The EMBO Journal vol.11 no.11 pp.3815-3820, 1992

Silent mitochondrial and active nuclear genes for subunit 2 of cytochrome *c* oxidase (*cox2*) in soybean: evidence for RNA-mediated gene transfer

Patrick S.Covello and Michael W.Gray[1]

Program in Evolutionary Biology, Canadian Institute for Advanced Research, Department of Biochemistry, Dalhousie University, Halifax, Nova Scotia B3H 4H7, Canada

[1]Corresponding author

Communicated by J.-H.Weil

In most plants and other eukaryotes investigated, the mitochondrial genome carries the gene encoding subunit 2 of cytochrome *c* oxidase (*cox2*). In this paper, we show that the previously reported mitochondrial *cox2* of soybean is actually silent, and that there is an expressed, single-copy, nucleus-encoded *cox2*. Molecular cloning and sequence analysis of *cox2* cDNA and genomic clones show that the soybean nuclear gene encodes an N-terminal extension that resembles a signal sequence for mitochondrial import and whose coding sequence is separated by an intron from that corresponding to mtDNA-encoded *cox2*. Comparison of soybean mitochondrial and nuclear *cox2* sequences clearly indicates that in an ancestor of soybean, *cox2* was transferred from the mitochondrion to the nucleus via a C-to-U edited RNA intermediate.

Key words: Glycine max/molecular evolution/reverse transcription/RNA editing/transit peptide

Introduction

Genes for mitochondrial and chloroplast proteins are encoded largely by nuclear DNA. Implicit in the endosymbiont hypothesis of organelle origins (Gray and Doolittle, 1982; Gray, 1983, 1989a) is the assumption that most of these genes were originally organellar but were transferred to the nucleus during subsequent evolution (Palmer, 1991; Gray, 1991). Although the bulk of this gene transfer appears to have occurred relatively early in organellar evolution (Palmer, 1985; Cavalier-Smith, 1987), data from the past decade suggest that DNA transfer in general and transfer of functional genes *per se* are ongoing processes (Palmer, 1991).

In fact, it has become clear that transfer of genetic information between cellular compartments can occur quite frequently. In yeast, it has been shown experimentally that plasmid DNA can move from mitochondria to the nucleus (Thorsness and Fox, 1990). Also, there are a number of examples of DNA sequences shared by more than one subcellular genome (Timmis and Scott, 1984), presumably as a result of evolutionarily recent transfer events. For example, mitochondrial sequences have been found in the nuclear DNA of yeast (Farrelly and Butow, 1983) and sea urchin (Jacobs *et al.*, 1983), whereas both nuclear (Schuster and Brennicke, 1987) and plastid (Stern and Lonsdale, 1982; Schuster and Brennicke, 1987) sequences are present in plant mitochondrial DNA (mtDNA). Some chloroplast-like tRNA sequences, encoded by the plant mitochondrial genome, are actually expressed as mature tRNA species in plant mitochondria (Maréchal *et al.*, 1985; Joyce and Gray, 1989).

Examination of the gene contents of the organelles of related species is also suggestive of gene transfers. For example, in *Saccharomyces cerevisiae*, the functional gene for subunit 9 of ATP synthase (*atp9*) is in the mitochondrion; in *Neurospora crassa* and *Aspergillus nidulans*, in contrast, there are silent mitochondrial, but expressed nuclear copies of *atp9* (van den Boogaart *et al.*, 1982; Brown *et al.*, 1985). In plants, there is now well documented evidence of the relatively recent transfer of the chloroplast genes *tufA* (Baldauf and Palmer, 1990) and *rpl22* (Gantt *et al.*, 1991) to the nucleus. Many other examples are known in which organellar gene contents differ slightly, but these have not been fully investigated (Gray, 1989b; Palmer, 1991).

Before functional transfer of an organellar gene to the nucleus can be achieved, a number of obstacles must be overcome. The gene must acquire the appropriate nuclear regulatory sequences for transcription, polyadenylation, etc., as well as a signal sequence to direct its product to the organelle. The mechanism(s) of transfer and the mode(s) of acquisition of such regulatory sequences are still open questions. Other barriers that might be expected to impede relocation of an organellar gene to the nucleus include deviations from the standard genetic code (common in non-plant mitochondrial systems; Jukes and Osawa, 1990) and the occurrence of RNA editing (Cattaneo, 1991) in organelles, including plant mitochondria (Covello and Gray, 1989; Gualberto *et al.*, 1989; Hiesel *et al.*, 1989) and chloroplasts (Hoch *et al.*, 1991; Kudla *et al.*, 1992). Circumstantial evidence that many transfers occur via RNA intermediates comes from the fact that transferred sequences can be identified with transcribed regions (Schuster and Brennicke, 1987).

Recently, Nugent and Palmer (1991) reported evidence for relocation of the gene encoding subunit 2 of cytochrome *c* oxidase (*cox2*) from mitochondria to the nucleus in an ancestor of cowpea, a legume. At the same time, we discovered this phenomenon in soybean, another legume, and in this paper we report the identification and characterization of a functional nuclear *cox2* in this plant. However, unlike the situation in cowpea (Nugent and Palmer, 1991), the corresponding mitochondrial gene in soybean has been retained. Comparison of this silent mitochondrial *cox2* with its active nuclear counterpart provides strong support for an RNA intermediate in the intracellular transfer of this gene.

Results

Sequence analysis of soybean nuclear cox2

In the course of reverse transcriptase sequencing of angiosperm mitochondrial mRNAs (Covello and Gray, 1990a), we found that we could not produce a sequencing

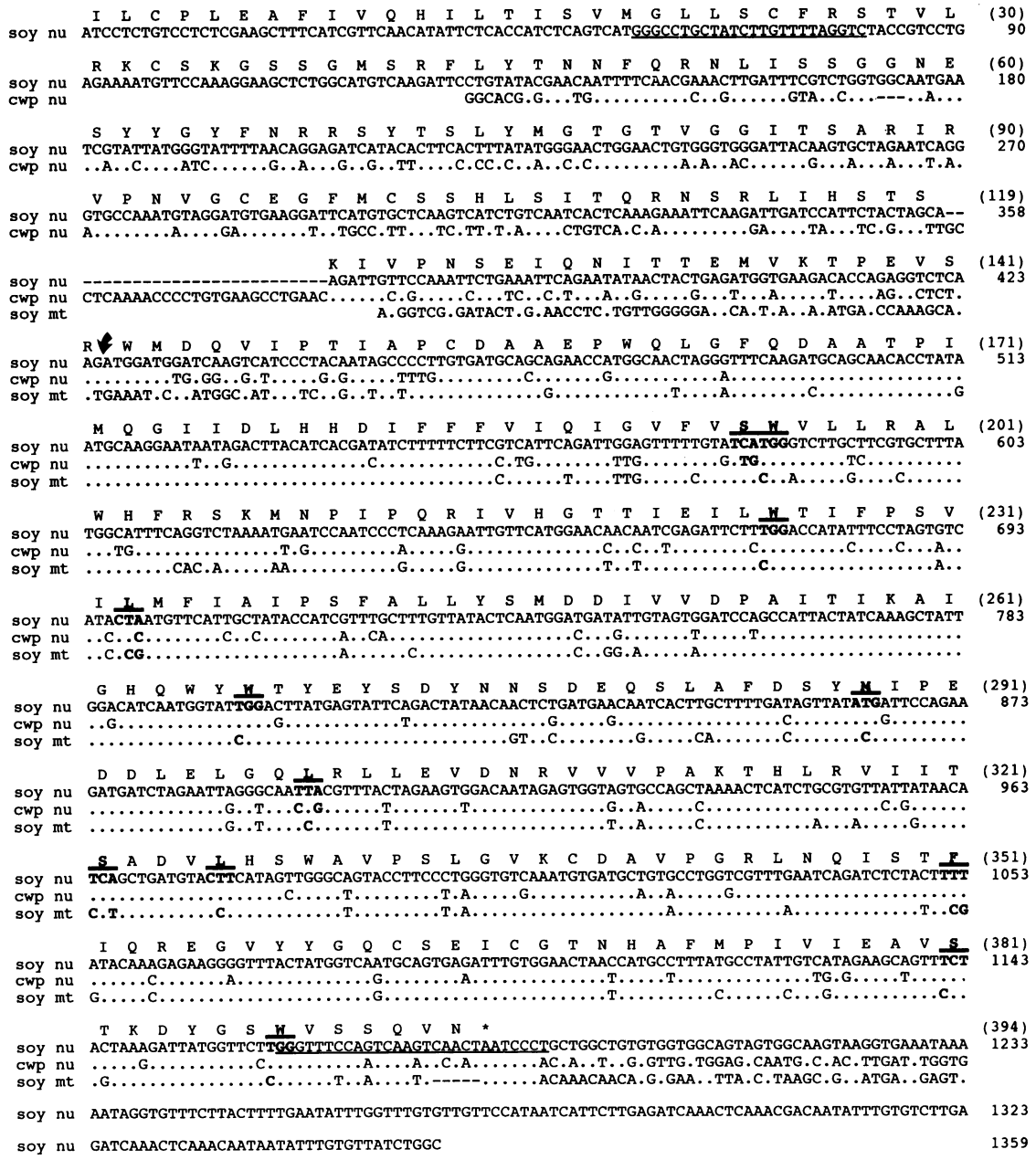

Fig. 1. Comparison of nucleus- and mitochondrion-encoded *cox2* sequences. The soybean nuclear *cox2* cDNA (soy nu; this paper) is shown below the corresponding predicted amino acid sequence. Cowpea nuclear (cwp nu; Nugent and Palmer, 1991) and soybean mitochondrial (soy mt; Grabau, 1987) *cox2* sequences are shown with dots indicating identity with the soybean nuclear sequence and dashes denoting gaps. Twelve out of 13 editing sites in pea *cox2* mRNA (Covello and Gray, 1990a) have a C at the homologous position in the soybean mitochondrial gene and are highlighted by horizontal bars. Note that of these 12 highlighted codons, 11 have a T in the two nuclear sequences. The arrow indicates the site of an intron in both the soybean and cowpea nuclear *cox2* sequences (corresponding to the putative start codon of the soybean mitochondrial sequence). Underlined sequences indicate the binding sites of the primers used for PCR (positions 57–80, primer 2; positions 1163–1190, primer 7) (see Materials and methods).

ladder for soybean mitochondrial *cox2* mRNA under experimental conditions that readily permitted sequencing of *cox2* mRNAs of pea and other species. A preliminary Northern blot (not shown) revealed that in soybean, *cox2* transcripts are abundant in poly(A)$^+$ RNA but not in mitochondrial RNA. These results indicated that *cox2* in soybean mtDNA, previously described by Grabau (1987), is actually a silent gene, and that a nuclear counterpart is the active *cox2* in this plant.

In order to confirm this indication of an active *cox2* in the soybean nucleus, we screened a cDNA library from this plant using a wheat *cox2* probe; four positive clones were isolated and subcloned. A 1.7 kb insert in one clone, p4.2, was sequenced in its entirety. Its 5′ portion was found to correspond to a 1.3 kb insert in a second clone, p4.5. The 3′ end of p4.2 beyond an oligo(A) stretch proved to be highly similar in sequence to the 3′ end of two wound-induced genes (*win1* and *win2*) of potato (Stanford *et al.*, 1989). The partial *win* sequence was fused to the *cox2* sequence immediately downstream of an oligo(A) stretch, strongly suggesting that the p4.2 insert resulted from artefactual ligation of *cox2* and *win* sequences during preparation of the cDNA clone bank. The terminal sequences and restriction map of p4.5 matched those of the *cox2* portion of p4.2 except that the p4.5 insert

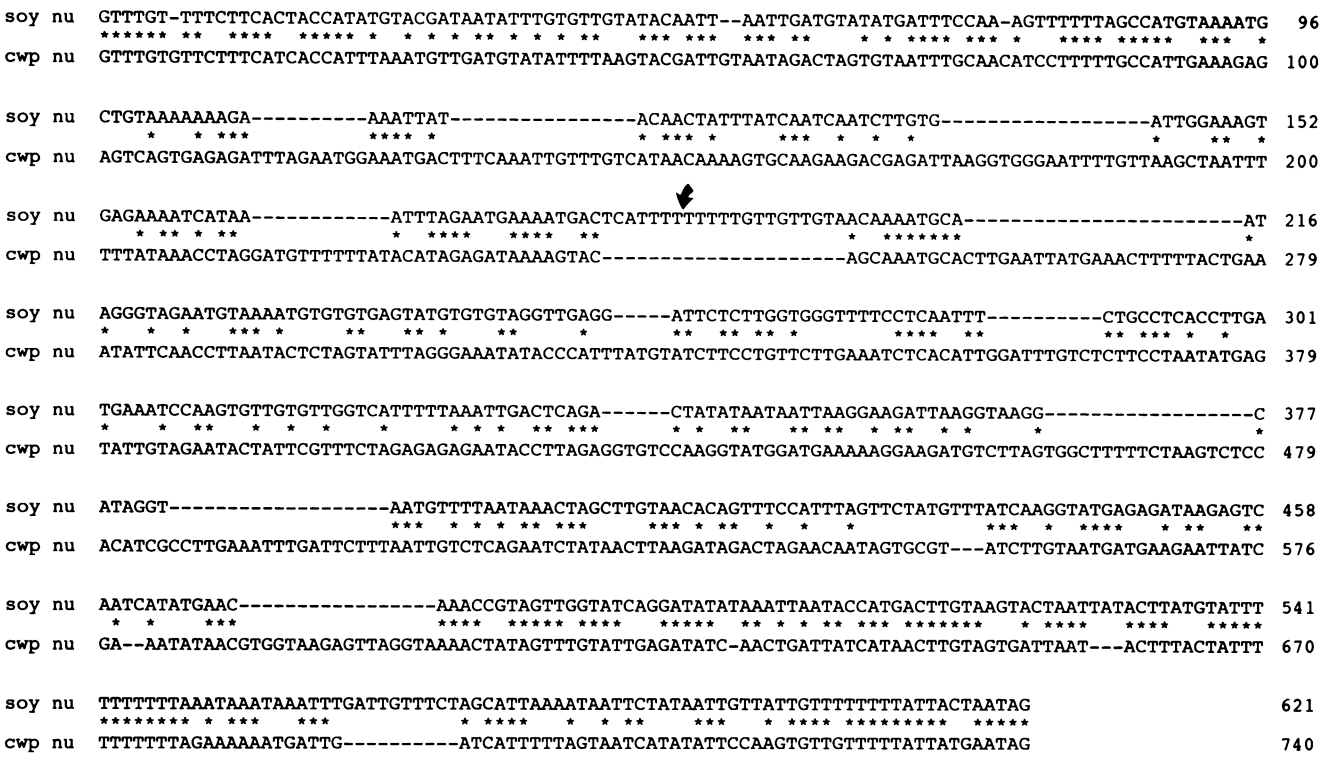

Fig. 2. Comparison of introns from nuclear *cox2* of soybean (this paper) and cowpea (Nugent and Palmer, 1991). The arrow indicates an additional T in three out of six soybean PCR clones. Alignment was performed using the CORPET program (Corpet, 1988) with a gap penalty of 20.

was shorter by 44 nt at its 5′ end and by 9 nt located immediately upstream of the oligo(A) sequence in p4.2. A third *cox2*-hybridizing clone was found to correspond to p4.5 in insert size and restriction site distribution, while the fourth positive clone appears to have been derived from contaminating mtDNA.

Comparison of the soybean cDNA sequence with related *cox2* sequences provides useful insights into the evolution of these genes. Figure 1 shows the sequence of the *cox2* portion of p4.2 in comparison with the previously published sequences for the soybean mitochondrial (Grabau, 1987) and cowpea nuclear (Nugent and Palmer, 1991) *cox2* genes. In the sequence corresponding to the coding region of soybean and other plant mitochondrial *cox2* genes (positions 424−1182), the three sequences are very similar, with the soybean nuclear gene sharing 87.5% and 86.7% nucleotide identity with the cowpea nuclear and soybean mitochondrial genes, respectively. The p4.2 cDNA is clearly identifiable as a *cox2* sequence and includes those codons specifying the highly conserved residues thought to be involved in copper binding and intermolecular interactions (Holm *et al.*, 1987; Covello and Gray, 1990b). Immediately downstream of the coding regions the similarities between the three genes drop sharply.

The data in Figure 1 provide a unique opportunity to compare two homologous protein-coding genes, in different compartments of the same cell, one of which is the evolutionary progenitor of the other. Such a comparison is of particular interest because *cox2* mRNA undergoes extensive C-to-U editing in both monocotyledons and dicotyledons, including pea (Covello and Gray, 1990a), which is closely related to soybean. We may therefore ask whether the soybean nuclear *cox2* sequence more closely resembles its silent mitochondrial counterpart or the sequence of an edited *cox2* transcript that might have been produced from the latter (once-functional) gene. The tendency for editing sites to be conserved among related species has been shown previously (Covello and Gray, 1990a). Good predictors of sites that would be expected to be edited in transcripts of any soybean mtDNA-encoded *cox2* are those sites that are edited in *cox2* mRNA of pea, the species closest to soybean for which editing has been examined. In pea *cox2* mRNA, there are 13 full C-to-U editing sites (Covello and Gray, 1990a). Of these, 12 have a C in the corresponding mtDNA sequence of soybean, i.e. sites requiring editing have largely been conserved in the silent mitochondrial *cox2*. The codons containing such positions are shown in bold and are overlined in Figure 1. It is notable that for all but one of these positions, there is a T in soybean nuclear *cox2*, i.e. soybean nucleus-encoded *cox2* resembles an edited *cox2* mRNA rather than the corresponding mtDNA sequence. This strongly suggests that the nuclear sequence was originally derived from an edited mitochondrial *cox2* mRNA (see Discussion).

One of the requirements of nucleus-encoded gene products destined for the mitochondrion is an N-terminal signal sequence mediating import (Horwich *et al.*, 1991). As in the case of the cowpea sequence (Nugent and Palmer, 1991), the soybean nuclear *cox2* reading frame extends 5′ from the position of the start codon of the mitochondrial gene, in this case as far as the beginning of the cDNA insert (for >400 nt). Primer extension analysis indicates that the 5′ end of the mRNA is positioned 26 nt upstream of the beginning (5′ end) of the p4.2 insert (data supplied to reviewers but not shown). Consequently, translation is likely to start at codon 19 or possibly further upstream, assuming that the first AUG codon is used. It has been suggested that translation of the cowpea nuclear *cox2* mRNA may start at

codon 91 (Nugent and Palmer, 1991); however, there is no AUG codon at this position in soybean. Moreover, the 5′ end of the cowpea *cox2* mRNA may actually extend well beyond the available cowpea cDNA sequence to a translation start site further upstream.

The 5′ end of soybean mitochondrial *cox2* potentially encodes a presequence containing signals for mitochondrial import and targeting to the inner membrane of the organelle. This presequence is rich in basic and hydroxylated residues and poor in negatively charged residues, features associated with signal sequences for mitochondrial import (Horwich *et al.*, 1991). The soybean *cox2* presequence is longer than a typical one and it remains to be determined what parts of it are important for import and for inner membrane targeting. In fact, the signals required for targeting of polypeptides to the inner membrane of mitochondria are not well understood and appear to vary from case to case (Horwich *et al.*, 1991). We also do not know what portion of the nuclear *cox2* sequence represents the mature (proteolytically processed) polypeptide, and it is possible that even plant mitochondrion-encoded COX2 polypeptides undergo some N-terminal processing. For example, the initiation codon of yeast *cox2* (also mtDNA-encoded) aligns within two amino acids of the putative *cox2* initiation codon in plant mitochondria (Covello and Gray. 1989); however, after translation, the yeast polypeptide is proteolytically processed to remove 15 N-terminal amino acids, such that the mature N-terminus corresponds to codon 155 in Figure 1 (Pratje *et al.*, 1983). It is notable that this position roughly marks the point of divergence (in the 5′ direction) of the three sequences shown in Figure 1.

In order to determine the structure of the nuclear gene and its identity with the cDNA sequence, polymerase chain reaction (PCR) products, obtained using primers 2 and 7 (see Materials and methods and Figure 1) and encompassing >90% of the *cox2* open reading frame, were cloned and sequenced. The results indicate the presence of a 621 bp intron at the same position as in cowpea (arrow in Figure 1). Interestingly, this position corresponds to the initiation codon of the soybean and other plant mtDNA-encoded *cox2* genes. Therefore, the intron effectively separates the putative signal sequence from the 'mature' sequence. The soybean nuclear *cox2* intron is compared with the corresponding cowpea intron in Figure 2. Both sequences have features typical of plant (dicot) nuclear introns (Goodall *et al.*, 1991), with splice sites in agreement with consensus sequences and with relatively high A+T contents. The ends of the introns comprise regions of ~70% nucleotide identity, while the central regions are more divergent. Three out of six PCR clones contained an additional T in a stretch of nine, as indicated in Figure 2. Whether this reflects genuine polymorphism or is a PCR artefact is not known.

Soybean cox2 is expressed as poly(A)⁺ RNA from a single-copy nuclear gene

When restriction-digested soybean nuclear DNA was probed with the full-length cDNA clone p4.5, two major bands were detected for three separate restriction enzymes (Figure 3, lanes marked '1'). Control experiments indicated that one band in each lane corresponds to mitochondrial *cox2* (data not shown); these bands did not appear when a probe containing only nuclear presequence was used (Figure 3, lanes labelled '2'). This level of contamination of purified

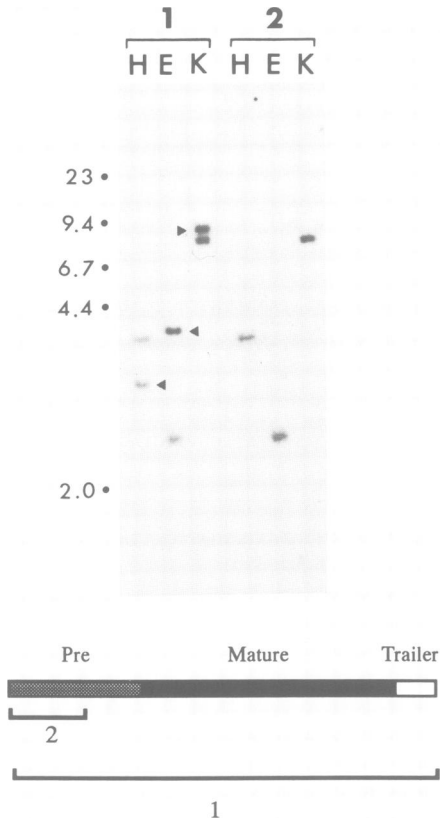

Fig. 3. Hybridization of *cox2* clones to soybean nuclear DNA. The figure shows a Southern blot of soybean nuclear DNA hydrolyzed with *Hin*dIII (H), *Eco*RI (E) or *Kpn*I (K). The same blot was probed with the inserts of cDNA clones p4.5 (lanes marked '1') and p4.2b-c (subclone encompassing positions 1−249 in Figure 1; lanes marked '2'). The sizes in kbp of DNA markers are indicated on the left. The positions of signals from soybean mtDNA are denoted by triangles.

soybean nuclear DNA by mtDNA is not unexpected, given that the copy number of plant mitochondrial genomes often exceeds 100 per cell (Lamppa and Bendich, 1984). The data therefore indicate that the gene from which the *cox2* cDNA sequence was derived is single copy. A different conclusion was reached by Nugent and Palmer (1991), whose hybridization results indicated multiple soybean nuclear sequences related to *cox2*. The difference may reflect their use of a mitochondrial rather than a nuclear *cox2* probe from soybean and/or different hybridization conditions.

When the soybean mitochondrial *cox2* was originally cloned and sequenced (Grabau, 1987), there was no indication that it would not be functional. Indeed, Northern analysis suggested that it was expressed, although the low intensity signals obtained led Grabau to conclude that 'the transcripts may not be very abundant'. In fact, discrete *cox2* transcripts are not detectable under Northern hybridization conditions in which the wheat mitochondrial *cox2* mRNA is readily seen (Figure 4, lanes 4 and 5). On the same blot, *cox1* transcripts are easily detectable in soybean mitochondrial RNA at a level comparable to that seen in wheat mitochondrial RNA (data not shown). Consistent with a nuclear location of the active soybean gene, soybean *cox2* is expressed as a 1.4 kb mRNA in a poly(A)⁺ fraction (Figure 4, lane 1), in agreement with previous work (Nugent and Palmer, 1991).

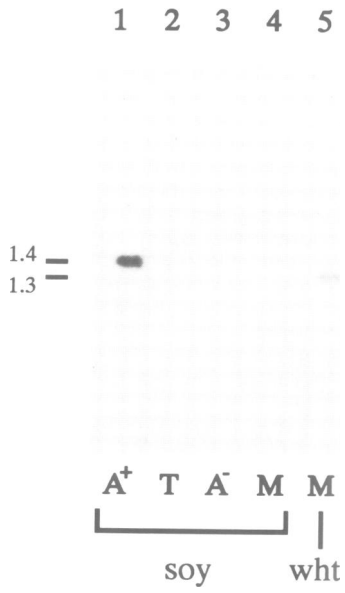

Fig. 4. Soybean *cox2* is expressed as poly(A)$^+$ RNA. Northern blot of 5 μg each of soybean poly(A)$^+$ RNA (lane 1), soybean total RNA (lane 2), soybean poly(A)$^-$ RNA (lane 3), soybean mitochondrial 1 M NaCl-insoluble RNA (lane 4) and wheat mitochondrial 1 M NaCl-insoluble RNA (lane 5) probed with cDNA clone p4.5. Estimated sizes in kbp of the soybean and wheat *cox2* transcripts are indicated on the left; these were based on RNA markers.

Discussion

As late as 1987, it was suggested that chloroplasts and mitochondria had undergone an initial 'burst of quantum evolution' that was now essentially complete (Cavalier-Smith, 1987). However, data appearing in the past decade suggest that the process of organelle-to-nucleus gene transfer is ongoing. The fact that more genes seem to have been transferred early in evolution may simply reflect the larger number of genes available to undergo this process at that time. Alternatively, there may be reasons why transfer of existing organellar genes is not favoured. It has been pointed out that the use of alternative genetic codes by most (non-plant) mitochondria would be expected to limit the transfer of functional genes to the nucleus (Gellissen and Michaelis, 1987).

Von Heijne (1986) has suggested that genes whose products contain hydrophobic domains may be prevented from transfer as a result of targeting difficulties. However, as reported here and by Nugent and Palmer (1991), the transfer of *cox2*, which encodes an integral membrane protein with a number of hydrophobic and putatively transmembrane domains, argues against this hypothesis.

The comparison of mitochondrial and nuclear copies of soybean *cox2* strongly suggests an RNA intermediate in the transfer process. This inference provides the most parsimonious explanation for the C-to-T substitutions at sites corresponding to edited sites in pea *cox2* mRNA, most of which were presumably also editable in the ancestor of soybean in which the *cox2* transfer occurred. Thus, the sequence of events appears to have been: (i) transcription and editing of a once-active mitochondrial *cox2*, followed by (ii) reverse transcription/transfer to the nucleus in an order that is not discernible; and (iii) incorporation into the nuclear genome of an edited version of the mitochondrial gene. Alternative explanations involving direct DNA transfer would require a number of independent substitutions in the transferred nuclear gene, a scenario that is much less plausible. RNA editing of the nuclear *cox2* transcript is ruled out by comparison of the genomic and cDNA sequences, which are identical. As discussed previously (Nugent and Palmer, 1991), it is not clear in which compartment reverse transcription would have occurred; mitochondrion, cytosol and nucleus are all reasonable candidates (Schuster and Brennicke, 1987).

Once transferred to the nucleus, *cox2* would have had to acquire regulatory and signal sequences in order to become functional. Indeed, this has obviously occurred. Sequence analysis shows the presence of a possible polyadenylation signal (AATAAA) at positions 1228−1233 (Figure 1) and two GT-rich regions downstream that may be important for polyadenylation (Mogen *et al.*, 1990). The long leader sequence of the nuclear *cox2* probably codes for an N-terminal extension that facilitates import of the COX2 polypeptide into the mitochondrion.

It is intriguing that soybean mtDNA retains a potentially functional but apparently silent copy of *cox2*. Because this gene is not expressed in etiolated hypocotyls, it may be a pseudogene. On the other hand, it shares 97% amino acid sequence identity with the active pea mitochondrial COX2 polypeptide and (allowing for editing) would encode all of the putative Cu_A-binding ligands (Holm *et al.*, 1987; Covello and Gray, 1990b). Thus, it may be expressed at other stages of development or in different tissues. The identification of any stage- or tissue-specific transcript would raise the issue of whether such a transcript undergoes editing in the same way as a bona fide *cox2* mRNA, or whether it specifies a protein of altered function. Alternatively, considering the low rate of nucleotide substitution in plant mtDNA (Wolfe *et al.*, 1987; Palmer, 1990), the soybean mtDNA-encoded *cox2* may have lost function recently enough that its sequence has not drifted significantly from other, active plant mitochondrial *cox2* genes. Curiously, the soybean *cox2* situation is mirrored by that of *atp9* in *Neurospora* and *Aspergillus*, in that the mitochondrial *atp9* appears to be silent but retains all of the conserved ATP9 structural features (van den Boogaart *et al.*, 1982; Brown *et al.*, 1985).

We may regard the soybean and cowpea *cox2* situations described by us and by Nugent and Palmer (1991), respectively, as representative of different stages in an evolutionary process of intergenomic information transfer. Models of organelle-to-nucleus gene transfer (Bogorad, 1975; Obar and Green, 1985) have invoked gene duplication and functional genetic redundancy as essential features of the transfer mechanism. The soybean *cox2* case is formally equivalent to Step F (Selection) in the model of Obar and Green (1985), in which the 'donor copy of the genetic element becomes nonfunctional ... [and] ... the recipient thus assumes full control of the genetic element and its product. The donor copy is now fully redundant: by definition, a pseudogene.' The cowpea *cox2* case (Nugent and Palmer, 1991) would represent the next step (G: Loss): 'The moribund (pseudogene) copy of the genetic element is altered beyond recognition by mutation or otherwise lost from the donor genome.' Pea, which reportedly has both nuclear and mitochondrial *cox2* sequences, with only the latter being expressed (Nugent and Palmer, 1991), presumably

represents an earlier stage in the gene transfer process. It will be interesting to examine other related legumes to determine whether some have active *cox2* genes in both compartments, an evolutionary stage that would be predicted by these gene transfer models.

Finally, the different subcellular locations of *cox2* in pea and non-legumes compared to the members of the Phaseoleae tribe raise questions about regulation of the biosynthesis of cytochrome oxidase and the nucleocytoplasmic interactions involved. Is there co-ordinated expression of mitochondrial and nuclear genes coding for COX subunits? If so, how does this differ among soybean, pea and related species?

Materials and methods

Nucleic acids

Nuclear DNA was obtained from 6 day etiolated soybean hypocotyls (*Glycine max*, var. Williams) as described by Jofuku and Goldberg (1988), except that a combination of eight layers of cheesecloth and one layer of Miracloth (Calbiochem, USA) was used to filter the tissue homogenate, and a single ultracentrifugation in CsCl solution was performed. Mitochondrial DNA and RNA were obtained in separate extractions from 6 day etiolated soybean hypocotyls (var. Maple Amber) and wheat embryos as described previously (Bonen and Gray, 1980; Covello and Gray, 1990a). Total cellular RNA was extracted from soybean (var. Maple Amber) hypocotyls using cetyltrimethylammonium bromide (CTAB) according to Taylor and Powell (1982). A portion of this RNA was separated into poly(A)$^+$ and poly(A)$^-$ fractions using an mRNA purification kit (Pharmacia, Canada) according to the manufacturer's instructions [with two rounds of oligo(dT) chromatography].

Cloning and sequencing

Clones were isolated from a soybean (var. Williams) cDNA library in λgt11 (Clontech, USA) using plaque lifts probed (Ausubel *et al.*, 1987) with a *SalI*–*Hin*dIII clone roughly corresponding to exon 2 of wheat *cox2* (Bonen *et al.*, 1984). λ DNA was prepared according to Helms *et al.* (1987) and used for subcloning into pBluescript II vectors (Stratagene, USA).

Genomic *cox2* DNA was amplified by PCR in a 100 μl volume containing 400 ng soybean (var. Williams) nuclear DNA; 20 pmol each of oligonucleotide primers 2 (GGGAATTCTATCTTGTTTTAGGTC) and 7 (GGGAATTCGTTGACTTGACTGGAAACCC), synthesized with a Cyclone Plus DNA synthesizer; Milligen/Biosearch, USA); 20 mM Tris–HCl (pH 8.3); 1.5 mM $MgCl_2$; 25 mM KCl; 0.05% Tween 20; 100 μg/ml gelatin; 50 μM each of dATP, dCTP, dGTP and dTTP; and 2 units *Taq* DNA polymerase. Thirty cycles of denaturation (96°C, 15 s), annealing (55°C, 30 s) and extension (72°C, 90 s) were performed. PCR products were cloned into the *Eco*RI site of pBluescript II KS+. Six PCR clones were pooled for double-stranded sequencing.

Plasmid clones were restriction-mapped and sequenced on both strands by either double- or single-stranded dideoxy methods using nested exonuclease III deletions or synthetic oligonucleotides (Ausubel *et al.*, 1987) and a Sequenase 2.0 kit (US Biochemicals, USA). The sequence of soybean nuclear *cox2* has been deposited in the EMBL Data Library under the accession number Z11980 and that for the soybean *win* homologue under accession number Z11977.

Hybridizations

Southern, Northern and plaque-lift hybridizations were performed with random-primed DNA or *in vitro*-transcribed RNA probes using standard methods (Ausubel *et al.*, 1987).

Acknowledgements

We thank Sandra Baldauf, Michael Coulthart, Jacqueline Nugent and Arlin Stoltzfus for helpful discussions. This work was supported by an operating grant (MT-4124) from the Medical Research Council of Canada to M.W.G. Fellowships from the Dalhousie Medical Research Foundation (P.S.C.) and the Canadian Institute for Advanced Research (M.W.G.) are also gratefully acknowledged.

References

Ausubel,F.M., Brent,R., Kingston,R.E., Moore,D.D., Seidman,J.G., Smith,J.A. and Struhl,K. (1987) *Current Protocols in Molecular Biology*. Wiley Interscience, New York.

Baldauf,S.L. and Palmer,J.D. (1990) *Nature*, **344**, 262–265.

Bogorad,L. (1975) *Science*, **188**, 891–898.

Bonen,L. and Gray,M.W. (1980) *Nucleic Acids Res.*, **8**, 319–335.

Bonen,L., Boer,P.H. and Gray,M.W. (1984) *EMBO J.*, **3**, 2531–2536.

Brown,T.A., Waring,R.B., Scazzocchio,C. and Davies,R.W. (1985) *Curr. Genet.*, **9**, 113–117.

Cattaneo,R. (1991) *Annu. Rev. Genet.*, **25**, 71–88.

Cavalier-Smith,T. (1987) *Ann. N.Y. Acad. Sci.*, **503**, 55–71.

Corpet,F. (1988) *Nucleic Acids Res.*, **16**, 10881–10890.

Covello,P.S. and Gray,M.W. (1989) *Nature*, **341**, 662–666.

Covello,P.S. and Gray,M.W. (1990a) *Nucleic Acids Res.*, **18**, 5189–5196.

Covello,P.S. and Gray,M.W. (1990b) *FEBS Lett.*, **268**, 5–7.

Farrelly,F. and Butow,R.A. (1983) *Nature*, **301**, 296–301.

Gantt,J.S., Baldauf,S.L., Calie,P.J., Weeden,N.F. and Palmer,J.D. (1991) *EMBO J.*, **10**, 3073–3078.

Gellissen,G. and Michaelis,G. (1987) *Ann. N.Y. Acad. Sci.*, **503**, 391–401.

Goodall,G.J., Kiss,T. and Filipowicz,W. (1991) *Oxford Surveys Plant Mol. Cell Biol.*, **7**, 255–296.

Grabau,E.A. (1987) *Curr. Genet.*, **11**, 287–293.

Gray,M.W. (1983) *BioScience*, **33**, 693–699.

Gray,M.W. (1989a) *Trends Genet.*, **5**, 294–299.

Gray,M.W. (1989b) *Annu. Rev. Cell Biol.*, **5**, 25–50.

Gray,M.W. (1991) In Bogorad,L. and Vasil,I.K. (eds), *The Molecular Biology of Plastids*. Academic Press, San Diego, pp. 303–330.

Gray,M.W. and Doolittle,W.F. (1982) *Microbiol. Rev.*, **46**, 1–42.

Gualberto,J.M., Lamattina,L., Bonnard,G., Weil,J.-H. and Grienenberger,J.-M. (1989) *Nature*, **341**, 660–662.

Helms,C., Dutchik,J.E. and Olson,M.V. (1987) *Methods Enzymol.*, **153**, 69–83.

Hiesel,R., Wissinger,B., Schuster,W. and Brennicke,A. (1989) *Science*, **246**, 1632–1634.

Hoch,B., Maier,R.M., Appel,K., Igloi,G.L. and Kössel,H. (1991) *Nature*, **353**, 178–180.

Holm,L., Saraste,M. and Wikström,M. (1987) *EMBO J.*, **6**, 2819–2823.

Horwich,A.L., Cheng,M., West,A. and Pollock,R.A. (1991) *Curr. Topics Microbiol. Immunol.*, **170**, 1–42.

Jacobs,H.T., Posakony,J.W., Grula,J.W., Roberts,J.W., Xin,J.-H., Britten,R.J. and Davidson,E.H. (1983) *J. Mol. Biol.*, **165**, 609–632.

Jofuku,K.D. and Goldberg,R.B. (1988) In Shaw,C.H. (ed.), *Plant Molecular Biology: A Practical Approach*. IRL Press, Oxford, pp. 37–66.

Joyce,P.B.M. and Gray,M.W. (1989) *Nucleic Acids Res.*, **17**, 5461–5476.

Jukes,T.H. and Osawa,S. (1990) *Experientia*, **46**, 1117–1126.

Kudla,J., Igloi,G.L., Metzlaff,M., Hagemann,R. and Kössel,H. (1992) *EMBO J.*, **11**, 1099–1103.

Lamppa,G.K. and Bendich,A.J. (1984) *Planta*, **162**, 463–468.

Maréchal,L., Guillemaut,P., Grienenberger,J.-M., Jeannin,G. and Weil,J.-H. (1985) *Nucleic Acids Res.*, **13**, 4411–4416.

Mogen,B.D., MacDonald,M.H., Graybosch,R. and Hunt,A.G. (1990) *Plant Cell*, **2**, 1261–1272.

Nugent,J.M. and Palmer,J.D. (1991) *Cell*, **66**, 473–481.

Obar,R. and Green,J. (1985) *J. Mol. Evol.*, **22**, 243–251.

Palmer,J.D. (1985) In MacIntyre,R.J. (ed.), *Monographs in Evolutionary Biology: Molecular Evolutionary Genetics*. Plenum Publishing Corporation, New York, pp. 131–240.

Palmer,J.D. (1990) *Trends Genet.*, **6**, 115–120.

Palmer,J.D. (1991) In Bogorad,L. and Vasil,I.K. (eds.), *The Molecular Biology of Plastids*. Academic Press, San Diego, pp. 5–53.

Pratje,E., Mannhaupt,G., Michaelis,G. and Beyreuther,K. (1983) *EMBO J.*, **2**, 1049–1054.

Schuster,W. and Brennicke,A. (1987) *EMBO J.*, **6**, 2857–2863.

Stanford,A., Bevan,M. and Northcote,D. (1989) *Mol. Gen. Genet.*, **215**, 200–208.

Stern,D.B. and Lonsdale,D.M. (1982) *Nature*, **299**, 698–702.

Taylor,B. and Powell,A. (1982) *BRL Focus*, **4(3)**, 4–6.

Thorsness,P.E. and Fox,T.D. (1990) *Nature*, **346**, 376–379.

Timmis,J.N. and Scott,N.S. (1984) *Trends Biochem. Sci.*, **9**, 271–273.

van den Boogaart,P., Samallo,J. and Agsteribbe,E. (1982) *Nature*, **298**, 187–189.

von Heijne,G. (1986) *FEBS Lett.*, **198**, 1–4.

Wolfe,K.H., Li,W.-H. and Sharp,P.M. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 9054–9058.

Received on May 20, 1992; revised on July 7, 1992