Plant Molecular Biology **37:** 225–234, 1998.
© 1998 *Kluwer Academic Publishers. Printed in Belgium.*

RNA editing in gymnosperms and its impact on the evolution of the mitochondrial *coxI* gene

Meng-Zhu Lu, Alfred E. Szmidt* and Xiao-Ru Wang
Department of Forest Genetics and Plant Physiology, Swedish University of Agricultural Sciences, 901 83 Umeå, Sweden (author for correspondence)*

Received 15 July 1997; accepted in revised form 8 December 1997

Key words: gymnosperms, mitochondria, molecular evolution, nucleotide substitution, RNA editing

Abstract

Sequence analysis of the mitochondrial *coxI* gene in eight gymnosperm species revealed a high rate of non-synonymous nucleotide substitutions with a strong (98%) predominance of C-T substitutions. Further analysis of the corresponding *coxI* cDNA sequences showed that all the non-synonymous C-T changes in the *coxI* genomic DNA sequences were eliminated by RNA editing resulting in nearly identical mRNA (amino acid) sequences among the species. Pronounced variation in the number and location of edited sites was found among species. Most species had a relatively large number of edited sites (from 25 to 34). However, no RNA editing of the *coxI* sequence was found in *Gingko biloba* or *Larix sibirica*. The sequence composition of the investigated *coxI* fragment suggests that the *coxI* gene in *G. biloba* and *L. sibirica* originated from edited mitochondrial *coxI* transcripts by reverse transcription followed by insertion into the nuclear genome or back into the mitochondrial genome. Our results also demonstrate that where there are a large number of edited sites, RNA editing can accelerate the divergence of nucleotide sequences among species.

Introduction

In terrestrial plants, mitochondrial and at least some of the chloroplast genetic information is modified by RNA editing, resulting in proteins that are different from those encoded by the corresponding genes in the genomic DNA [1, 11, 24]. In both organelles, specific cytidines (Cs) in the primary transcripts are changed to uridines (Us) in the mature mRNAs. This results in critical changes in codons, leading to amino-acid sequences that are more highly conserved in evolution than those encoded in the genomic DNA [1]. The correction of the ‘false’ genomic DNA may release the selection pressure on the DNA sequence and accelerate the rate of nucleotide substitutions. This may be particularly important for species that are believed to harbour many edited sites such as some gymnosperms [8, 12, 14]. However, there is little experimental evidence as yet to support this hypothesis.

It has been postulated that the structure of plant mitochondrial DNA evolves rapidly, but the gene sequences evolve very slowly [9, 20, 26]. However, the suggestion has been based solely on analysis of the genomic DNA sequences from angiosperm species, and little is known about the rates of divergence at the cDNA level. Even less is known on this subject in gymnosperms. More recent studies on RNA editing of various mitochondrial gene products have suggested that RNA editing occurs more frequently in gymnosperms than in angiosperms [8, 12, 14]. These results imply that the evolution of mitochondrial DNA may be more strongly affected by RNA editing in gymnosperms than in angiosperms. However, these conjectures are based upon observations of a very limited number of taxa. They do not, therefore, allow for reliable deductions concerning the intensity of RNA editing and its sig-

The nucleotide sequence data reported will appear in the EMBL Nucleotide Sequence Database under the accession numbers AJ000351 (*Larix sibirica*), AJ000352 (*Ginkgo biloba*), AJ000353 (*Taxus baccata*), AJ000354 (*Pinus sylvestris*), AJ000355 (*Pinus sibirica*), AJ000356 (*Picea abies*) and AJ000357 (*Juniperus procera*).

nificance for mitochondrial DNA evolution in gymnosperms.

Various lines of evidence suggest the possibility of the transfer of mitochondrial sequences to the nucleus or back to the mitochondria [4, 7, 19]. Such sequences, if derived from edited mitochondrial transcripts, would be expected to lack edited sites. So far, however, the empirical evidence for unedited mitochondrial genes is limited to data from a very few angiosperm species and it is unclear whether they also occur in gymnosperms.

In the present study, we surveyed both genomic DNA and cDNA of the *coxI* gene in seven gymnosperm species from four families: *Ginkgo biloba* L. (Ginkgoaceae), *Taxus baccata* L. (Taxaceae), *Pinus sylvestris* L., *Pinus sibirica* Loud. & Mayr, *Picea abies* L. & Karst., *Larix sibirica* Münchn. & Ledeb. (Pinaceae), and *Juniperus procera* Hochst. (Cuppressaceae). Our specific goal was to determine the amount and distribution of RNA editing in different species and to evaluate the effect of RNA editing on the sequence evolution of the *coxI* gene.

Materials and methods

Plant material

The plant species included in this study are listed in Table 1. Leaves of *G. biloba* and needles of *T. baccata*, *P. sylvestris*, *P. sibirica*, *P. abies*, *L. sibirica*, *L. gmelinii*, *L. laricina*, *L. leptolepis* and *J. procera* were collected from documented individuals of each species in the Kórnik Arboretum, Poland, the Hørsholm Arboretum, Denmark and Arboretum Norr, Sweden. The samples were stored at −80 °C until DNA and mRNA extraction.

DNA and mRNA isolation

Total DNA was extracted from needles or leaves by the CTAB method described by Doyle and Doyle [6]. mRNA was purified using a QuickPrep Micro mRNA purification kit (Pharmacia) according to instructions provided by the manufacturer. A 20 μl portion of the final mRNA eluate was treated with 2 U RQ1 RNase-free DNase (Promega) in 50 μl of 1× reverse transcription buffer (Promega) at 37 °C for 1 h, followed by phenol-chloroform extraction to remove the DNase.

DNA and cDNA amplification and sequencing

CoxI primers designed by Glaubitz and Carlson [8] were used to amplify the *coxI* sequence for each species (Table 2). The PCR conditions for *coxI* amplification from genomic DNA were described in our previous study [25]. *CoxI* cDNA was amplified from the mRNA by reverse transcription PCR (Access RT-PCR System, Promega) according to the manufacturer's instructions. To confirm the RT-PCR products were not amplified from possible genomic DNA contaminants, the RT-PCR components in the reaction mixture were used as controls, without reverse transcriptase. Amplified products were first examined by resolving the products in 1.5% agarose gels in 0.04 M Tris-acetate, 0.001 M EDTA buffer pH 7.5 and then sequenced directly using a Sequenase PCR Product Sequencing Kit (Amersham) following the manufacturer's recommendations. The two primers used in the sequencing (Table 2), which were designed according to the sequences published by Glaubitz and Carlson [8], covered the middle part of the *coxI* gene. Four additional pairs of primers were designed for further examination of the *coxI* sequences of *L. sibirica* (pL1, pL2, Table 2) and *G. biloba* (pG1, pG2, Table 2), conforming to the sequence obtained in this study and the sequences published by Sper-Whitis *et al.* [23]. The positions of these primers on the *coxI* gene are presented in Table 2. The amplification products obtained with these primers were subsequently sequenced as described above.

Data analysis

To assess the sequence variation among gymnosperms and angiosperms more fully, the sequences obtained in this study for the gymnosperms were compared with sequences available from the EMBL DNA database: *coxI* genomic DNA and cDNA sequences from *Thuja plicata* [8], and *coxI* genomic DNA sequences from eight angiosperm species (Table 1). The genomic DNA and cDNA sequences were aligned using the Clustal V program [13]. Edited sites were inferred by comparison of the aligned *coxI* genomic DNA and cDNA sequences.

The numbers of synonymous (L_S) and non-synonymous (L_A) sites, the rate of nucleotide substitutions per synonymous (K_S) and per non-synonymous (K_A) site and their standard deviations (SD) were calculated for both the genomic DNA and cDNA sequences, employing the method proposed by Comeron [3] using the KESTIM program developed by the

Table 1. Species included in the present study

	Family	Genus	Species	Source
Gymnosperm	Ginkgoaceae	*Ginkgo*	*G. biloba* L.	Kórnik Arboretum, Poland
	Taxoaceae	*Taxus*	*T. baccata* L.	Hørsholm Arboretum, Denmark
	Cuppressaceae	*Juniperus*	*J. procera* Hochst.	Hørsholm Arboretum, Denmark
		Thuja	*T. plicata*	Glaubitz and Carlson, 1992
	Pinaceae	*Larix*	*L. sibirica* Münchn. & Ledeb.	Arboretum Norr, Umeå, Sweden
		Larix	*L. gmelinii* Rupr. & Rupr.	Arboretum Norr, Umeå, Sweden
		Larix	*L. leptolepis* Sieb. & Zucc & Gord.	Arboretum Norr, Umeå, Sweden
		Larix	*L. laricina* Du Roi & K. & Koch	Arboretum Norr, Umeå, Sweden
		Pinus	*P. sylvestris* L.	Arboretum Norr, Umeå, Sweden
		Pinus	*P. sibirica* Loud. & Mayr	Arboretum Norr, Umeå, Sweden
		Picea	*P. abies* L. & Karst.	Arboretum Norr, Umeå, Sweden
Angiosperm	Fabaceae	*Pisum*	*P. sativum*	EMBL, X14409
	Onagraceae	*Oenothera*	*O. berteriana*	EMBL, X05465
	Chenopodiaceae	Beta	*B. vulgaris*	EMBL, X57693
	Brassicaceae	*Arabidopsis*	*A. thaliana*	EMBL, X94583
	Poaceae	*Triticum*	*T. aestivum*	EMBL, X56186
		Sorghum	*S. bicolor*	EMBL, M14453
		Oryza	*O. sativa*	EMBL, X15990
		Zea	*Z. mays*	EMBL, X02660

Table 2. Primers used in the study.

	Upstream ($5' \rightarrow 3'$)	Position	Downstream ($5' \rightarrow 3'$)	Position
Amplification	TTATTATCACTTCCGGTACT		AGCATCTGGATAATCTGG	
Sequencing	ATGCCATGATCAGTATTGGTG	179–189	CATGCAGAGCAATGTCTAGC	441–460
pL1[a]	GTGTATATTCTCATTCTGCCT	85–105	AAAAATGAATCCTACAGCAAA	373–393
pL2[b]	GTGCATATTCCCATTCCGCCC	85–105	AAAGATGGACCCTGCAGCAGG	373–393
pG1[a]	CTGCCTGGATTCGGTATCATT	100–120	GCAATGTCTAGCCCAGAATTT	432–452
pG2[b]	CTGCCTGGATTCGGTATCATA	100–120	GCAATGTCTAGCCCAGAATTA	432–452

[a] Primers designed for *L. sibirica* and *G. biloba coxI* sequences obtained in this study.
[b] Primers designed for *Larix* spp. and*G. biloba coxI* sequences published by Sper-Whitis *et al.* [23].
The numbering of the positions is in accordance with the sequences presented in Figure 1. Potentially mismatched nucleotides are underlined.

same author. This method represents a modification of previous methods [17, 18, 21] by separating the twofold degenerate sites into two types: only transitional (2S-fold) and transversional (2V-fold) substitutions are synonymous respectively. Subsequently, a two-parameter correcting method [15] for multiple substitutions at a given site was applied, using the overall frequency of observed synonymous transversions to estimate the number of synonymous substitutions, both transversional and transitional, per site. The estimates for the means of K_S and K_A and their standard deviations were based on 1000 replicates for each pair-wise sequence combination. The average values of L_S, L_A, K_S, K_A and SD were calculated separately for the genomic DNA and cDNA sequences according to Wolfe *et al.* [26]. In this paper we define the synonymous and nonsynonymous substitution for cDNA in the classical terms as they are for genomic DNA.

Results

Amplification and sequencing of the coxI *genomic DNA and cDNA*

The expected 750 bp *coxI* fragment [8] was amplified from both genomic DNA and cDNA of all species. A 648 bp *coxI* partial sequence for each investigated

species was obtained by directly sequencing the amplification products (Figure 1). The *coxI* genomic DNA sequences of the four *Larix* species were found to be identical. Thus, only the sequence obtained from *L. sibirica* is presented (Figure 1). Surprisingly, the *coxI* sequences for the four *Larix* species and *G. biloba* obtained in this study were different from the corresponding sequences published by Sper-Whitis *et al.* [23]. To investigate this discrepancy, four additional pairs of primers (pL1, pL2, pG1, pG2, Table 2) were designed for *Larix* species and *G. biloba*, using the *coxI* sequences obtained in this study and those published by Sper-Whitis *et al.* [23]. The differences between the two pairs of primers for each species were mainly located at the 3′ ends (Table 2), which are critical for the amplification of the target fragments [16]. PCR amplification of *coxI* genomic DNA from the four *Larix* species using the pL1 primer pair, specific to the *coxI* sequences obtained in this study, produced a 309 bp fragment, as expected. In contrast, the pL2 primer pair, specific to the *Larix* spp. *coxI* sequence published by Sper-Whitis *et al.* [23], failed to yield any amplification products for the four *Larix* species analysed in this study. The use of lower annealing temperatures in PCR did not help to amplify the *Larix coxI* sequence reported by Sper-Whitis *et al.* [23], resulting instead in apparently non-specific amplification of several fragments. The amplification results obtained with the pL1 and pL2 primers for *L. sibirica* and *L. gmelinii* are presented in Figure 2. The products amplified by the pL1 primers were further sequenced and found to be identical with the *coxI* sequence we found in all the four *Larix* species. Incidentally, the pL2 primers showed a considerable homology to the *P. abies coxI* sequence and amplified the corresponding 309 bp *coxI* fragment from this species. This indicates that the lack of amplified product found using this pair of primers for the *Larix* species was due not to faulty primers, but to their failure to bind to the template sequence.

The pG1 primer pair, based on the *G. biloba coxI* sequence obtained in this study, strongly produced a 353 bp fragment from extracts of *G. biloba*. However, the pG2 primer pair corresponding to the *coxI* sequence published by Sper-Whitis *et al.* [23] yielded a fragment of this size very weakly (Figure 2). The poor amplification was probably due to 3′ A-A mismatches [16] between the corresponding regions on the *G. biloba coxI* template and the pG2 primers (Table 2 and Figure 1). The poorly amplified pG2 product was subsequently concentrated and sequenced. The sequence was found to be identical to that obtained with the pG1 primers, but different from the *G. biloba coxI* sequence reported by Sper-Whitis *et al.* [23]. The pG1 and pG2 primers were also used to amplify *coxI* sequences from *P. abies*. Again, the expected 353 bp fragment was strongly amplified using the pG1 primers, but very weakly amplified using the pG2 primers (Figure 2).

RNA editing of coxI

Comparison of the eight gymnosperm *coxI* cDNA sequences with their corresponding genomic DNA sequences revealed a total of 67 edited sites (Figure 1). While no edited sites were found in the *G. biloba* and *L. sibirica* cDNA, as many as 34 sites were present in that of *P. abies*. The observed 67 edited sites were distributed among 56 codons. The editing changes the genomic amino acid information stored in these codons, and only six of the changes are synonymous. Consequently, after editing, the investigated species differed by only one amino acid due to a G-T substitution (Figure 1). Interestingly, the *coxI* sequences of *G. biloba* and *L. sibirica* had Ts at all the non-synonymous edited sites found in the other species (Figure 1). Species belonging to the same family generally had similar locations and numbers of edited sites. For example, *J. procera* and *T. plicata* from the family Cupressaceae were very similar in these respects, as were *P. abies*, *P. sibirica* and *P. sylvestris* (though not *L. sibirica)*, from the family Pinaceae (Figure 1). In contrast, considerable variation in the number and location of edited sites was found among species from different families (Figure 1).

Nucleotide substitutions among coxI *sequences*

A summary of nucleotide substitutions observed in the investigated species is presented in Table 3. C-T type substitutions accounted for 84% of all substitutions detected in this study. Most of these substitutions were found at the first and second positions of the codons. However, for the angiosperm species, most of the C-T substitutions were found at the third codon position (Table 3). In total, 98% of the non-synonymous substitutions found in gymnosperms were of the C-T type. The C-T type transition was much more frequent than the A-G type transition (12.4:1). Nearly all substitutions at the third codon position in the genomic DNA sequences were preserved in the cDNA sequences but 85% and 100% of the substitutions at the first and second codon positions, respectively, were eliminated by RNA editing (Table 3).

```
      F   N   T   T   F   F   D   P   A   G   G   G   D   P   I   L   Y   Q   H   L   F   W   F   F   G   H   P   E   V   Y
LS   TTT AAT ACA ACC TTT TTC GAT CCT GCT GGA GGG GGA GAC CCA ATA TTA TAC CAG CAT CTC TTT TGG TTC TTC GGT CAT CCA GAG GTG TAT   90
GB   *** *** *** *** *** **T *** *** *** *** *** *** *** **G *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** ***
PSY  *** *** *** *** *** *** *** *** *** *** *** *** *** **G *** *** *** *** *** *** *** C** *** *C* *** *** *** *** *** ***
PSI  *** *** *** *** *** *** *** *** *** *** *** *** *** **G *** *** *** *** *** *** *** C** *** *C* *** *** *** *** *** ***
PA   *** *** *** *** *** *** *** *** *** *** *** *** *** **G *** *** *** *** *** *** *** C** *** *C* *** *** *** *** *** C**
JP   *** *** *** *** C** *** *** *** *** *** *** *** *** **G **C *** *** *** *** *C* CC* *** *** *** **C *** *** *** *** ***
TP   *** *** *** *** C** *** *** *** *** *** *** *** *** **G **C *** *** *** *** *C* CC* *** *** *** **C *** *** *** *** ***
TB   *** *** *** *** CC* *** *** *** *** *** *** *** *** **G **C *** *** *** *** *C* CC* *** *** C** **C *** *** *** *** ***

      I   L   I   L   P   G   F   G   I   I   S   H   I   V   S   T   F   S   G   K   P   V   F   G   Y   L   G   M   V   Y
LS   ATT CTC ATT CTG CCT GGA TTC GGT ATC ATT AGT CAT ATC GTA TCG ACT TTT TCG GGA AAA CCA GTA TTC GGG TAT CTA GGC ATG GTT TAT   180
GB   *** *** *** *** **C *** *** *** *** *** *** *** *** *** **T *** *** *** *** *** **G *** *** *** *** *** *** *** *** ***
PSY  *** *C* *** *** **C *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** **G *** *** *** *** *C* *** *** *** ***
PSI  *** *C* *** *** **C *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** **G *** *** *** *** *C* *** *** *** ***
PA   *** *C* *** *C* **C *** *** *** *** *** *** *** *** *** *** *** *** C** *** *** **G *** *** *** *** *C* *** *** *** ***
JP   *** T*G *** *** *** *** *** *** *** *** **C *** *** *** *** *** *** C** *** *** **G **C *** *** *** **G *** *** *** ***
TP   *** T*G *** *C* *** *** *** *** *** *** **C *** *** *** *** *** *** C** *** *** **G **C *** *** *** **G *** *** *** ***
TB   **C **G *** *C* *** *** *** *** **C *** **C *** *** *** *** *** *** *** *** *** **G **C *** *** *** **G *** *** *** ***

      A   M   I   S   I   G   V   L   G   F   L   V   W   A   H   H   M   F   T   V   G   L   D   V   D   T   R   A   Y   F
LS   GCC ATG ATC AGT ATT GGT GTT CTT GGA TTT CTT GTT TGG GCT CAT CAT ATG TTT ACT GTG GGC TTA GAC GTT GAT ACG CGT GCT TAC TTT   270
GB   *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** ***
PSY  *** *** *** *** *** *** *** *C* *** *** *C* *** C** *** *** *** *** *** *** *** *** *C* *** *** *** *** *** *** C** *C*
PSI  *** *** *** *** *** *** *** *C* *** *** *C* *** C** *** *** *** *** *** *** *** *** *C* *** *** *** *** *** *** C** *C*
PA   *** *** *** *** *** *** *** *C* *** *** *** *** C** *** *** *** *** *** *** *** *** *C* *** *** *** *** *** *** C** *C*
JP   *** *** **A *** *** *** *** *** *** **C *** *** *** *** *** *** *C* *** *** *** *** CC* *** *** *** *** *** *** C** *C*
TP   *** *** **A *** *** *** *** *** *** **C *** *** *** *** *** *** *C* *** *** *** *** CC* *** *** *** *** *** *** C** *C*
TB   *** *** *** *** *C* *** *** *** *** **C **C **C *** *** *** **C *** **C *** *** *** C** *** *** *** *** *** *** C** *C*

      T   A   A   T   M   I   I   A   V   P   T   G   I   K   I   F   S   W   I   A   T   M   W   G   G   S   I   R   Y   K
LS   ACC GCA GCT ACC ATG ATC ATA GCT GTC CCT ACT GGA ATC AAA ATC TTT AGT TGG ATC GCT ACC ATG TGG GGA GGT TCG ATA CGA TAC AAA   360
GB   *** *** *** *** *** *** *** *** *** **C *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** ***
PSY  *** *** *** *** *** *** *** *** *** **C *** *** *** *** *** *** *** C** *** *** *** *** *** *** *** *** *** *** *** ***
PSI  *** *** *** *** *** *** *** *** *** **C *** *** *** *** *** *** *** C** *** *** *** *** *** *** *** *** *** *** *** ***
PA   *** *** *** *** *** *** *** *** *** **C *** *** *** *** *** *C* *** C** *** *** *** *** *** *** *** *** *** *** *** ***
JP   *** *** *** *** *** *** *C* *** *** **C *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** ***
TP   *** *** *** *** *** *** *C* *** *** **C *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** ***
TB   *** *** *** *** *** *** *** *** *** **C *** *** *** *** **A *** *** *** *** *** *** *** *** *** *** *** *** *** *** ***

      T   P   M   L   F   A   V   G   F   I   F   L   F   T   I   G   G   L   T   G   I   V   L   A   N   S   G   L   D   I
LS   ACA CCC ATG TTA TTT GCT GTA GGA TTC ATT TTT CTG TTC ACC ATA GGA GGA CTC ACT GGA ATA GTC CTG GCA AAT TCT GGG CTA GAC ATT   450
GB   *** *** *** *** *** *** *** **G *CT **C *** T** *** *** *** **G *** *** **C *** *** *** *** *** *** *** *** *** *** ***
PSY  *** *** *** *** C** *** *C* **G *C* **C *** *C* *C* *** *** *** *** *** *** *** *** *** *** *** *** C** *** *** *** ***
PSI  *** *** *** *** C** *** *C* **G *C* **C *** *C* *C* *** *** *** *** *** *** *** *** *** *** *** *** C** *** *** *** ***
PA   *** *** *** *** *C* *** *C* **G *C* **C *** *C* *C* *** *** *** *** *** *** *** *** *** *** *** *** C** *** *** *** ***
JP   *** **T *** CC* *C* *** *** **G *C* **C C** *** *C* *** *** *** *** **T *** *** *** *** *** **G *** **C *** *** *** ***
TP   *** **T *** CC* *C* *** *** **G *C* **C C** *** *C* *** *** *** *** **T *** *** *** *** *** **G *** **C *** *** *** ***
TB   *** **T *** CC* *C* *** *** **G *C* **C *** *** *C* *** *** *** *** *CT *** *** *** *** *C* **G *** C*C *** *** *** ***

      A   L   H   D   T   Y   Y   V   V   A   H   F   H   Y   V   L   S   M   G   A   V   F   A   L   F   A   G   F   Y   F
LS   GCT CTG CAT GAT ACT TAT TAT GTG GTT GCA CAT TTC CAT TAT GTA CTT TCT ATG GGA GCC GTT TTT GCT TTA TTT GCA GGA TTT TAC TTC   540
GB   *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** **T
PSY  *** *** *** *** *** C** *** *** *** *** *** *C* *** *** *** *** *** *** *** *** *** *C* *** *C* *C* *** *** *** C** ***
PSI  *** *** *** *** *** C** *** *** *** *** *** *C* *** *** *** *** *** *** *** *** *** *C* *** *C* *C* *** *** *** C** ***
PA   *** *** *** *** *** C** *** *** *** *** *** *C* *** *** *** *** *** *** *** *** *** *C* *** *C* *C* *** *** *C* C** ***
JP   *** *C* *** *** *** C** **C *** *** *** *** *C* *** *** *** *** C*C *** *** *** *** **C *** *** **C *** *** *** *** ***
TP   *** *C* *** *** *** C** **C *** *** *** *** *C* *** *** *** *** C*C *** *** *** *** **C *** *** **C *** *** *** *** ***
TB   *** *** *** *** *** C** **C *** *** *** *** CC* *** *** *** **C C*C *** *** *** *** **C *** *C* **C *** *** *** *** *C*

      W   V   G   K   I   S   G   R   T   Y   P   E   T   L   G   Q   I   H   F   W   I   T   F  F/V  G   V   N   L   T   F
LS   TGG GTG GGT AAA ATT TCT GGT CGA ACA TAC CCT GAA ACT TTA GGT CAA ATC CAT TTT TGG ATC ACT TTT TTC GGG GTT AAT TTG ACC TTC   630
GB   *** *** *** *** **C *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** *** ***
PSY  C** *** *** *** **C C** *** *** *** *** *** *** *** *** *** *** *** *** C** C** *** *** C** *** *** **G *** *C* *** ***
PSI  C** *** *** *** **C C** *** *** *** *** *** *** *** *** *** *** *** *** C** C** *** *** C** *** *** **G *** *C* *** ***
PA   C** *** *** *** **C C** *** *** *** *** *** *** *** *** *** *** *** *** C** C** *** *** C** *** *** **G *** *C* *** ***
JP   C** *** *** **G **C *** *** *** *** *** **C *** *** *** *** *** **T *** *** C** *** *** **C G** *** **G *** *** *** CC*
TP   C** *** **G **G **C *** *** *** *** *** **C *** *** *** *** *** **T *** *** C** *** *** **C G** *** **G *** *** *** CC*
TB   C** *** *** **G **C *** *** *** *** *** **C *** *** *C* *** *** *** *** *** C** *** *** **C G** *** **G *** *** *** C**

      F   P   M   H   F   L
LS   TTT CCC ATG CAT TTC TTG                                                                                                  648
GB   *** *** *** *** *** ***
PSY  *** *** *** *** *** ***
PSI  *** *** *** *** *** ***
PA   *** *** *** *** *** ***
JP   *** *** *** *** *** ***
TP   *** *** *** *** *** ***
TB   *** *** *** *** *** *C*
```

Figure 1. Nucleotide and amino acid *coxI* genomic DNA sequences and RNA edited sites (boxed) in the gymnosperm species. The *L. sibirica* sequence is shown in entirety, and only the nucleotides that differed from it are given for other species. The amino acids affected by RNA editing are marked in bold, underlined characters. The only amino acid substitution that remained after RNA editing is shown by giving the alternative residue. LS, *L. sibirica*; GB, *G. biloba*; PSY, *P. sylvestris*; PSI, *P. sibirica*; PA, *P. abies*; JP, *J. procera*; TP, *Thuja plicata*; TB, *T. baccata*.

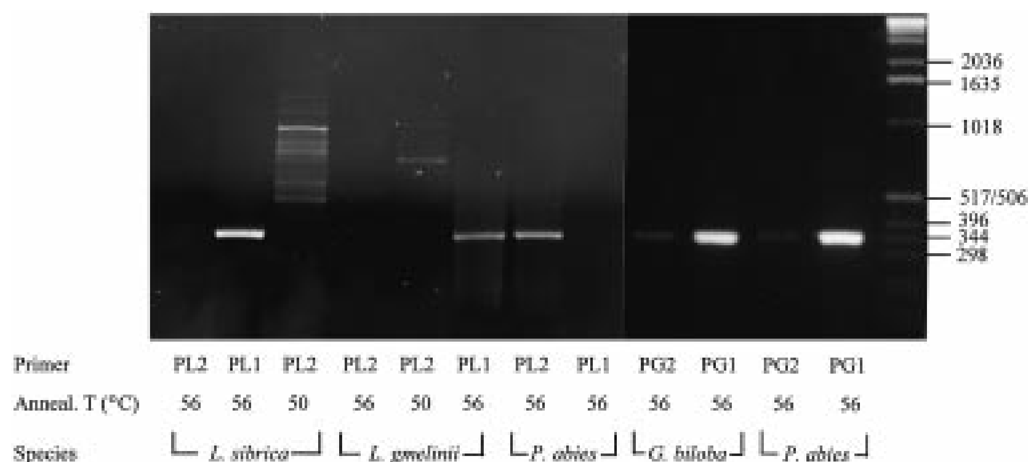

Figure 2. The amplification of *coxI* fragments using different *coxI* primers.

Table 3. The number of different types of nucleotide substitutions at different *cox1* codon positions.

	Genomic DNA				cDNA			
	first	second	third	total	first	second	third	total
Gymnosperms								
C-T	26	34	27	87	4[a]	0	26	30
T-A	0	0	0	0	0	0	0	0
T-G	1	0	3	4	1	0	3	4
A-G	0	0	7	7	0	0	7	7
A-C	0	0	4	4	0	0	4	4
C-G	0	0	1	1	0	0	1	1
Sum	27	34	42	103	5	0	41	46
Angiosperms								
C-T	5	7	22	34				
Other	2	3	36	41				
Sum	7	10	58	75				

[a] All substitutions are synonymous substitutions.

A summary of the estimated *coxI* nucleotide substitutions amongst gymnosperms and angiosperms is presented in Table 4. The numbers of synonymous (L_S) and non-synonymous (L_A) sites were very similar for the two groups of plants. At the genomic DNA level, gymnosperms had fewer synonymous substitutions per site (K_S) than angiosperms. However, the number of non-synonymous substitutions per site (K_A) was nearly 5 times higher in gymnosperms than in angiosperms. As a result of RNA editing, the number of non-synonymous substitutions per site decreased by a factor of 32 among the *cox*I cDNAs of gymnosperm species.

Discussion

Effect of RNA editing on codon and amino acid composition

As noted by Covello and Gray [5], RNA editing is essential to preserve the function of amino acid sequences for those sequences that harbor editable sites. Among the 216 codons analysed in our study, 56 (26%) contained edited sites. For 50 of these codons (23%), RNA editing changed the amino acid information from that specified by the genomic DNA. As a result, the investigated taxa differed by only a single amino acid at the protein level. These results provide

Table 4. The estimated number of synonymous (K_S) and non-synonymous (K_A) substitutions per site for the investigated species.

	L_S	K_S	SD_s	L_A	K_A	SD_A
Gymnosperms						
genomic DNA	211.2	0.104	0.023	521.1	0.064	0.011
cDNA	211.0	0.108	0.023	531.3	0.002	0.001
Angiosperms						
genomic DNA	211.5	0.158	0.033	532.5	0.013	0.004

L_S and L_A are the estimated numbers of synonymous and non-synonymous sites respectively. SD: standard deviation.

further evidence for the biological importance of RNA editing for ensuring the production of functional proteins. Thus, although RNA editing accelerates genomic DNA evolution in gymnosperms by allowing the accumulation of C-T transitions, the final gene products of these species are nearly the same. The biological significance of these nucleotide substitutions is still largely unclear. It is possible that they represent an evolutionary accident [2], or they increase the efficiency of transcription or regulate gene expressions [1, 11].

In gymnosperms, except for a single T-G substitution in the *coxI* sequences, all observed nucleotide substitutions at the first and second codon position were C-T transitions. However, apart from editing sites, genomic DNA and predicted amino acid sequences among gymnosperms *coxI* genes are highly conserved. In contrast, among the angiosperms included in this study, all the nucleotide substitutions were observed distinctly more frequently at the third codon position than at the first and second codon positions. Thus, the observed RNA editing and editing-related C-T transitions are more prevalent in gymnosperms than in angiosperms. This difference between these two groups of plants is likely to represent a combined effect of general sequence divergence, predominant in angiosperms, and editing related sequence divergence, predominant in gymnosperms. However, cDNA data for the angiosperms are necessary to determine the proportion of edited sites in the total number of sites that differ among taxa compared and thus the predominant pattern of sequence divergence.

The efficiency of the protective function of RNA editing depends on the position in the codon of the edited sites. As most substitutions at the third codon position are synonymous, the occurrence of edited sites at this position would have little biological significance. On the other hand, by 'correcting' usually non-synonymous C-T substitutions at the first and second codon positions, RNA editing can be regarded as an important step in decoding genetic information and therefore preserving the function of the coding sequences. The observed predominance of edited sites at the first and second codon positions found in this and other studies is consistent with this view.

Rate of nucleotide substitution / RNA editing stimulates divergence

The number of non-synonymous substitutions per site (K_A) of *coxI* genomic DNA was nearly five times higher in the gymnosperms studied than in the angiosperms. The higher value of K_A in gymnosperm species was mainly due to the high number of non-synonymous C-T substitutions observed, which accounted for 98% of all non-synonymous substitutions detected in our study. This is clearly reflected in the low value of K_A for gymnosperms derived from the cDNA analysis, since RNA editing removes most of the C-T non-synonymous substitutions. In contrast, the rate of synonymous substitutions, K_S, was slightly higher in the angiosperms than in the gymnosperms analysed in this study. Our values of K_S for the angiosperm *coxI* gene are lower than those reported by Wolfe *et al.* [26], probably because these authors analysed a different set of species.

To find out how much RNA editing can affect the rate of sequence divergence among species, it is necessary to determine how many nucleotides differ among species at edited sites. We found a total of 103 nucleotide substitutions among gymnosperm species compared in this study. As much as 67% of these substitutions were found at edited sites. In other words, the majority of sequence divergence among the investigated species was associated with edited sites that are

insensitive to selection. This result indicates that in the presence of a large number of 'protected' nucleotides, RNA editing can contribute to an acceleration of sequence divergence among species by allowing for free, random accumulation of C-T substitutions. The extent of this acceleration will depend on the number and location of edited sites and is likely to vary among species. Rapid accumulation of nucleotide substitutions would therefore lead to more rapid sequence divergence of genomic DNA than of cDNA, which may give rise to biased phylogenetic inferences if they are based solely on genomic DNA analysis (Lu, in preparation).

The Larix *and* Ginkgo coxI *genes originated from edited transcripts*

To date, all gymnosperm mitochondrial genes studied have been found to contain edited sites [10 and references therein]. Furthermore, earlier studies on mitochondrial genes in gymnosperms have led to suggestions that mitochondrial DNA is more highly edited in gymnosperms than in angiosperms [8, 12, 14]. Surprisingly, however, we found that the 648 bp *coxI* sequence is not edited in *L. sibirica* and *G. biloba*. Furthermore, taking into account the fact that the *coxI* genomic sequence in *L. sibirica* was found to be shared by three other *Larix* species, it is likely that the *coxI* gene is not edited in these taxa either.

It has been postulated that since some mitochondrial genes produce partially edited transcripts, it is not possible to detect all the edited sites by sequencing only a limited number of cDNA clones [11]. In our opinion, however, this does not account for the lack of editing of the *coxI* gene product we found in *G. biloba* and *L. sibirica*. Direct sequencing of PCR products should be able to detect such partially edited transcripts present at a reasonably high frequency, which would appear as simultaneous presence of co-migrating bands in the T and C lanes on the sequencing gel [22]. In addition, if a site is edited at such a low frequency that it can not be detected on the sequencing gel, it is questionable that it should be regarded as an edited site, unless it can be proven to exist in the mRNA (by hybridization or allele-specific oligo PCR). Furthermore, the presence of Ts at all the potential non-synonymous edited sites in the two species suggests that they do not require editing. We therefore believe that our findings genuinely demonstrate the absence of RNA editing of the *coxI* sequence in *G. biloba*, *L. sibirica* and, very likely, the other three *Larix* species analysed in this study.

The observed sequence composition of *coxI* in *G. biloba*, and *L. sibirica* provides interesting clues for the possible origin of this gene in these taxa. In both species, the *coxI* sequences have Ts at nearly all the potentially edited sites. Therefore, they more closely resemble the edited mitochondrial transcripts than the unedited genomic sequences of other species. In fact, the *G. biloba* and *L. sibirica coxI* genomic DNAs have nearly the same sequences as the *coxI* cDNA of other species analysed in this study. This suggests that the *coxI* genes are transcribed in these two species, and that they originated from edited mitochondrial *coxI* transcripts by reverse transcription, followed by insertion into the nuclear genome, or back into the mitochondrial genome. Alternatively, the observed lack of edited sites on *coxI* gene in *G. biloba* and *L. sibirica* represents the ancestral states of the *coxI* gene (i.e., most of the edited sites were already Ts in their ancestral sequence). Therefore there is no need for the editing. However, this explanation would require additional assumptions: (1) the two species lack the capacity for T to C substitution; or (2) the T to C changes occurred but were eliminated by selection due to the lack of RNA editing. The facts that 9 of the 14 nucleotide differences found between *G. biloba* and *L. sibirica* are of the C-T type, and that RNA editing has been found on other mitochondrial genes in both species [12, Lu, unpublished results] and on *coxI* in their closely related species (e.g. *Pinus* and *Picea*), do not particularly favour this suggestion. Clearly, more data is needed to establish the evolutionary history of these two *coxI* sequences.

The sequence discrepancy

Our results show that the *Larix* spp. *coxI* sequence reported by Sper-Whitis *et al.* [23] was not present in our material. Unfortunately, Sper-Whitis *et al.* [23] did not provide the full name of the *Larix* species included in their study. Therefore, we can not establish whether their material represented one of the species included in our study. Interestingly, however, all but three of the 32 nucleotide differences between the *Larix coxI* sequences found in our study and the one reported by Sper-Whitis *et al.* [23] involved C and T nucleotides. It is therefore possible that the 'original' *coxI* gene has been lost or become silent in some *Larix* species. By using low annealing temperatures (47 °C) in the amplification of *coxI*, Sper-Whitis *et al.* [23] may have been able to amplify such a silent sequence, similar to that reported for soybean by Covello and Gray

[4]. This possibility could also explain the difference between the *coxI* sequence found in this study and the sequence reported by Sper-Whitis *et al.* [23]. It does not explain, however, the differences in the *G. biloba coxI* sequences obtained by us and that found by Sper-Whitis *et al.* [23]. Unlike the *Larix* sequences, most of these differences (22 out of 38) involve A and G nucleotides. We believe that the most likely explanation for this discrepancy is a sampling error. Several lines of evidence can be invoked to support this explanation. First, the plant material analysed in our study was collected from documented sources, but the *G. biloba* material studied by Sper-Whitis *et al.* [23] apparently came from undocumented sources. Second, our *coxI* sequence of *G. biloba* shows 99% homology to the *Cabomba* spp. sequence reported by Sper-Whitis *et al*. [23] which also lacks edited sites. It is therefore possible that the *Cabomba* spp. *coxI* sequence published by these authors represents the *G. biloba* sequence.

Variation of RNA editing among genes and species

The variation in RNA editing of the *coxI* gene product among species found in this study is striking. As many as 34 RNA edited sites were found in *P. abies*. However, we found no RNA editing in two species: *G. biloba* and *L. sibirica*. The reverse pattern has been reported for the *coxIII* gene by Hiesel *et al.* [12] who found more edited sites in *G. biloba* [21] than in *P. abies* [16]. In addition, we have found six edited sites on a 240 bp *coxIII* fragment from *L. sibirica* (Lu, unpublished results). All these results provide additional evidence that there is pronounced variation in the frequency of RNA editing both among species and among genes within species [11 and references therein]. Furthermore, the observed lack of RNA editing in *G. biloba* and *L. sibirica* suggests that intensive editing is not necessarily a typical feature of gymnosperms.

Acknowledgements

We thank the Staff of the Hørsholm Arboretum, Denmark and the Kórnik Arboretum, Poland for generous help in material collection. We are also grateful to Roger Granbom for excellent technical assistance and to Dr John Blackwell for language corrections and valuable comments. This work was supported by grants from the Swedish National Research Council (NFR), Swedish Council for Agricultural and forestry Research (SJFR), the Kempe Foundation, the CF Lundströms Council and Carl Tryggers Research Council.

References

1. Araya A, Begu D, Litvak S: RNA editing in plants. Physiol Plant 91: 543–550 (1994).
2. Borner GV, Paabo S: Evolutionary fixation of RNA editing. Nature 383: 225 (1996).
3. Comeron JM: A method for estimating the numbers of synonymous and nonsynonymous substitutions per site. J Mol Evol 41: 1152–1159 (1995).
4. Covello PS, Gray MW: 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 (1992).
5. Covello PS, Gray MW: On the evolution of RNA editing. Trends Genet 9: 256–268 (1993).
6. Doyle JJ, Doyle JL: A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull 19: 11–15 (1987).
7. Geiss KT, Abbas GM, Makaroff CA: Intron loss from the NADH dehydrogenase subunit 4 gene of lettuce mitochondrial DNA: evidence for homologous recombination of a cDNA intermediate. Mol Gen Genet 243: 97–105 (1994).
8. Glaubitz JC, Carlson JE: RNA editing in the mitochondria of a conifer. Curr Genet 22: 163–165 (1992).
9. Gray MW: Origin and evolution of mitochondrial DNA. Annu Rev Cell Biol 5: 25–50 (1989).
10. Gray MW: RNA editing in plant organelles: a fertile field. Proc Natl Acad Sci USA 93: 8157–8159 (1996).
11. Hanson MR, Sutton CA, Lu BW: Plant organelle gene expression: altered by RNA editing. Trends Plant Sci 1: 57–64 (1996).
12. Hiesel R, Combettes B, Brennicke A: Evidence for RNA editing in mitochondria of all major groups of lands plants except the Bryophyta. Proc Natl Acad Sci USA 91: 629–633 (1994).
13. Higgins DG, Bleasby AJ, Fuchs R: CLUSTAL V: improved software for multiple sequence alignment. CABIOS 8: 189–191 (1992).
14. Karpinska B, Karpinski S, Hällgren JE: The genes encoding subunit 3 of NADH dehydrogenase and ribosomal protein S12 are co-transcribed and edited in *Pinus sylvestris* (L.) mitochondria. Curr Genet 28: 423–428 (1995).
15. Kimura M: A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16: 111–120 (1980).
16. Kwok S, Kellog DE, McKinney N, Spasic D, Goda L, Levenson C, Snisky JJ: Effects of primer-template mismatches on the polymerase chain reaction: human immunodeficiency virus type 1 model studies. Nucl Acids Res 18: 999–1005 (1990).
17. Li W-H: Unbiased estimation of the rates of synonymous and nonsynonymous substitution. J Mol Evol 36: 96–99 (1993).
18. Li W-H, Wu C-I, Luo C-C: A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. Mol Biol Evol 2: 150–174 (1985).
19. Nugent JM, Palmer JD: RNA-mediated transfer of gene *coxII* from the mitochondrion to the nucleus during flowering plant evolution. Cell 66: 473–481 (1991).
20. Palmer JD: Mitochondrial DNA in plant systematics: applications and limitations. In: Soltis PS, Soltis DE, Doyle JJ (eds)

Molecular Systematics of Plants, Chapman and Hall, New York, London (1992).

21. Pamilo P, and Bianchi NO: Evolution of the *Zfx* and *Zfy* genes: rates and independence between the genes. Mol Biol Evol 10: 271–281 (1993).
22. Perrotta G, Regina TMR, Ceci LR, Quagliariello C: Conservation of the organization of the mitochondrial *nad3* and *rps12* genes in evolutionarily distant angiosperms. Mol Gen Genet 251: 326–337 (1996).
23. Sper-Whitis GL, Moody JL, Vaughn JC: Universality of mitochondrial RNA editing in cytochrome-c oxidase subunitI (*coxI*) among the land plants. Biochim Biophys Acta 1307: 301–308 (1996).
24. Wakasugi T, Hirose T, Horihata M, Tsudzuki T, Kössel H, Sugiura M: Creation of a novel protein coding region at the RNA level in black pine chloroplasts: the pattern of RNA editing in the gymnosperm chloroplast is different from that in angiosperms. Proc Natl Acad Sci USA 93: 8766–8770 (1996).
25. Wang X-R, Szmidt AE, and Lu M-Z: Genetic evidence for the presence of cytoplasmic DNA in pollen and gametophytes and maternal inheritance of mitochondrial DNA in *Pinus*. Forest Genet 3: 37–44 (1996).
26. Wolfe KH, Li W-H, Sharp PM: Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proc Natl Acad Sci USA 84: 9054–9058 (1987).