

# Intracellular gene transfer in action: Dual transcription and multiple silencings of nuclear and mitochondrial *cox2* genes in legumes

Keith L. Adams<sup>\*†</sup>, Keming Song<sup>\*†‡</sup>, Philip G. Roessler<sup>†</sup>, Jacqueline M. Nugent<sup>†§</sup>, Jane L. Doyle<sup>¶</sup>, Jeff J. Doyle<sup>¶</sup>, and Jeffrey D. Palmer<sup>\*||</sup>

<sup>†</sup>Department of Biology, Indiana University, Bloomington, IN 47405; and <sup>¶</sup>L. H. Bailey Hortorium, 462 Mann Library Building, Cornell University, Ithaca, NY 14853

Communicated by C. S. Levings III, North Carolina State University, Raleigh, NC, October 7, 1999 (received for review June 1, 1999)

The respiratory gene *cox2*, normally present in the mitochondrion, was previously shown to have been functionally transferred to the nucleus during flowering plant evolution, possibly during the diversification of legumes. To search for novel intermediate stages in the process of intracellular gene transfer and to assess the evolutionary timing and frequency of *cox2* transfer, activation, and inactivation, we examined nuclear and mitochondrial (mt) *cox2* presence and expression in over 25 legume genera and mt *cox2* presence in 392 genera. Transfer and activation of *cox2* appear to have occurred during recent legume evolution, more recently than previously inferred. Many intermediate stages of the gene transfer process are represented by *cox2* genes in the studied legumes. Nine legumes contain intact copies of both nuclear and mt *cox2*, although transcripts could not be detected for some of these genes. Both *cox2* genes are transcribed in seven legumes that are phylogenetically interspersed with species displaying only nuclear or mt *cox2* expression. Inactivation of *cox2* in each genome has taken place multiple times and in a variety of ways, including loss of detectable transcripts or transcript editing and partial to complete gene loss. Phylogenetic evidence shows about the same number (3–5) of separate inactivations of nuclear and mt *cox2*, suggesting that there is no selective advantage for a mt vs. nuclear location of *cox2* in plants. The current distribution of *cox2* presence and expression between the nucleus and mitochondrion in the studied legumes is probably the result of chance mutations silencing either *cox2* gene.

Most genes encoding mitochondrial (mt) and chloroplast proteins are located in the nucleus, many of these having been transferred from the organellar genomes. Most of this intracellular gene transfer probably occurred early during organellar evolution (reviewed in ref. 1). A number of reports have shown that physical movement of organellar DNA to the nuclear genome is an ongoing process. Although numerous examples of gene fragments of organellar origin have been found in the nuclear genomes of a diverse range of eukaryotes (ref. 2 and references therein), few cases of recent gene transfers resulting in activation of the transferred gene have been documented. Most of these recent evolutionary transfers involve mt genes in plants (e.g., refs. 3–9); in contrast, the numerous sequenced mt genomes of animals all contain the same set of 13 (or occasionally 12) protein genes (10).

Previously characterized cases of gene transfer have revealed several intermediate stages in the process of intracellular gene transfer. In most cases, physical transfer of plant mt genes has been inferred to occur by an RNA intermediate because the nuclear copy resembles an edited mt mRNA rather than an unedited mt gene (e.g., refs. 3–5). Acquisition of regulatory elements and a mt targeting sequence is critical for nuclear gene activation. This has been inferred to have occurred by recombination with a duplicated targeting sequence from a preexisting nuclear gene encoding a mt protein (6), by alternative splicing of a targeting sequence shared with a preexisting mitochondrial-targeted gene (8, 9), and by exon shuffling from a nuclear gene

encoding a cytosolic protein and gain of mt targeting function (11). After nuclear gene activation, both the nuclear and mt copies should be expressed, at least transiently, but no such examples have been reported previously in plants. Inactivation (silencing) and loss have been reported only for the organellar copy but not the nuclear copy, suggesting that inactivation of the organellar gene is favored.

The respiratory gene *cox2*, encoding subunit 2 of cytochrome oxidase, was functionally transferred to the nucleus during flowering plant evolution (3). Functional nuclear genes for *cox2* have been characterized in two legumes, one of which (*Vigna unguiculata*; cowpea) lacks the gene in the mt genome and the other of which (*Glycine max*; soybean) retains an apparently silent copy in the mitochondrion (3, 4). The transfer of *cox2* to the nucleus may be a relatively recent evolutionary event and provides an opportunity to identify and study additional intermediate stages of the transfer process by taking a comparative approach. In the current report, we document the first examples of inactivation of a once-active nuclear gene of recent organellar origin and of transcription of both nuclear and mt copies of a transferred gene in the same plant. We show that there have been about equal numbers of separate inactivations of nuclear and mt *cox2* genes among the studied legumes, suggesting that there is no selective advantage for a mt or nuclear location of *cox2* in plants.

## Materials and Methods

For plant voucher information, see ref. 12; for complete genus and species names, see supplemental data on the PNAS web site, www.pnas.org. Total cellular DNA or RNA was extracted as described (12, 13); slot blot DNA hybridizations were performed as described (12). All RNAs were treated twice with RNase-free DNase (Promega) before reverse transcription and once before use in Northern hybridizations. Poly(A)<sup>+</sup> and poly(A)<sup>–</sup> RNAs were fractionated by using the poly(A)Ttract kit (Promega). RNAs were electrophoresed through 1.0% gels containing formaldehyde and then were transferred to Immobilon (Millipore) membranes by standard blotting procedures. The membranes were prehybridized and hybridized at 58–68°C in 5× standard saline citrate (SSC), 0.1% SDS, 50 mM Tris (pH 8.0), 10 mM EDTA, 2× Denhardt's solution, and 5% dextran sulfate (Pharmacia). After hybridization, filters were washed twice in 0.5%

Abbreviations: mt, mitochondrial; RT, reverse transcription.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. AF181930–AF181949 and AF182779–AF182791).

\*K.L.A. and K.S. contributed equally to this paper.

†Current address: Sigma Chemical Co., 3300 South Second Street, St. Louis, MO 63118.

‡Current address: Department of Biology, NUI Maynooth, Co. Kildare, Ireland.

§To whom reprint requests should be addressed. E-mail: jpalmer@bio.indiana.edu

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

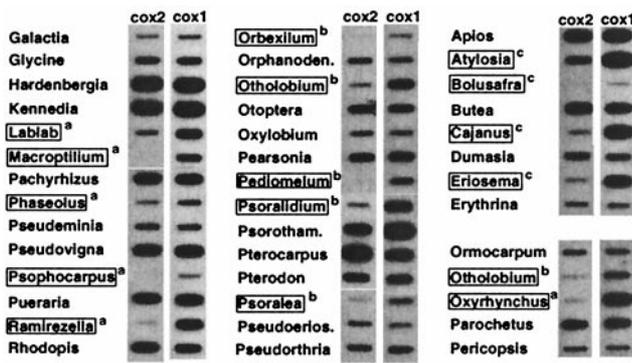


Fig. 1. Presence of *cox2* in mt DNA assayed by slot blot hybridization. Shown are hybridizations of mt *cox2* and *cox1* probes from soybean to 41 of 392 legume DNAs examined. Boxes designate legumes with diminished or undetectable *cox2* hybridization relative to the *cox1* control: a, Phaseolinae species (left, and also bottom right), b, Psoraleeae species (middle, and also bottom right), and c, Cajaninae species (top right).

SDS and either 2×, 0.5×, or 0.25× SSC at the hybridization temperature. Hybridization probes were made by <sup>32</sup>P-labeling using random oligonucleotide primers.

Nuclear and mt *cox2* coding region sequences were isolated by PCR using various pairs of primers (see supplemental material at www.pnas.org). PCR was performed by using 20–50 ng of total cellular DNA in 10- to 50- $\mu$ l reactions, with 0.8–1 mM MgCl<sub>2</sub>, 1 mM each dNTP, 2  $\mu$ M of each primer, and *Taq* polymerase for 30–35 cycles using an Idaho air thermal cycler. Denaturation was at 92°C for 10 sec, annealing at 50–54°C for 20 sec, and extension at 72°C for 1–1.5 min. Reverse transcription was performed at 37°C for 60 min, using 1–2  $\mu$ g of total cellular RNA, 200 units of Moloney murine leukemia virus reverse transcriptase (Life Technologies, Grand Island, NY), 5  $\mu$ M random hexamers, 2.5 mM each dNTP, and 20 units of RNasin (Promega). Controls with all reagents except reverse transcriptase were used to detect contaminating DNA. The cDNAs were treated with RNase A to remove residual RNA, and then PCR was performed as above. 5' rapid amplification of cDNA ends was performed by using a gene-specific primer for reverse transcription, followed by addition of a poly(C) tail using polynucleotide kinase, and finally PCR using an anchor poly(G) forward primer and a gene-specific reverse primer. PCR and reverse transcription (RT)-PCR products were cloned by using the TA cloning kit (Invitrogen), or were sequenced directly, after gel purification using the Gene-Clean kit (Bio 101). Nucleotide sequences were determined for both strands using Applied Biosystems and Li-Cor (Lincoln, NE) automated DNA sequencers.

## Results

**Mitochondrial *cox2* Gene Losses in Legumes.** We began this study by searching for mt *cox2* gene losses among legumes. Finding loss of mt *cox2* in a certain species would suggest the presence of *cox2* in the nucleus. A slot blot hybridization survey of total DNAs from 392 legume genera (listed in ref. 12) was performed to detect mt *cox2* losses. After hybridization with a soybean mt *cox2* probe, the blots were stripped and subsequently hybridized with a probe for *cox1*, a gene “universally” located in the mt genome, as a loading control. Comparison of the *cox2* and *cox1* hybridizations, taking into account unequal DNA loadings, revealed comparable levels of hybridization to most DNAs, a sampling of which are shown in Fig. 1. However, hybridization of the *cox2* probe was weaker than *cox1* hybridization in three groups of legumes (Fig. 1): (i) many genera from the Phaseoleae subtribe Phaseolinae, (ii) all genera sampled from the tribe Psoraleeae, and (iii) some genera from the Phaseoleae subtribe Cajaninae.

We infer that mt *cox2* is partially to completely absent in each of these plants.

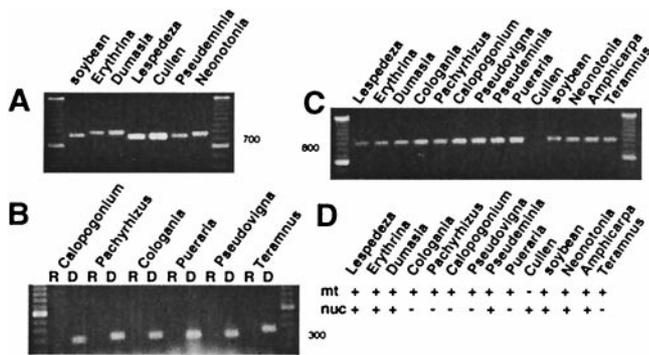
Hybridization of the *cox2* probe to *Phaseolus* DNA was only slightly weaker than *cox1* hybridization, suggesting that much of the *cox2* gene is present. Efforts to PCR amplify the entire mt *cox2* coding region from several species within the Phaseolinae (including *Phaseolus*; see next paragraph), Psoraleeae, and Cajaninae failed, further suggesting the lack of an intact gene. Based on phylogenetic criteria described later, it is likely that mt *cox2* gene inactivation occurred independently during the evolution of the Phaseolinae, Psoraleeae, and Cajaninae. We were surprised by the small number of plants lacking an intact copy of mt *cox2*, and that they all belong to a single legume lineage (Phaseoleae and allies; ref. 14). However, the DNA hybridization surveys would not detect pseudogenes that are mostly intact, and thus there could be mt *cox2* pseudogenes in other legumes.

**Intact and Disrupted Nuclear and Mitochondrial *cox2* Genes.** To search for previously undescribed intermediate stages of the gene transfer process, and to study the evolutionary dynamics of *cox2* gene inactivation and loss, we isolated and sequenced mt and nuclear *cox2* genes from a number of legumes within the Phaseoleae lineage and assayed transcription of each gene. The entire mt *cox2* coding region was PCR-amplified, using primers located in the 5' and 3' untranslated regions, and was sequenced from 20 genera within the Phaseoleae lineage and 13 genera from related lineages that potentially contain an intact mt *cox2* (as judged by DNA hybridization), to determine whether the genes are indeed intact or are disrupted by molecular lesions. The sequences from all examined species, except *Erythrina*, were completely intact. *Erythrina* mt *cox2* contains a deletion of 5 bp, at positions 494 to 498 compared with soybean, that disrupts the reading frame. Attempts to isolate an entire mt *cox2* gene from common bean failed. We were only able to isolate a 360-bp fragment at the 3' end of *cox2*, suggesting that the gene is not intact; this result is consistent with the hybridization results in Fig. 1 and shows that mt pseudogenes could be detected by the DNA hybridizations.

Nuclear *cox2* genes and/or cDNAs were isolated and sequenced from 20 legume genera by PCR, RT-PCR, and 5' rapid amplification of cDNA ends, using several primer combinations. All sequenced nuclear *cox2* genes contain a mt targeting sequence that is homologous to the targeting sequences of nuclear *cox2* from cowpea and soybean and of similar length to the soybean sequence (4). The cowpea targeting sequence is thus longer than previously reported (3). All examined nuclear *cox2* genes also contain an intron located at an identical position between the transit peptide and mature coding regions of the gene. The size of this intron, although generally between 700 and 800 bp, ranged from 117 bp in *Atylosia* and *Eriosema* to  $\approx$ 1,200 bp in *Erythrina*. Two copies of nuclear *cox2* were discovered only in *Erythrina*, a polyploid genus.

Nuclear *cox2* is intact in all but three studied legumes. There are insertions in the targeting sequence of nuclear *cox2* in *Cologania lemonii* (4 bp) and *Teramnus uncinatus* (2 bp) that disrupt the reading frame. Nuclear *cox2* in *Pueraria phaseoloides* contains a 22-bp deletion at the 5' splice site of the intron, which probably disrupts splicing. Interestingly, *Teramnus labialis* and *Pueraria lobata* do not contain corresponding molecular lesions, suggesting that the insertion in *T. uncinatus* and deletion in *P. phaseoloides* occurred very recently.

We were unable to isolate nuclear *cox2* sequences from any examined species outside the lineage containing Phaseoleae and allies and from four examined species within this lineage (*Canavalia*, *Galactia*, *Tephrosia*, and *Clitoria*), suggesting that these legumes may not contain nuclear *cox2* (data not shown). Pea does not appear to contain a nuclear *cox2*, contrary to the previous inference based on Southern hybridization (3). How-

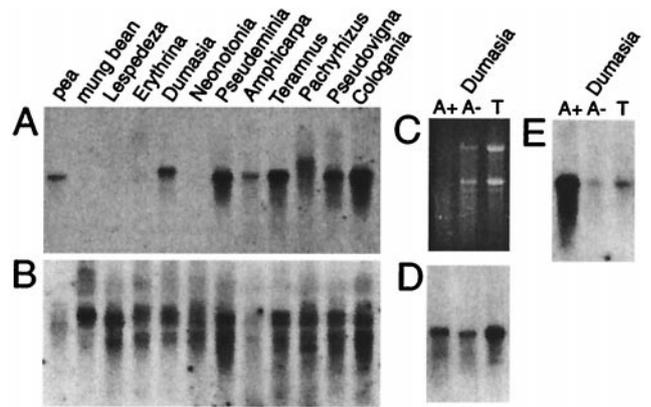


**Fig. 2.** Detection of mt and nuclear *cox2* transcripts by RT-PCR. (A) Agarose gel showing RT-PCR products generated by using primers that amplify almost all of the nuclear *cox2* targeting sequence plus about half of the mature coding region (700–790 bp). Sizing marker is a 100-bp ladder. (B) RT-PCR and PCR products obtained by using primers that amplify the nuclear *cox2* targeting sequence. RT-PCR reactions are indicated by "R" whereas control PCRs using genomic DNA as template are indicated by "D." (C) RT-PCR products generated by using primers that amplify the entire mt *cox2* coding sequence ( $\approx$ 885 bp). (D) Summary of RT-PCR data, indicating the presence (+) or absence (-) of detectable transcripts in the mitochondrion (mt) or nucleus (nuc) of the indicated species.

ever, we cannot rule out the possibility that a highly divergent nuclear *cox2* gene fragment exists in pea or other legumes for which we were unable to isolate a nuclear *cox2* gene.

**Transcribed and Silenced Nuclear and Mitochondrial *cox2* Genes.** Mt and nuclear *cox2* transcription was assayed by a combination of RT-PCR and Northern hybridizations in 13 legumes that contain both *cox2* genes. RT-PCR was particularly useful to detect transcripts present at low levels, but was not performed in a quantitative manner, whereas Northern hybridizations were useful to determine whether transcripts were abundant.

RT-PCR was performed by using multiple primer combinations to detect nuclear *cox2* transcripts. In the first assay, the forward primer was located in the targeting sequence and the reverse primer was located within the mature coding region. Soybean was used as a positive control; nuclear *cox2* in soybean was previously shown to be transcribed (3, 4). This assay produced nuclear *cox2* cDNA fragments of 700–790 bp shown in all seven lanes of Fig. 2A and also a 550 bp fragment from *Amphicarpa* (data not shown), with the sizes depending on the exact location of the primers and the lengths of insertions and deletions within the targeting sequence. These products are all of the correct size for spliced nuclear *cox2* cDNAs from the respective legumes, and they were all sequenced to verify nuclear *cox2* cDNA amplification from the correct species. In the second assay, RT-PCR was performed for six other species by using primer pairs that were located within the targeting sequence. Fig. 2B shows that amplification of nuclear *cox2* was successful in each case when DNA was used as the template, as indicated by the bands of  $\approx$ 320–420 bp in each lane, but not when cDNA was used as template. The gel was blotted and hybridized to a *cox2* probe to verify lack of *cox2* amplification in lanes without bands (data not shown), and the PCR products were sequenced to verify their identity. Multiple independent RT-PCR reactions using varying amounts of cDNA template were performed to confirm the results. Mt *cox2* and *rps10* cDNAs were readily obtained from each species (see below), indicating that the reverse transcription reactions were successful. To confirm the lack of nuclear *cox2* transcripts in leaves of the above six species, RT-PCR was also performed by using multiple reverse primers located in the mature coding region, but no nuclear *cox2* cDNAs were amplified (data not shown). The results of the RT-PCR



**Fig. 3.** Levels of *cox2* transcripts as determined by Northern hybridizations. (A) High stringency hybridization of a mt *cox2* cDNA probe from *Amphicarpa* to various total RNAs. (B) Rehybridization with a mt *cox1* probe from soybean. (C) RNA gel used for hybridizations in D and E. (D) Hybridization of a mt *cox2* cDNA probe from *Amphicarpa*. (E) Rehybridization with an  $\alpha$ -tubulin probe from soybean.

assays indicate that steady-state transcripts of nuclear *cox2* are not present and suggest that the gene is not transcribed, in young leaves of *Pachyrhizus*, *Calopogonium*, *Cologania*, *Pueraria*, *Pseudovigna*, and *Teramnus*. Lack of detectable nuclear *cox2* transcripts was further confirmed by Northern hybridizations using fractionated poly(A)<sup>+</sup> and poly(A)<sup>-</sup> RNAs from *Pueraria*, *Pachyrhizus*, *Teramnus*, and *Pseudovigna* (data not shown).

Presence or absence of mt *cox2* transcripts was assayed by using primers located in the 5' and 3' untranslated regions. Fig. 2C shows that a band of  $\approx$ 885 bp, the expected size for a mt *cox2* cDNA, is present in all lanes except that of *Cullen*. Lack of mt *cox2* cDNA amplification from *Cullen* is consistent with results of the slot blot DNA hybridizations showing partial to complete gene loss in each examined species within the Psoraleae (Fig. 1). All of the mt *cox2* RT-PCR products were cloned, and multiple clones were sequenced to verify cDNA amplification from the indicated species and to determine whether mt *cox2* transcripts were properly edited. C-to-U RNA editing of transcripts from most plant mt protein genes is necessary to produce mRNAs that encode the correct amino acid sequence (reviewed in ref. 15). Transcripts of all examined species, except soybean, were C-to-U edited at 13–15 sites. In soybean, only one cDNA clone of seven sequenced showed any RNA editing, but at only two sites. These results indicate that mt *cox2* is effectively a pseudogene in soybean, consistent with previous observations (3, 4), but is potentially functional in all other examined species with intact *cox2* genes.

The abundance of mt *cox2* transcripts was assayed by Northern blots. Total RNAs were hybridized at high stringency (68°C) with a mt *cox2* cDNA probe and were washed at high stringency (68°C, 0.25 $\times$  SSC) to detect only mt *cox2* transcripts. Pea and mung bean were used as positive and negative controls, respectively. A mt *cox2* cDNA probe hybridized strongly to transcripts of  $\approx$ 1,200 bases from *Dumasia*, *Pseudeminia*, *Teramnus*, *Pachyrhizus*, *Pseudovigna*, and *Cologania* (Fig. 3A) and *Calopogonium* (not shown) whereas somewhat weaker hybridization was detected in the under-loaded pea and *Amphicarpa* lanes. We were surprised to find that the *cox2* probe hybridized very weakly, if at all, to RNAs from *Lespedeza*, *Erythrina*, and *Neonotonia* because fully edited mt transcripts were detected by RT-PCR for all three species. A *cox1* control probe hybridized well to all RNAs except *Amphicarpa* and pea (Fig. 3B), indicating that ample RNA was present in most lanes. Thus, mt *cox2* transcripts are not very abundant in leaves of *Lespedeza*, *Neo-*

*notonia*, and *Erythrina*. Mt *cox2* transcripts in *Lespedeza* and *Neonotonia* could potentially make functional proteins, but their low abundance suggests that most COX2 protein in leaf cells of these plants is produced from the nuclear *cox2* gene.

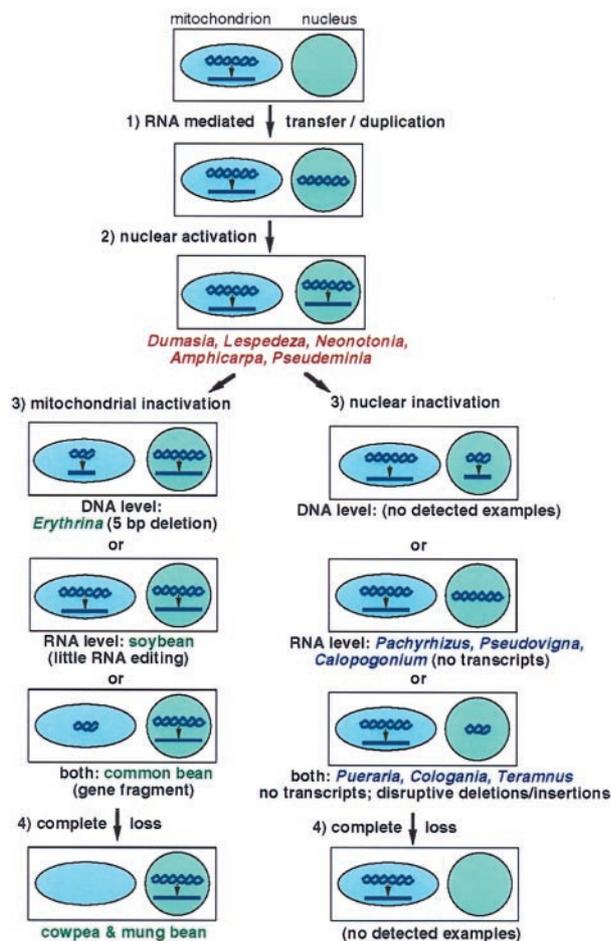
Considering that mt *cox2* is transcribed at low levels in *Neonotonia* and *Lespedeza* but at relatively high levels in *Dumasia*, *Amphicarpa*, and *Pseudeminia*, we wanted to determine whether nuclear *cox2* transcripts in the latter three species are abundant. *Cox2* transcription in *Dumasia*, *Amphicarpa*, and *Pseudeminia* was assayed by Northern hybridizations using fractionated poly(A)<sup>+</sup> and poly(A)<sup>-</sup> RNAs. Most nuclear mRNAs are polyadenylated whereas most plant mt transcripts are not. A mt *cox2* cDNA probe was used because it would be expected to bind approximately equally to nuclear *cox2* transcripts from each species. This probe hybridized strongly to RNAs in all three *Dumasia* lanes, with hybridization in the poly(A)<sup>+</sup> lane being stronger and the band being slightly larger than in the poly(A)<sup>-</sup> lane (Fig. 3D). An  $\alpha$ -tubulin probe hybridized very strongly to the *Dumasia* poly(A)<sup>+</sup> lane, showing that enriched poly(A)<sup>+</sup> RNA is indeed present (Fig. 3E). Levels of contaminating poly(A)<sup>-</sup> RNA, and thus mt *cox2* transcripts, are low in the *Dumasia* poly(A)<sup>+</sup> RNA lane, as seen by the lack of rRNA bands on the ethidium bromide stained gel (Fig. 3C). The *cox2* hybridization signal in the poly(A)<sup>+</sup> lane of *Dumasia* must therefore reflect the presence of nuclear *cox2* transcripts rather than contaminating mt transcripts. Thus, *cox2* transcripts from both compartments are relatively abundant in *Dumasia* leaves. Hybridization of the *cox2* probe was strong to total and poly(A)<sup>-</sup> RNAs from *Amphicarpa* and *Pseudeminia* but very weak to poly(A)<sup>+</sup> RNAs despite strong hybridization of the  $\alpha$ -tubulin probe to the poly(A)<sup>+</sup> lanes (data not shown). Thus, nuclear *cox2* transcripts are present at low levels in *Amphicarpa* and *Pseudeminia*.

To further examine the phylogenetic distribution of nuclear *cox2* expression within the Phaseoleae lineage and related species, Northern blots were performed by using fractionated poly(A)<sup>+</sup> and poly(A)<sup>-</sup> RNAs from *Galactia*, *Canavalia*, *Clitoria*, and *Tephrosia* (all part of the Phaseoleae lineage), along with *Calpurnia* and *Crotalaria*. Hybridization to the *cox2* probe was seen only in poly(A)<sup>-</sup> lanes of each species, although there was strong hybridization of a probe for the nuclear glycine decarboxylase H-protein gene to all poly(A)<sup>+</sup> lanes (data not shown). We infer that nuclear *cox2* is not expressed and, based on PCR experiments discussed above, may be absent in these six legumes. Thus, nuclear *cox2* expression appears to be restricted to a single group of legumes within the Phaseoleae lineage (14).

## Discussion

**Intermediate Stages of Gene Transfer Represented by Legume *cox2* Genes.** Through comparative study of *cox2* gene presence and expression in legumes, we have caught a rare glimpse of the intracellular gene transfer process in action, an opportunity facilitated by the very recent evolutionary transfer of *cox2* to the nucleus. A wide range of intermediate stages in the gene transfer process is represented by nuclear and mt *cox2* genes in the studied legumes (Fig. 4). The physical transfer of mt *cox2* to the nuclear genome involved an edited RNA intermediate, although the cellular location of the reverse transcription step is unknown (3, 4). All of the many nuclear *cox2* genes examined in this study contain a homologous intron located at the same position between the (also homologous) targeting sequence and the mature coding region of the gene; this is consistent with the earlier suggestion (3) that the transferred gene was activated by an exon-shuffling-like process, obtaining its mt targeting sequence and regulatory elements from another, as yet uncharacterized gene.

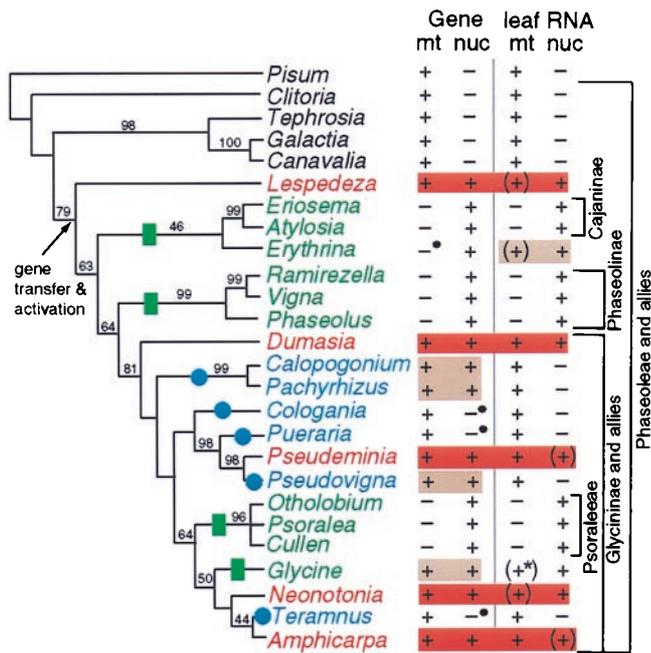
Once nuclear *cox2* was activated, a state of dual intact and transcribed *cox2* genes was established; this is a relictual stage



**Fig. 4.** A model of the intermediate stages of gene transfer represented by nuclear and mt *cox2* in different legumes. Disrupted or fragmented *cox2* genes are indicated by a shorter double helix compared with intact genes. Transcription is signified by an arrow and dark line. Names: red indicates dual transcription, green indicates silenced mt *cox2*, and blue indicates inactivated nuclear *cox2*; these colors correlate with Fig. 5. See text for explanations of the “DNA level,” “RNA level,” and “both” inactivation categories.

that persists in only five, phylogenetically scattered legumes (Fig. 5) of the many examined in this study. Dual transcription of intact genes in both the nucleus and an organelle has been reported only for *atp9* in *Neurospora crassa* (16). Nuclear and mt *cox2* transcripts are both abundant in *Dumasia*, and it is entirely possible that both transcripts are expressed as proteins, perhaps making the genes functionally redundant. In contrast, transcript levels for either mt or nuclear *cox2* are very low in leaves of *Lespedeza*, *Neonotonia*, *Amphicarpa*, and *Pseudeminia*. One of the *cox2* genes in each of these four species may be in the process of being silenced. Alternatively, the *cox2* transcripts that are present in low levels in leaves may be present at high levels in other tissues, organs, or developmental stages in these plants; such differential expression could provide a selective rationale for retaining active copies of both genes.

Inactivation (silencing) of either nuclear or mt *cox2* has occurred multiple times and in a variety of ways (Fig. 4). We have grouped the types of inactivation into two classes: “DNA level” inactivation refers to genes containing mutations within the exons or introns that disrupt the reading frame or intron splicing. “RNA level” inactivation refers to lack of detectable steady-state transcripts or of RNA editing and may reflect mutations in either *cox2* regulatory elements or mutations that disrupt trans-acting



**Fig. 5.** Summary of *cox2* gene distribution and transcript data in a phylogenetic context. The left two columns indicate the presence (+) or absence (-) of an intact *cox2* gene in the mitochondrion (mt) or nucleus (nu) of the indicated species. Bullets indicate genes containing small insertions or deletions that disrupt the reading frame or intron splicing. The right two columns indicate the presence (+) or absence (-) of detectable mt and nuclear *cox2* transcripts in young leaves. Parentheses indicate transcripts present at low levels; the asterisk indicates transcripts that are not properly edited. Red shading highlights dual (nuclear and mt) transcription and proper processing of dual intact *cox2* genes in a given plant; tan indicates dual intact genes or dual transcription, but not both. Nuclear *cox2* transcripts in *Eriosema*, *Alyosia*, *Ramirezella*, *Psoralea*, and *Otholobium* were not assayed and are inferred from the partial to complete absence of mt *cox2* in each plant. Green rectangles and names indicate inactivation of mt *cox2* whereas blue circles and names indicate inactivation of nuclear *cox2*. The phylogenetic tree is one of three equally parsimonious trees obtained from parsimony analysis of a data set consisting of 2,154 bp of two chloroplast gene sequences (*rbcL* and *ndhF*) and 557 chloroplast restriction sites. See supplemental material at [www.pnas.org](http://www.pnas.org) for details of this and other analyses. Bootstrap values above 40% are shown.

factors necessary for proper gene expression. DNA, but not RNA, level inactivation (i.e., a transcribed pseudogene) is exemplified by mt *cox2* in *Erythrina*. RNA, but not DNA, level inactivation (i.e., a silent intact gene) is represented by nuclear *cox2* in *Pachyrhizus*, *Calopogonium*, and *Pseudovigna*, at least in young leaves. Silencing at the level of RNA processing is exemplified by mt *cox2* in soybean, whose transcripts are virtually unedited, rendering them unable to produce a functional protein. Inactivation at both the DNA and RNA levels (i.e., a silent pseudogene) is represented by nuclear *cox2* in *Cologania lemonii*, *Pueraria phaseoloides*, and *Teramnus uncinatus*, and mt *cox2* in common bean. The final step in the gene transfer process, complete gene loss, is exemplified by mt *cox2* in cowpea and mung bean (3), along with most other surveyed members of the genus *Vigna* and four *Phaseolus* species (data not shown).

A necessary limitation of this study, given its broad phylogenetic scope, is that transcription was assayed primarily in young green leaves. To look for cases of differential expression of *cox2* genes, we have initiated a study to examine other tissues, organs, and developmental stages of a few legumes having intact copies of the *cox2* gene in both compartments. Soybean mt *cox2* transcripts are not detected by Northern hybridizations, and thus are not abundant, in hypocotyls, flowers, roots, apical meristems,

primary leaves, and three ages of secondary leaves (data not shown). Thus, it is unlikely that differential expression occurs in soybean.

**Timing of *cox2* Transfer and Activation in the Nucleus.** Three lines of evidence suggest that *cox2* gene transfer occurred once during the recent evolution of the legume subfamily Papilionoideae, probably during the evolution of the Phaseoleae lineage (Fig. 5). First, DNA hybridizations of 392 legume genera revealed mt *cox2* losses in only three groups, all belonging to a single lineage (Phaseoleae and allies). Second, we were unable to isolate nuclear *cox2* sequences from any species outside the Phaseoleae lineage and from four examined species that are basal within the Phaseoleae lineage (*Canavalia*, *Galactia*, *Tephrosia*, and *Clitoria*). Those species found to contain nuclear *cox2* are all members of a well supported monophyletic group (79–96% bootstrap support; see Fig. 5 and supplemental material at [www.pnas.org](http://www.pnas.org)). Third, Northern hybridizations and RT-PCR experiments indicated that nuclear *cox2* expression appears to be restricted to a subset of species within the Phaseoleae lineage. Taken together, these data suggest that transfer of *cox2* occurred much more recently than previously inferred (3, 17). We attribute these previous, seemingly erroneous inferences to phylogenetic artifacts (long branches going deep) stemming from the  $\approx 10$ -fold higher mutation rate (18) of nuclear *cox2* compared with mt *cox2*. The last common ancestor of all legumes known to have nuclear *cox2* also had an “active” nuclear *cox2* gene (i.e., gene transfer and activation are phylogenetically coincident among the studied legumes; Fig. 5). Therefore, it is likely that nuclear *cox2* activation occurred soon after transfer, before the initially inactive nuclear gene could accumulate mutations that would cause it to encode a nonfunctional protein product.

**Multiple Inactivations of Both Mitochondrial and Nuclear *cox2*.** We placed the *cox2* presence/absence and expression data in a phylogenetic context to infer the number of losses and silencings of mt and nuclear *cox2* that have occurred among the studied legumes. Based on phylogenetic evidence (Fig. 5; supplemental material at [www.pnas.org](http://www.pnas.org)), we infer that mt *cox2* has been silenced approximately four times during the evolution of the studied legumes. The 5-bp deletion in mt *cox2* of *Erythrina* could be independent of the partial-to-complete gene loss in *Alyosia* and *Eriosema*; conversely, the mt *cox2* loss in the Psoraleae could be the result of the same inactivation event as for soybean, given the low bootstrap support for soybean grouping with *Neonotonia*, *Amphicarpa*, and *Teramnus* (see supplemental material). Although other examples of silencing and partial-to-complete loss of the mt gene after gene transfer have been reported (e.g., refs. 3, 5, and 7), all have been restricted to one loss per transfer event.

We were surprised to discover that nuclear *cox2* has also been inactivated several times. Five separate inactivations are inferred in Fig. 5. An unlikely alternative is that nuclear *cox2* was reactivated in *Pseudeminia* after inactivation in a common ancestor of *Pseudeminia*, *Cologania*, *Pueraria*, and *Pseudovigna*; this would call for fewer inactivations. Overall, both nuclear and mt *cox2* appear to have been silenced from three to five times among the studied legumes, and nuclear *cox2* might have been reactivated once.

Our finding of inactivation of a previously active nuclear copy of a recently transferred gene establishes that either copy of a gene can be inactivated after the intermediate stage of dual expression, consistent with the gene transfer ratchet model recently proposed by Doolittle (19). Why have all previously reported gene inactivations, after recent gene transfer and nuclear activation, been restricted to the mt copy? Most examples of recent gene transfers were discovered by finding a disrupted or missing mt gene in one plant, or by obtaining a

nuclear sequence from one plant, and then studying the functional nuclear copy of this gene from the same plant only. A comparative phylogenetic approach might reveal that the nuclear copy of other recently transferred organelle genes has become inactivated in one or more species related to the single plant studied so far.

**Selection, Chance, and Directionality of Gene Transfer.** Previous hypotheses have proposed selective advantages for organelle genes to be relocated to the nucleus. One proposed advantage is the release from the effects of Muller's ratchet, the accumulation of deleterious mutations, and increased genetic load within asexual populations, attributable to the lack of recombination (20). Evidence for Muller's ratchet can be seen in the tRNAs of animal mt genomes (21). However, the much lower mutation rate in plant organelle genomes than in the nuclear genome (18) would seem to mitigate, if not entirely offset, the effects of Muller's ratchet (20). Other proposed selective advantages of organelle genes being encoded in the nucleus include relief from free-radical induced mutations (22), which again seems unlikely for plants with their low organellar mutation rates, and improved regulatory control in the nucleus (23).

Finding multiple inactivations of nuclear and mt *cox2* suggests that there is no selective advantage for COX2 to be encoded in the nucleus versus the mitochondrion in plants. We surmise that a state of functional redundancy was established once nuclear *cox2* became expressed, such that the current distribution of the *cox2* gene and its expression among legumes is largely the result of chance mutations silencing one *cox2* gene or the other. Perhaps there is no selective advantage for other mt genes to be transferred to the nucleus in plants and other eukaryotes. Alternatively, there may be selective advantages for some mt

genes, but not others, to be transferred to the nucleus. Also, selective pressures to relocate genes may have changed over time. For instance, Cavalier-Smith (24) has speculated that selection for organellar economy and efficiency was a major force driving the massive transfer of genes early in organellar evolution but became unimportant once organellar genomes were stripped of most of their genes.

We are left with a paradox: For *cox2*, at least, there seem to be no selective forces determining which compartment's gene is silenced and which is retained. Yet, over time it is clear that gene transfer has occurred in a pervasively unidirectional manner, from organelles to the nucleus, possibly according to the gene transfer ratchet model of Doolittle (19). Perhaps the unidirectional flow of genes to the nucleus is largely neutral, driven by mechanistic forces such as the general propensity for organellar DNA to escape from the organelle and be taken up by the nucleus. In this regard, it is notable that experimental studies in yeast have shown remarkably high rates ( $\approx 2 \times 10^{-5}$  per cell per generation) of "escape" of engineered plasmids from the mitochondrion to the nucleus, with the reverse process being undetectable (23, 25). In conclusion, functional gene transfer has been strikingly unidirectional; why this is so is unclear, and mechanistic rather than selective pressures may be of primary importance.

We thank H. Deiderick, W. Fischer, and P. Keeling for critical reading of the manuscript and E. O'Reilly for amplifying and sequencing mt *cox2* from common bean. This work was supported by U.S. Department of Agriculture National Needs Fellowship 95-38420-2214 to K.L.A., National Science Foundation Grant DEB-9614984 to J.J.D., and National Institutes of Health Grant R01 GM-35087 to J.D.P.

- Gray, M. W. (1992) *Int. Rev. Cytol.* **141**, 233–357.
- Blanchard, J. L. & Schmidt, G. W. (1995) *J. Mol. Evol.* **41**, 397–406.
- Nugent, J. M. & Palmer, J. D. (1991) *Cell* **66**, 473–481.
- Covello, P. S. & Gray, M. W. (1992) *EMBO J.* **22**, 3815–3820.
- Wischmann, C. & Schuster, W. (1995) *FEBS Lett.* **375**, 152–156.
- Kadowaki, K., Kubo, N., Ozawa, K. & Hirai, A. (1996) *EMBO J.* **15**, 6652–6661.
- Sanchez, H., Fester, T., Kloska, S., Schroder, W. & Schuster, W. (1996) *EMBO J.* **15**, 2138–2149.
- Figuerola, P., Gómez, I., Holuigue, L., Araya, A. & Jordana, X. (1999) *Plant J.* **18**, 601–609.
- Kubo, N., Harada, K., Hirai, A. & Kadowaki, K. (1999) *Proc. Natl. Acad. Sci. USA* **96**, 9207–9211.
- Beagley, C. T., Okimoto, R. & Wolstenholme, D. R. (1998) *Genetics* **148**, 1091–1108.
- Long, M., de Souza, S., J., Rosenberg, C. & Gilbert, W. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 7727–7731.
- Doyle, J. J., Doyle, J. L. & Palmer, J. D. (1995) *Syst. Bot.* **20**, 272–294.
- Schultz, D. J., Craig, R., Cox-Foster, D. L., Mumma, R. O. & Medford, J. I. (1994) *Plant Mol. Biol. Rep.* **12**, 310–316.
- Doyle, J. J., Doyle, J. L., Ballenger, J. A., Dickson, E. E., Kajita, T. & Ohashi, T. (1997) *Am. J. Bot.* **84**, 541–554.
- Maier, R. M., Zeltz, P., Kössel, H., Bonnard, G., Gualberto, J. M. & Grienberger, J. M. (1996) *Plant Mol. Biol.* **32**, 343–365.
- Bittner-Eddy, P., Monroy, A. F. & Brambl, R. (1994) *J. Mol. Biol.* **235**, 881–897.
- Laroché, J., Li, P., Maggia, L. & Bousquet, J. (1997) *Proc. Natl. Acad. Sci. USA* **94**, 5722–5727.
- Wolfe, K. H., Li, W.-H. & Sharp, P. M. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 9054–9058.
- Doolittle, W. F. (1998) *Trends Genet.* **14**, 307–311.
- Martin, W. & Herrmann, R. G. (1998) *Plant Physiol.* **118**, 9–17.
- Lynch, M. (1996) *Mol. Biol. Evol.* **13**, 209–220.
- Allen, J. F. & Raven, J. A. (1996) *J. Mol. Evol.* **42**, 482–492.
- Thorsness, P. E. & Weber, E. R. (1996) *Int. Rev. Cytol.* **165**, 207–233.
- Cavalier-Smith, T. (1987) *Ann. NY Acad. Sci.* **503**, 55–71.
- Thorsness, P. E. & Fox, T. D. (1990) *Nature (London)* **346**, 376–379.