ′).

MatK would be the only known "generalized" maturase for group II introns, with a role beyond splicing of its host intron (Ems et al. 1995). Additional indirect evidence for this idea came from the observation that in barley several mutations disrupting chloroplast protein synthesis also block splicing of a subset of group II introns, most of which are IIA in structure, as is *trnK*I1 (Hess et al. 1994; Hübschmann, Hess, and Börner 1996; Vogel et al. 1997; Vogel, Börner, and Hess 1999). Because MatK is the only chloroplast-encoded protein with a putative role in group II intron splicing, it is the obvious candidate to affect splicing of the other introns, although it remains formally possible that an uncharacterized chloroplast-encoded protein is involved. Another related notable speculation is that a generalized maturase may be encoded in the nucleus. Angiosperm nuclear genomes contain four group II intron maturase-related genes (nMat-1a, nMat-1b, nMat-2a, and nMat-2b), which are not associated with an intron structure; these ORFs are not closely related to *matK* (Mohr and Lambowitz 2003).

Previously, we proposed a model for the evolution of group II introns, termed the retroelement ancestor hypothesis, which predicts that the ancestor of all known group II introns was a retroelement in bacteria (Toor, Hausner, and Zimmerly 2001). A similar idea was also described by Fontaine et al. (1997). In our hypothesis, a mobile, RT-encoding ribozyme is proposed to have migrated from bacteria to mitochondria and chloroplasts, where the ORF was sometimes lost and the RNA structure sometimes degenerated. This model rationalizes the observation that many plant group II introns do not encode ORFs, have dubious RNA foldings, and that no higher plant group II introns have been shown to self-splice in vitro (Michel and Ferat 1995; Ostheimer et al. 2003). According to this scenario, plant group II introns have lost autonomous ribozyme activity and rely on host-splicing factors to rescue the defects, as well as to regulate the splicing process. Several such host-splicing factors have been characterized in plant chloroplasts (Jenkins, Kulhanek, and Barkan 1997; Perron, Goldschmidt-Clermont, and Rochaix 1999; Jenkins and Barkan 2001; Rivier, Goldschmidt-Clermont, and Rochaix 2001; Till et al. 2001; Ostheimer et al. 2003).

To examine the issue of RNA structural degeneration during plant evolution, we investigated RNAs of a single intron across plants to follow the structural changes and to test the idea that plant group II introns are derived from bacterial-like introns with "standard" RNA structures, which then degenerated. The *trnK* intron was chosen because of its presence from algae to higher plants and the availability of sequences from many organisms. Moreover, it encodes an ORF that can be used to link the introns to other ORF-containing introns. The combined RNA structural data support the existence of an ancestor with standard group II intron RNA features, which subsequently accumulated deviations and degenerations throughout the molecule during plant evolution. We suggest that the patterns of RNA evolution seen here will apply to other group II introns in plants.

Materials and Methods

Sequence Acquisition

Genomic DNA was prepared from *Equisetum arvense* as described (Hausner et al. 1999). The *trnK*I1 sequence was polymerase chain reaction (PCR) amplified using standard methods with the primers 5′-GGGTTGCTAACTCAACGGTAG-3′ and 5′-GGTTGCCCGGGACTCGAACCCGGAACTCGTCGG-3′, followed by cloning and sequencing (GenBank accession number AY348551). Charophyte sequences were obtained from isolates F140C (*Chara connivens*; GenBank accession number AY170442), F146 (*Nitella opaca*; GenBank accession number AY170449), and F138 (*Tolypella nidifica*; GenBank accession number AY170450) as previously described (Sanders, Karol, and McCourt 2003). Other *trnK* sequences were downloaded directly from GenBank with the following accession numbers: *Amborella trichopoda* NC_005086; *Anthoceros formosae* AB086179; *Arabidopsis thaliana* NC_000932; *Atractylodes koreana* AB008760; *Chaetosphaeridium globosum* AF494278; *Cycas panzhihuaensis* AF143440; *Lotus japonicus* NC_002694; *Marchantia polymorpha* X04465; *Nicotiana tabacum* NC_001879; *Nymphaea alba* NC_006050; *Oryza sativa* AF148650; *Pellia borealis* AF238498; *Physcomitrella patens* NC_005087; *Pinus thunbergii* D11467; *Plagiomnium insigne* AY522573; *Podocarpus macrophyllus* AF228111; *Porella baueri* AY168653; *Psilotum nudum* AP004638; *Sphaghum inundatum* AY342156; *Torreya grandis* AF228108; *Zea mays* ZMA86563.

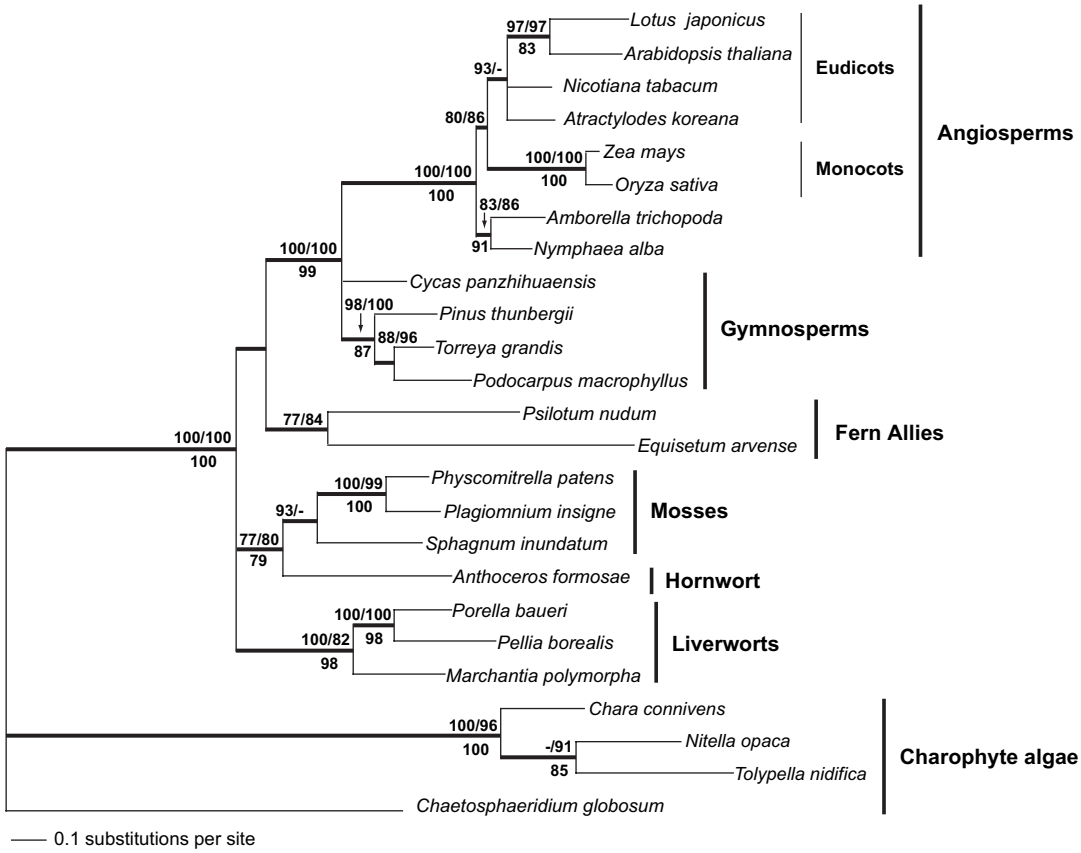

FIG. 2.—Phylogenetic relationships among *matK* ORFs. Twenty-five representative *matK* ORFs were aligned along their lengths with ambiguously aligned positions excluded (424 amino acids total). Their phylogeny was inferred from Bayesian analysis (JTT + Γ + I model) and was outgroup rooted with the *Chaetosphaeridium* sequence. Thick branches denote ≥95% posterior probability in the Bayesian inference. Bootstrap values for 2,000 replicates are also shown for maximum likelihood analysis (JTT + Γ + I model) (top left number), minimum evolution (top right), and unweighted maximum parsimony (bottom). Only bootstrap values ≥75% are indicated. The phylogram agrees with known plant phylogenies, as expected.

Phylogenetic Analysis

Sequences were aligned initially with the ClustalX program (Thompson et al. 1997) and manually refined. For figure 2, the final data set after ambiguously aligned positions were removed was 424 amino acids, 376 of which were parsimony informative (PI). ProtTest was used in part to identify the best-fit model for subsequent analyses (Abascal, Zardoya, and Posada 2005) [Jones-Taylor-Thornton {JTT} + Γ + I model ($\alpha = 2.35$; $I = 0.04$) (Jones, Taylor, and Thornton 1992)]. Metropolis-coupled Markov chain Monte Carlo Bayesian analysis was performed using the program MrBayes V3.0b4 (Ronquist and Huelsenbeck 2003). This analysis was initiated from a random starting tree, and four chains were run simultaneously for 2,000,000 generations, with trees sampled every 100 generations. After discarding the first 100,000 trees ("burn-in"), posterior probabilities were computed from the remaining trees. Bootstrap replicates (2,000 each) were generated using three different phylogenetic methods. Minimum evolution values were obtained using the programs SEQBOOT, PROTDIST, and FITCH within the PHYLIP 3.63 package (Felsenstein 2004). Maximum likelihood values were generated using PHYML with four rate categories and optimization of tree topology, branch lengths, and rate parameters (Guindon and Gascuel 2003). Finally, unweighted maximum parsimony values were obtained using PAUP* V4.0b10 (Swofford 2003), using a heuristic search method with random sequence addition starting trees (10 rounds) and Tree Bisection-Reconnection branch rearrangements. Gaps were treated as missing data.

For figure 3, a second data set was assembled with the same *matK* sequences plus representatives from major classes of group II introns. These sequences were aligned across RT subdomains 5–7 and domain X, and after unalignable positions were removed, 143 amino acids remained (139 PI). Phylogenetic reconstruction was as described above, with the exception of the evolutionary model. Due to the highly divergent data set and small number of characters, a relatively simple model of evolution was chosen (JTT). It has previously been shown that complicated models (e.g., identified by likelihood ratio tests) used with highly divergent data sets can result in incorrect topologies (Posada and Crandall 2001*a*; Piontkivska 2004). In addition, the ability of model selection tests to choose the correct model is compromised when, as is the case here, the number of characters is small (Posada and Crandall 2001*b*). Bayesian and maximum parsimony analyses were performed as described above, and the NEIGHBOR program (PHYLIP3.63 package) was used to obtain a second set of bootstrap values. Bacterial class C was chosen as an

outgroup because it has the most divergent RNA structure. Rooting the tree with other classes of ORFs did not affect the placement of *matK* ORFs within the ML lineage.

RNA Secondary Structure Folding

Intron RNAs were folded using MFOLD (Zuker 2003). Initial secondary structure calculations were progressively refined by the addition of secondary structure constraints to produce agreement with consensus structures (Michel, Umesono, and Ozeki 1989; Toor, Hausner, and Zimmerly 2001). Structures were also refined by detailed comparisons with relatives, with the requirement that alignable sequence in two introns be folded identically. The *Arabidopsis* structure proved problematic in the ID(iii) region and was resolved by comparison with four related angiosperm sequences not otherwise used in this article (not shown). Remaining uncertainties after extensive comparisons are listed in the legend of Supplementary Data Figure 1 (Supplementary Material online).

Results and Discussion

Initially, we collected *trnK* intron and *matK* ORF sequences from representative plants to sample the diversity of *trnK* sequence. Thousands of *matK* sequences are present in the databases, but many entries contain only ORF sequence and lack the flanking intron RNA sequence. Among full-length *trnK* intron sequences, there was good representation among higher plants; however, the only early-branching plant representatives were the charophyte alga *C. globosum* and the liverwort *M. polymorpha*. To increase data for primitive representatives, we PCR amplified and sequenced *trnK*I1 from *E. arvense* (horsetail fern; GenBank accession number AY348551) and also from three additional charophytes (*C. connivens*, *N. opaca*, and *T. nidifica*; GenBank accession numbers AY170442, AY170449, and AY170450, respectively). Subsequently, *trnK* sequences for two other liverworts, three mosses, and a hornwort were reported to GenBank. The final data set of *trnK* sequences represents the spectrum of plant diversity and comprises 25 introns, including four charophyte algae (*C. globosum*, *C. connivens*, *N. opaca*, and *T. nidifica*), three liverworts (*M. polymorpha*, *P. baueri*, and *P. borealis*), a hornwort (*A. formosae*,) three mosses (*S. inundatum*, *P. patens*, and *P. insigne*), two fern allies (*P. nudum* and *E. arvense*), four conifers (*P. thunbergii*, *P. macrophyllus*, *T. grandis*, and *C. panzhihuaensis*), two early-branching angiosperms (*A. trichopoda* and *N. alba*), two monocots (*O. sativa* and *Z. mays*), and four eudicots (*A. thaliana*, *N. tabacum*, *Lotus japonica*, and *A. koreana*). All the sequences were complete except for three of the conifers and *Tolypella*, which were missing some or all of the RNA domain VI.

It is of interest to examine which organisms lack *trnK* or its intron, as this gives information about intron gain and loss. As previously noted, the earliest branching plants to contain *trnK*I1 are charophyte algae (Turmel, Otis, and Lemieux 2002; Sanders, Karol, and McCourt 2003), while earlier branching algae (red algae, glaucophytes, and green algae) contain only *trnK* but no intron (*Chlorella* GenBank accession number NC_005353; *Chlamydomonas* GenBank accession number NC_001865; and *Nephroselmis* GenBank accession number NC_000927). Even the early-branching streptophyte *Mesostigma* lacks the intron (GenBank accession number NC_002186), suggesting that a group II intron inserted into the *trnK* gene of a primitive streptophyte after divergence of chlorophytes and *Mesostigma*. In all instances, the intron-encoded ORF has the degenerate features of *matK* ORFs diagrammed in figure 1(*A*).

A second observation is that no RNA structure is associated with the *matK* sequence of the fern *Adiantum*, based on the lack of a domain 5 and other diagnostic features, nor is there a *trnK* gene in the genome (GenBank accession number NC_004766). The retention of *matK* without the *trnK* gene resembles the situation of *Epifagus* and strengthens the argument for a second function for the MatK protein, which is most likely facilitation of splicing other group II introns retained in the genome.

To reconstruct the evolutionary history of the *trnK* intron, we first analyzed the *matK* ORFs phylogenetically to establish a framework for later comparison with RNA structures. Amino acid sequences were aligned, with ambiguously aligned positions excluded, for a data set of 424 amino acids. The resulting phylogeny (fig. 2) is in agreement with known plant phylogenetic relationships and indicates strict vertical inheritance as expected (Savolainen and Chase 2003).

A more informative analysis examined the relationship between *matK* ORFs and other group II intron ORFs, as it has implications for the origin of the *trnK* intron. The ORF alignment included representatives from all major intron subgroups and included sequence from RT domains 5–7 and X, again with unalignable positions excluded, for a final data set of 143 amino acids. Phylogenetic analyses indicated that *matK* ORFS fall within the ML (mitochondrial-like) class of introns (fig. 3*A*). (It should be noted that the ML class is not only a phylogenetic clade of intron ORFs that are predominant in fungal mitochondria but also contains bacterial members; *matK* is the only known chloroplast member in this class.) The support placing *matK* within the ML class is moderate in figure 3(*A*) because of few characters and high sequence divergence. However, the conclusion is further supported by sequence identity within domain X shared by ML class and *matK* ORFs but not other classes (fig. 3*B*). This sequence conservation was previously observed between *matK* and fungal mitochondrial introns (Mohr, Perlman, and Lambowitz 1993), while domain X of the chloroplast-like (CL) clade has been noted to be divergent from the ML clade (Zimmerly, Hausner, and Wu 2001). Additional support placing *matK* in the ML clade comes from the shared A1 RNA secondary structures of ML and *trnK* introns, while other classes have B or C class RNA structures (Michel, Umesono, and Ozeki 1989; Toor, Hausner, and Zimmerly 2001). Together, one can firmly conclude that the *trnK* family of introns descended from an RT-encoding intron of the ML class with an A1 RNA structure. However, the phylogenetic resolution in figure 3(*A*) and in other analyses (not shown) were not sufficient to suggest close relatives, and so one cannot distinguish if the *trnK*I1 ancestor came from a bacterial, mitochondrial, or unknown chloroplast source.

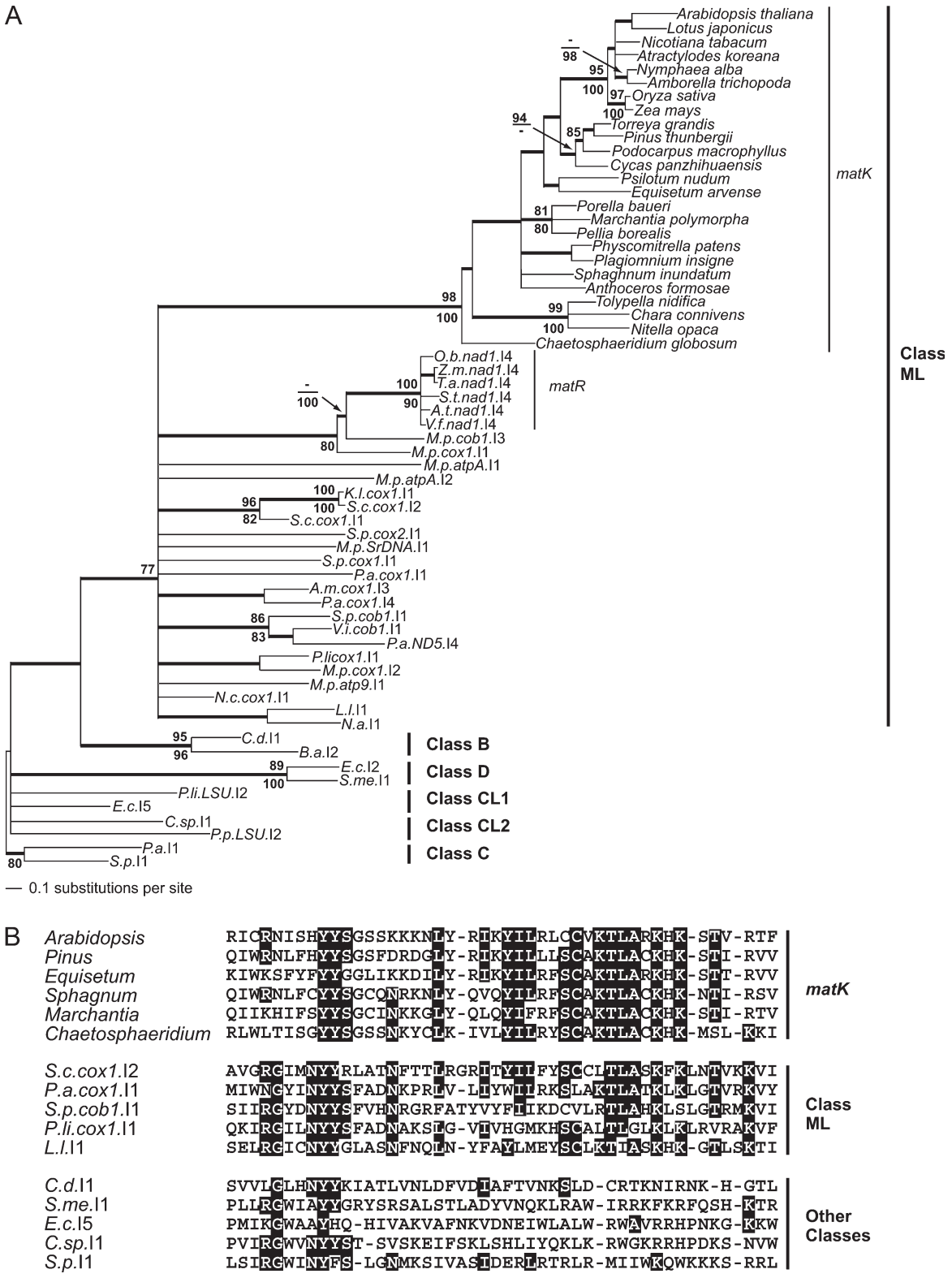

FIG. 3.—Phylogenetic relationship of *matK* ORFs with respect to other group II intron ORFs. (*A*) RT subdomains 5–7 and domain X were aligned for *matK* and ORFs of representative introns of Classes B, C, D, CL1, and CL2 (bacterial classes B, C, and D and CL classes 1 and 2; see Zimmerly, Hausner, and Wu [2001] for class definitions). The phylogeny shown is the most likely tree retrieved from Bayesian analysis (JTT model). Thick lines denote Bayesian posterior probabilities ≥95%. Bootstrap values for 2,000 replicates are shown for neighbor-joining (JTT model; top) and unweighted maximum parsimony (bottom). The tree was outgroup rooted with ORFs from class C. Species abbreviations and GenBank accession numbers are as follows: *O.b.*: *Oenothera berteriana*, M63034; *Z.m.*: *Zea mays*, U09987; *T.a.*: *Triticum aestivum*, X57965; *S.t.*: *Solanum tuberosum*, AJ003130; *A.t.*: *Arabidopsis thaliana*, X98300; *V.f.*: *Vicia faba*, M30176; *M.p.*: *Marchantia polymorpha*, M68929; *K.l.*: *Kluyveromyces lactis*, X57546; *S.c.*: *Saccharomyces cerevisiae*, V00694; *S.p.*: *Schizosaccharomyces pombe*, X54421 (*cob1*I1), AJ251292 (*cox1*I1), and AJ251293 (*cox2*I1); *P.a.*: *Podospora anserina*, X55026

Together, the information supports the following model for the origin of the *trnK* intron and the degeneration of its ORF (fig. 4). The distant ancestor is predicted to be an ML class intron encoding a RT ORF and being competent for mobility, and this ancestor intron might have come from any source (bacterial, mitochondrial, or chloroplast). The ancestor intron likely invaded the *trnK* gene in streptophytes after the divergence of chlorophytes and *Mesostigma*, placing the origin of the *trnK* intron and MatK protein at approximately 1200–800 MYA (Yoon et al. 2004). Sometime after the initial insertion but before divergence of other derived charophytes, the intron lost its mobility properties by acquiring drastic mutations in RT subdomains 0–4 and En and more modest mutations in RT subdomains 5–7 and X. It is possible that this degeneration coincided with the second function of MatK as a generalized maturase (Ems et al. 1995; Vogel, Börner, and Hess 1999). Consistent with its putative function as a maturase, the entire length of the ORF is under selective pressure (Young and dePamphilis 2000), in agreement with experiments showing that maturase function relies on the entire RT domain in addition to the X domain (Cui et al. 2004). Subsequent to ORF degeneration, the *trnK* intron was inherited only vertically, because its mobility functions were lost.

In order to examine evolution of the intron RNA structures, the *trnK*I1 sequences were folded into secondary structures, based on agreement with consensus structures for IIA1 introns and comparative support (Michel, Umesono, and Ozeki 1989; Toor, Hausner, and Zimmerly 2001) (see *Methods*). Structure models of individual introns can be found in Supplementary Data Figure 1 (Supplementary Material online). Although one might expect the foldings to be straightforward because the introns are homologous and several structures are published (Michel, Umesono, and Ozeki 1989), it was in fact difficult to arrive at refined structures, due to many sequence changes, numerous deletions and lengthy insertions, and substantial structural variation. Consequently, it was critical to use extensive sequence comparisons among related *trnK* sequences to arrive at accurate foldings. It was found that with careful comparisons, regions initially appearing irregular could be refolded into fairly conventional structures with minor irregularities. This complication underscores the general difficulty of folding plant organellar introns into accurate secondary structures, in contrast to bacterial group II introns, and shows the necessity of detailed structural comparisons for plant introns. Despite these comparisons, some uncertainties remained, which are listed in the legend of Supplementary Data Figure 1 (Supplementary Material online). In all cases, the uncertainties correspond to deviations for which no folding could be found to agree with the consensus structure or with closely related introns, and we predict that these regions deviate from the standard structure, even if the modeled pairings are not completely accurate.

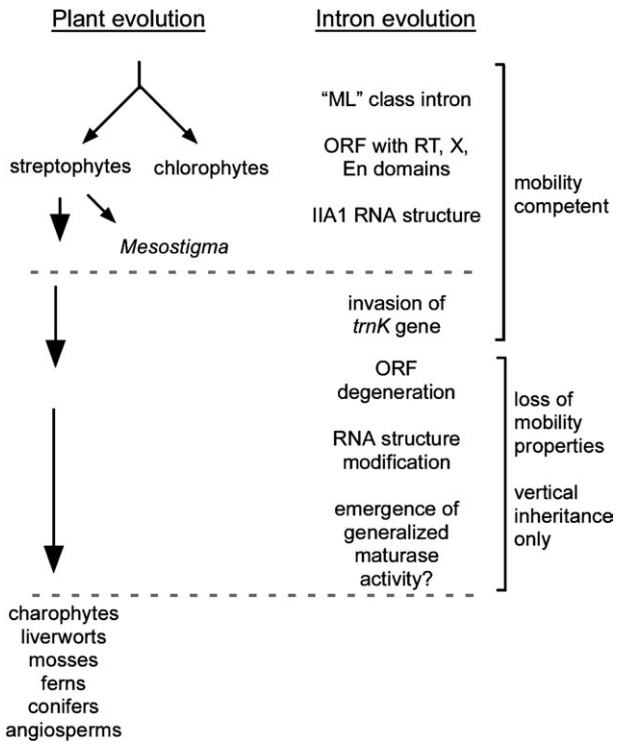

FIG. 4.—Model of *trnK*I1 origin and ORF evolution. The ancestor of *trnK* introns is predicted to be a mobile intron of the ML clade having an ORF with RT, X, and En domains and a typical A1 secondary structure. The intron probably invaded the *trnK* gene in streptophytes after the divergence of *Mesostigma*. Following ORF degeneration and RNA structural changes, the intron lost mobility properties and was inherited only vertically.

The final secondary structures were compared with each other to analyze their differences. First, consensus structures were made for each subgroup of *trnK* introns (e.g., mosses, eudicots), with the consensus structure representing 100% identity of homologous positions in the RNA structures. Further consensus structures were made for combined subgroups until a consensus structure was obtained for all *trnK* introns. Comparisons among the consensus structures and among individual introns showed many deviations, which are listed in table 1. The changes can be categorized according to how many introns contain them, which suggests the timing of acquisition. One type of deviation is found in isolated examples, and suggests a relatively recent change. For example, there is a deletion of domain ID(ii)1 in *Arabidopsis* and a deletion of domain IC2 in *Porella* which can be predicted to have occurred recently because they are found in only one example (table 1).

←

(*cox1*I1, *cox1*I4, and ND5I4); *A.m.*: *Allomyces macrogynus*, U41288; *V.i.*: *Venturia inaequalis*, AF004559; *P.li.*: *Pylaiella littoralis*, Z72500 (*cox1*I1) and Z48620 (LSUI2); *N.c.*: *Neurospora crassa*, X14669; *L.l.*: *Lactococcus lactis*, U50902; *N.a.*: *Novosphingobium aromaticivorans*, AF079317; *C.d.*: *Clostridium difficile*, AF333235; *B.a.*: *Bacillus anthracis*, AF065404; *E.c.*: *E. coli*, X77508 (*E.c.*I2) and AF074613 (*E.c.*I5); *S.me.*: *Sinorhizobium meliloti*, Y11597; *C.sp.*: *Calothrix* sp., X71404; *P.p.*: *Pseudomonas putida*, AF101076; *P.a.*: *Pseudomonas alcaligenes*, U77945; *S.p.*: *Streptococcus pneumoniae*, AF030367. (*B*) Alignment of the domain X region. The most conserved region within domain X was aligned for representative classes, and there is clear alignment between *matK* and ML class ORFs but not ORFs of other classes. Black shading indicates ≥50% identity.

Table 1
Deviations Within *trnK*I1 RNA Structures Compared to the Predicted ML Class Ancestor

	Wheel Sequence					IB																		
	5′ End	J1/2	J2/3	J5/6	3′ End	I(i) Stem	α-α′	IC1 (ii)	θ Motif	IC2 Size	ID(ii)1	ID3(ii)*	EBS1 Sym-metry	ID(i)/ I(ii) Spacer	D2 η Motif	D2 Stem	D3 Size	D3 Internal Loop Motif	D4	D5 Stem	D5 Bulge	D5 Loop	D6	Extra bp in D6**	Other
ML class consensus	GUGCG	GA	CGGA	GG**	AC	4 bp	~6 bp	6 bp	GAAA		1-5 bp				GNRA					9bp +5 bp	AC	GAAA			
Chaetosphaeridium		UUA	AAGA	G	AU	2 bp	0	2 nt B		65	3 bp		sh	1	Loss***		67	Loss		1′ MP	AU		1′ MP	+	Loss of ID(iii)2
Chara		AUA	AAGA		AU	2 bp	3	2 1-nt B's		39	un-paired			7	Loss***	1 bMP	193	Loss		1″ MP	AU			–	>200 nt insertion in D6
Nitella		ACA	AAGA		AU	1 bp	3 bp	5 bp	Loss***	39	1 nt B	1 MP		7	Loss***	1 bMP	77	Loss		1″ MP	AU			–	
Tolypella		AUAGA	AAGA	ND	ND	0 bp	3 bp	2′ MP	Loss***	33	3 nt B	2 MP		4	Loss***		75	Loss		2 MP	GU		ND	ND	
Marchantia		AUG	GAGA		AU	1′ MP	UU MP	2″ MP		57		1 MP	sh	3	Loss***		47	Loss			AA	AAAA		–	
Pellia		GUG	GAGA		AU	1″ MP	UU MP	1 MP	AAAA + dMP	63		4 MP	sh	2	Loss***		55	Loss	1 bMP		AA			–	ID(iv) mispaired
Porella		GUG	AAGA		AU	1″ MP	UU MP	1 MP		8		3 MP		2	Loss***		56	Loss	1′ MP		AA			–	
Sphagnum		GUG	AAGA	G	AU	3 bp		3 MP	AAAA	48		1 MP	sh	4	Loss***		147	Loss			AA			+	
Physcomitrella		GUG	GAGA	G	AU	3 bp		3 MP		16		2 MP		5	Loss***		86	Loss	1 bMP		AA	AAAA	1′ MP	+	
Plagiomnium		UUG	GAGA	G	AU	3 bp		3 MP		15		2 MP		4	Loss***		71	Loss	1″ MP		AA	AAAA	1′ MP	+	
Anthoceros		UUG	AAGA	G	AU	3 bp	UU MP	1 nt B	GGAAA	167		1 MP	sh	7	Loss***		186	Loss		1‴ MP	AA			+	
Equisetum		UUG	AAGA	G	AU		3 bp	5	1 bMP	Loss***	138	3 nt B opp		sh	9	Loss***		56	Loss	1‴ MP	AA		1′ MP	+	
Psilotum		UUG	GAGA	G	AU	2 MP			5 bp	Loss***	131	1 nt B opp		sh	22	Loss***		156	Loss	1‴ MP	AA		1′ MP	+	
Pinus		GUA	GAGA		AU		1′ nt B	5	3 bp	Loss***	193	2 MP	3 MP	sh	3	Loss***		135	Loss		AA	GAGA	1′ MP	–	~50 nt insertion in ID(iii); Irr zeta
Podocarpus		GUA	GAGA		ND		1′ nt B	5	3 bp	Loss***	157		3 MP	sh	6	Loss***		108	Loss		AA		ND	ND	
Torreya		GUA	GAGA		ND		2 MP	5	3 bp	Loss***	161		5 MP	sh	7	Loss***		111	Loss		AA		ND	ND	
Cycas		AUA	GAGA		ND		1″ nt B	2	3 bp	Loss***	171	1 MP	3 MP	sh	5	Loss***		130	Loss	3 MP	AA		ND	ND	
Nymphaea		GUA	AAGA		AU	3 bp	UU MP	4	3 bp	Loss***	107		2 MP	sh	31 nt SL	Loss***		158	Loss	1‴ MP	1‴ MP	AA			–	
Amborella		GUA	AAGA	AG	AU	3 bp	UU MP	4	3 bp	Loss***	103		2 dMP	sh	30 nt SL	Loss***		170	Loss		1‴ MP	AA			–	Irr ID(iii)
Zea	GUAAG	AUA	AAGA		AU	3 bp		4	3 bp	Loss***	52		1 MP	sh	34 nt SL	Loss***		150	Loss		1‴ MP	AA			–	
Oryzae		AUA	AAGA		AU	3 bp		4	3 bp	Loss***	60		1 MP	sh	32 nt SL	Loss***		162	Loss		1‴ MP	AA			–	
Atractylodes		AUA	AAGA		AU	3 bp	UU MP	4	3 bp	Loss***	99		3 dMP	sh	19 nt SL	Loss***	1 MP	205	Loss	1⁗ MP	1‴ MP	AA			–	
Lotus	GCGCG	AUA	AAGA		AU	3 bp	UU MP	4	4 bp	Loss***	92		1 MP	sh	31 nt SL	Loss***		196	Loss		1‴ MP	AA			–	MP in 1D(iii)
Arabidopsis		AUA	AAGA		AU	3 bp		4	3 bp	Loss***	100	Loss	1 MP	sh	24 nt SL	Loss***		169	Loss		1‴ MP	AA			–	
Nicotiana		AUA	AAGA		AU	3 bp	UU MP	4	2 bp	Loss***	88		1 MP	sh	21 nt SL	Loss***		174	Loss		1‴ MP	AA			–	

Abbreviations: MP, mispair; bMP, mispair at the base of a stem; dMP, mispair on the distal side of a stem; B, bulge; B opp, bulge in the opposite strand; SL, stem-loop; Irr, irregular folding; ND, no data; sh, shifted symmetry in the EBS1 loop (see Supplementary Data Figure 1; Michel, Umesono, and Ozeki 1989; Toor, Hausner, and Zimmerly 2001).

Distinct mispairs and bulges in different positions are denoted by one to five ticks.

* All mispairs in this column are distinct.

** The missing nucleotide in J5/6 coincides with an extra bp in D6. These structures could be redrawn with a predicted less stable GG in J5/6 combined with an unusual loop at the bulged A in D6. In either case, the structure would deviate from the mitochondrial consensus.

*** It is not possible to exclude that a theta-like or eta-like interaction is supported by the diverged sequence.

In contrast, the loss of the α-α′ pairing in *Chaetosphaeridium* cannot be dated because no close relatives are available. A second category of deviation is general to a phylogenetic subgrouping, indicating that a change occurred in a common ancestor. This type of deviation includes the loss of the θ motif in vascular plants and the gain of a stem-loop in angiosperms between ID(i) and I(ii). Third, some RNA structural features are shared among all *trnK* intron RNAs, but differ from the consensus of ML introns, suggesting that these features were present in the common ancestor of *trnK* introns. Examples include the loss of the D2 tetraloop with the η-η′ interaction and a γ-γ′ pairing of AU, whereas most ML introns have a GC pair. Finally, in addition to substantial changes in sequences and motifs, there are a multitude of minor changes that are observed in both conserved and nonconserved regions, such as mispairs, small insertions, or minor sequence changes (see Supplementary Data Figures 1 and 2, Supplementary Material online). It should be noted that C to U RNA editing may correct some of these deviations and restore pairing to a more conventional structure, as has been shown in the case of domain 6 of the wheat *nad1*I4 intron (Farré and Araya 1999); however, this possibility would apply to only a small proportion of the single mispair deviations and not to the more numerous indels or extended mispairings.

Based on these differences, we can reconstruct the evolution of the intron RNA structure during plant evolution. This reconstruction is summarized briefly in figure 5 and in detail in Figure 2 in Supplementary Data (Supplementary Material online). The distant ancestor is predicted to be a typical mobile group II intron of the ML clade, having a standard A1 RNA structure and self-splicing activity (fig. 5*A*). The last common ancestor of *trnK* introns had an RNA structure with the following modest modifications from the ML consensus: the η-η′ and β-β′ interactions were lost, as was the internal loop motif of domain 3; the γ-γ′ pairing was A-U; and the base pair at the base of domain 5 was A-U, whereas most ML introns have a G-C pair (fig. 5*B*). Among the early-branching charophytes, there were many significant deviations from the standard structure that suggest compromised RNA structure. The *Chaetosphaeridium* intron lost the α-α′ pairing and the ID(iii)2 motif. The *Chara* intron acquired weakened I(i) and α-α′ pairings, lost stable pairing for ID(ii)1, gained two mispairs in domain 5, and gained large insertions in domains 3 and 6. In *Tolypella*, all pairing in domain I(i) was lost, while both *Tolypella* and *Nitella* lost the θ-θ′ interaction. Together, these changes suggest a compromised structure compared to the ancestral intron, with the changes being acquired independently in each charophyte lineage.

In land plants, different deviations occurred compared to charophytes (fig. 5*C*). As a group, land plant *trnK* introns acquired an AA in place of the conserved AC bulge in domain 5. The introns also diverged individually, for example, by losing domain IC2 (*Porella*, *Physcomitrella*, and *Plagiomnium*), gaining a large insertion in IC2 (*Anthoceros*) or domain 3 (*Sphagnum*, *Anthoceros*), losing pairing in ID(iv) (*Pellia*), or gaining a mispair in domain 5 (*Anthoceros*). As a group, the main change among vascular plants was the loss of the θ-θ′ interaction (fig. 5*D*), while there were many individual changes. Hornwort, fern allies, and conifers obtained substantial insertions in IC2, while all except *Equisetum* obtained a large insertion in domain 3. *Pinus* gained an insertion within ID(iii), and *Psilotum* gained an insertion between 1D(i) and I(ii). Among the clade of angiosperms, the intron gained a conserved A-A mispair in domain 5 and an insertion of a stem-loop between ID(i) and I(ii) (fig. 5*E*). Again, individual changes are seen, which occurred recently, including a change in the conserved GUGYG sequence to GUAAG in *Zea* and loss of ID(ii)1 in *Arabidopsis*. It should be noted that all *trnK* introns except four had a variety of mispairs in ID3(ii) (EBS1 stem) but in different places (table 1). Such mispairs are not common in other group II introns, and the variety seen for the *trnK* introns suggests a need for flexibility in positioning EBS1. We also note that the earliest branching introns (charophytes) appear more defective than introns in higher plants, a generalization that seems to apply to other introns as well (R. Olson, I. Johnson, S. Zimmerly, unpublished data).

Together, the data illustrate that RNA structural deviations accumulate during plant evolution, from a predicted ancestor that resembles a self-splicing intron. While some changes introduce new structures (e.g., the ID(i)/I(ii) stem in angiosperms), most changes are losses of motifs associated with ribozyme activity. A second major type of change comes from insertions or deletions in peripheral regions; however, the newly inserted sequences do not form conserved structures that might contribute to ribozyme function (e.g., the D3 insertion in conifers cannot be folded into a common structure); hence, we argue that motif loss is the prevailing trend during *trnK* intron evolution. Mispairs within stem structures are also very common among individual introns, even within the highly conserved domain 5. While it is impossible to conclude the biochemical consequences of all the individual changes, it is hard to imagine that cumulatively they do not compromise ribozyme function of the *trnK* intron, considering that many changes are in conserved regions and important motifs (e.g., mispairs in domain 5; loss of α-α′, κ-κ′, θ-θ′, and η-η′ interactions). This conclusion is supported by mutational studies of some of these motifs in other introns. Mispairing of the α-α′ interaction in the yeast aI5γ intron blocked splicing (Harris-Kerr, Zhang, and Peebles 1993); point mutations in the highly conserved residues of the κ motif also blocked splicing (Boudvillain and Pyle 1998); and mutations in the θ motif of the yeast aI5γ and *cox1*I1 introns decreased efficiency of splicing, while mutations of the η motif had modest effects (Costa et al. 1997).

The pattern of RNA deviations for the *trnK* intron agrees with previous studies on the evolution of the *petD* intron in angiosperms and gymnosperms (Löhne and Borsch 2005). Mutations did not accumulate uniformly throughout the intron during evolution but were concentrated in domains II, III, and IV (less critical regions). Domains Ic and II were the most variable in sequence, with hotspots occurring in loops rather than stems. The *rpl16* intron of Myoporaceae similarly showed heterogeneous substitutions, with highest levels in domain 2 (Kelchner 2002).

The pattern of ongoing degeneration of the *trnK*I1 RNA structure during plant evolution is consistent with

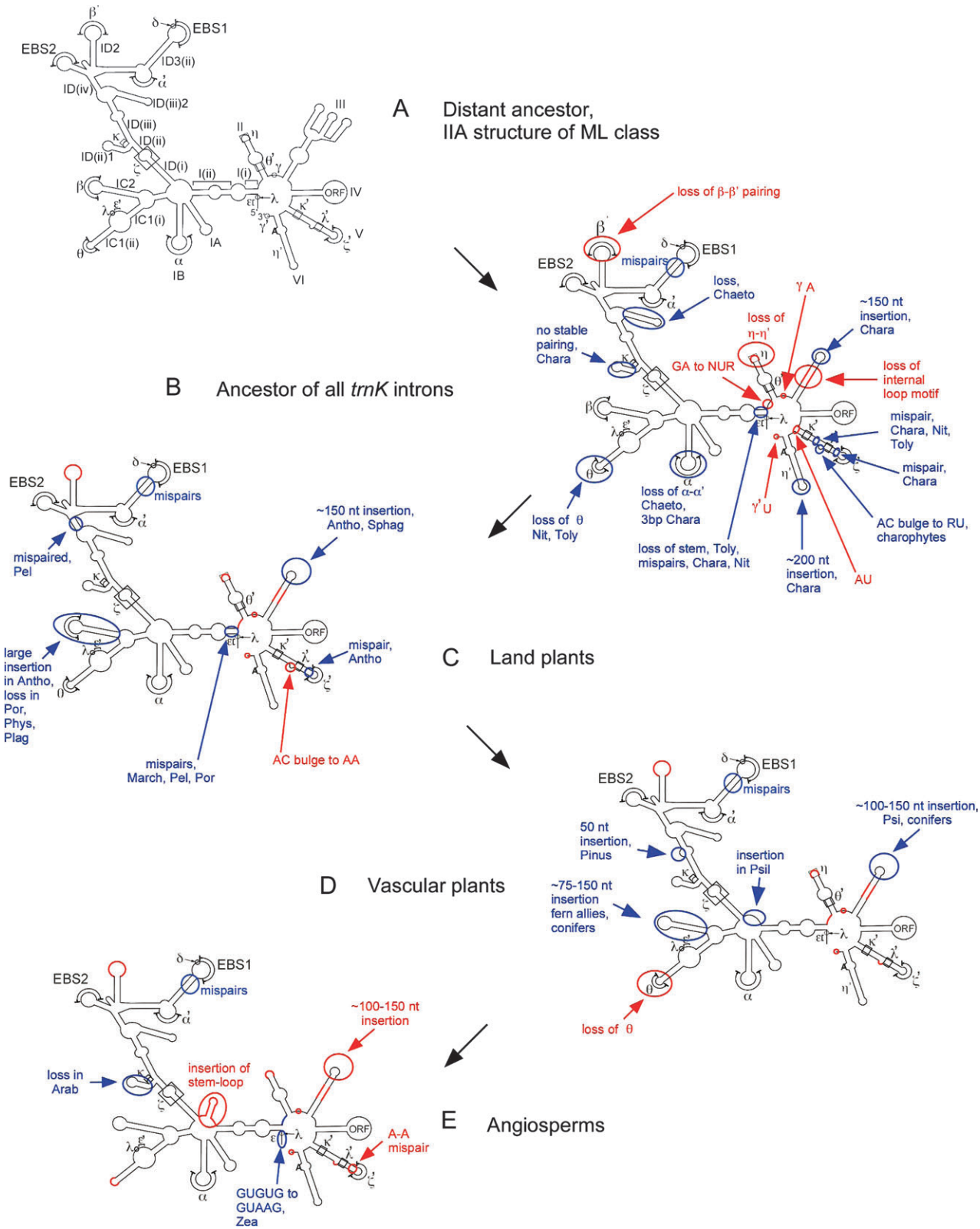

FIG. 5.—Summary of evolution of the *trnK*I1 secondary structure. Folded introns were analyzed for differences, and an evolutionary progression was reconstructed as described in the text. The brief summary here outlines the most important changes, while a detailed depiction can be found in Supplementary Data Figure 2 (Supplementary Material online). Features in blue are changes found for an individual intron or localized subsets of introns. Features in red indicate differences retained by all descendant introns. A comparison of the panels illustrates an accumulation of deviations from the ancestral, putatively self-splicing structure. Abbreviations are as follows: Chaeto, *Chaetosphaeridium globosum*; Chara, *Chara connivens*; Nit, *Nitella opaca*; Toly, *Tolypella nidifica*; March, *Marchantia polymorpha*; Pel, *Pellia borealis*; Por, *Porella baueri*; Sphag, *Sphagnum inundatum*; Phys, *Physcomitrella patens*; Plag, *Plagiomnium insigne*; Antho, *Anthoceros formosae*; Equi, *Equisetum arvense*; Psi, *Psilotum nudum*; Arab, *Arabidopsis thaliana*; and Zea, *Zea mays*.

the scenario that the self-splicing function of group II introns came to require host-splicing factors in organelles to alleviate deficiencies in ribozyme function (Lambowitz et al. 1999; Toor, Hausner, and Zimmerly 2001; Barkan 2004). Splicing factors appear to have been adapted independently in different organisms, often from proteins with other known functions (Lambowitz et al. 1999; Barkan 2004). It follows that as splicing factors emerged in organelles to alleviate ribozyme defects, further RNA structure degeneration would be tolerated, until eventually the splicing reaction would be under the control of the host cell. If this scenario is true, it is interesting to consider that patterns of intron structure degeneration may correspond to the specific splicing factors available. Introns dependent on different factors may consequently have distinct patterns of RNA structure irregularities.

To test whether the observed patterns of RNA structure evolution extend to other group II introns, we examined the *nad1*I4 (*nad1*i728) intron of plant mitochondria, which encodes the MatR protein. In some respects, this intron is the mitochondrial equivalent of the *trnK* intron because it is the only intron in its organelle to encode an ORF. The intron is found in all land plants diverging after liverworts (Qiu et al. 1998; Qiu and Palmer 2004), and like MatK, the function of MatR is not completely clear. The ORFs have two insertions totaling nearly 300 amino acids, located between domains 4 and 5, and 7 and X, and the ORFs are also missing domains 0–1 (Zimmerly, Hausner, and Wu 2001). It is tempting to suggest that the function of MatR might also be as a generalized splicing factor because it is the only potential maturase encoded in the mitochondrion, but there is no specific evidence.

The RNAs of the five angiosperm *nad1*I4 introns were folded, along with *Sphagnum* and *Notothylas* (hornwort) introns (Supplementary Data Figure 3, Supplementary Material online). All introns were found to have lost domain IC2, to contain mispairings in the epsilon motif region, and to have acquired inserted nucleotides in the single-stranded flanks of I(i), together suggesting that these deviations were present in the *nad1*I4 ancestor (ancient changes). Deviations for individual introns were also found (recent changes) and include the loss of pairing within 1D2 in *Notothylas*, severe mispairing in I(i) of *Sphagnum*, and IB irregularities in angiosperm structures. Thus, the *nad1*I4 structures support our hypothesis of ongoing RNA structural degeneration for plastid introns.

We also considered additional introns in mitochondria and chloroplasts that do not encode ORFs. The introns *trnA*I1 and *trnI*I1 are highly conserved in chloroplasts, and the introns were folded and compared for *Coleochaete orbicularis* (green alga) and tobacco (not shown). The *C.o. trnA*I1 and *C.o. trnI*I1 lack the motifs ID(ii)1 and IA, respectively, while tobacco have intact motifs, indicating loss of the features in *C. orbicularis* since their divergence. In plant mitochondria, the related introns *Anthoceros punctatus nad5*I1 and *M. polymorpha cox1*I2 were folded and found to differ by minor indels and by a mispair in D5 (*M.p. cox1*I2) and disruptive bulge in D6 (*M.p. cox1*I2; not shown), again indicating degeneration in liverwort since their common ancestor. Similarly, the homologous introns of liverwort *nad2*I1 and *Arabidopsis nad2*I3 were compared. The *Arabidopsis* intron differs by a large insertion in D6 and by mispairs in D5 and other stems in the structure (not shown), which represents a deviation in *Arabidopsis* since the split of the liverwort and *Arabidopsis* ancestors. Therefore, within multiple intron families across plants, and in both organelles, there are clear examples of both shared (ancient) and sporadic (recent) deviations from a presumed ancestral structure, suggesting that the patterns of ongoing RNA structural degeneration seen for *trnK*I1 are general to plants.

Overall, the picture of RNA structure evolution that emerges in this study is consistent with our previous model for evolution of group II introns (Toor, Hausner, and Zimmerly 2001). Specifically, we predicted that organellar group II introns are degenerate versions of mobile bacterial introns, in which the introns often lost their ORFs and the RNA structures frequently degenerated in multiple ways. While this study does not address the issue of ORF loss, it does provide clear evidence for ongoing RNA structural degeneration in a chloroplast intron, which we argue is a general phenomenon for introns in both organelles. An interesting question is this: given the many degenerated forms of introns in plant organelles, why are there so many copies (~20 each in mitochondria and chloroplasts)? Was there an ancient explosion of mobile introns in organelles followed by massive loss of ORFs? Or are the degenerated introns still mobile in some way, perhaps with the help of cellular factors?

Supplementary Material

Supplementary Data Figures 1–3 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

This work was supported by Natural Sciences and Engineering Research Council grant 203717-98 to S.Z., NSERC grant 238289-01 to G.H., and National Science Foundation (NSF) grants DEB9407606 and DEB9978117 to R.M.M. Salary support for S.Z. was from Alberta Heritage Foundation for Medical Research. In addition, G.H. acknowledges the University Research Grants Committee (University of Calgary) for early support of this work. This material is based in part upon work supported while R.M.M. served at the NSF. Any opinion findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF.

Literature Cited

Abascal, F., R. Zardoya, and D. Posada. 2005. ProtTest: selection of best-fit models of protein evolution. Bioinformatics **21**: 2104–2105.

Barkan, A. 2004. Intron splicing in plant organelles. Pp. 281–308 *in* H. Daniell and C. Chase, eds. Molecular biology and biotechnology of plant organelles. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Bonen, L., and J. Vogel. 2001. The ins and outs of group II introns. Trends Genet. **17**:322–331.

Boudvillain, M., and A. M. Pyle. 1998. Defining functional groups, core structural features and inter-domain tertiary contacts essential for group II intron self-splicing: a NAIM analysis. EMBO J. **17**:7091–7104.

Costa, M., E. Dème, A. Jacquier, and F. Michel. 1997. Multiple tertiary interactions involving domain II of group II self-splicing introns. J. Mol. Biol. **267**:520–536.

Cui, X., M. Matsuura, Q. Wang, H. Ma, and A. M. Lambowitz. 2004. A group II intron-encoded maturase functions preferentially in cis and requires both the reverse transcriptase and X domains to promote RNA splicing. J. Mol. Biol. **340**:211–231.

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 Mol. Biol. **29**:721–733.

Farré, J. C., 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 Mol. Biol. **40**:959–967.

Felsenstein, J. 2004. PHYLIP (phylogeny inference package). Version 3.6. Distributed by the author, Department of Genome Sciences, University of Washington, Seattle.

Fontaine, J. M., D. Goux, B. Kloareg, and S. Loiseaux-de Goër. 1997. The reverse-transcriptase-like proteins encoded by group II introns in the mitochondrial genome of the brown alga *Pylaiella littoralis* belong to two different lineages which apparently coevolved with the group II ribosyme lineages. J. Mol. Evol. **44**:33–42.

Guindon, S., and O. Gascuel. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. **52**:696–704.

Harris-Kerr, C. L., M. Zhang, and C. L. Peebles. 1993. The phylogenetically predicted base-pairing interaction between alpha and alpha′ is required for group II splicing in vitro. Proc. Natl. Acad. Sci. USA **90**:10658–10662.

Hausner, G., K. Y. Rashid, E. O. Kenaschuk, and J. D. Procunier. 1999. The development of co-dominant PCR/RFLP based markers for the flax rust resistance alleles at the L locus. Genome **42**:1–8.

Hess, W. R., B. Hoch, P. Zeltz, T. Hübschmann, H. Kossel, and T. Börner. 1994. Inefficient *rpl2* splicing in barley mutants with ribosome-deficient plastids. Plant Cell **6**:1455–1465.

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

Hübschmann, T., W. R. Hess, and T. Börner. 1996. Impaired splicing of the *rps12* transcript in ribosome-deficient plastids. Plant Mol. Biol. **30**:109–123.

Jenkins, B. D., and A. Barkan. 2001. Recruitment of a peptidyl-tRNA hydrolase as a facilitator of group II intron splicing in chloroplasts. EMBO J. **20**:872–879.

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.

Jones, D. T., W. R. Taylor, and J. M. Thornton. 1992. The rapid generation of mutation data matrices from protein sequences. Comput. Appl. Biosci. **8**:275–282.

Kelchner, S. A. 2002. Group II introns as phylogenetic tools: structure, function, and evolutionary constraints. Am. J. Bot. **89**:1651–1669.

Lambowitz, A. M., M. G. Caprara, S. Zimmerly, and P. S. Perlman. 1999. Group I and group II ribozymes as RNPs: clues to the past and guides to the future. Pp. 451–485 *in* R. F. Gesteland, T. R. Cech, and J. F. Atkins, eds. The RNA world. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

Lambowitz, A. M., and S. Zimmerly. 2004. Mobile group II introns. Annu. Rev. Genet. **38**:1–35.

Löhne, C., and T. Borsch. 2005. Molecular evolution and phylogenetic utility of the *petD* group II intron: a case study in basal angiosperms. Mol. Biol. Evol. **22**:317–332.

Michel, F., and J. L. Ferat. 1995. Structure and activities of group II introns. Annu. Rev. Biochem. **64**:435–461.

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

Mohr, G., and A. M. Lambowitz. 2003. Putative proteins related to group II intron reverse transcriptase/maturases are encoded by nuclear genes in higher plants. Nucleic Acids Res. **31**:647–652.

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 Res. **21**:4991–4997.

Ostheimer, G. J., R. Williams-Carrier, S. Belcher, E. Osborne, J. Gierke, and A. Barkan. 2003. Group II intron splicing factors derived by diversification of an ancient RNA-binding domain. EMBO J. **22**:3919–3929.

Perron, K., M. Goldschmidt-Clermont, and J. D. Rochaix. 1999. A factor related to pseudouridine synthases is required for chloroplast group II intron trans-splicing in *Chlamydomonas reinhardtii*. EMBO J. **18**:6481–6490.

Piontkivska, H. 2004. Efficiencies of maximum likelihood methods of phylogenetic inferences when different substitution models are used. Mol. Phylogenet. Evol. **31**:865–873.

Posada, D., and K. Crandall. 2001*a*. Simple (wrong) models for complex trees: a case from retroviridae. Mol. Biol. Evol. **18**:271–275.

———. 2001*b*. Selecting the best-fit model of nucleotide substitution. Syst. Biol. **50**:580–601.

Qin, P. Z., and A. M. Pyle. 1998. The architectural organization and mechanistic function of group II intron structural elements. Curr. Opin. Struct. Biol. **8**:301–308.

Qiu, Y. L., Y. Cho, J. C. Cox, and J. D. Palmer. 1998. The gain of three mitochondrial introns identifies liverworts as the earliest land plants. Nature **394**:671–674.

Qiu, Y. L., and J. D. Palmer. 2004. Many independent origins of trans-splicing of a plant mitochondrial group II intron. J. Mol. Evol. **59**:80–89.

Rivier, C., M. Goldschmidt-Clermont, and J. D. Rochaix. 2001. Identification of an RNA-protein complex involved in chloroplast group II intron trans-splicing in *Chlamydomonas reinhardtii*. EMBO J. **20**:1765–1773.

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

San Filippo, J., and A. M. Lambowitz. 2002. Characterization of the C-terminal DNA-binding/DNA endonuclease region of a group II intron-encoded protein. J. Mol. Biol. **324**:933–951.

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

Savolainen, V., and M. W. Chase. 2003. A decade of progress in plant molecular phylogenetics. Trends Genet. **19**:717–724.

Swofford, D. L. 2003. PAUP*: phylogenetic analysis using parsimony (*and other methods). Version4.0b8. Sinauer Associates, Sunderland, Mass.

Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins. 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. **25**:4876–4882.

Till, B., C. Schmitz-Linneweber, R. Williams-Carrier, and A. Barkan. 2001. CRS1 is a novel group II intron splicing

factor that was derived from a domain of ancient origin. RNA **7**:1227–1238.

Toor, N., G. Hausner, and S. Zimmerly. 2001. Coevolution of group II intron RNA structures with their intron-encoded reverse transcriptases. RNA **7**:1142–1152.

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

Vogel, J., T. Börner, and W. R. Hess. 1999. Comparative analysis of splicing of the complete set of chloroplast group II introns in three higher plant mutants. Nucleic Acids Res. **27**: 3866–3874.

Vogel, J., T. Hübschmann, 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 splice factor. J. Mol. Biol. **270**:179–187.

Wolfe, K. H., C. W. Morden, S. C. Ems, and J. D. Palmer. 1992. Rapid evolution of the plastid translational apparatus in a nonphotosynthetic plant: loss or accelerated sequence evolution of tRNA and ribosomal protein genes. J. Mol. Evol. **35**: 304–317.

Yoon, H. S., J. D. Hackett, C. Ciniglia, G. Pinto, and D. Bhattacharya. 2004. A molecular timeline for the origin of photosynthetic eukaryotes. Mol. Biol. Evol. **21**:809–818.

Young, N. D., and C. W. dePamphilis. 2000. Purifying selection detected in the plastid gene *matK* and flanking ribozyme regions within a group II intron of nonphotosynthetic plants. Mol. Biol. Evol. **17**:1933–1941.

Zimmerly, S., G. Hausner, and X. Wu. 2001. Phylogenetic relationships among group II intron ORFs. Nucleic Acids Res. **29**:1238–1250.

Zuker, M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. **31**:3406–3415.

Franz Lang, Associate Editor

Accepted October 10, 2005