Euphytica **118:** 9–18, 2001.

Effects of RNA editing on the *coxI* evolution and phylogeny reconstruction

Alfred E. Szmidt[1*], Meng-Zhu Lu[2] & Xiao-Ru Wang[1]
[1]*Department of Forest Genetics and Plant Physiology, Swedish University of Agricultural Sciences, 901 83, Umeå, Sweden;* [2] *Forest Research Institute, Chinese Academy of Forestry, Beijing 100091, China; (*author for correspondence)*

Received 18 October 1999; accepted 31 May 2000

Key words: coxI, evolution, gymnosperms, phylogeny, RNA editing

Summary

CoxI genomic and cDNA sequences from gymnosperms and angiosperms were used to study the effects of RNA editing on gene evolution and phylogeny reconstruction. In six gymnosperms harboring edited *coxI* gene the number of nucleotide substitutions at 1st, 2nd and 3rd codon positions was similar. In contrast, in angiosperms, the number of nucleotide substitutions at 1st and 2nd codon positions was much lower than at the 3rd. The *coxI* gene in long-lived gymnosperms evolved much faster than in annual angiosperms. This accelerated rate of nucleotide substitution in gymnosperms is due to accumulation of T-C substitutions at edited sites that can randomly appear at all three codon positions. Editing predominantly occurred at 1st and 2nd codon positions as a result of selection against nonsynonymous T-C substitutions and other types of mutations. The tree topologies for the investigated species based on genomic DNA data were in concordance with their taxonomic positions. The trees based on polymorphic edited sites agreed with trees derived from complete sequence information. This indicates that edited sites are reliable sources of phylogenetic information especially for species that contain many edited sites. However, the fast evolution rate of *coxI* gene in gymnosperms has caused the long branches in the phylogenetic trees. The inclusion of the species with a processed paralog i.e., edited form of the *coxI* gene, affected the topology of phylogenetic trees, especially when the taxon with a processed paralog was closely related to the other species analyzed.

Introduction

Slow rate of evolution has made plant mitochondrial gene sequences not useful for phylogenetic studies at lower taxonomic levels (Laroche et al., 1995). Consequently, there are very few studies addressing variation of mitochondrial coding regions among closely related species. In plants, mitochondrial and chloroplast genetic information is modified by RNA editing resulting in mRNAs that are different from those encoded by the corresponding genes in the genomic (g) DNA (Araya et al., 1994; Freyer et al., 1997; Hanson et al., 1996). In both organelles, specific cytidines (C's) in the primary transcripts are changed to uridines (U's) in the mature mRNAs (Maier et al., 1996; Steinhauser et al., 1999). The occurrence of editing has raised concerns about the effect of this phenomenon on the evolution of edited sequences and their usefulness for phylogenetic reconstruction (Bowe & dePamphilis, 1996; Hiesel et al., 1994; Lu et al., 1998; Pesole et al., 1996).

Recently, we have sequenced and characterized genomic and cDNAs of the mitochondrial *coxI* gene in several closely related gymnosperm species (Lu et al., 1998). Most species included in our study possessed a large number of edited sites ranging between 25 and 34 in a 648 base pairs (bp) region of the *coxI* gene. We also found that the *coxI* gene in gymnosperms evolves nearly five times faster than in angiosperms. In two taxa, *Gingko biloba* and *Larix sibirica*, the *coxI* gene did not contain any edited sites and resembled an edited transcript (Lu et al., 1998). Such sequences are usually regarded as processed paralogs and have been found to affect phylogeny reconstruction (Bowe

& dePamphilis, 1995; 1996). Taking into account the presence of closely related species, the large number of edited sites and the presence of two putative paralogs, our *coxI* data are particularly useful for analysis of the potential effects of RNA editing on the evolution of *coxI* gene and on phylogeny reconstruction.

Three phylogenies of land plants have been constructed using mitochondrial genomic and cDNA sequences (Bowe & dePamphilis, 1996; Hiesel et al., 1994; Pesole et al., 1996). Although these studies agreed that mitochondrial sequences would be useful for phylogenetic analyses, Hiesel et al. (1994) have pointed out that the differences introduced by RNA editing might distort evolutionary similarities of mitochondrial genomic sequences. On the other hand, Bowe & dePamphilis (1996) have argued that cDNAs may provide fewer informative sites and therefore less resolution on a phylogenetic tree. The previous studies on evolution of edited mitochondrial genes included distant species and very few conifers (Bowe & dePamphilis, 1996; Hiesel et al., 1994; Laroche et al., 1997; Pesole et al., 1996). Therefore, it is uncertain how RNA editing would affect phylogenetic analysis of closely related taxa in this group of plants. Furthermore, the number of edited sites in the previously investigated mitochondrial genes was relatively low, which further limits the detailed evaluation of RNA editing effects upon gene evolution and phylogenetic reconstruction.

In the present study, we analyzed genomic DNA and cDNA sequences of the *coxI* gene from representatives of several lineages of gymnosperms and angiosperms. We address the following questions: (*i*) what are the patterns of evolution of highly edited *coxI* sequences in gymnosperms? (*ii*) can mitochondrial genomic and cDNA sequences resolve phylogenetic relationships at low taxonomic levels? (*iii*) do edited sites provide the same phylogenetic information as the sites that are not affected by RNA editing? (*iv*) what are the effects of a faster evolution of the *coxI* gene in gymnosperms than in angiosperms upon phylogeny? (*v*) what are the effects of processed paralogous sequences upon phylogeny?

Material and methods

The DNA and mRNA isolation, the PCR and RT-PCR amplification of *coxI* fragments from genomic DNA and mRNA, and the sequencing of the PCR products were described in our previous study (Lu et al., 1998). The genomic and cDNA sequences of the *coxI* gene used in the present study came from eight gymnosperm species belonging to four different families: Ginkgoaceae – *Ginkgo biloba* (EMBL accession # AJ000352); Pinaceae – *Pinus sylvestris* (AJ000354), *Pinus sibirica* (AJ000355), *Picea abies* (AJ000356), *Larix sibirica* (AJ000351); Taxaceae – *Taxus baccata* (AJ000353); and Cuppressaceae – *Juniperus procera* (AJ000357), *Thuja plicata* (X64833) (Glaubitz & Carlson, 1992; Lu et al., 1998). In addition, *coxI* genomic DNA sequences from four angiosperm species: *Pisum sativum* (X14409), *Oenothera berteriana* (X05465), *Oryza sativa* (X15990), *Zea mays* (X02660), and from *Marchantia polymorpha* (M68929) retrieved from EMBL database were included in the analysis. The cDNA sequences of the four angiosperm species and *M. polymorpha* were obtained by replacing all the non-synonymous T-C substitutions in the genomic sequences with Ts, for encoding the consensus amino acids (Covello & Gray, 1990).

To investigate the effect of the putative processed paralogs in *L. sibirica* and *G. biloba* as well as the effect of edited sites upon the phylogeny reconstruction we analysed 11 different data sets (Table 1). The sets I, III, V and VII included genomic *coxI* DNA sequences for gymnosperm and angiosperm species and differed by the presence or absence of *G. biloba* and/or *L. sibirica*. The sets II, IV, VI and VIII had the same species composition but included *coxI* cDNA sequences. At some edited sites both edited and unedited C's were present in different species. To discriminate between these two situations, the edited C's at these sites were coded as a fifth nucleotide. To better explore the effects of RNA editing on phylogeny reconstruction analysis was also carried out using three additional data sets comprising only six gymnosperm species that possessed edited sites (Table 1). First of these three sets (IX), comprised complete genomic *coxI* sequences. The next data set (X) comprised only the edited sites of the *coxI* gene. As before, the edited Cs were coded as a fifth nucleotide. Finally, the last data set (XI) contained unedited portion of the *coxI* gene.

To assess the sequence dynamics in gymnosperms and angiosperms we used the *coxI* data for six gymnosperms possessing edited sites and four angiosperms. Comparative analysis of the patterns of nucleotide substitutions was carried out separately for gymnosperms and angiosperms as well as using *M. polymorpha* as a reference taxon. Calculations were made

Table 1. Data sets analyzed in the present study

Data Set	I	II	III	IV	V	VI	VII	VIII	IX	X[1]	XI[2]
DNA	gDNA	cDNA	gDNA	cDNA	gDNA	cDNA	gDNA	cDNA	gDNA	gDNA	gDNA
P. sylvestris	+	+	+	+	+	+	+	+	+	+	+
P. sibirica	+	+	+	+	+	+	+	+	+	+	+
P. abies	+	+	+	+	+	+	+	+	+	+	+
L. sibirica	+	+	+	+							
T. plicata	+	+	+	+	+	+	+	+	+	+	+
J. procera	+	+	+	+	+	+	+	+	+	+	+
T. baccata	+	+	+	+	+	+	+	+	+	+	+
G. biloba	+	+			+	+					
P. sativum	+	+	+	+	+	+	+	+			
O. berteriana	+	+	+	+	+	+	+	+			
Z. mays	+	+	+	+	+	+	+	+			
O. sativa	+	+	+	+	+	+	+	+			
M. polymorpha	+	+	+	+	+	+	+	+	+	+	+

[1] Edited sites.
[2] Unedited sites.

using the DAMBE (Xia, 1999) and the MEGA programs (Kumar et al., 1993).

Parsimony analysis was performed using the computer program PAUP 3.1.1 (Swofford, 1993) using Branch and Bound searches with furthest addition sequence. The autapomorphic characters were ignored in all analyses. The *M. polymorpha* was used as an outgroup species. To examine the confidence intervals for the clades obtained for all data sets, bootstrap values (Felsenstein, 1985) were calculated. The bootstrap analysis involved 1000 replicates, and furthest sequence addition. To evaluate the strength of the parsimony result, we calculated the consistency index (CI) (Kluge & Farris, 1969), retention index (RI) (Swofford, 1993) and decay index (Bremer, 1988; Donoghue et al., 1992). In addition, trees were also constructed based on pairwise distances (Kimura, 1980) using the neighbor-joining method of Saitou & Nei (1987). The trees were constructed using MEGA program (Kumar et al., 1993).

Results

Characteristics of genomic and cDNA sequences

There were distinct differences between gymnosperms and angiosperms with respect to the number of nucleotide substitutions at individual codon positions and the effect of editing. In six gymnosperms harboring edited *coxI* gene the number of nucleotide substitutions at first, second and third positions of the codons was similar (182, 222 and 227 respectively). In contrast, in angiosperms, the number of nucleotide substitutions at first and second positions of the codons was much lower (27 and 30 respectively) than at the third codon position (118). Furthermore, the two groups differed with respect to the type of substitutions at the first codon position (Table 2). In gymnosperms, the majority of substitutions at the first codon position involved T-C transition and T-G transversion. Most (147/173) of the T-C substitutions were eliminated by editing. In angiosperms, however, all the observed substitutions at the first codon position were G-A transitions and C-A transversions. At the second

Table 2. Number and type of nucleotide substitutions at individual codon positions in genomic and cDNA (in parentheses) sequences of the *coxI* gene in gymnosperms and angiosperms

Codon position		Gymnosperms			Angiosperms		
		A	G	C	A	G	C
1st	G	0			13		
	C	0	0		3	0	
	T	0	9	173 (26)	0	0	11 (0)
2nd	G	0			3		
	C	0	0		0	0	
	T	0	0	222 (0)	0	3	24 (0)
3rd	G	27			16		
	C	31	9		18	8	
	T	0	5	155 (151)	30	3	43 (43)

Table 3. Characteristics of *coxI* sequences and phylogenetic trees. NT = number of most parsimonious trees, NS = number of steps, CI = consistency index; RI = retention index

Data set	Variable sites	Informative sites	Edited sites	Unedited sites	NT	NS	CI	RI
I	187	116	56	65	1	184	0.810	0.869
II	129	70	–	–	5	121	0.727	0.781
III	186	113	56	62	1	176	0.830	0.883
IV	128	66	–	–	2	110	0.755	0.810
V	186	115	56	62	1	181	0.818	0.874
VI	128	68	–	–	2	115	0.739	0.795
VII	185	112	56	59	1	172	0.843	0.889
VIII	127	64	–	–	2	104	0.769	0.825
IX	160	63	45	22	1	84	0.893	0.917
X	67	42	42	–	1	56	0.875	0.900
XI	93	21	–	21	1	28	0.929	0.949

codon position all observed substitutions in gymnosperms involved T-C transition that was eliminated by editing. In angiosperms, most substitutions at the second codon position also involved C-T transition that was eliminated by editing. At third codon position, the gymnosperms showed total lack of the T-A transversion, which was very frequent in angiosperms (30/107, 28%). In both groups, the most frequent substitution at the third codon position was the T-C transition. However, only four out of 155, and none out of 43 T-C substitutions at the third codon position were eliminated by editing in gymnosperms and angiosperms respectively.

In gymnosperms, all the edited sites at first and second positions of the codons contained either T or edited C. In contrast, some of the edited sites at third codon position contained T, as well as both edited and unedited C. However, all these differences were synonymous. In angiosperms edited sites were found only at the first and second codon positions. The unedited *coxI* sequences from *L. sibirica* and *G. biloba* had Ts at all the non-synonymous edited sites found in other species.

Of the 216 codons analyzed in this study, 56 (25,9%) codons contained edited sites in gymnosperms while only nine (4.2%) edited codons were found in angiosperms. For the six gymnosperms, excluding *G. biloba* and *L. sibirica*, we observed a total of 179 editing events, mostly occurring at the second (94, 52.5%) and first codon positions (71, 39.7%). Only fourteen edited sites (7.8%) were observed at the third codon position. In angiosperms, there were 35 edited sites of which 11 (32%) and 24 (68%) occurred at first and second positions of the codons respectively. In gymnosperms, the ts/tv ratio was 10.69 for gDNA and 3.77 for cDNA. In angiosperms the corresponding ratio was 1.69 and 1.15 respectively. When the species were compared to *M. polymorpha coxI* gene the ts/tv ratio was 2.28 and 1.52 for gymnosperms and 1.25 and 1.20 for angiosperms genomic and cDNA, respectively.

Effects of RNA editing and processed paralogs on phylogeny reconstruction

The number of variable and informative nucleotides in the 648 bp *coxI* region for individual data sets analyzed in our study are given in Table 3. When all 13 species were considered (data set I), the number of informative nucleotides was 116, and nearly half of them (56) included edited sites. At the cDNA level (data set II), the number of informative sites in the investigated 13 species was reduced to 70. When only six gymnosperms and *M. polymorpha* were included in our analysis (data set IX), the number of informative sites was 63 and as many as 45 (ca. 70%) of them were edited. In other words, the majority of the observed sequence divergence among gymnosperms was associated with nucleotides affected by RNA editing.

When both processed paralogs from *L. sibirica* and *G. biloba* were included in the analysis of genomic DNA (data set I), a single most parsimonious tree requiring 184 steps (CI = 0.810, RI = 0.869) was found (Table 3). The two species formed separate basal branches in the gymnosperm clade (Figure 1A). The tree topology for the remaining species agreed with the generally accepted taxonomic relationships inferred

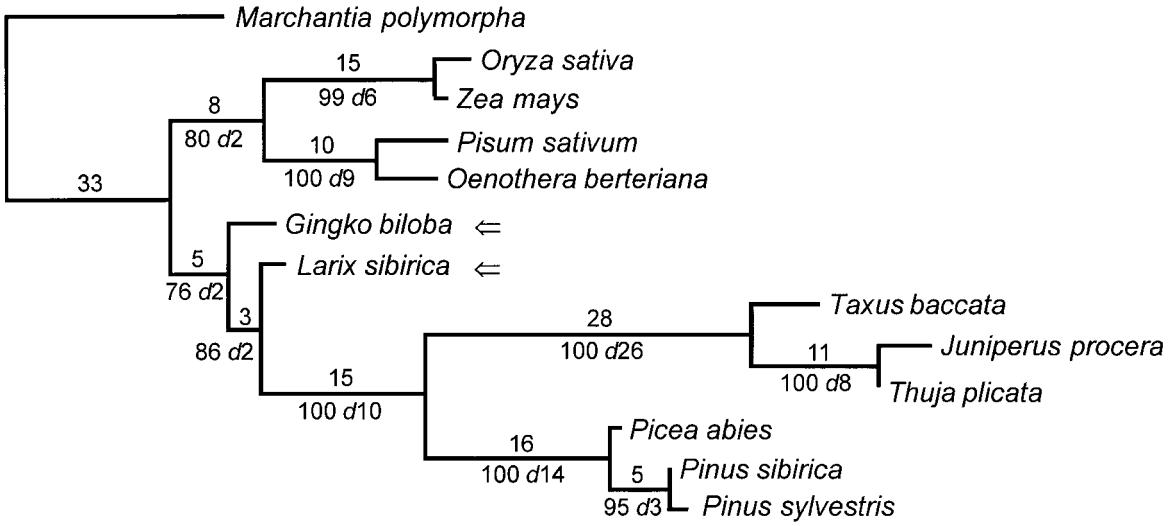

Figure 1A. Single most parsimonious tree found in analysis of *coxI* genomic DNA for gymnosperm and angiosperm species using data set I; refer to Table 3 for more information. Branch length (above) and bootstrap values with corresponding decay indices (*d*) (below). Species with paralogs are marked with arrows.

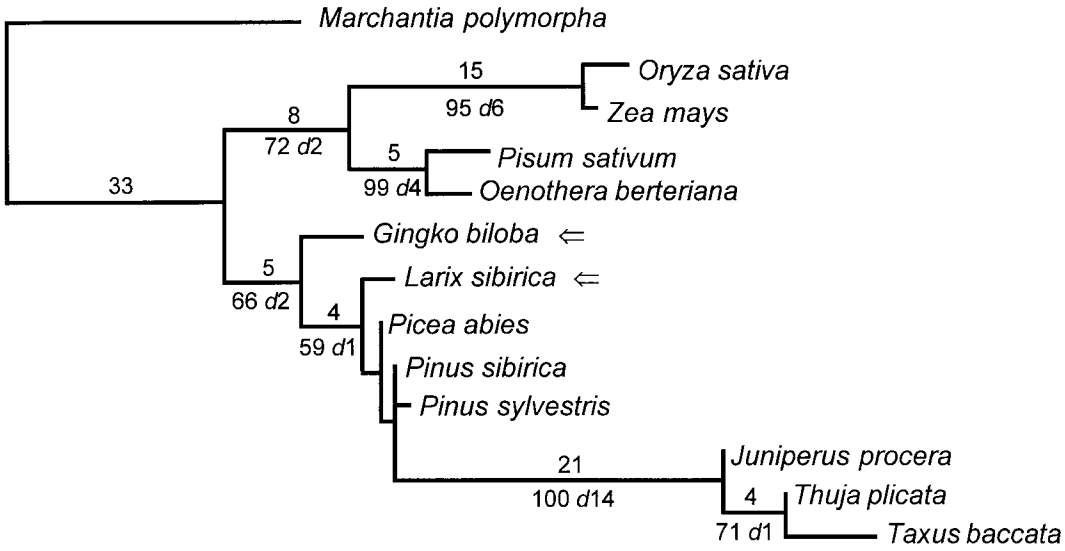

Figure 1B. One of the five most parsimonious trees found in analysis of *coxI* cDNA for gymnosperm and angiosperm species using data set II; refer to Table 2 for more information. Branch length (above) and bootstrap values with corresponding decay indices (*d*) (below). Species with paralogs are marked with arrows.

from morphological and other molecular characters (Chase et al., 1993; Stefanovic et al., 1998). A monophyletic gymnosperm clade, a monocot clade and a dicot clade were clearly formed and separated. The species from the same family were grouped together, and different families were well separated within each clade. For instance, *P. sylvestris*, *P. sibirica* and *P. abies* from Pinaceae family, *J. procera* and *T. plicata* from Cuppressaceae family, formed separate clades. *Taxus baccata* from Taxaceae family was positioned closer to Cuppressaceae species than to the Pinaceae. All clades were supported by high bootstrap and decay index values. Compared to the angiosperm clade, the gymnosperm clade had longer branches in the phylogenetic tree (Figure 1A). For instance, the branch of the clade including *T. baccata*, *J. procera* and *T.*

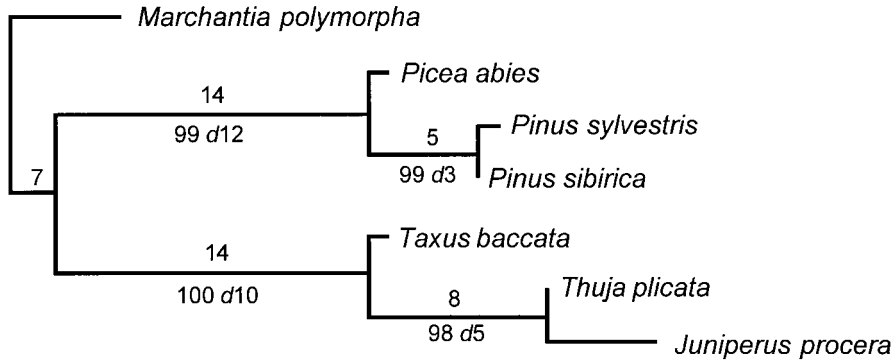

Figure 2. Single most parsimonious tree found in analyses of *coxI* genomic DNA for gymnosperms using data set X; refer to Table 3 for more information. Branch length and bootstrap values with the corresponding decay indices (*d*) are given above and below the nodes respectively.

plicata was nearly two times longer than the branch of the dicot clade.

One of the five most parsimonious trees found in analysis of data set II containing cDNA *coxI* sequences from 13 species is presented in Figure 1B. The trees required fewer steps and had lower CI and RI values than the corresponding tree based on genomic DNA (Table 3). On two of these five trees *L. sibirica* and *G. biloba* again formed separate basal branches in the gymnosperm clade. On the remaining three trees *L. sibirica* branched with *P. abies* occupying different positions in the clade containing the remaining species from the Pinaceae family (trees not shown). Furthermore, for the remaining 11 species, only the main clades representing gymnosperms, dicots and monocots were resolved and in concordance with their taxonomic positions. Individual gymnosperm species from the Pinaceae family were not resolved at all, and the position of *J. procera* and *T. baccata* from the Cupressaceae and Taxaceae respectively, did not agree with their taxonomic relationships (Figure 1B).

When *G. biloba* was excluded from the analysis of genomic DNA (data set III), again a single most parsimonious tree was found (Table 3). *Larix sibirica* formed a separate but relatively short basal branch outside the clade containing other species from the Pinaceae family (tree not shown). Similarly, when *L. sibirica* was replaced by *G. biloba* (data set V) one most parsimonious tree was found (Table 3). Similar to *L. sibirica*, *G. biloba* formed a separate relatively short branch in the gymnosperm clade (tree not shown). The clustering of other taxa remained unaffected. Analysis of the corresponding data sets containing cDNA sequences (sets IV and VI) each produced two most parsimonious trees (Table 3). The topology of these trees (trees not shown) was nearly the same as the topology of cDNA trees obtained with all 13 species (Figure 1B). When both *G. biloba* and *L. sibirica* were excluded from analysis (data sets VII, VIII), parsimony search with genomic DNA sequences yielded a single most parsimonious tree requiring 172 steps (CI = 0.843, RI = 0.889), whereas two trees each requiring 104 steps (CI = 0.769, RI = 0.825) were found with cDNA sequences. All these trees were identical in topology to corresponding genomic and cDNA trees respectively obtained with all 13 species (Figures 1A and B) and are therefore not presented.

Parsimony analysis including genomic *coxI* sequence for six gymnosperm species and *M. polymorpha* (data set IX) gave a single most parsimonious tree requiring 84 steps (CI = 0.893, RI = 0.917). The overall order of individual species was the same as in analysis including both gymnosperm and angiosperm taxa and individual clades were supported by high bootstrap and decay index values (tree not shown). When only edited sites were used (data set X) we found a single most parsimonious tree (Figure 2, 56 steps, CI = 0.875, RI = 0.900). The topology of this tree was identical with the topology of the tree obtained with the entire genomic *coxI* sequence. The taxa were split into two distinct clusters supported by high bootstrap and decay index values. The first cluster comprised all three species from the Pinaceae family and was further split into two groups corresponding to the genus *Pinus* and *Picea* respectively. The second cluster included species from the Cupressaceae and Taxaceae families. Also in this cluster the two species from the same family (Cupressaceae) were grouped together. The topology of a single tree recovered from analysis of unedited sites was similar, however, the members of the Pinaceae family were not resolved (tree not shown). Results obtained using the neighbor-

joining method were nearly identical to those revealed by parsimony analysis and therefore are not presented.

Discussion

Effect of editing on coxI evolution

A common feature of most coding regions studied so far is a lower rate of substitution at first and second codon position than at the third (Li, 1997). Contrary to this expectation, the edited genomic *coxI* sequences in gymnosperms analysed in our study had very similar rate of substitution at all three positions of the codons. Similar result was reported for sequences of two other edited mitochondrial genes (Pesole et al., 1996). Therefore, it has been suggested that the evolutionary analysis of edited mitochondrial genes can be performed by grouping together first, second, and third codon positions, which, evolve following comparable dynamics (Pesole et al., 1996). However, we found that the *coxI* gene in angiosperm species analyzed in our study had much fewer nucleotide substitutions at the first and second codon positions than at the third. It thus appears that this advantage applies only to intensively edited sequences.

Several studies on mitochondrial sequences have suggested that generation time can be a factor that may account for substitution rate heterogeneity (Bousquet et al., 1992; Britten, 1986; Gaut et al., 1992; Laroche et al., 1997; Stephan & Langley, 1992; Wu & Li, 1985). Indeed, woody perennials which have longer generation times compared with annual taxa, were found to exhibit lower numbers of nucleotide substitutions per site (Laroche et al., 1995; Laroche et al., 1997). However, the previous studies of substitution rates in mitochondrial genes did not consider the potential effects of editing. The gymnosperms analyzed in our study are typical representatives of woody perennials with very long generation time. Therefore, they were expected to exhibit slower substitution rate when compared to annual angiosperms. Contrary to this expectation, we found that the *coxI* gene in long-lived gymnosperms is evolving much faster than in angiosperms which have much shorter generation time. In our opinion, the observed accelerated rate of nucleotide substitution on the *coxI* gene in gymnosperms is due to random, free accumulation of T-C substitutions at edited sites (Lu et al., 1998). This suggestion is clearly supported by the fact that most sequence divergence among gymnosperms included in our study was due to T-C transitions that were 'preserved' by editing. Similar 'quasi-neutral' mode of evolution was reported for additional two intensively edited mitochondrial genes (Pesole et al., 1996). It thus appears that accelerated sequence evolution caused by an intensive editing can override slowdown caused by long generation time.

In the recently studied mitochondrial *nad3* and *rps12* genes, there were several cases where both edited and unedited Cs were found at first codon position in different species (Pesole et al., 1996). All except one of these cases were associated with the two codons (CTG and CTA) where the editing of the C nucleotide at first codon position did not result in the amino acid change. On the other hand, the *coxI* gene in gymnosperms analyzed in our study had no unedited Cs at the first codon position. Unlike *nad3* and *rps12* genes, in all *coxI* codons possessing edited Cs at first codon position the lack of editing would lead to amino acid change. The same pattern could be discerned at the third codon position in the *nad3* and *rps12* genes studied by Pesole et al. (1996) and in the *coxI* gene analyzed in our study. Namely, edited and unedited Cs were found only in codons where editing of C does not result in the amino acid change. The occurrence of such redundant synonymous editing indicates that edited Cs can randomly appear at all three codon positions, and that their predominance at first and second codon positions is due to 'purifying' effect of selection against nonsynonymous T-C substitutions and other types of mutations. This suggestion is further supported by the observed absence of redundant editing at first codon position on the *coxI* gene in gymnosperms and by the absence of unedited Cs at the second position of edited codons in gymnosperms and angiosperms.

In the recent survey of 15 mitochondrial sequences in plants, Laroche et al. (1997) have found that transversions occurred more frequently than transitions. Similar result was reported by Wolfe (1987), who observed that transitions made up $<50\%$ of the substitutions in the plant mitochondrial genome. In contrast, we found that in the *coxI* gene sequence within gymnosperms transitions accounted for more than 90% and 79% of substitutions in genomic and cDNA, respectively. This ratio was lower but still larger than one when *M. polymorpha* was used as a reference taxon. Furthermore, we observed pronounced differences between gymnosperms and angiosperms with respect to the number and type of nucleotide substitutions at individual positions of the codons. For

instance, in gymnosperms the most frequent substitution at first codon position was T-C transition and T-G transversion. In contrast, in angiosperms, all the observed substitutions at first codon position involved G-A transitions and C-A transversions. These results provide further evidence for a considerable heterogeneity of sequence dynamics among different mitochondrial genes and plant lineages.

The effects of processed paralogs on phylogeny reconstruction

Several studies have demonstrated the occurrence of the so called 'edited' paralogs (Covello & Gray, 1992; Nugent & Palmer, 1991), which represent the edited forms of mitochondrial genes. Such sequences are not orthologous to the other mitochondrial DNA sequences and tend to cause problems in phylogeny reconstruction (Bowe & dePamphilis, 1996). The *coxI* gene in *G. biloba* and *L. sibirica* included in the present analysis gave some new information about this issue. Similar to results reported by Bowe & dePamphilis (1996), inclusion of these putative paralogs has affected the topology of both genomic and cDNA trees. This was particularly evident for *L. sibirica*, which sometimes appeared outside the clade containing other members of the Pinaceae family. However, in contrast to the previous results, where the processed paralogs gave long branches, the branches for *G. biloba* and *L. sibirica* were relatively short. This discrepancy is probably associated with the different evolutionary histories of the processed paralogs used in our analysis. The paralogs used by Bowe & dePamphilis (1996) were located in the nuclear genome, which evolved faster than their counterparts located in mitochondrial genome. The genomic location of the *coxI* sequence of *G. biloba* is not known. However, Southern analysis of *Larix* species with probe specific for the *coxI* gene revealed maternal inheritance of the hybridizing fragments (DeVerno et al., 1993). This result confirms that at least in *Larix* the *coxI* gene is located in the mitochondrial genome. The relatively short branches obtained with *coxI* sequences from *G. biloba* and *L. sibirica* suggest that they evolve slowly. At least two hypotheses can be put forward to explain this result. First, it is possible that in both *G. biloba* and *L. sibirica* the *coxI* gene is located in the mitochondrial genome and the slow rate of evolution is associated with the lack of edited sites. Second, the *coxI* gene in *G. biloba* is located in the nuclear genome but the transfer occurred only recently. It has been suggested that if a processed paralog is a result of a recent event, or if reinsertion was into the same genomic compartment as the original sequence, its detection through analysis of tree topology may be difficult if possible at all (Bowe & dePamphilis, 1996). Our present results lend additional support to this suggestion. Without the prior knowledge of the *coxI* sequence in *G. biloba*, detection of this particular paralog in our trees would not be possible as the placement of this taxon did not deviate from its expected position.

Phylogenetic informativeness of edited sites

Bowe & dePamphilis (1996) and Pesole et al. (1996) have studied the behavior of genomic and cDNA sequences in phylogenetic analysis. Both studies have suggested that cDNAs may provide fewer informative sites and therefore less resolution in a phylogenetic tree. Results from our present comparative analysis of genomic and cDNA *coxI* sequences provide further support for this suggestion. Analysis of genomic *coxI* sequences from 11 gymnosperm and angiosperm taxa gave expected topology and better statistical support for all resolved nodes than cDNA sequences.

Previous analyses of cDNA from unrelated species have produced fully resolved trees (Hiesel et al., 1994; Bowe & dePamphilis, 1996). Our present results suggest, however, that the cDNA sequences appear to be of little use for resolving tree topologies at lower taxonomic levels. The cDNA trees obtained in our study were not only poorly resolved at the family level, but also some species (*T. baccata* and *J. procera*) were not placed in their expected position on the tree. Furthermore, while all analyses using genomic sequences yielded only a single most parsimonious tree, the number of trees found with cDNA sequences was higher and ranged from one to five.

The previous studies of the effects of RNA editing upon phylogeny reconstruction involved comparison between cDNA and genomic DNA sequences that included both edited and unedited polymorphic sites (Bowe & dePamphilis, 1996; Hiesel et al., 1994; Pesole et al., 1996). In our opinion, this approach provides only indirect information about the effect of the editing upon tree topology as it includes combined effect of both edited and unedited sites. In the present study, we analyzed the direct impact of edited sites upon phylogeny reconstruction by inclusion of edited sites only. We found that it was possible to recover correct topology when only the edited inform-

ative sites were included in the analysis. This result provides further evidence that edited sites not only contain important phylogenetic information, but also that in taxa with a large number of edited sites, such as some gymnosperms, these sites are critical for good tree resolution especially at lower taxonomic levels.

Bowe & dePamphilis (1996) have considered potential problems that may arise when phylogenetic analysis is carried out using sequences with many and only a few edited sites. The authors concluded that T for C substitutions associated with edited sites are not any less reliable than other types of substitutions. In our opinion, this suggestion will be valid for most phylogenetic comparisons. However, we found that the faster evolution of the *coxI* gene in gymnosperms than in angiosperms has caused generally longer tree branches in the former group of taxa, although the general tree topology remained intact. This indicates that simultaneous analysis of sequences containing variable number of edited sites may sometimes present a problem in estimation of the divergence time.

Acknowledgements

This work was supported by grants from the Carl-Tryggers Research Council to XRW and AES, the Swedish Council for Agriculture and Forest Research (SJFR) and the Swedish International Development Agency Department for Research Cooperation (SIDA-SAREC) to AES.

References

Araya, A., D. Begu & S. Litvak, 1994. RNA editing in plants. Physiol Plantarum 91: 543–550.

Bousquet, J., S.H. Strauss, A.H. Doerksen & R.A. Price, 1992. Extensive variation in evolutionary rate of *rbcL* gene sequences among seed plants. Proc Natl Acad Sci USA 89: 7844–7848.

Bowe, L.M. & C.W. dePamphilis, 1995. Phylogenetic significance of RNA editing. Am J Bot 82: 116.

Bowe, L.M. & C.W. dePamphilis, 1996. Effects of RNA editing and gene processing on phylogenetic reconstruction. Mol Bio Evo 13: 1159–1166.

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

Britten, R.J., 1986. Rates of DNA sequence evolution differ between taxonomic groups. Science 231: 1393–1398.

Chase, M.W., et al., 1993. Phylogenetics of seed plants – An analysis of nucleotide sequences from the plastid gene *rbc*L. Ann Mo Bot Gard 80: 528–580.

Covello, P.S. & M.W. Gray, 1990. Differences in editing at homologous sites in messenger RNAs from angiosperm mitochondria. Nucleic Acids Res 18: 5189–5196.

Covello, P.S. & M.W. Gray, 1992. Silent mitochondrial and active nuclear genes for subunit 2 of cytochrome c oxidase (cox2) in soybean: evidence for RNA-mediated gene transfer. EMBO J 11: 3815–3820.

DeVerno, L.L., P.J. Charest & L. Bonen, 1993. Inheritance of mitochondrial DNA in the conifer *Larix*. Theor Appl Genet 86: 383–388.

Donoghue, M.J., R.G. Olmstead, J.F. Smith & J.D. Palmer, 1992. Phylogenetic relationships of *Dipsacales* based on rbcL sequences. Ann Mo Bot Gard 79: 333–345.

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

Freyer, R., M.-C. Kiefer-Meyer & H. Kössel, 1997. Occurrence of plastid RNA editing in all major lineages of land plants. Proc Natl Acad Sci USA 94: 6285–6290.

Gaut, B.S., S.V. Muse, W.D. Clark & M.T. Clegg, 1992. Relative rates of nucleotide substitution at the rbcL locus of monocotyledonous plants. J Mol Evo 35: 292–303.

Glaubitz, J.C. & J.E. Carlson, 1992. RNA editing in the mitochondria of a conifer. Curr Genet 22: 163–165.

Hanson, M.R., C.A. Sutton & B.W. Lu, 1996. Plant organelle gene expression: Altered by RNA editing. Trends Plant Sci 1: 57–64.

Hiesel, R., B. Combettes & A. Brennicke, 1994. Evidence for RNA editing in mitochondria of all major groups of lands plants except the Bryophyta. Proc Natl Acad Sci USA 91: 629–633.

Kimura, M., 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16: 111–120.

Kluge, A.G. & J.S. Farris, 1969. Quantitative phyletics and the evolution of Anurans. Syst Zool 18: 1–32.

Kumar, S., K. Tamura & M. Nei, 1993. MEGA: Molecular Evolutionary Genetics Analysis, version 1.01, The Pennsylvania State University.

Laroche, J., P. Li & J. Bousquet, 1995. Mitochondrial DNA and monocot-dicot divergence time. Mol Biol Evol 12: 1151–1156.

Laroche, J., P. Li, L. Maggia, & J. Bousquet, 1997. Molecular evolution of angiosperm mitochondrial introns and exons. Proc Natl Acad Sci USA 94: 5722–5727.

Li, W.-H., 1997. Molecular evolution. Sinauer Associates, Sunderland, Massachusetts.

Lu, M.-Z., A.E. Szmidt, &X.-R. Wang, 1998. RNA editing in gymnosperms and its impact on the evolution of the mitochondrial *cox* I gene. Plant Molec Biol 37: 225–234.

Maier, R.M., P. Zeltz, H. Kössel, G. Bonnard, J.M. Gualberto & J.M. Grienenberger, 1996. RNA editing in plant mitochondria and chloroplasts. Plant Molec Biol 32: 343–365.

Nugent, J.M. & J.D. Palmer, 1991. RNA -Mediated transfer of gene coxII from the mitochondrion to the nucleus during flowering plant evolution. Cell 66: 473–481.

Pesole, G., L.R. Ceci, C. Gissi, C. Saccone & C. Quagliariello, 1996. Evolution of the *nad3-rps12* gene cluster in angiosperm mitochondria: comparison of edited and unedited sequences. J Mol Evol 43: 447–452.

Saitou, N. & M. Nei, 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol Biol Evo 4: 406–425.

Stefanovic, S., M. Jager, J. Deutsch, J. Broutin & M. Masselot, 1998. Phylogenetic relationships of conifers inferred from partial 28S rRNA gene sequences. Am J Bot 85: 688–697.

Steinhauser, S., S. Beckert, I. Capesius, O. Malek & V. Knoop, 1999. Plant mitochondrial RNA editing. J Mol Evol 48: 303–312.

Stephan, W. & C.H. Langley, 1992. Evolutionary consequences of DNA mismatch inhibited repair opportunity. Genetics 132: 567–574.

Swofford, D.L., 1993. PAUP: Phylogenetic analysis using parsimony, version 3.1.1, Champain, IL. Computer program distributed by the Illinois Natural History Survey.

Wolfe, K.H., W.-H Li, & P.M. Sharp, 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proc Natl Acad Sci USA 84: 9054–9058.

Wu, C.-I. & W.-H. Li, 1985. Evidence for higher rates of nucleotide substitutions in rodents than in man. Proc Natl Acad Sci USA 82: 1741–1745.

Xia, X-H., 1999. DAMBE 3.56 (Software Package for Data Analysis in Molecular Biology and Evolution). Department of Ecology and Biodiversity, University of Hong Kong, Hong Kong.