Multiple Losses and Transfers to the Nucleus of Two Mitochondrial Succinate Dehydrogenase Genes During Angiosperm Evolution

Keith L. Adams, Monica Rosenblueth,1 Yin-Long Qiu2 and Jeffrey D. Palmer

Department of Biology, Indiana University, Bloomington, Indiana 47405

Manuscript received December 21, 2000
Accepted for publication April 17, 2001

ABSTRACT

Unlike in animals, the functional transfer of mitochondrial genes to the nucleus is an ongoing process in plants. All but one of the previously reported transfers in angiosperms involve ribosomal protein genes. Here we report frequent transfer of two respiratory genes, sdh3 and sdh4 (encoding subunits 3 and 4 of succinate dehydrogenase), and we also show that these genes are present and expressed in the mitochondria of diverse angiosperms. Southern hybridization surveys reveal that sdh3 and sdh4 have been lost from the mitochondrion about 40 and 19 times, respectively, among the 280 angiosperm genera examined. Transferred, functional copies of sdh3 and sdh4 were characterized from the nucleus in four and three angiosperm families, respectively. The mitochondrial targeting presequences of two sdh3 genes are derived from preexisting genes for anciently transferred mitochondrial proteins. On the basis of the unique presequences of the nuclear genes and the recent mitochondrial gene losses, we infer that each of the seven nuclear sdh3 and sdh4 genes was derived from a separate transfer to the nucleus. These results strengthen the hypothesis that angiosperms are experiencing a recent evolutionary surge of mitochondrial gene transfer to the nucleus and reveal that this surge includes certain respiratory genes in addition to ribosomal protein genes.

Most mitochondrial genes were lost or transferred to the nucleus early in eukaryotic evolution, relatively soon after the endosymbiotic origin of the mitochondrion (Gray 1992; Gray et al. 1999). Although functional gene transfer has ceased in animals (Boore 1999) and perhaps some other eukaryotes, gene transfer to the nucleus is an ongoing process in plants (reviewed in Palmer et al. 2000). Characterization of recent cases of gene transfer in angiosperms provides an opportunity to address the frequency, mechanisms, and intermediates of a process of fundamental importance to the genetic coevolution of the eukaryotic cell. The process of functional gene transfer involves several steps (reviewed in Brennicke et al. 1993; Martin and Herrmann 1998): reverse transcription of a mitochondrial mRNA (e.g., Nugent and Palmer 1991; Covello and Gray 1992; Wischmann and Schuster 1995; Kobayashi et al. 1997), movement to the nucleus (see Thorsness and Weber 1996), chromosomal integration (Blanchard and Schmidt 1996; Ricchetti et al. 1999), gain of a nuclear promoter and other regulatory elements, usually the gain of a mitochondrial targeting presequence for routing of the protein to the mitochondrion, and silencing or loss of the original mitochondrial gene.

The frequency of recent transfers of mitochondrial genes to the nucleus is relatively unexplored. Although several mitochondrial genes have been reported to have been transferred to the nucleus in one or two groups of angiosperms (reviewed in Palmer et al. 2000), the frequency of parallel transfers of the same gene during angiosperm evolution is only beginning to be elucidated. We recently found evidence for many independent transfers of the mitochondrial ribosomal protein gene rps10 to the nucleus during recent angiosperm evolution (Adams et al. 2000). Some of the nuclear rps10 genes lacked mitochondrial targeting presequences, bypassing a normal requirement for gene activation and possibly facilitating many transfers of this gene. It is an open question as to whether rps10 is exceptional with its repeated and recent transfers, or whether other mitochondrial genes have also been transferred many times in angiosperms.

All but one of the reported cases of gene transfer in angiosperms involve ribosomal protein genes. The only documented transfer of a respiratory gene is that of cox2 (encoding cytochrome oxidase subunit 2) in legumes (Nugent and Palmer 1991; Covello and Gray 1992; Adams et al. 1999). This raises the possibility that ribosomal protein genes might be easier to functionally relocalize in the nucleus than respiratory genes in angiosperms. Alternatively, additional cases of respiratory gene transfer may be waiting to be discovered.

In plants there are five respiratory gene complexes
in the mitochondrial inner membrane that have some subunits encoded by the nucleus and some by the mitochondrion. The succinate dehydrogenase complex (complex II) plays a role in both the tricarboxylic acid cycle and electron transfer to ubiquinone. The complex contains four subunits: SDH1 is a flavin protein, SDH2 is an iron-sulfur protein, SDH3 is a small integral membrane apocytochrome, and SDH4 is a small membrane-anchoring protein. Sdh1 is not present in the mitochondrion of any examined eukaryote (Lange et al. 1999), suggesting a very ancient transfer of this gene to the nucleus. Genes for the other three subunits were discovered in mitochondrial genomes only in 1996 (Burger et al. 1996) and are currently known to be present in but a few disparately related mitochondrial genomes (Gray et al. 1998). Sdh2 is not present in any characterized plant mitochondrial genome, sdh3 is mitochondrially located in the liverwort Marchantia polymorpha (Odai et al. 1992) but not in any characterized vascular plant, and sdh4, also found in Marchantia mitochondrial DNA, is present as a pseudogene in a few angiosperms (Giege et al. 1998). All four sdh genes are located in the nucleus in all examined animals and fungi (Boore 1999; Lange et al. 1999).

In this study we show that sdh3 and sdh4 are present and expressed in the mitochondrion of diverse angiosperms, but that both genes have also been frequently lost from the mitochondrial genome during recent angiosperm evolution. We present evidence for multiple recent transfers of both sdh3 and sdh4 to the nucleus, with some transfers highlighting potential mechanisms for the acquisition of mitochondrial targeting presequences.

MATERIALS AND METHODS

Nucleic acid extractions and hybridizations: DNA and RNA extractions were performed as previously described (Qiu et al. 1998; Adams et al. 1999). DNAs were digested with BamHI, electrophoresed, and blotted onto nylon membranes (Immobilon) using standard procedures. DNAs included the 277 used in Adams et al. (2000), plus cotton, tomato, and soybean. Membranes were prehybridized and hybridized at 60° 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, Piscataway, NJ). After hybridization, filters were washed twice for 45 min in 0.5% SDS and 2× SSC at 60°. Hybridization probes were made by 32P-labeling PCR products from tomato sdh3, Arabidopsis mitochondrially sdh4 (Giege et al. 1998), or Arabidopsis nad9 (Unsedl et al. 1997) using random oligonucleotide primers. The sdh3 and nad9 probes contained the entire open reading frame. The sdh4 probe omitted the first 80 bp because of small pseudogene fragments corresponding to this region in soybean and some grasses.

Gene isolation and sequencing: Mitochondrial sdh3 and sdh4 genomic sequences were isolated by PCR using pairs of the following primers, designed on the basis of conserved regions of angiosperm cox3/sdh4 genes or the tomato mitochondrial sdh3 sequence: sdh3 F1 (5′-CCCATTATCTTACATATCAGTCT-3′); sdh3 R2 (5′-AATCCGAAAAATACCCGATA-3′); sdh3 R3 (5′-CAC AGGCATGTATCACTTT-3′); sdh4 F1, located in cox3 (5′-GAGAGRAGCATGGTT-3′); sdh4 R1 (5′-GAATTCGATCCATTAGGTT-3′). PCR was performed by using 20 ng of total cellular DNA in 10-μl reactions, with 0.8 mM MgCl2, 1 mM each dNTP, 2 μM of each primer, and Taq polymerase for 30 cycles using an Idaho air thermal cycler. Denaturation was at 94° for 10 sec, annealing was at 50° for 10 sec, and extension was at 72° for 1 min. Mitochondrial sdh3 and sdh4 cDNA sequences were isolated by RT-PCR using pairs of the above primers; reverse transcription was performed as previously described (Adams et al. 1999), and PCR was done as above. Genomic DNA sequences of nuclear sdh3 were isolated by PCR using 50 ng of DNA and the following primers: cotton sdh3 (for: 5′-TGAGACTTCTTACATATCAGTCT-3′; rev: 5′-TGACTCAGATTCATATCACTTT-3′); soybean sdh3 (for: 5′-CGATCAATTCTCCGAGTACAAAG-3′; rev: AGTTCCACCTGATTCAACGTA-3′). PCR conditions were similar to those above, except that annealing temperatures were 48° and 54°, respectively. Inverse PCR was performed by digesting 2 μg of DNA with RsaI, self-ligating the fragments using 800 units of DNA ligase (New England Biolabs, Beverly, MA) in 400-μl volumes at 12° overnight, and performing PCR using gene-specific primers in inverse orientations.

PCR products were sequenced directly, and RT-PCR products were cloned into the TA cloning vector (Invitrogen, San Diego) followed by sequencing of multiple clones. All sequencing was done on both strands using an ABI 377 or 3700 DNA sequencer.

Sequence alignments were performed using Genetics Computer Group’s Print/program and refined by eye. Transmembrane segments were predicted using TMPred (Hofmann and Stoffel 1993). Hydrophylic analysis, including local hydrophylicity and mesohydrophylic, was done using MITO-PROT (Claros and Vincens 1996).

Sequence accession numbers: Sequences determined in this study have been assigned the following GenBank numbers: tomato sdh3 (AF362730), Podophyllum sdh3 (AF362731), Oxalis sdh3 (AF362732), Gymnocalus sdh3 (AF362733), cotton mitochondrial sdh3 (AF362734), tomato sdh4 (AF362735), Podophyllum sdh4 (AF362736), Gymnocalus sdh4 (AF362737), Euphorbia sdh4 (AF362738), cotton sdh4 (AF362739), tobacco nuclear sdh3 genomic (AF362739), and soybean nuclear sdh3 genomic (AF362740).

The following nuclear sdh3 expressed sequence tagged (EST) sequences were utilized in this study: cotton (AI727357, AI727171, and AI726398), soybean (AW509894, AW506629, AW892930, and BE611137), Medicago (AL366685, AW777565, and AL388295). Lotus (AV423073, AV43748), maize (BE050093, AE94541, AW076327, and AW258051), wheat (BE443463, BE606408, and BE443636), barley (BE437666 and BE421665), Sorghum (BG464906, BG464704), rice (C52905, AU063694, C98132, and D45345), and Arabidopsis (AV544146). The following nuclear sdh4 EST sequences were utilized in this study: Arabidopsis (AV553901), soybean (AI756274, AW423419, AI443575, and BG157730), Medicago (BE240253), Lotus (AV412486 and AV420169), rice (C25392, C28430, and C28680), and Alfalfa (AI009987), maize (AW062093 and AW562838), barley (AW98-2649, BE558940, and BF266472), and wheat (BF209714, BF496958, BE424770, and BE499448).

Nuclear sdh3 and sdh4 genomic sequences from Arabidopsis have the following GenBank accession numbers and gene numbers: sdh3 on chromosome 4, AL021811, part of gene F10M6.150; sdh3 on chromosome 5, AL353994, gene F17L14.210; sdh4, AC006418, part of gene AL2g46510. Nuclear sdh3 and sdh4 genomic sequences from rice were obtained from Mor-anto’s rice genome sequence (http://www.rice-research.org) and are available in sequence contigs OSM128526, OSM-1351, and OSM14077. The following non-sdh3 or sdh4 sequences are included in

K. L. Adams et al.
RESULTS

Sdh3 and sdh4 are present and expressed in some angiosperm mitochondrial genomes: Previously, three protein genes (sdh3, rpl6, and rps8) present in Marchantia mitochondrial DNA had not been found in angiosperm mitochondrial genomes, even upon complete sequencing of Arabidopsis and sugar beet (Beta vulgaris) mitochondrial genomes (Unseld et al. 1997; T. Kubo et al. 2000) and despite extensive study of mitochondrial genes in a few other angiosperms. One hypothesis was that these genes were transferred to the nucleus and lost from the mitochondrion before the emergence of angiosperms.

To identify nuclear-encoded homologs of sdh3 in angiosperms, we used tBLASTx searches of the National Center for Biotechnology Information (NCBI) sequence databases with the Marchantia sdh3 as a query sequence. An sdh3 EST was identified from tomato (Lycopersicon esculentum or Solanum esculentum) that contained no upstream extension of the open reading frame (ORF) that might serve as a mitochondrial targeting presequence. The genomic sdh3 sequence, obtained by PCR amplification, was identical to the EST sequence except at one site that has a C in the gene and a T in the cDNA. We suspected that this might represent an RNA editing site. Extensive C-to-U RNA editing occurs in the mitochondrion, but not the nucleus, of angiosperms (reviewed in Maier et al. 1996); thus the gene might still be located in the mitochondrion in tomato. Sequencing of three RT-PCR clones of tomato sdh3 revealed a total of three RNA editing sites (Figure 1). Thus, we conclude that sdh3 is indeed present and expressed in the mitochondrion of tomato.

To examine whether sdh3 is a mitochondrial gene in other angiosperms, sdh3 genomic and cDNA sequences were determined from three diverse eudicots (Figure 1): Oxalis (wood sorrel), the legume Gymnocladus (Kentucky coffee tree), and Podophyllum (mayapple). In each case, the sdh3 ORF is intact and subject to C-to-U RNA editing, suggesting that in each plant a functional sdh3 gene is located in the mitochondrion. In addition, a mitochondrial sdh3 pseudogene sequence was isolated from cotton (Gossypium) that contains a small deletion at a highly conserved region of eukaryotic sdh3 genes.

Sdh4 was first identified in angiosperm mitochondrial DNA by Gieger et al. (1998) as a pseudogene downstream of cox3 in Arabidopsis, sunflower (Helianthus), and Oenothera, and as a small gene fragment in soybean (Glycine), broad bean (Vicia), maize (Zea), rice (Oryza), and wheat (Triticum). We discovered a sdh4 pseudogene in the mitochondrial genome of sugar beet that was not identified in the report describing the complete sequence of this genome (T. Kubo et al. 2000). Sdh4 from Arabidopsis, Oenothera, sugar beet, and sunflower contains a highly conserved sequence of ~327 bp, but the sequences completely diverge from each other downstream of this region.

To determine if sdh4 is intact, transcribed, and RNA edited in selected angiosperms, we sequenced genomic DNA and cDNAs from the highly conserved region of sdh4 from tomato, Gymnocladus, Euphorbia (poinsettia), and Podophyllum, representing four diverse groups of eudicots. Each cDNA sequence is intact (from the initiation codon located at the end of cox3) and RNA edited at several sites (data not shown; see MATERIALS AND METHODS for GenBank numbers), suggesting that functional mitochondrial sdh4 genes are present in these four angiosperms. An intact mitochondrial sdh4 gene was also sequenced from cotton, although transcription was not assayed.

Many losses of sdh3 and sdh4 from angiosperm mitochondrial DNA: The presence of sdh3 in the mitochondrion of five angiosperms, but lack of the gene in the mitochondrion of Arabidopsis and sugar beet (Unseld et al. 1997; T. Kubo et al. 2000), suggests that it might have a sporadic distribution in angiosperm mitochondria. To comprehensively survey the presence/absence of sdh3 and sdh4 in angiosperm mitochondria, a Southern blot hybridization survey was performed using DNAs from 280 diverse genera of angiosperms, representing 169 families. Such a survey was possible because of the very low nucleotide substitution rate of angiosperm mitochondrial genes (Wolfe et al. 1987; Lacroche et al. 1997). Mitochondrial gene loss was inferred if there was no detectable hybridization on an overexposed autoradiograph against two layers of controls: good hybridization to the DNA in question using other probes (e.g., nad9) and good hybridization to other DNAs with the probe in question. This inference assumes that mitochondrial genes are in high copy number relative to nuclear genes of single or low copy number. The sdh3 probe failed to hybridize to 87 angiosperm DNAs, while the sdh4 probe failed to hybridize to 32 angiosperm DNAs (e.g., Figure 2).

When the hybridization data were plotted on a phylogenetic tree of the surveyed species, a total of 40 separate losses of sdh3 and 19 separate losses of sdh4 were inferred (Figure 3). Sdh3 losses were broadly distributed across dicots and encompassed most monocots (middle left). Sdh4 losses were concentrated in the monocots and no losses were detected in basal angiosperms (lower left). Most of the losses are limited in phylogenetic depth to a single family and occurred recently.

Our blot surveys will not detect mitochondrial pseudogenes unless much or all of the probe region is missing, and thus there are probably even more losses of functional sdh3 and sdh4 genes. Conversely, we probably
Figure 1.—Alignment of sdh3 cDNA sequences from angiosperm mitochondria. Dots indicate nucleotides identical to tomato. Gaps inserted to improve alignment are indicated by dashes. Bases for all C-to-T RNA-edited nucleotides are shown in boldface type.

have overestimated the number of functional gene losses in phylogenetic groups where multiple losses were inferred among closely related species. One example is in the Caryophyllales (Figure 3, top left). If the sdh3 hybridization signals in Spinacia and Celosia represent sdh3 pseudogenes instead of intact genes, then there would be one functional loss of sdh3 rather than three. Most notably, the 10 losses of sdh3 mapped onto monocots in Figure 3 reduce to a single functional loss if one invokes mitochondrial pseudogenes in the 8 (out of 40) monocots (excluding the basal monocot Acorus) for which sdh3 hybridization was detected. If all such cases are taken into account, the inferred number of losses of sdh3 and sdh4 would be reduced to 27 and 16, respectively. Regardless of the exact number of functional losses, our blot surveys indicate that sdh3 and sdh4 genes have been lost from the mitochondrion many times during angiosperm evolution.

Sdh3 has been transferred to the nucleus in four an-giosperm families: gain of mitochondrial presequences from preexisting genes: The absence of sdh3 from the mitochondrion of many angiosperms, including Arabidopsis, suggests that it has been transferred to the nucleus. We identified sdh3 sequences on Arabidopsis nuclear chromosomes 4 and 5 (MAYER et al. 1999; TABATA et al. 2000) by searches of the NCBI databases (see MATERIALS AND METHODS for accession numbers), although neither sequence was annotated as sdh3. While our manuscript was in preparation, the copy of sdh3 on chromosome 5 was identified by others as being a gene of recent mitochondrial origin for a putative protein (ARABIDOPSIS GENOME INITIATIVE 2000). The two sdh3 sequences are 99% identical, suggesting a very recent duplication. Each sequence contains an upstream extension of the open reading frame, much of which is homologous to, and shares two intron positions with, the mitochondrial targeting presequence of the anciently transferred gene for the mitochondrial chaperone HSP70 (Figure 4). Thus, sdh3 has gained a presumptive mitochondrial targeting presequence, and perhaps upstream cis-regulatory elements, from a preexisting gene for a mitochondrial protein. Interestingly, the next gene upstream of sdh3 on chromosome 5 is a full-length and potentially functional copy of hsp70; the implications of this are explored in the DISCUSSION. At least two full-length copies of hsp70 are present in the Arabidopsis genome.

Given that mitochondrial sdh3 is a pseudogene in cotton (Gossypium hirsutum), we predicted that the functional gene had been transferred to the nucleus. Nuclear sdh3 cDNAs were identified from several cotton EST sequences from the NCBI databases (see MATERIALS and METHODS for accession numbers), and the genomic sdh3 sequence was obtained by PCR amplification and sequencing. Cotton sdh3 has an upstream extension during angiosperm evolution. Sdh3 has been transferred to the nucleus in four angiosperm families: gain of mitochondrial presequences from preexisting genes: The absence of sdh3 from the mitochondrion of many angiosperms, including Arabidopsis, suggests that it has been transferred to the nucleus. We identified sdh3 sequences on Arabidopsis nuclear chromosomes 4 and 5 (MAYER et al. 1999; TABATA et al. 2000) by searches of the NCBI databases (see MATERIALS AND METHODS for accession numbers), although neither sequence was annotated as sdh3. While our manuscript was in preparation, the copy of sdh3 on chromosome 5 was identified by others as being a gene of recent mitochondrial origin for a putative protein (ARABIDOPSIS GENOME INITIATIVE 2000). The two sdh3 sequences are 99% identical, suggesting a very recent duplication. Each sequence contains an upstream extension of the open reading frame, much of which is homologous to, and shares two intron positions with, the mitochondrial targeting presequence of the anciently transferred gene for the heat-shock protein HSP22 (Figure 4). Thus, sdh3 has gained a presumptive mitochondrial targeting presequence, and perhaps upstream cis-regulatory elements, from a preexisting gene for a mitochondrial protein. Interestingly, the next gene upstream of sdh3 on chromosome 5 is a full-length and potentially functional copy of hsp70; the implications of this are explored in the DISCUSSION. At least two full-length copies of hsp70 are present in the Arabidopsis genome.

Given that mitochondrial sdh3 is a pseudogene in cotton (Gossypium hirsutum), we predicted that the functional gene had been transferred to the nucleus. Nuclear sdh3 cDNAs were identified from several cotton EST sequences from the NCBI databases (see MATERIALS and METHODS for accession numbers), and the genomic sdh3 sequence was obtained by PCR amplification and sequencing. Cotton sdh3 has an upstream extension of the open reading frame that is homologous to, and shares an intron position with, the 5’ end of the anciently transferred gene for the heat-shock protein HSP22 (Figure 4). Thus, sdh3 has gained a presumptive mitochondrial targeting presequence, and perhaps upstream cis-regulatory elements, from a preexisting gene for a mitochondrial protein. Interestingly, the next gene upstream of sdh3 on chromosome 5 is a full-length and potentially functional copy of hsp70; the implications of this are explored in the DISCUSSION. At least two full-length copies of hsp70 are present in the Arabidopsis genome.

Given that mitochondrial sdh3 is a pseudogene in cotton (Gossypium hirsutum), we predicted that the functional gene had been transferred to the nucleus. Nuclear sdh3 cDNAs were identified from several cotton EST sequences from the NCBI databases (see MATERIALS and METHODS for accession numbers), and the genomic sdh3 sequence was obtained by PCR amplification and sequencing. Cotton sdh3 has an upstream extension of the open reading frame that is homologous to, and shares an intron position with, the 5’ end of the anciently transferred gene for the heat-shock protein HSP22 (Figure 4). Thus, sdh3 has gained a presumptive mitochondrial targeting presequence, and perhaps upstream cis-regulatory elements, from a preexisting gene for a mitochondrial protein. Interestingly, the next gene upstream of sdh3 on chromosome 5 is a full-length and potentially functional copy of hsp70; the implications of this are explored in the DISCUSSION. At least two full-length copies of hsp70 are present in the Arabidopsis genome.

Given that mitochondrial sdh3 is a pseudogene in cotton (Gossypium hirsutum), we predicted that the functional gene had been transferred to the nucleus. Nuclear sdh3 cDNAs were identified from several cotton EST sequences from the NCBI databases (see MATERIALS and METHODS for accession numbers), and the genomic sdh3 sequence was obtained by PCR amplification and sequencing. Cotton sdh3 has an upstream extension of the open reading frame that is homologous to, and shares an intron position with, the 5’ end of the anciently transferred gene for the heat-shock protein HSP22 (Figure 4). Thus, sdh3 has gained a presumptive mitochondrial targeting presequence, and perhaps upstream cis-regulatory elements, from a preexisting gene for a mitochondrial protein. Interestingly, the next gene upstream of sdh3 on chromosome 5 is a full-length and potentially functional copy of hsp70; the implications of this are explored in the DISCUSSION. At least two full-length copies of hsp70 are present in the Arabidopsis genome.

Given that mitochondrial sdh3 is a pseudogene in cotton (Gossypium hirsutum), we predicted that the functional gene had been transferred to the nucleus. Nuclear sdh3 cDNAs were identified from several cotton EST sequences from the NCBI databases (see MATERIALS and METHODS for accession numbers), and the genomic sdh3 sequence was obtained by PCR amplification and sequencing. Cotton sdh3 has an upstream extension of the open reading frame that is homologous to, and shares an intron position with, the 5’ end of the anciently transferred gene for the heat-shock protein HSP22 (Figure 4). Thus, sdh3 has gained a presumptive mitochondrial targeting presequence, and perhaps upstream cis-regulatory elements, from a preexisting gene for a mitochondrial protein. Interestingly, the next gene upstream of sdh3 on chromosome 5 is a full-length and potentially functional copy of hsp70; the implications of this are explored in the DISCUSSION. At least two full-length copies of hsp70 are present in the Arabidopsis genome.
of soybean (Glycine), Medicago, and Vigna (Figure 3, top right), we predicted that the gene had been transferred to the nucleus in a common ancestor of these legumes. Searches of NCBI databases revealed multiple sdh3 ESTs from soybean, Medicago trunculata, and Lotus japonicus (see MATERIALS AND METHODS for accession numbers). One intron was revealed upon PCR amplification and sequencing of most of the genomic sequence of soybean sdh3. Soybean SDH3 is predicted to be a mitochondrial protein by MITOPROT (Claros and Vincens 1996), Predotar version 0.5 (http://www.inra.fr/Internet/Produits/Predotar), and TargetP (Emanuelsson et al. 2000). The first 31 amino acids of soybean SDH3 are predicted by MITOPROT to be a mitochondrial targeting presequence, but this region has no similarity to any sequences currently in the NCBI databases. The sequence corresponding to amino acids 82–155 is similar to, and shares an intron position with, a region at the 3’ end of the gene for cytoplasmic aconitase (Figure 4). It is not known if the aconitase region serves any function.

Compared to sdh3 genes in the liverwort Marchantia, in all other examined eukaryotes and in α-proteobacteria, all sequenced nuclear and mitochondrial sdh3 genes from angiosperms are missing the region corresponding to the third transmembrane segment, located at the carboxy terminus of the protein. This suggests that angiosperm SDH3 proteins do not need this region of the protein for anchoring to the inner mitochondrial membrane, unlike in yeast where deletion of this region results in considerably reduced growth (Oyedotun and Lemire 1999). Each of the nuclear SDH3 proteins has a decreased local hydrophobicity compared to tomato mitochondrial SDH3 (data not shown), and both Arabidopsis and soybean SDH3 have a decreased mesohydrophobicity (i.e., highest average hydrophobicity 60–80 amino acids). Similar reductions in hydrophobicity were noted for the transferred cytochrome oxidase subunit 3 (cox3) gene in two green algae (Pérez-Martínez et al. 2000). The implications of these reductions in hydrophobicity are discussed later.

All members of the grass family (Poaceae) that were surveyed by Southern blot hybridization lack sdh3 in the mitochondrion (Figure 3, middle left). Searches of the Monsanto rice genome sequence database (http://www.rice-genome.org) revealed two copies of nuclear sdh3 that each contain one intron in the coding region. One copy also contains an intron in the 5’ untranslated region, raising the possibility of regulatory element acquisition by exon shuffling; it has not been determined if the second copy has this intron. Nuclear sdh3 was also identified in maize, wheat, barley (Hordeum), and Sorghum by searches of the NCBI EST databases (see MATERIALS AND METHODS for accession numbers). Each sequence contains a 5’ extension of the open reading frame of ~83 amino acids that might include a mitochondrial presequence, although the prediction scores were low. Each of the sdh3 genes from the grasses contains a shorter sdh3-homologous region than in other eukaryotes. All grass SDH3 sequences are 42 amino acids shorter at the carboxy terminus than Arabidopsis SDH3 and are missing two of the three transmembrane segments that are present in non-angiosperms. Considering that sdh3 has been lost from the mitochondrion of maize, wheat, and barley (Figure 3), that nuclear sdh3 is transcribed in five grasses representing three major tribes, and that no other sdh3-homologous sequences have been discovered in the grasses, the truncated sdh3 may code for all or part of the functional SDH3 protein in the grasses.

*sdh4* has been transferred to the nucleus in three angiosperm families: Our Southern blot hybridizations
Figure 3.—Many losses of mitochondrial sdh3 and sdh4 among 280 surveyed angiosperms. Columns of horizontal ovals and rectangles indicate lack of hybridization on the Southern blots (as in Figure 2). Vertical ovals on the branches of the tree indicate losses of sdh3; vertical rectangles indicate losses of sdh4. Boldface lettering indicates species from which sdh3 or sdh4 was discovered in the nucleus (see next section of results). The phylogenetic tree is based largely on the strict consensus trees from Soltis et al. (1999, 2000), although other studies were also consulted (see http://www.bio.indiana.edu/~palmerlab).
revealed several losses of sdh4 from the mitochondrion of angiosperms, including the legumes soybean and Medicago. The loss of sdh4 from the mitochondrion of some legumes is probably recent, as suggested by the presence and expression of sdh4 in the mitochondrion of another legume, Gymnocladus (Figure 1). Searches of NCBI EST databases revealed sdh4 genes in the nucleus of soybean, Medicago, and Lotus (see materials and methods for accession numbers). The soybean sdh4 gene product is predicted by MITOPROT, Predotar, and TargetP to be targeted to the mitochondrion and contains a putative presequence of 82 amino acids (Figure 5). The presequence has no similarity to any genes in the NCBI databases.

To determine whether sdh4 has been transferred to the nucleus of Arabidopsis, in which a pseudogene exists in the mitochondrion, searches of the NCBI databases were performed. Sdh4 is present on chromosome 2 (Lin et al. 1999), but was annotated as part of a transcription factor. The Arabidopsis nuclear sdh4 has a putative presequence of 55 amino acids (predicted by MITOPROT; Figure 5) that has no similarity to any sequences currently in NCBI’s databases nor to the soybean presequence (Figure 5), suggesting acquisition from different sources. Sdh4 is located between genes encoding a putative bHLH transcription factor and a putative cellular apoptosis susceptibility protein. Considering the microsynteny in nuclear gene order present in many angiosperm families, including the Brassicaceae (Acarkan et al. 2000; O’Neil and Bancroft 2000), this location may represent the point of insertion into the nuclear genome immediately following transfer, although subsequent genomic rearrangement is also possible.

On the basis of Southern hybridizations (Figure 3), sdh4 was inferred to be absent from the mitochondrion of all four examined grasses: maize, wheat, barley, and Dendrocalamus. Sdh4 was discovered in the nucleus of rice, maize, wheat, and barley by searches of NCBI EST databases and the Monsanto rice genome database. Rice sdh4 is predicted by MITOPROT, Predotar, and TargetP to be targeted to the mitochondrion and has a putative presequence of 93 amino acids (Figure 5) that has no similarity to any sequences in the above databases. There are two introns in rice sdh4, one of which is within the
mitochondrial presequence. The next genes upstream of sdh4 in rice encode a putative glutaredoxin and a putative gamma glutamyltransferase; neither of these genes is located in the vicinity of sdh4 in Arabidopsis. The next genes downstream of rice sdh4 could not be determined.

**DISCUSSION**

**Perspectives on the origin of mitochondrial targeting presequences:** An important step in the activation of most newly transferred mitochondrial genes is the acquisition of a mitochondrial targeting presequence. Both Arabidopsis sdh3 and cotton sdh3 have presequences that were derived from preexisting genes for mitochondrial proteins (hsp70 and hsp22, respectively). Clues as to how the Arabidopsis sdh3 presequence was acquired are available. The next gene upstream of sdh3 on chromosome 5 of Arabidopsis is an intact copy of hsp70. Tandem duplications of genes along Arabidopsis chromosomes have been documented (e.g., Lin et al. 1999; Mayer et al. 1999); one example is a tandem duplicate of an hsp70-related gene located on chromosome 1. We propose that sdh3 became associated with a tandem duplicate of hsp70 on chromosome 5 by a recombinational or insertional event leading to functional activation (Figure 6). Subsequently, a region of almost 3 kb, containing sdh3 and a portion of hsp70, was apparently duplicated and translocated to chromosome 4. Upstream of sdh3 on chromosome 4 is a region of 605 bp with high similarity to the 3′ end of hsp70, but an entire hsp70 gene is not present. No other genes in the vicinity of sdh3 are shared on chromosomes 5 and 4, suggesting a duplication and translocation of ~3 kb that involved only the genes sdh3 and part of hsp70. This duplication/translocation appears to have been very recent because the sequences on each chromosome are 99% identical.

**Arabidopsis** sdh3 and cotton sdh3 add to the growing number of identified cases of presequence acquisition from preexisting genes for mitochondrial proteins. Mechanisms of presequence acquisition, when deduced, vary. One route to presequence acquisition is insertion into the host gene. Rps10 in carrot was inserted into the coding region of hsp22 and essentially parasitized the host gene (Adams et al. 2000), whereas rps14 in the grasses was inserted into an intron of sdh2 and the two genes are co-expressed by alternative splicing (Figure 5). SDH4 presequences from Arabidopsis and soybean cannot be for nonmitochondrial proteins. It is now apparent from the Arabidopsis genome sequence that the presequence acquisition from preexisting genes for mitochondrial proteins by unknown mechanisms. It is intriguing, but probably coincidental, that Arabidopsis and Fuchsia species acquired their presequences from other genes for mitochondrial proteins. Additional potential source of presequences is genes in the vicinity of sdh4 and part of hsp70. This duplication/translocation appears to have been very recent because the sequences on each chromosome are 99% identical.

**Figure 5.—Predicted mitochondrial presequences from angiosperm nuclear SDH4 proteins.** (Top) Unaligned SDH4 presequences. SDH4 presequences from Arabidopsis and soybean were force aligned using no gap penalties (middle) or low gap penalties (bottom).
gene is derived from the three terminal exons of cytosolic GapC, probably by exon shuffling (Braun et al. 1992; Wegener and Schmitz 1993; Long et al. 1996). It is hypothesized that the GapC-derived region serves as a targeting region (Braun et al. 1992; Long et al. 1996), although this has not been experimentally verified. A second copy of the gene for cytochrome c1, also present in potato (Braun et al. 1992), although the published sequence is incomplete. The sequence at the 5’ end of this gene has recently become available from an EST sequencing project (accession nos. BE922586 and AW906022), revealing a putative presequence that is homologous to the cytochrome c1 presequence from Arabidopsis and several other angiosperms. Thus, potato has two cytochrome c1 genes with different putative presequences, probably derived by gene duplication followed by acquisition of the GapC-derived region by one copy. Whether these genes have redundant functions or whether there is differential expression is not known, although the non-GapC copy is the major form in potato tubers (Braun et al. 1992).

Mitochondrial presequences vary considerably in primary sequence, but overall tertiary structural features, such as the ability to form an amphiphilic α-helix, appear to be conserved (reviewed in Glaser et al. 1998). The sequence flexibility of presequences is highlighted by the finding that ~2.5% of Escherichia coli clones generated in a shotgun screen exhibited mitochondrial targeting activity when added to a truncated yeast gene for cytochrome oxidase subunit 4 (Baker and Schatz 1987). Thus, gain of mitochondrial targeting ability by certain sequences de novo is not difficult to envision, at least for some mitochondrial proteins. The mitochondrial presequences of many recently transferred genes have no similarity to any genes currently in sequence databases. Considering that almost all of the genome of Arabidopsis has been sequenced (Arabidopsis Genome Initiative 2000), it is tempting to speculate that the presequences of unknown origin found in sdh4, rps10 (Wischmann and Schuster 1995), and rps14 (Figuerola et al. 1999b) are not derived from other genes, but instead acquired this function de novo upon association with genes transferred from the mitochondrion.

Finally, a few recently transferred rps10 genes have become activated in the nucleus without gaining a mitochondrial targeting presequence (Adams et al. 2000; N. Kubo et al. 2000), thereby bypassing a typical requirement for gene activation. It will be interesting to determine what regions of these proteins are responsible for targeting.

Multiple separate transfers of sdh3 and sdh4 to the nucleus during recent angiosperm evolution: The nuclear sdh3 genes from Arabidopsis, cotton, legumes, and grasses could all result from the same transfer to the nucleus, or they could be derived from as many as four separate transfers. The recent losses of sdh3 from the mitochondrion of the surveyed Brassicaceae, legumes (Fabaceae), and grasses (Poaceae), as revealed by our Southern blot hybridizations (Figure 3), and the presence of a mitochondrial sdh3 gene in cotton and the legume Gymnocladus suggest separate and recent transfers of this gene to the nucleus in each family. The nuclear sdh3 genes from the three rosid groups (Arabidopsis, cotton, and legumes) have presequences that are derived from different sources: hsp70, hsp22, and an unknown source, respectively. Thus, it is likely that these three presequences were acquired during separate gene activation events and probably also separate gene transfers. Given that nuclear sdh3 genes in the rosids are likely the result of separate and recent transfers, the gene transfer in grasses is probably separate too.

The losses of sdh4 from the mitochondrion of Arabi-
Adams and soybean appear to be very recent, as inferred by the Southern blot survey and mitochondrial sdh4 sequencing. Arabidopsis and soybean have presequences that appear to be nonhomologous (Figure 5); therefore these probably were acquired during separate gene activations. We hypothesize that the nuclear sdh4 genes from Arabidopsis and the legumes are the result of two separate transfers to the nucleus. The mitochondrial loss of sdh4 and its transfer to the nucleus in the grasses are also inferred to be recent, as judged by Southern hybridizations. If the sdh4 genes of Arabidopsis and soybean, both rosids, were indeed transferred independently of one another, then by extension sdh4 in the grasses is also probably the result of a separate transfer.

Although all of the available data point to seven separate transfers of sdh3 and sdh4 to the nucleus, a single transfer of each gene early in angiosperm evolution cannot be ruled out. The many recent losses of sdh3 and sdh4 from the mitochondrion are more difficult to explain by a single transfer. In contrast, we have detected relatively ancient losses (early during eudicot evolution) of the mitochondrial rps2 and rps11 genes (K. L. Adams, Y.-L. Qiu and J. D. Palmer, unpublished data) that may have accompanied a single transfer of each of these two genes to the nucleus. To explain the unique mitochondrial presequences by a single transfer and activation, at least three scenarios can be envisioned. First, following a single common and relatively ancient transfer and presequence acquisition, the current presequences might have been derived recently by genomic rearrangements or reinsertion of cDNAs. Such frequent presequence “switching” seems quite unlikely, however, considering that presequence switching is virtually unprecedented among the many characterized anciently transferred genes (Glaser et al. 1998). The only documented example of presequence switching involves one copy of cytochrome c1 in potato (see previous section). Second, the nonhomologous presequences could be derived by separate gene activations occurring tens of millions of years after a single common gene transfer. Although little is known about the relative timing of gene transfer and activation, in the case of cox2 transfer in legumes, transfer and activation were approximately coupled in time (Adams et al. 1999). It is also difficult to imagine that a gene would remain inactive in the nucleus for such a long period of time without becoming a pseudogene or acquiring mutations detrimental to protein function. Third, relatively soon after a single common transfer, there might have been multiple gene duplications preceding activation, with each copy activated separately using a presequence from a different source. The case of rps11 in rice provides precedent for gene transfer followed by duplication and acquisition of a different presequence by the two duplicates (Kadowaki et al. 1996). Considering only those plants with characterized nuclear sdh3 genes, each of four descendant lineages (legumes, Brassicaceae, grasses, and cot-
cluding factors that depress or elevate the incidence of transfer of different genes in diverse lineages.

What might account for a high rate of transfer of sdh3 and sdh4 in angiosperms relative to other respiratory genes? Some highly hydrophobic membrane proteins are difficult to import into organelles and to insert correctly into the appropriate membrane (Popot and de Vitry 1990; Claros et al. 1995); this characteristic has been invoked as a potential limitation to the successful functional transfer of such respiratory genes to the nucleus (Borst 1977; Thornsness and Weber 1996; Palmer 1997; Doolittle 1998; Lang et al. 1999; McFadden 1999; Palmer et al. 2000). In addition it has been hypothesized that hydrophobic mitochondrial proteins might be misrouted to the endoplasmic reticulum if synthesized in the cytosol (von Heijne 1987). Reduction in hydrophobicity has been suggested to have facilitated transfer of the cox3 gene to the nucleus in two green algae (Pérez-Martínez et al. 2000). Functional transfer of sdh3 in angiosperms might have been facilitated by loss of one of three transmembrane segments in a common ancestor of eudicots and monocots, along with reductions in hydrophobicity of the proteins encoded by some transferred sdh3 genes. By making SDH3 proteins less hydrophobic, these changes might make it easier to import into mitochondria and sort to the correct location. In this regard, it is noteworthy that in angiosperms sdh3 has been lost from the mitochondrion, and presumably transferred to the nucleus, more often than sdh4, for which all examined angiosperm sequences retain all three transmembrane segments.

We thank various EST sequencing projects, the Monsanto (a unit of Pharmacia) rice-research.org program, and the Arabidopsis genome sequencing project for providing genomic or EST data for this project. K.L.A. was supported by Floyd and Ogg fellowships from Indiana University and M.R. received a postdoctoral fellowship from DGAPA (Dirección General de Asuntos del Personal Académico). Funding for this work came from National Institutes of Health grant GM-53087 to J.D.P.

LITERATURE CITED


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


Grohmann, L., A. Brennicke and W. Schuster, 1992 The mitochondrial gene encoding ribosomal protein S12 has been transferred to the nuclear genome in Oenothera. Nucleic Acids Res. 20: 5641–5646.


Kobayashi, Y., V. Knopf, H. Fukuzawa, A. Brennicke and K. Ohya, 1997 Interorganellar gene transfer in bryophytes: the func-


Communicating editor: K. J. Newton