J Mol Evol (1995) 41:397–406

JOURNAL OF MOLECULAR EVOLUTION

Pervasive Migration of Organellar DNA to the Nucleus in Plants

Jeffrey L. Blanchard, Gregory W. Schmidt

Department of Botany, University of Georgia, Athens, GA 30602, USA

Received: 30 September 1994 / Accepted: 5 January 1995

Abstract. A surprisingly large number of plant nuclear DNA sequences inferred to be remnants of chloroplast and mitochondrial DNA migration events were detected through computer-assisted database searches. Nineteen independent organellar DNA insertions, with a median size of 117 bp (range of 38 to >785 bp), occur in the proximity of 15 nuclear genes. One fragment appears to have been passed through a RNA intermediate, based on the presence of an edited version of the mitochondrial gene in the nucleus. Tandemly arranged fragments from disparate regions of organellar genomes and from different organellar genomes indicate that the fragments joined together from an intracellular pool of RNA and/or DNA before they integrated into the nuclear genome. Comparisons of integrated sequences to genes lacking the insertions, as well as the occurrence of coligated fragments, support a model of random integration by end joining. All transferred sequences were found in noncoding regions, but the positioning of organellar-derived DNA in introns, as well as regions 5′ and 3′ to nuclear genes, suggests that the random integration of organellar DNA has the potential to influence gene expression patterns. A semiquantitative estimate was performed on the amount of organellar DNA being transferred and assimilated into the nucleus. Based on this database survey, we estimate that 3–7% of the plant nuclear genomic sequence files contain organellar-derived DNA. The timing and the magnitude of genetic flux to the nuclear genome suggest that random integration is a substantial and ongoing process for creating sequence variation.

Key words: Chloroplast — Mitochondrion — Nucleus — Gene transfer — DNA migration — mRNA editing — Illegitimate recombination — Nonhomologous recombination — Random integration — End joining — Genomic evolution

Correspondence to: J.L. Blanchard

Introduction

The endosymbiotic origins of chloroplast and mitochondria have become a central concept of cellular evolution (Margulis 1993). The reduced coding capacity of the genomes of these semiautonomous organelles relative to eubacterial genomes reflects the loss of some genes and the transfer of other genes from the endosymbiont to the host nucleus. (See Palmer 1991; and Gray 1992, for reviews.) The transfer of genes is an ongoing evolutionary process, as evidenced by variations in the coding capacity of organellar genomes and instances where some chloroplast and mitochondrial genes have moved relatively recently to the nuclear genome (van den Boogaart et al. 1982; Baldauf and Palmer 1990; Nugent and Palmer 1991; Gantt et al. 1991; Grohmann et al. 1992).

Based on DNA hybridization studies, nuclei of animals and insects appear to contain several hundred copies of mitochondrial DNA sequences coexisting in mitochondria (Gellissen and Michaelis 1987, and reference therein). Similarly, plant nuclei harbor many copies of chloroplast DNA (Timmis and Scott 1983). These copies of organellar DNA have been found within introns (Pichersky and Tanksley 1988) and flanking regions of nuclear genes (Bernatzky et al. 1989; Knoop and Brennicke 1991; Hua et al. 1993; Knoop and Brennicke 1994) or as transcribed nuclear pseudogenes (Koch et al. 1989; Shay et al. 1991).

The movement of organellar DNA has been suggested to occur through RNA intermediates (Schuster and Bren-

nicke 1987). This hypothesis is supported by two instances in which the nuclear sequence appears to have been derived from the reverse transcription of a mitochondrial RNA that has undergone cytosine-to-uracil editing (Nugent and Palmer 1991; Grohmann et al. 1992). However, regions of organellar DNA ranging in size to over 18 kb (Ayliffe and Timmis 1992a) which contain introns and flanking sequences also have been identified in the nucleus (Cheung and Scott 1989; Pichersky et al. 1991; Knoop and Brennicke 1991; Ayliffe and Timmis 1992a,b; Sun and Callis 1993). Hence, DNA fragments escaping from the organelle may be directly assimilated into the nucleus without passage through a RNA intermediate.

Recently, mutants of *Saccharomyces cerevisiae* with elevated rates of mitochondrial DNA transfer to the nucleus have been isolated (Thorsness and Fox 1993). Phenotypic characterization indicates that some of these mutations are in proteins which may play a role in mitochondrial division, perhaps causing membranes to leak nucleic acids (Thorsness et al. 1993). This finding supports a hypothesis that the transfer of mitochondrial DNA is triggered by stress conditions which lead to a disruption of the mitochondrial membrane (Hadler et al. 1983). There have also been suggestions that transfer may be facilitated by specific membrane transporters such as that postulated for import of tRNAs into plant mitochondria (Schuster and Brennicke 1987).

Analysis of the sequences surrounding organellar DNA insertion sites has lead to various proposals implicating the involvement of homologous recombination (Pichersky 1991), transposable elements or short dispersed elements (Gellissen et al. 1983; Farrelly and Butow 1983; Ossario et al. 1991; Zullo et al. 1991), and nonhomologous or illegitimate recombination (Sun and Callis 1993). Since no consistent pattern has been found near the sequence junctions, it is not apparent whether any or all of these mechanisms are utilized.

We have designed a search strategy using the GenBank and EMBL databases to locate chloroplast and mitochondrial DNA sequences which have transposed into the nuclear genome of plants. This larger sampling of genomic sequences provides a broader perspective on the migration phenomenon. The data presented here indicate an extensive infiltration of organellar DNA into plant nuclear genomes. We detected 19 probable examples of organellar DNA migration. Fifteen of these regions were located in nuclear genes. The implications of these data for mechanisms involved in DNA transfer and for the influence of organellar DNA on genomic evolution are discussed.

Materials and Methods

Computer Resources. Computer programs were accessed through the BioSciences Computational Resource at the University of Georgia. The GenBank (version 82.0) and EMBL (version 37.0) databases were queried with the algorithms in BLASTN (National Center for Biotechnology, National Institutes of Health, Bethesda, MD; Altschul et al. 1990) and FASTA (University of Wisconsin, Madison, Genetics computer Group (GCG) version 7.0; Deveraux et al. 1984) search programs. The percent identity between the nuclear and organellar sequences was calculated from pairwise alignment using FASTA. Sequence alignments were conducted using PILEUP (GCG) and GENALIGN (IntelliGenetics, Mountain View, CA). Parsimony analysis was conducted using the programs contained in PAUP version 3.1.1 (Swofford 1993).

Identification of Chloroplast- and Mitochondrial-Derived DNA in Nuclear Sequences. The rice and tobacco chloroplast genomes and plant mitochondrial sequences from petunia, tobacco, wheat, and primrose were retrieved from the GenBank and EMBL DNA sequence libraries. The chloroplast genomes were subsequently divided into 10-kb files with 100-nucleotide overlaps. The catalogue of sequence files was then sequentially used to probe GenBank and EMBL libraries using the algorithms in BLASTN and FASTA. Sequences were collected which met the following selection criteria: (1) lengths greater than 45 nucleotides and (2) at least 80% identity to chloroplast or mitochondrial DNA when between 50 and 100 nucleotides and at least 70% identity when greater than 100 nucleotides. Segments with long stretches of adenine and/or thymine were not reported nor were cDNA sequences due to the frequent formation of chimeric constructs during the cloning process. These cutoff values generally resulted in probabilities of finding a sequence because of random identity of greater than $1/10^{10}$ using BLASTN. The validity of these cutoff values was further supported by queries of GenBank and EMBL with the tobacco and rice chloroplast genomes reversed to read in the 3′ and 5′ direction. These negative controls did not identify any random matches above the cutoff values.

Results

How Pervasive Is Intercompartmental DNA Migration?

Computer searches using the assembled catalogue of chloroplast and mitochondrial sequences detected 28 previously unidentified regions of chloroplast and mitochondrial DNA in nuclear sequences from nine different plant genera (Table 1). The insertions vary from 38 bp to over 785 bp with a median value of 117 bp and are derived from disparate regions of the organellar genomes. Many of the organellar DNA insertions are located in the vicinity of nuclear genes (Fig. 1). The insertions occur in introns, a 3′ transcribed region, and near-transcribed regions. Some sequences listed in Table 1 and depicted in Fig. 1 were previously described or reported and are included in our tabulation to emphasize the scope of organellar DNA transfer into the nucleus. Three of the fragments are found near previously described cases of DNA migration including a second mitochondrial DNA fragment adjacent to the *nad1* fragment in the *Arabidopsis* polyubiquitin gene (Sun and Callis 1993). The other two mitochondrial sequences are found within five other organellar DNA insertions reported by Fukuchi et al. (1991) and are included in Table 1 because this region of the rice genome appears to be completely composed of organellar or plasmid DNA in-

Table 1. Plant nuclear DNA sequences similar to chloroplast or mitochondrial DNA

Plant species[a]	Sequence description[b]	Locus	Position[c]	Size	% Identity	Origin of segment[d]
*A. thaliana**	Enhancer region	AT127ENH	1–233[e]	>234	92[f]	cp rpoB
*A. thaliana**	Enhancer region	AT12QENH	1–223[e]	>234	72[f]	cp ndhK-ndhJ
*A. thaliana**	Acyl carrier protein	ATACPA1	1512–1590	79	85[f]	cp ndhl
A. thaliana	Polyubiquitin (1)	ATHPOLYUBQ	1133–1729 (≠1)	596	92[f]	mt nad1 + intron
*A. thaliana**	Polyubiquitin	ATHPOLYUBQ	1735–1868 (≠2)	134	71[f]	mt 5′ of trnfM
*B. vulgaris**	Satellite DNA	BPSATDNAB	1–119 (≠1)[e]	>119	70[f]	cp accD
*B. vulgaris**	Satellite DNA	BPSATDNAB	120–243 (≠2)	124	90[f]	cp ndhB intron
*E. gunnii**	Cinnamyl-alcohol dehydrogenase	EGCAD	2805–2885	81	95[f]	cp ycf2
D. biflourus	Seed lectin (2)	DBISLEC	761–881	123	87	mt nad5 intron
*G. hirsutum**	Late embryogenesis abundant	GHLEA34	214–292	79	84[f]	cp ycf2
L. gibba	Negatively phytochrome regulated (3)	LGNPR2	1882–2065	184	79[f]	cp rpoB
*L. esculentum**	Phytoene desaturase	LEPDSG	5362–5431	70	86[f]	cp ycf1
*L. esculentum**	Threonine deaminase	TOMILV1B	4183–4239	57	81[f]	cp ycf2
*L. esculentum**	Abscisic acid (ABA) regulated	TOMLE25	249–465	117	71[f]	cp rpl23
*L. esculentum**	Prosystemin	TOMPRO	3060–3176	117	70	mt coxII intron
*L. esculentum**	Rubisco small subunit-3a	S44160	203–422	220	76[f]	cp rpoC1
*L. esculentum**	Phytoene synthase	LEGTOM5	2916–3055 (≠1)	140	79[f]	cp ycf1
*L. esculentum**	Phytoene synthase	LEGTOM5	3133–3321 (≠2)	189	79[f]	cp ycf1
L. esculentum	Chlorophyll a/b binding protein-7 (4)	LECAB7	2426–2521 ≠1)	96	86[f]	cp ndhK
L. esculentum	Chlorophyll a/b binding protein-7 (4)	LECAB7	2818–2939 (≠2)	122	89[f]	cp ndhK
M. crystallinum	Phosphoenol pyruvate (PEP) carboxylase (5)	MCPPC1A	1–620[e]	>621	90[f]	cp psaB
N. alata	Self-incompatibility protein-s2 (6)	NAU08860	1105–1160	56	96	mt unknown
N. tabacum	Oxygen evolving complex-2A (7)	NTOEE2AG	1–97[e]	>97	96	mt nad1 intron
N. tabacum	Basic beta-1,3-glucanse (8)	TOBGLA13B	4286–4359	74	84	mt nad1 intron
N. tabacum	Basic beta-1,3-glucanse (8)	TOBGLB13B	4507–4579	73	88	mt nad1 intron
N. tabacum	Basic beta-1,3-glucanse (8)	TOBGLUCA	4374–4448	74	85	mt nad1 intron
N. tabacum	Acidic beta-1,3-glucanse (8)	TOBGLUCB	1918–2103	186	72	mt nad1 intron
*O. sativa**	RNA polymerase pseudogene	RICRPSBP	1–785[e] (≠1)	>786	93	mt ctDNA in mt
O. sativa	RNA polymerase pseudogene	RICRPBSP	2115–2413 (≠2)	299	80	cp rpoC2
O. sativa	Homologous to B4 plasmid (9)	OSB4P411	148–195 (≠1)	44	100	mt plasmid B4
O. sativa	Homologous to B4 plasmid (9)	OSB4P411	245–337 (≠2)	83	100	mt coxII
*O. sativa**	Homologous to B4 plasmid	OSB4P411	335–411 (≠3)	77	99	mt plasmid B2
*O. sativa**	Homologous to B4 plasmid	OSB4P411	412–458 (≠4)	47	100	mt nad7 intron
O. sativa	Homologous to B4 plasmid (9)	OSB4P411	456–501 (≠5)	46	98	mt plasmid B4
O. sativa	Homologous to B4 plasmid (9)	OSB4P411	560–600 (≠6)	41	100	mt nad1 intron
O. sativa	Homologous to B4 plasmid (9)	OSB4P411	597–635 (≠7)	39	92	cp 5′ of psaA
O. sativa	Homologous to B4 plasmid (9)	OSB4P411	636–666 (≠8)	31	100	mt cyt b
O. sativa	Homologous to B4 plasmid (9)	OSB4P411	660–740 (≠9)	81	94	mt plasmid B4
*O. sativa**	Homologous to B4 plasmid	OSB4P411	740–777 (≠10)	38	100	mt nad2
O. sativa	Homologous to B4 plasmid (9)	OSB4P411	852–928 (≠11)	77	100	mt rps 12 intron
*P. hydrida**	Chalcone synthase	PHCHSB	3098–3186	89	81	mt atp9
*P. inflata**	Self-incompatibility protein-s3	PETS3ALLE	303–360	53	96	mt 5′ of atp9
*S. comersonii**	Osmotinlike protein	X72927	2109–2243	135	74[f]	cp atpA
*S. tuberosum**	Cytochrome C1	S66866	2614–2816	203	82[f]	cp ycf3 intron
*S. tuberosum**	Patatin	POTPATA	3434–3508	75	80[f]	cp rpL2 intron
*S. tuberosum**	Patatin	STPATG1	5083–5157	75	<80[f]	cp rpl2 intron
*S. tuberosum**	Patatin pseudogene	STPATP1	6339–6423	85	<80[f]	cp rpl2 intron
*S. tuberosum**	Patatin pseudogene	STPATP2	5624–5703	80	<80[f]	cp rpl2 intron
*S. tuberosum**	Patatin	STPATG	3925–4147[g]	80	<80[f]	cp rpl2 intron
*Z. mays**	Responsive to ABA	ZMRAB17G	1654–1719 (≠1)	138	100	cp ndhK
*Z. mays**	Responsive to ABA	ZMRAB17G	1791–1833 (≠2)	43	100	cp ndhK
*Z. mays**	Responsive to ABA	ZMRAB17G	1829–1874 (≠3)	46	100	cp ndhK
*Z. mays**	Responsive to ABA	ZMRAB17G	1873–1973 (≠4)[e]	>114	98	cp rpoC2

[a] Plant species (common name): *Arabidopsis thaliana* (arabidopsis); *Beta vulgaris* (sugar beet); *Eucalyptus gunnii* (eucalyptus); *Dolichos biflourus* (bean); *Gossypium hirsutum* (cotton); *Lemna gibba* (duckweed); *Lycopersicon esculentum* (tomato); *Mesembryanthemum crystallinum* (ice plant); *Nicotiana tabacum* (tobacco); *Oryza sativa* (rice); *Petunia hybrida* (petunia); *Petunia inflata* (petunia); *Solanum comersonii* (potato); *Solanum tuberosum* (potato); *Zea mays* (corn). Asterisks indicate previously unreported sequences similar to organellar DNA identified in this study.

[b] References noting similarity to organellar DNA: (1) Sun and Callis 1993; (2) Knoop and Brennicke 1991; (3) Okubara et al. 1993; (4) Pichersky and Tanksley 1988; (5) Cushman and Bonhert 1992; (6) Bernatzky et al. 1989; (7) Hua et al. 1993; (8) Knoop and Brennicke 1994; (9) Fukuchi et al. 1991

[c] According to the author's numbering in the sequence file

[d] cp, chloroplast; mt, mitochondrial; rpo, RNA polymerase; ndh, NADH dehydrogenase; ycf, open reading frame; cox, cytochrome oxidase; psa, photosystem I; nad, NADH dehydrogenase; atp, Fo-F1 ATPase complex; rpl, large-subunit ribosomal protein; rps, small riposomal subunit; cyt b, cytochrome b; acc, acetyl-CoA carboxylase

[e] The length of the insertion may be longer since the similarity terminates at the end of the sequence reported to GenBank or EMBL.

[f] The % identity was determined by comparison with tobacco cpDNA or the closest available relative, since the cpDNA of this plant has not been sequenced

[g] This region has been split by a subsequent insertion event from 3965–4119

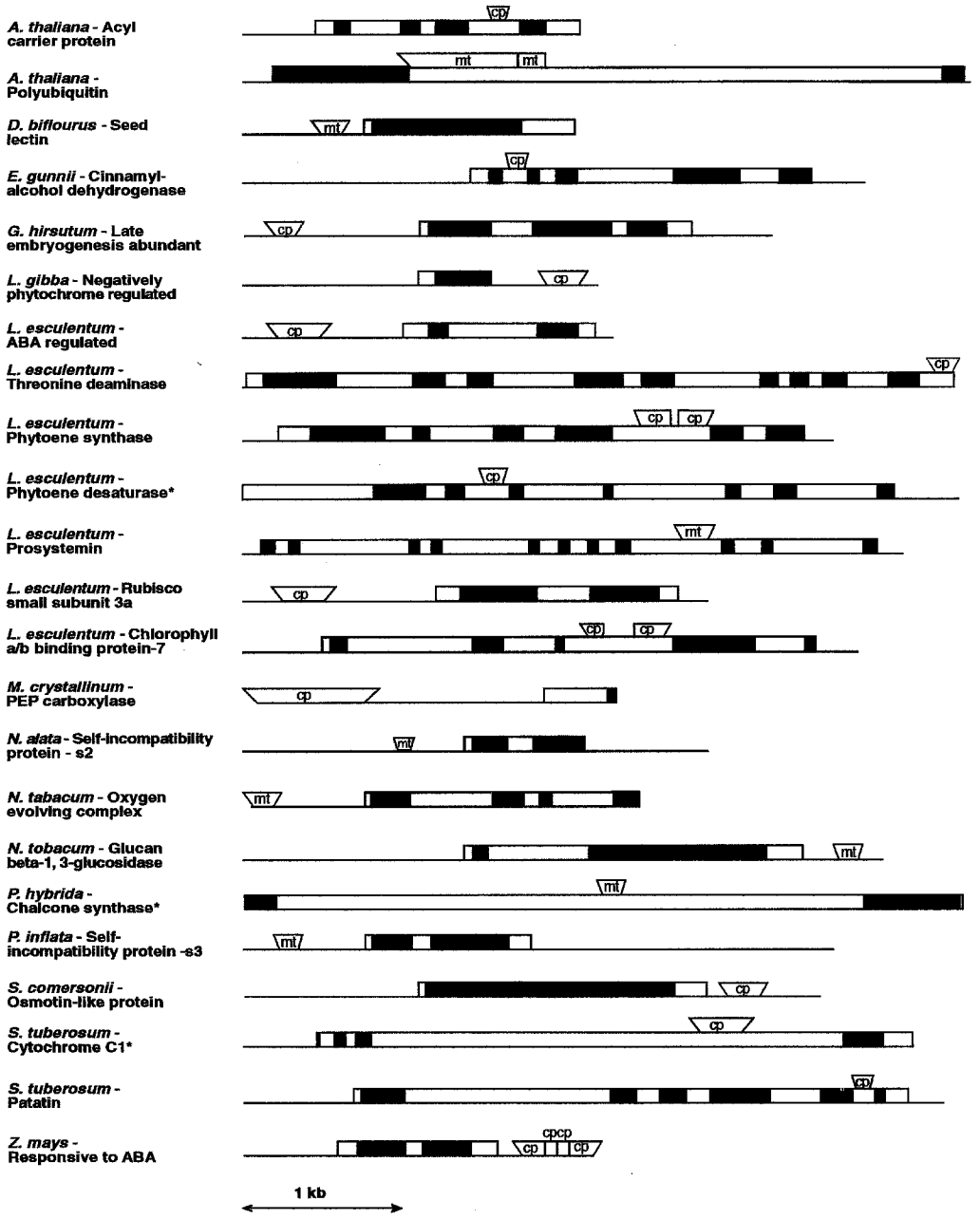


Fig. 1. Location of the organellar DNA insertion with respect to the nuclear genes. Nontranscribed regions are represented by a *line*. Transcribed regions are represented by a *box*, with *clear boxes* indicating untranslated regions and introns and *block boxes* indicating protein coding regions. The cp- and mt-labeled trapezoids represent chloroplast- and mitochondrial-derived DNA. Only one representative each from patatin and glucan β-1,3-glucosidase gene is shown since the insertions are found in the same position in each member. *Asterisks* indicate that only portions of these genes are shown. Transcription start and stop, sites as well as the coding regions, are derived from the sequences files or from the literature cited in the sequence files.

sertions. A fragment of the mitochondrial B2 plasmid was also detected in this file, bringing the total number of chloroplast, mitochondrial, and plasmid fragments in this 971-bp sequence to 11, and there may be additional fragments from regions of the mitochondrial genome which have not been sequenced. Additional points of note in Table 1: other mitochondrial plasmid sequences are present in the nuclear genome of rice (Fukuchi et al. 1991) but are not included in the table; the sequence in BPSATDNAB is in the reverse orientation; similarity between mitochondrial DNA and NTOEE2AG extends 17 bp beyond what is reported in the database entry (Hua et al. 1993); and a region of similarity at the 3′ end of the tobacco beta-1,3-glucanase gene is extended beyond that reported by Knoop and Brennicke (1994). Functional nuclear genes which were previously organellar encoded are not included in Table 1 but should also be considered to be examples of organellar DNA transfer (Baldauf and Palmer 1990; Nugent and Palmer 1991; Gantt et al. 1991; Grohmann et al. 1992).

(a)

```
nu s3 allele,Pet    TTGAACATCT CAATGTCCAA GATAAAAAGA ACGAGCGGAA GAATGGACGA GGC
mt 5' of atp9,Pet   .......... .......... .......... ....C..... ...G..... ...
mt 5' of atp9,Tob   .......... .......... .......... ....C..... ...G..... ...
mt 5' of atp9,Rap   A.TCCA.GTG AG........ ...T...G.. .....G.T.. ....C..... G..
mt 5' of atp9,Pri   A.TCCA.GTG AG...C.... ..G....G.. .....G.A.. .....A.GA. TC.
mt 5' of atp9,Bea   A.TT.A.GTG AG........ ..G....G.. .....GTTTT ...GC..... .AA
```

(b)

```
nu rbcs3a,Tom   GATTTCGTGT GA--TTGCTT GGTAAATAAG TTGATTATTT CGAACGTTCT
cp rpoC1,Tob    .........A ..CTC..... ..C...CG.. .C.......C A.G.......
cp rpoC1,Spi    .........A ..CT..A... ......CGG. .C.......C ..G......C
cp rpoC1,Mai    .G.....C.A ..CTC..... ..G...CGG. .C.....C.C G.GG......
cp rpoC1,Ric    .G.....C.A A.CTC..... ..G...CGG. .C.....C.C G.GG......
cp rpoC1,Liv    .A.......A A.AT..A... ..A...AGG. .........C A.GTA.A...

nu rbcs3a,Tom   GTTATTATCG TTGATTCTTG GTTT--ATTA CACCAATGTG GATTGCTACG
cp rpoC1,Tob    ..C...G... .G.G.C...C AC..TC.... ..T.G..... ......CT..
cp rpoC1,Spi    ..C...G... .A.GCC...C AC..TC.... ....G..... ......CT..
cp rpoC1,Mai    ..C...G.T. .G.G.C...C .C..TC.... ..T....... ....A.CT.T
cp rpoC1,Ric    ..C...G.T. .G.G.C...C AC..TC.... ..T....... .G..A.CT.T
cp rpoC1,Liv    ......G.A. .A.GCC...T TC.ACC...G ..T....... ....A.CT..

nu rbcs3a,Tom   TGACATAGCG GTAAAACTTT TTCGTACATT TGTAATTCGT -ATCTAATTA
cp rpoC1,Tob    ...A.....A A..G...... .C.AG..... .......... GG........
cp rpoC1,Spi    ...A..T..A A..G.G..C. .C.AG..... .......... GG........
cp rpoC1,Mai    A..G.....A A....G.... ...AGCT... .........C G..T....C.
cp rpoC1,Ric    A..G.....A A....G.... ...AGCT... ....\....C G..T....C.
cp rpoC1,Liv    C..A..G..A A..G...... ...AAG.... ...T.....A GG...T...G

nu rbcs3a,Tom   GACAACATCA ATCTT-GCTT CTTGGGGTCG CTAAGAGAAA AATTCGAGAA
cp rpoC1,Tob    .........T TG...C.AAC A.A..A..T. .......T.. ..........
cp rpoC1,Spi    .G.......T TG...C.AAC A.A..A..T. .....C.T.. ......G...
cp rpoC1,Mai    CGA...G.GC TA...CTAA. G.CA..A.T. ....A..G.. ....T.G...
cp rpoC1,Ric    CGA...G.GC TA...CTAA. G..A..A.T. ....A..G.. ....T.G...
cp rpoC1,Liv    ...G.A..TT TG..CCTAA. C.GA.A.CA. .G..A.CT.T G....A.A.T

nu rbcs3a,Tom   AAAAAATCCA TTATACAGGA AACTAC
cp rpoC1,Tob    ...G.GC.G. ..G.TTG... ..-.C.
cp rpoC1,Spi    ...G.GC.G. ..G..TG.A. ..-...
cp rpoC1,Mai    ..GG..C... ..G..TG... ..-...
cp rpoC1,Ric    ..GG..C... ..G..TG... ..-...
cp rpoC1,Liv    ...G..C.T. ....TTG.A. ..-...
```

Fig. 2. Nucleotide sequence alignments of nuclear regions identified through computer searches with organellar DNA. **a** Alignment of a region 5′ of the nuclear *Petunia inflata* self-incompatibility (s3) protein (PETS3ALLE) with mitochondrial DNA. **b** Alignment of a region 5′ of the nuclear *Lycopersicon esculentum* Rubisco small-subunit protein (S44160) with chloroplast DNA. *Periods* indicate identical nucleotides and gaps in the alignment are indicated by *dashes. Abbreviations:* nuclear (*nu*), mitochondrial (*mt*), chloroplast (*cp*), *Petunia* (*Pet*), tobacco (*Tob*), spinach (*Spi*), maize (*Mai*), rice (*Ric*), liverwort (*Liv*), bean (*Bea*), rapeseed (*Rap*), primrose (*Pri*), RNA polymerase (*rpo*), F0-F1 ATPase complex (*atp*). GenBank accession numbers for chloroplast and mitochondrial DNA sequences: *rpoC1,* Tob (Z00044); *rpoC1,* Spi (M55297); *rpoC1,* Mai (X17318); *rpoC1,* Ric (X15901); *rpoC1,* Liv (X04465); 5′ of *apt9,* Pet (X05808); 5′ of *atp9,* Tob (X04019); 5′ of *atp9,* Bea (X072736); 5′ of *apt9,* Rap (D13696); 5′ of *atp9,* Pri (X15765).

Most of the sequences exhibit additional characteristics in support of the hypothesis that they are derived from the intercompartmental transfer of DNA. When a nuclear DNA sequence is aligned with chloroplast or mitochondrial DNA, the nuclear sequence is always most similar to organellar DNA from its closest relatives as in the two examples shown in Fig. 2. Hence, a region of the nuclear *Petunia* s3 allele is most similar to *Petunia* mtDNA and less similar to mitochondrial DNA from more distantly related plants (Fig. 2a). Also, a portion of the tomato Rubisco small subunit gene is most similar to cpDNA of tobacco (which is in the same family as tomato) and less similar to more distantly related plants (Fig. 2b). From these sequence alignments it is possible to identify nucleotides which may have changed after the organellar DNA transposed into the nucleus and nucleotide insertion and deletion events which disrupt the original organellar reading frame. The existence of such sequence alterations supports the hypothesis that the sequences are the result of former DNA migration events and that they are not the result of cloning artifacts.

With these 19 newly identified transfer events and the ten other insertions which have been identified serendipitously (see Table 1) there currently are 29 independent transfer events included in 35 different sequence database entries. The directed studies of Cheung and Scott (1989), Pichersky et al. (1991), and Ayliffe and Timmis (1992a,b), in which genomic clones were identified by screening with chloroplast DNA probes, are not included in this figure because they would bias the random assessment of the sequence files. To estimate the quantity of plant nuclear genomic sequences we inventoried the GenBank and EMBL libraries. The majority of plant sequences are from cDNAs or bona fide organellar DNA with only 500 to 1,000 plant genomic sequence entries present. This is in line with counts of monocot genomic sequences (Bureau and Wessler 1994). Based on our estimate, 3–7% of the plant genomic sequence entries contain organellar-derived DNA.

Intercompartmental DNA Transfer Through an RNA Intermediate

Of the fragments listed in Table 1, only the short fragment of a mitochondrial ATPase gene located in a *Petunia hybrida* chalcone synthase intron contains RNA editing sites. All four RNA editing sites in the transferred mitochondrial DNA have been converted from cytosine to uracil (Fig. 3), indicating the transfer pathway involved RNA. The transfer appears to have occurred after the expansion of the chalcone synthase gene family in

```
Petunia nuDNA     TAT GCT ATT TTG GGAT TTT GCT TTA ACC GAA ACT ATT GCC TTG
Petunia mtDNA     ... ..C ... ... ..-C ... ... C.. ... ... G.. ... ..A .C.
Petunia mtRNA     ... ..C ... ... ..-C ... ... C.. ... ... G.. ... ..A ...
mt atp9 protein   Tyr Ala Ile Leu Gly  Phe Ala Leu Thr Glu Ala Ile Ala Leu

Petunia nuDNA     TTT TCC TTA ATG ATG GTC TTC TTA ATC CTC CTT GTC TTT TAG
Petunia mtDNA     ... G.. CC. ... ... .C. ..T ..G ... TCA T.C ..A ..C C.A
Petunia mtRNA     ... G.. C.. ... ... .C. ..T ..G ... T.A T.C ..A ..C ..A
mt atp9 protein   Phe Ala Leu Met Met Ala Phe Leu Ile Leu Phe Val Phe STOP
```

Fig. 3. A sequence alignment of a portion of a nuclear *Petunia* chalcone synthase gene (PHCHSB) with the *Petunia* mitochondrial *atp9* gene (X05808) and edited *Petunia* mitochondrial *atp9* mRNA (mtRNA; Wintz and Hanson 1991). The uracil bases in the mtRNA sequence have been replaced by thymine (*T*) for graphic clarity. The corresponding region of the mitochondrial (*mt*) *atp9* protein is shown for reference. The *boxed* cytosine (*C*) residues in the mtDNA sequence undergo editing in the mRNA transcripts to uracil resulting in three amino acid changes and the creation of a stop codon.

Petunia, since many other plant genera only contain one chalcone synthase gene (Dangl et al. 1989) and because the mitochondrial sequence is found in only one of eight *Petunia* chalcone synthase genes (Koes et al. 1989).

Random Integration by End Joining

A chloroplast DNA fragment was identified in an *Eucalyptus gunnii* cinnamyl-alcohol dehydrogenase gene but is not found in the *Eucalyptus boitryoides* gene (Fig. 4a). Eight of nine nucleotides at the 5′ end of the chloroplast fragment are identical to the *E. boitryoides* gene. A shorter region of overlap is also evident between a chloroplast fragment found in only one of two tandemly duplicated acyl carrier protein genes (Fig. 4b). If the *acpa2* gene is inferred to represent the ancestral gene arrangement, then the insertion appears to have accompanied the deletion of ~22 bp.

Because of suggestions that transposable elements may facilitate the movement of organellar DNA (Gellissen et al. 1983; Ossario et al. 1991; Zullo et al. 1991), we translated the boundaries of the insertions in all six reading frames and used these amino acid sequences to query the database. We could not detect significant similarity to any known transposable elements. We also could not find any other patterns in or near the nuclear/organellar DNA junctions that would suggest a mechanism by which the organellar DNA was fragmented or assimilated into the nucleus.

DNA Migration from the Chloroplast to Mitochondria Then to Nuclei

A mitochondrial DNA segment is found near a pseudogene for RNA polymerase (Fig. 5a). While the RNA polymerase pseudogene appears to originate from chloroplast DNA, the mitochondrial fragment appears to be derived from a region of the rice mitochondrial genome which contains a segment of chloroplast DNA (Nakazono and Hirai 1993). The probable sequence of events involved the integration of chloroplast DNA into the mitochondrial genome (integration event #1) followed by the migration of a mitochondrial fragment containing a portion of the chloroplast DNA segment into the nuclear genome near the chloroplast RNA polymerase–derived fragment (integration event #2). Alignment of the nuclear sequence with the chloroplast and mitochondrial fragments (Fig. 5b) shows that the nuclear and mitochondrial sequences share a deletion and numerous nucleotides which are not present in the chloroplast sequence. The mitochondrial origin of this sequence is further supported by the continued similarity to mitochondrial DNA upstream of the site of the chloroplast DNA insertion into the mitochondrial genome and by ingroup analysis using PAUP (data not shown).

Timing of the Transfer Events

In most instances, it is not possible to determine if the organellar-derived DNA fragments are the result of primary insertions or are derived from fragmentation or duplication of previous insertions. For example, six nuclear sequences were identified which have multiple regions similar to cpDNA and mtDNA. A region 3′ of the maize *rab17* gene harbors three tandemly arranged fragments derived from the chloroplast *ndhK* gene and one fragment derived from *rpoC2* (Fig. 6). This organization could have arisen from the concatenation of cpDNA or by tandem duplication of an original *ndhK* insertion. Because the fragments are joined end to end, the occurrence of successive insertions within a short time span is unlikely. The multiple fragments in the *Arabidopsis* polyubiquitin gene, in the beet satellite DNA, and in the "homologous to B4" section of rice nuclear DNA are from disparate regions of organellar genomes or from different genomes and appear to have undergone ligation prior to integration. In the tomato phytoene synthase gene, it appears that a single insertion event was followed by duplication of the chloroplast sequence with the 5′ stretch of nuclear DNA, resulting in the present structure found in the phytoene synthase intron (Fig. 7). The chloroplast- and mitochondrial-derived fragments described in Fig. 6 are separated by 1.3 kb, and it is not clear if they are derived from separate insertion events or were integrated into the nuclear genome along with the 1.3-kb region.

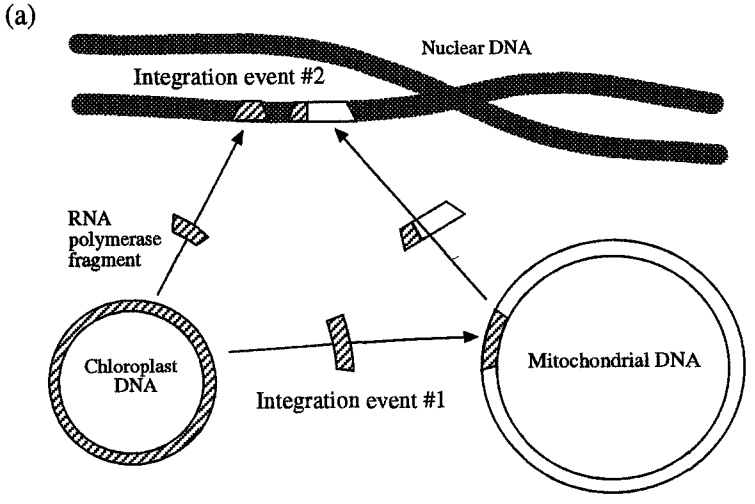
```
(a) egcad    TCTGTTTTTCTTGTTCgATCAATTACGCATTAATCAATATTCGATTATTGGTCTGAGGTTAT
    egccad   ..............a...----------------------------------------

    egcad    CAACAAAAAAGATTTGATCTAAGTCACTTCCAACAAATGC
    egccad   ----------------------------..........

(b) atacpa1  TGTTACT--GATACAggcaaatatataggacatactcaaacacatacttcacaagcaatgca
    atacpa2  .......ga......----------------------------------------------

    atacpa1  tttatcaaaTtaaatTcAaaaTgGaTtcgacgaacgTgTGcACaTACTCATG
    atacpa2  ---------.actca.g.ttc.t.t.atattattgt.a..g..g........
```

Fig. 4. Alignment of an intron from (**a**) cinnamyl-alcohol dehydrogenase genes from *Eucalyptus gunnii* (*egcad*) and *Eucalyptus botryoides* (*egccad*) and an intron from (**b**) two duplicated *Arabidopsis thaliana* acyl carrier protein genes (*atacpa1* and *atacpa2*). A third *Arabidposis* acyl carrier protein gene (*atacp*) shares little similarity in this intron. The *box* encloses the region similar to chloroplast DNA.



(b)

```
nuDNA  CAGACTAATT CGATTCTATA GAATAGATAA AAGATATGG- CTT-CTGCTC
mtDNA  .......... .......... A......... T.....▲.G ...T......
cpDNA  T.TCAA.TCG ATCCCTATGC TCCA.ATGG. T.AGAGA..T A..T......

nuDNA  ------AATT GTCACC-ATT CCTTCAGCTT -CACA-CCAG AGATAATCAA
mtDNA  ------.... ...T.TT.. .....T.... G...A..... .....T....
cpDNA  AGCCCA.... C..T.TTT.. .....T.T.. GA...A.A.. .....GAA..
```

Fig. 5. Sequential movement of a region of chloroplast DNA to mitochondria and then to the nucleus. **a** Diagram of the inferred sequence of events. The chloroplast sequence first moved to the mitochondrial genome (integration event #1) and then moved to the nuclear genome along with a portion of the mitochondrial genome (integration event #2). *Striped* regions represent chloroplast DNA and *clear* regions represent mitochondrial DNA. **b** Alignment of a portion of the nuclear sequence (nuDNA) 5′ of RNA polymerase pseudogene (RICRPBSP) with rice chloroplast (*cp*) DNA (X15901) and mitochondrial (*mt*) DNA (D13112). The *shaded boxes* indicate characters shared between the mitochondrial and nuclear sequences. The *arrow* indicates the boundary of the chloroplast insertion into the mitochondrial genome.

In two instances organellar DNA insertions are found in more than one member of a multigene family. A segment of chloroplast DNA is found in the same intron in all five potato patatin genes of the class 1 family. (See Table 1.) Therefore, it appears that the insertion in this gene family occurred prior to the gene amplification events with a subsequent insertion into the STPATG gene. Similarly, a segment of mitochondrial DNA transposed into the 3′-flanking region of the tobacco basic β-1,3-glucanase gene and was duplicated along with the coding regions. A fragment from a different intron of the same mitochondrial gene is found near the same position in the acid β-1,3-glucanase gene as in the basic β-1,3-glucanase genes.

Discussion

Effects of Intercompartmental DNA Transfer on Genomic Composition and Nuclear Gene Structure

None of the 28 nuclear DNA segments derived from organellar DNA that are identified in this study contain a complete organellar gene. The majority of the segments are short enough and/or have degenerated sufficiently from the original organellar sequence so that they would be difficult to detect using conventional DNA hybridization methods. The computer search strategy developed here allows the criteria for a potential match to be de-

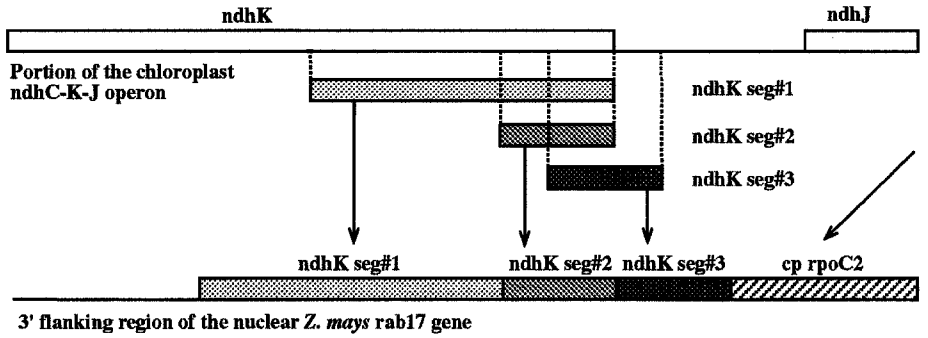


Fig. 6. An illustration of the origin of the three segments (*segs*) of the chloroplast (*cp*) NADH dehydrogenase (*ndhk*) gene found in the nuclear *Zea mays rab17* gene (ZMRAB17G). A segment derived from the chloroplast RNA polymerase (*rpoC2*) gene is derived from different region of the chloroplast genome.

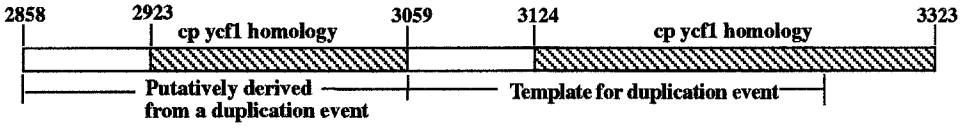


Fig. 7. An illustration of the duplication event which occurred after the insertion of a region from chloroplast (*cp*) open reading frame (*ycf1*) into the *L. esculentum* phytoene synthase gene. *Striped boxes* represent the region homologous to tobacco chloroplast *ycf1*. *Clear boxes* represent a region of unknown origin which was duplicated along with *ycf1*.

termined by a variety of parameters. Although we were able to locate 28 new fragments which appear to be derived from DNA migration events, there might be other fragments that are shorter or more divergent. Also, more sequences might be detected if a broader range of chloroplast and mitochondrial sequences were used to probe the sequence libraries.

The transfer events are probably recent based on the insertions in gene families which have been duplicated only within a particular plant genus. A recent movement is also supported by the overall high level of sequence similarity (70–100%) between the organellar and nuclear segments. Although molecular rates of change can vary widely, it seems probable that these insertions occurred within the last 80 million years, based on estimates for the divergence of major lineages of higher dicots (Olmstead et al. 1992) and on molecular rate estimates (Wolfe et al. 1989). Since the majority of the insertions appear to be silent, they are probably free from selective constraints and over time mutationally drift until the sequence is no longer recognizable as organellar-derived DNA.

The location of the insertions in introns and flanking regions indicates that the impact of intercompartmental DNA transfer has not been limited to the functional transfer of genes to the nucleus. Eleven fragments are within nuclear transcripts and 13 others were found in nontranscribed flanking regions. In addition there are many other organellar-derived regions whose location or proximity to nuclear genes is unknown. Unfortunately, the influence of these insertions on gene expression has not yet been experimentally assayed. However, two organellar-derived regions, *Arabidopsis* enhancer 12-7 and 12-Q (Table 1), confer cell-specific expression to reporter genes in transgenic plants (Ott and Chua 1990). Constructs with deletions of the organellar-derived region need to be tested to determine the role, if any, these insertions have in gene expression.

Our estimate that 3–7% of the plant genomic sequence files contain organellar-derived DNA is a rough approximation of its abundance in the nuclear genome. It does not quantitate the percent of organellar DNA per nuclear DNA nor does it give an indication of the amount of organellar-derived DNA in intergenic regions. Further study of long tracts of chromosomal DNA will be needed to provide more robust quantitation of the phenomenon. Our estimate does indicate that, based on the magnitude of DNA transfer, this is an important process for creating sequence variation in plants. Although no mitochondrial sequences have been reported in nuclear genes from nonplant species, application of a similar computer search strategy has identified examples of this assimilatory process in other organisms (Blanchard and Schmidt, in preparation).

Mechanisms of Intercompartmental DNA Transfer

The temporal order of events in chloroplast and mitochondrial mRNA processing often involves editing of primary RNA transcripts followed by splicing and subsequent processing to monocistronic mRNA (reviewed by Gray and Covello 1993). Accumulation of unedited and unspliced transcripts along with fully mature RNA (Gray and Covello 1993) indicates that RNA-based transfer of organellar sequences potentially could occur

during various steps of transcript processing. The fragment of a mitochondrial *atp9* gene found in the *Petunia* chalcone synthase gene is the third case of intercompartmental DNA transfer involving mitochondrial DNA which has been shown to undergo RNA editing. In this case, as in the previous examples (Nugent and Palmer 1991; Grohmann et al. 1992), the nuclear sequence seems to be derived from an edited mitochondrial transcript. The 12 fragments of organellar introns which are found in the nucleus indicate that if the transfer event in all cases was via an RNA intermediate, then the template was not fully processed RNA. Although the presence of organellar introns is typically interpreted as evidence for DNA-based transfer (Cheung and Scott 1989; Pichersky et al. 1991; Knoop and Brennicke 1991; Ayliffe and Timmis 1992a,b; Sun and Callis 1993), based on the limited data set of edited sequences, it appears that RNA-based transfer from mitochondria may be more common than previously assumed on the basis of transferred introns. No examples were identified in this study in which a migrant organellar sequence had undergone intron removal, nor are there any examples in the literature for the movement of a spliced RNA to the nucleus.

The lack of potential editing sites in any of the transferred chloroplast sequences precludes us from resolving whether chloroplast DNA transfer also occurs through RNA intermediates. To determine if RNA-based transfer of chloroplast sequences to the mitochondrial genome has occurred, we analyzed fragments of chloroplast sequences in the rice mitochondrial genome (Nakazono and Hirai, 1993) for potential RNA editing sites (Blanchard and Schmidt, results not shown). None of the putative RNA editing sites had undergone editing before the transfer of these fragments to the mitochondrial genome, supporting conclusions that these transfers are solely DNA-based (Stern and Lonsdale 1982; Nakazono and Hirai 1993).

As elaborated by Sun and Callis (1993), some characteristics of the insertion events (ligation of fragments, and short deletions at the insertion site) are reminiscent of "foreign" DNA assimilation in transfected animal cells. The duplication of the chloroplast *ycf1* fragment along with a stretch of nuclear DNA (Fig. 7) is also similar to rearrangements which accompany the integration of foreign DNA (Roth and Wilson 1988). The integration of a mitochondrial DNA fragment into a seed lectin gene (Knoop and Brennicke 1991) as well as the integration of chloroplast DNA into the cinnamyl-alcohol dehydrogenase and acyl carrier protein genes (Fig. 4) may have involved nucleotide complementarity between nuclear and organellar DNA. Pichersky (1990) hypothesized that heteroduplex formation at the DNA ends may play a role in the integration of organellar DNA. Short regions of overlap exist between some fragments in the 3′ end of the maize *rab17* gene in between fragments in the "homologous to B4 plasmid" sequence. However, some of these fragments, as well as fragments in the beet satellite DNA and arabidopsis polyubiquitin gene, do not have overlapping sequences. Therefore, the nuclear capture of many of these transferred organellar sequences does not appear be dependent on sequence homology and may simply be a result of random integration via end-joining processes used in repairing DNA breaks. Random integration is described by Roth and Wilson (1988) as a subset of illegitimate recombination events in which DNA ends can join together regardless of whether complimentarity is present at the ends. Such a general mechanism of DNA assimilation would encompass the previous models (Sun and Callis 1993; Pichersky 1990).

Roth and Wilson (1988) speculate that "efficient end joining is a well spring of evolutionary opportunity." End joining need not be limited to the coligation of fragments from disparate regions of organellar genomes and from different organellar and plasmid genomes; theoretically it could also occur with fragments of nuclear genomic DNA, reversed transcribed RNA, viral DNA, transposable elements, and other sources of extrachromosomal DNA. Nonetheless, from the many examples of organellar DNA recruitment into nuclear chromosomes presented in this report, intercompartmental transfer via end-joining sequences from either DNA or RNA sources emerges as a potentially important mechanism for generating genomic diversity.

Acknowledgments. We would like to thank Shawn White and Tom Bureau for their instruction in using the computer programs and for their helpful suggestions and comments. We would also like to thank Mary Hagen, Shawn White, David Geiser, Matthew Hare, Elizabeth Bachman, Robert Price, and the two anonymous reviewers for their critical reading of this manuscript. This study was supported by a grant from DOE to G.W.S.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Ayliffe MA, Timmis JN (1992a) Plastid DNA sequence homologies in the tobacco nuclear genome. Mol Gen Genet 236:105–112

Ayliffe MA, Timmis JN (1992b) Tobacco nuclear DNA contains long tracts of homology to chloroplast DNA. Theor Appl Genet 85:229–238

Baldauf SL, Palmer JD (1990) Evolutionary transfer of the chloroplast *tufA* gene to the nucleus. Nature 344:262–265

Bernatzky R, Mau S-L, Clarke AE (1989) A nuclear sequence associated with self-incompatibility in *Nicotiana alata* has homology with mitochondrial DNA. Theor Appl Genet 77:320–324

Bureau TE, Wessler SR (1994) Mobile inverted repeat elements of the *Tourist* family are associated with the genes of many cereal grasses. Proc Natl Acad Sci USA 91:1411–1415

Cheung WY, Scott NS (1989) A contiguous sequence in spinach nuclear DNA is homologous to three separated sequences in chloroplast DNA. Theor Appl Genet 77:625–633

Cushman JC, Bohnert HJ (1992) Salt stress alters A/T-rich DNA-binding factor interactions within the phosphoenolpyruvate carboxylase promoter from *Mesembryanthemum crystallinum.* Plant Mol Biol 20:411–424

Dangl JL, Hahlbrock K, Schell J (1989) Regulation and structure of

chalcone synthase genes. Cell culture and somatic cell genetics of plants, vol 6: Genetics of plants. Academic Press, San Diego

Devereaux J, Haeberli P, Smithies O (1984) A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res 12: 387–395

Farrelly F, Butow RA (1983) Mitochondrial genes in the yeast nuclear genome. Nature 301:296–301

Fukuchi M, Shikanai T, Kossykh VG, Yamada Y (1991) Analysis of nuclear sequences homologous to the B4 plasmid-like DNA of rice mitochondria: evidence for sequence transfer from mitochondria to nuclei. Curr Genet 20:487–494

Gantt JS, Baldauf SL, Calie PJ, Weeden NF, Palmer JD (1991) Transfer of *rpl22* to the nucleus greatly preceded its loss from the chloroplast and involved the gain of an intron. EMBO J 10:3073–3078

Gellissen G, Bradfield JY, White BN, Wyatt GR (1983) Mitochondrial DNA sequences in the nuclear genome of a locust. Nature 301: 631–634

Gellissen G, Michaelis G (1987) Gene transfer: mitochondria to nucleus. Ann N Y Acad Sci 503:391–401

Gray MW (1992) The endosymbiotic hypothesis revisited. Int Rev Cytol 141:233–357

Gray MW, Covello PS (1993) RNA editing in plant mitochondria and chloroplasts. FASEB J 7:64–71

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

Hadler HI, Dimitrijevic B, Mahalingam R (1983) Mitochondrial DNA and nuclear DNA from normal rat liver have a common sequence. Proc Natl Acad Sci USA 80:6495–6499

Hua S, Dube SK, Kung S (1993) Sequence similarity in nuclear and mitochondrial gene regions in plants. J Plant Biochem Biotech 2:71–73

Knoop V, Brennicke A (1991) A mitochondrial intron sequence in the 5′-flanking region of a plant nuclear lectin gene. Curr Genet 20: 423–425

Knoop V, Brennicke A (1994) Promiscuous mitochondrial group II intron sequences in plant nuclear genomes. J Mol Evol 39:144–150

Koch I, Hofschneider PH, Koshy R (1989) Expression of a hepatitis B virus transcript containing fused mitochondrial-like domains in human heptoma cells. Virology 170:591–594

Koes RE, Spelt CE, van Den Elzen PJM, Mol JNM (1989) Cloning and molecular characterization of the chalcone synthase multigene family of *Petunia hybrida.* Gene 81:245–257

Margulis L (1993) Symbiosis in cell evolution: microbial communities in the Archeon and Protorozoic eons. Freeman, New York

Nakazono M, Hirai A (1993) Identification of the entire set of transferred chloroplast DNA sequences in the mitochondrial genome of rice. Mol Gen Genet 236:341–346

Nugent JM, Palmer JD (1991) RNA-mediated transfer of the gene *coxII* from the mitochondrion to the nucleus during flowering plant evolution. Cell 66:473–481

Okubara PA, Williams SA, Doxsee RA, Tobin EM (1993) Analysis of genes negatively regulated by phytochrome action in *Lemna gibba* and identification of a promoter region required for phytochrome responsiveness. Plant Physiol 101:915–924

Olmstead RG, Michaels HJ, Scott KM, Palmer JD (1992) Monophyly of the Asteridae and identification of their major lineages inferred from DNA sequences of *rbcL.* Ann Missouri Botanical Gardens 79:249–265

Ossario PN, Sibley DL, Boothroyd JC (1991) Mitochondrial-like DNA sequences flanked by direct and inverted repeats in the nuclear genome of *Toxoplasma gondii.* J Mol Biol 22:525–536

Ott RW, Chua NH (1990) Enhancer sequences from *Arabidopsis thaliana* obtained by library transformation of *Nicotiana tabacum.* Mol Gen Genet 223:169–179

Palmer JD (1991) Plastid chromosomes: structure and evolution. In: Bogorad L, Vasil IK (ed) Cell culture and somatic cell genetics of plants, vol 7A: The molecular biology of plastids. Academic Press, San Diego, pp 5–53

Pichersky E (1990) Nomad DNA—a model for movement and duplication of DNA sequences in plant genomes. Plant Mol Biol 15: 437–448

Pichersky E, Logsdon Jr JM, McGrath JM, Stasys RA (1991) Fragments of plastid DNA in the nuclear genome of tomato: prevalence, chromosomal location, and possible mechanism of integration. Mol Gen Genet 225:453–458

Pichersky E, Tanksley SD (1988) Chloroplast DNA sequences integrated into an intron of a tomato nuclear gene. Mol Gen Genet 215:65–68

Roth D, Wilson (1988) Illegitimate recombination in mammalian cells. In: Kucherlapati R, Smith G (eds) Genetic recombination. American Society for Microbiology, Washington, DC

Schuster W, Brennicke A (1987) Plastid, nuclear and reverse transcriptase sequences in the mitochondrial genome of *Oenothera:* is genetic information transferred between organelles via RNA? EMBO J 6:2857–2863

Shay JW, Baba T, Zhan Q, Kamimura N, Cuthbert JA (1991) HeLaTG cells have mitochondrial DNA inserted into the *c-myc* oncogene. Oncogene 6:1869–1874

Stern DB, Lonsdale DM (1982) Mitochondrial and chloroplast genomes of maize and have a 12-kilobase DNA sequence in common. Nature 299:698–702

Sun C-W, Callis J (1993) Recent stable insertion of mitochondrial DNA into an *Arabidopsis* polyubiquitin gene by nonhomologous recombination. Plant Cell 5:97–107

Swofford DL (1993) Phylogenetic analysis using parsimony (PAUP), version 3.1.1. Illinois Natural History Survey, Champaign, IL

Thorsness PE, Fox TD (1993) Nuclear mutations in *Saccharomyces cerevisiae* that affect the escape of DNA from mitochondrial to the nucleus. Genetics 134:21–28

Thorsness PE, White KH, Fox TD (1993) Inactivation of *YME1,* a member of the *ftsH-SEC18-PAS1-CDC48* family of putative ATPase-encoding genes, causes increased escape of DNA from mitochondria in *Saccharomyces cerevisiae.* Mol Cell Biol 13:5418–5426

Timmis JN, Scott NS (1983) Sequence homology between spinach nuclear and chloroplast genomes. Nature 305:65–67

van den Boogaart P, Samallo J, Agsterbibe E (1982) Similar genes for a mitochondrial ATPase subunit in the nuclear and mitochondrial genomes of *Neurospora crassa.* Nature 298:187–189

Wintz H, Hanson M (1991) A termination codon is created by RNA editing in the petunia mitochondrial *atp9* gene transcript. Curr Genet 19:61–64

Wolfe KH, Sharp PM, Li W-H (1989) Rates of synonymous substitution in plant nuclear genes. J Mol Evol 29:208–211

Zullo S, Leang CS, Slighton JL, Hadler HI, Eisenstadt JM (1991) Mitochondrial D-loop sequences are integrated in the rat nuclear genome. J Mol Biol 221:1223–1235