How do mitochondrial genes get into the nucleus?

Katrin Henze and William Martin

It is well known that genes from chloroplasts and mitochondria were transferred to the nucleus many times during plant evolution. But in what form do the transferred genes physically make that intracellular journey – as RNA, as cDNA, as pieces of organelle DNA, or as whole organelle chromosomes? Current views focus upon cDNA as the vehicle, based upon some examples from plants. But other mechanisms, involving direct transfer of DNA from organelle chromosomes, could also account for the available data. Direct DNA transfer, rather than cDNA-mediated transfer, does occur today, and it probably prevailed during the early phases of organelle evolution.

The evolution of plastids and mitochondria entailed the transfer of genes from organelles to the nucleus. Study of this topic goes back about 20 years, beginning with the discovery of recombined mitochondrial chromosome segments in nuclear DNA[1,2]. Examples of DNA transfer from plant mitochondria were subsequently found[3,4] that suggest some genes might be transferred to the nucleus as cDNA (i.e. as double-stranded reverse-transcribed mitochondrial mRNA).

The evidence supporting the view that cDNA might be involved primarily comes from the findings that integrated nuclear copies of genes that stem from plant mitochondria often lack the introns that their mitochondrial copies in related species possess. In addition, the nuclear copies are sometimes more similar to edited mitochondrial mRNAs than they are to the genes encoded in the mitochondrial DNA itself[3–6]. Furthermore, genes from plant mitochondrial DNA have been transferred to the nucleus multiple times in independent lineages, a good example being the *rps10* gene that encodes a protein of the mitochondrial ribosome. Some years ago, Volker Knoop showed that the *rps10* gene is present in the mitochondrial genome of some flowering plants (angiosperms), but located in the nucleus of others[5]. A recent paper extended those findings, reporting *rps10* in the nucleus of numerous other angiosperm lineages[6] and underscoring the view that gene transfer from organelles to the nucleus is an active, prevalent and ongoing evolutionary process.

These developments raise the important question: how do genes actually relocate from organelles to the nucleus during evolution? Here we consider the evidence that mitochondrial genes are transferred to the nucleus through reverse transcription of spliced and edited mitochondrial mRNA into cDNA, and we examine whether other mechanisms involving direct transfer of chromosomal DNA might also account for the same observations. A particularly striking example of direct DNA transfer – the massive 620-kb DNA chunk of mitochondrial DNA in the nucleus of *Arabidopsis* – features prominently in this discussion.

K. Henze
W. Martin*
Institute of Botany III, Universität Düsseldorf, Universitätsstraße 1, D-40225 Düsseldorf, Germany.
*e-mail: w.martin@uni-duesseldorf.de

Editing

RNA editing is widespread in plant mitochondria. It usually involves the conversion of uracil residues in the mitochondrial primary RNA transcripts into cytosine residues (U→C editing) so that the proper amino acid is specified by the respective codon in the resulting mRNA, although C→U editing also occurs[7]. The biochemical mechanisms of plant mitochondrial editing are still not known, and many questions remain about the evolutionary dynamics of mitochondrial editing across different plant lineages and even across genes within the same mitochondrion[7,8]. For example, in the sequenced *Arabidopsis* mitochondrial genome[9] there are 441 C→U editing sites[8]. These sites have a density of about 14 per kb in protein-coding regions, one per kb in introns, 0.5 per kb in 5′- and 3′-untranslated regions, and none in tRNAs, rRNAs and noncoding regions[8]. But even within these coding regions, editing density ranges from no sites in the 1.5-kb *cox1* gene to 39 sites in the 620-bp *ccb2* gene[8]. No obvious consensus sequences are associated with edited sites[8], and worse, the phylogenetic distribution of mitochondrial editing across various land-plant lineages is highly erratic, showing both gene-specific and lineage-specific patterns.

If mitochondrial copies of a gene contain edited sites when the nuclear copies of the same gene lack them, does this mean that the gene must have been transferred through an mRNA intermediate (Fig. 1a)? No, not really, because it is also possible that the gene was transferred at a time when the sites in question were not edited (Fig. 1b). In the plant kingdom, mitochondrial editing has not been found in algae, so we can assume it to be an invention of the higher plants, where it is reasonably (but not uniformly) widespread. If a gene was transferred to the nucleus in one plant lineage before mitochondrial editing evolved but remained in the organelle in other lineages where editing arose, the nuclear copy would appear more similar to an edited transcript than to the remaining mitochondrial copies at the edited sites (Fig. 1b). However, the same observation is taken as evidence for cDNA-mediated transfer.

Because editing shows very erratic patterns of taxon-specific occurrence even among higher-plant lineages[7,8], either editing (and/or specific edited sites) arose independently in many higher-plant lineages, or editing was universal among ancestral

http://tig.trends.com 0168–9525/01/$ – see front matter PII: S0168-9525(01)02312-5

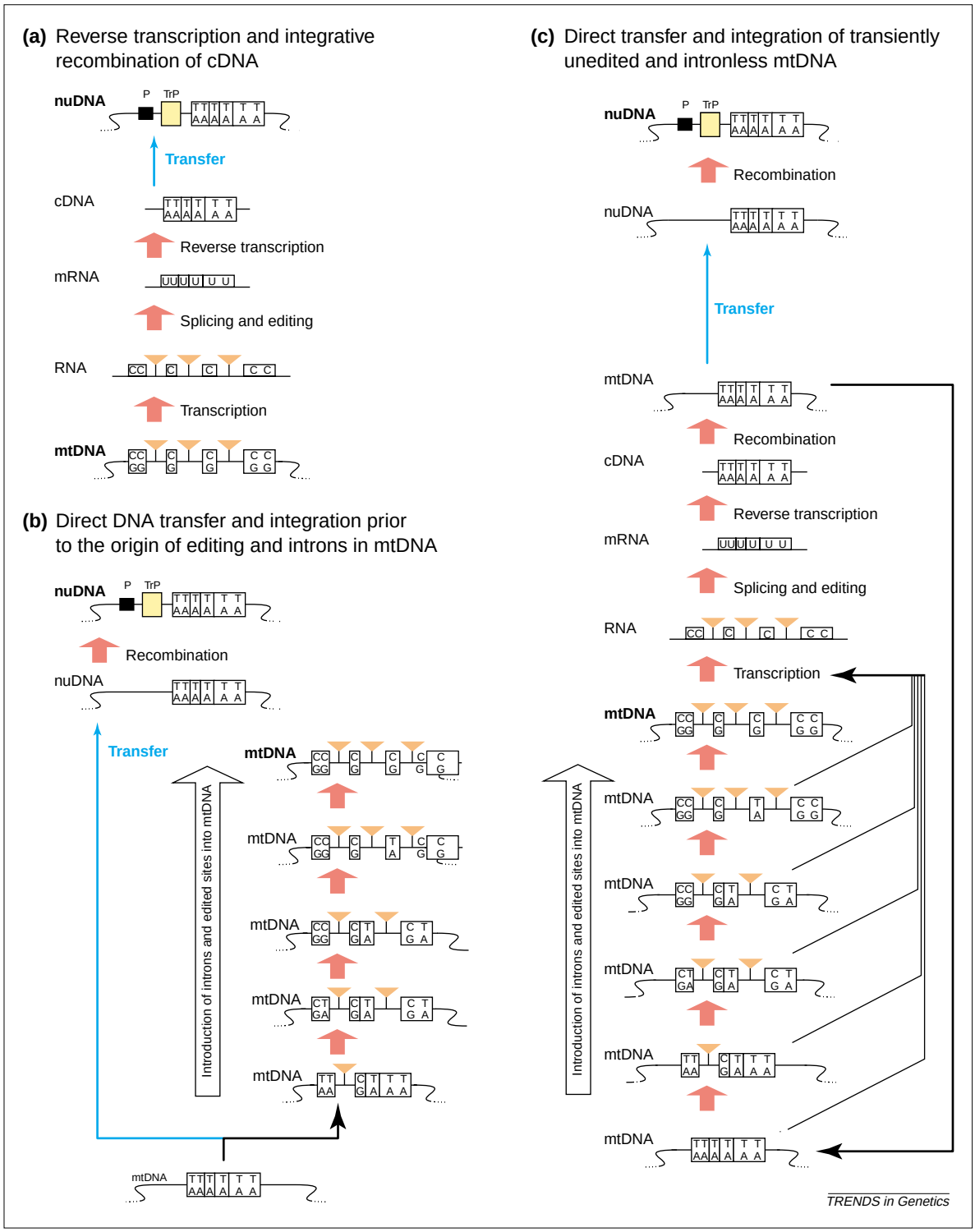

Fig. 1. Integration of mitochondrial genes into the nuclear DNA: mechanisms that appear as though a cDNA intermediate was involved. (a) A cDNA-mediated model entailing reverse transcription of a spliced and edited mitochondrial transcript that becomes integrated in the nucleus. (b) A mechanism involving direct transfer of mtDNA, whereby editing and introns arise in the mitochondrion later. (c) A mechanism involving direct transfer of mtDNA, whereby editing and introns in the mitochondrion come and go during evolution. Boldface lettering indicates the nuclear DNA (nuDNA) and mitochondrial DNA (mtDNA) that would exist today, and hence be analysed in the laboratory (the observation). Note that, although the mechanisms and underlying processes are different, the observations are identical in all three panels. C→U edited sites and introns (orange triangles) are indicated. Note that in panel (c), any of the mtDNA intermediates could be transferred, but that the topmost in the panel would have the highest likelihood of giving rise to a nuclear gene with a fuctional gene product. Recent findings provide strong evidence that direct transfer of mtDNA as indicated in (b) and (c) is quite common in higher plants, given that a whole mitochondrial genome was found integrated into in the *Arabidopsis* nuclear chromosomes.

higher-plant lineages, but some process exists that removes edited sites from mitochondrial DNA (Fig. 1c).

One way to remove editing from mitochondrial DNA would be reverse transcription of edited mRNA and recombination back into the mitochondrial genome, which could proceed through homologous recombination in the organelle (Fig. 1c). Another way would be U→C edits in addition to typical C→U edits, and this process is found in some lower land-plant mitochondria[7]. Under the influence of editing, reverse transcription and homologous recombination of mitochondrial transcripts, or with the help of reverse editing, a hypothetical evolutionary equilibrium of the edited state versus the unedited state for a given site in the mitochondrial DNA would result (Fig. 1c). In this case, direct transfer of mitochondrial DNA without a cDNA intermediate could easily occur. Furthermore, although direct DNA transfer of any intermediate

state could occur, only transfers of the unedited state would have a chance of successfully substituting the function of the mitochondrial copy, because genes with edited sites could not be decoded by the nuclear–cytoplasmic gene-expression machinery. Clearly, this scenario requires the existence of mechanisms that remove edited sites in mitochondrial DNA.

Both in the case of transfer of a gene before the origin of editing (Fig. 1b) and in the case of transfer of transiently unedited genes during evolution (Fig. 1c), a sequence comparison would show that the nuclear copy of mitochondrial origin (designated in boldface type as nuDNA in Fig. 1) is more similar to the edited transcript than to a mitochondrial gene that possesses edited sites (boldface mtDNA in Fig. 1). Until now, this type of observation has been taken as evidence that the transfer mechanism of plant mitochondrial DNA to the nucleus proceeded through a cDNA intermediate (Fig. 1a). But Fig. 1b and Fig. 1c both show mechanisms of gene transfer that produce exactly the same observation even though they do not involve the transfer of a cDNA, rather bulk DNA is transferred from the mitochondrial genome.

Introns

If the mitochondrial copy of a gene contains introns when the nuclear copies of the same gene lack them, does this mean that the gene must have been transferred through an mRNA intermediate? Not necessarily, because it is also possible that the gene was transferred at a time when it did not possess the intron. There is evidence that plant mitochondrial introns can be mobile, in particular the *rps10* intron, which is clearly related to both the second intron of the mitochondrial *cox3* gene and the intron in the gene for 26S ribosomal RNA of *Marchantia*[5,10]. The *rps10* intron is missing in the mtDNA of some plant lineages[1], including those recently surveyed, where it even cropped up in one mtDNA lineage (carrot) without the flanking exons[6]. Here again, it is possible that transfer to the nucleus could have occurred without a cDNA intermediate. It is even conceivable that the *rps10* intron was removed after transfer to the nucleus, perhaps in the same manner that the normal GT–AG introns of eukaryotes are thought to have been lost; that is, through splicing of the nuclear transcript, reverse transcription and recombination in the nuclear DNA[11]. In the case of self-splicing introns, this would not require any auxiliary factors, so splicing of a transcript generated outside the confines of the mitochondrion would, in principle, be possible. As in the case of editing, it seems that introns do not provide compelling evidence that the transfer mechanisms involve cDNA rather than direct DNA transfer.

Evidence for direct DNA transfer

Is there any good, hard evidence for direct wholesale transfer of mitochondrial DNA to the nucleus that does not involve a cDNA intermediate? Yes. By far the most eye-opening organelle-to-nucleus transfer yet observed is an entire mitochondrial genome that cropped up on chromosome 2 of the *Arabidopsis* genome[12,13]. This chunk of mitochondrial DNA in the *Arabidopsis* nucleus –referred to here as *c2mtch* (for chromosome 2 mitochondrial-DNA chunk) – was initially estimated[12] to comprise fully 75% of the 366924-bp *Arabidopsis* mitochondrial genome[9]. However, a recent more-detailed study[13] using *in situ* hybdridization showed that *c2mtch* in fact represents an entire mitochondrial genome – and then some! It seems that *Arabidopsis* contig-assembly computers missed a detail or two, and that *c2mtch* is not 270 kb, but ~620 kb long. This is longer than the *Arabidopsis* mtDNA itself, the additional length coming from internal duplications of large segments of this complete mtDNA genome, which might have occurred while the molecule was still in the mitochondrion[13]. Clearly, *c2mtch* was inserted (recombined) into chromosome 2 as a single contiguous piece of mitochondrial chromosomal DNA[12,13] – introns, tRNAs, noncoding regions and all – indicating that it simply stems from a lysed mitochondrion and got 'tangled up' in the nucleus. In fact, organelle lysis and direct DNA transfer – in theory and in practice – will very efficiently drive gene transfer to the nucleus over evolutionary time[14,15]. The *c2mtch* insertion is a striking demonstration that whole organelle genomes (with all of their genes) can integrate into the nucleus.

Furthermore, *c2mtch* is 99% identical at the nucleotide level to the authentic mitochondrial genome, suggesting that transfer was recent[12]. Using a rough molecular-clock guesstimate, if we assume that the nuclear synonymous nucleotide substitution rate in plants is about 5×10^{-9} per site per year[16], then a 1% divergence to the mitochondrial copy indicates that the transfer of *c2mtch* occurred only ~2 million years ago. Of course, *c2mtch* is only one data point and might be unique, or similar events could have occurred in other higher-plant lineages as well – data from other plant genomes will show whether that is true. If we consider *c2mtch* as being roughly representative of transfer from organelles to the nucleus, then it would seem that an entire plant mitochondrial genome gets transferred to the nucleus at the rate of about once every 2 million years – per plant species. That would be the kind of continuous flux of bulk DNA from organelles to the nucleus that would work in favour of the mechanisms shown in Fig. 1b and, particularly, Fig. 1c; recombination in the nucleus could supply promoters, transit peptides and the rest.

If both editing and intron insertion in plant mitochondrial DNA are indeed reversible processes (and we are positing that they are), then bulk transfer of the type documented in *Arabidopsis*

c2mtch would also encompass (at some points in evolutionary time) the transfer of transient states in the mitochondrial genome. (In other words, a gene in the mitochondrion that is edited and intron-containing today might have been unedited and intronless, say, 50 million years ago, and edited and intron-containing 100 million years ago, etc.) Hence it would be impossible to distinguish what state was transferred (with or without intron, with or without edited sites) by looking at contemporary plant mitochondrial DNA.

Recombination

Transfer events of the type revealed by *c2mtch* provide an inexhaustible source of genetic material pouring into nuclear chromosomes – the starting material for recombination, mutation and new functions. Is there evidence for nuclear recombination involving DNA derived from the mitochondrion? Yes. Kadowaki *et al.*[17] reported very clear examples for the mitochondrial ribosomal protein, Rps11. Rice has two recently duplicated nuclear genes for Rps11 (*Rps11-1* and *Rps11-2*), in addition to an *Rps11* pseudogene in the mitochondrion. Both nuclear copies have N-terminal transit peptides to direct the protein to the organelle. In *Rps11-1*, part of the transit peptide was acquired from the nuclear gene for mitochondrial AtpB (a component of the ATPase) through recombination, such that part of the same transit peptide is found on two different nuclear genes, *Rps11-1* and *AtpB*. The transit peptide of *Rps11-2* was taken from the nuclear gene for mitochondrial cytochrome *c* oxidase subunit Vb, *coxVB* (Ref. 17). Long *et al.*[18] found that the transit peptide for mitochondrial cytochrome *c* in potato was acquired by nuclear exon shuffling between the transferred gene for cytochrome *c* and a gene for glyceraldehyde-3-phosphate dehydrogenase. Similarly, recombination between the gene encoding Rps10 and genes encoding the heat shock proteins Hsp22 and Hsp70 was found in the more recent study[2].

One of the most bizarre examples of Nature's ingenuity in recombining a mitochondrial gene involves the mitochondrial ribosomal protein Rps14 in rice[19] and maize[20]. This mitochondrial gene (*rps14*) has recombined into the intron of the nuclear gene *SDHB*, which encodes the β-subunit of succinate dehydrogenase – a mitochondrial protein containing a transit peptide. To add to the complexity, *SDHB* is itself a nuclear gene of mitochondrial origin and is even encoded in some mitochondrial genomes still[21]. To make matters worse, through alternative splicing, the same transit peptide encoded by a single nuclear locus is used by two different mitochondrial proteins, Rps14 and SdhB, in rice and maize[20,21].

Finally, Kubo and colleagues[22] found a striking example in the rice genome that is convincing evidence in favour of the view that relocation of mitochondrial genes to the nucleus involves the transfer of bulk DNA and recombination, not cDNA. The study found a piece of DNA from the rice mitochondrion containing a fragment of the *rps19* coding region, a fragment of the *rps3* coding region and a segment of the group II intron in the *rps3* gene. This DNA has made its way to the nucleus and recombined into the *3′* region of a nuclear gene for a vacuolar ATPase[22]. The hybrid gene is rather strongly expressed as an mRNA, but it encodes nothing but exquisite junk[22] – half of the ATPase-coding region is missing, the mitochondrial intron and coding sequences are expressed in the antisense orientation, and there is an unedited editing site left in the mtDNA-derived sequence[22]. Clearly, as in the case of *Arabidopsis c2mtch*, recombination involving a piece of rice mtDNA, transferred as bulk DNA, not as cDNA, was at work here[22].

Summing up – looking back in time

There are plenty of interesting things to look for in the *Arabidopsis* genome[23]. The first search for bits of organelle DNA on the other four chromosomes was less fruitful than the search of chromosome 2 because, in addition to the massive 620-kb *c2mtch* insert[12,13], only 11 other insertions of mitochondrial DNA (totalling 7 kb) and 17 chloroplast insertions (totalling 11 kb and including an intron) were found using BLAST-based analyses[23]. But broad-scale phylogenetic estimates indicate that several hundred (perhaps even several thousand) active and expressed *Arabidopsis* genes are acquisitions from chloroplasts alone[24].

Much current thinking about the mechanisms of gene transfer from organelles to the nucleus focuses on cDNA-mediated processes, perhaps because some interesting examples suggest that this might be true. And it might very well be true in some cases but, as outlined here, alternative mechanisms can account for the same observations where cDNA is invoked. Well in advance of findings implicating cDNA[3–6], evidence for direct DNA transfer from organelles to the nucleus was clear[1,2]. The presence of an entire mitochondrial DNA genome in the *Arabidopsis* nucleus[12,13] (*c2mtch*) is a reminder that direct DNA transfer certainly does occur and – all things considered – probably constitutes the general rule to which evolutionarily recent cases of cDNA-mediated transfer in some higher plants are exceptions.

After all, the majority of gene transfer from organelles to the nucleus occurred during the early stages of the integration of these organelles into the cytosol of their host[25,26]. Both editing[7] and the spread of mitochondrial introns[27] are comparatively recent developments in the overall course of plant evolution. Thus, there must have been a time in life's history when neither organellar editing nor organellar introns existed. During that phase of evolution, the transfer of bulk DNA and activation of genes through

recombination to yield promoter- and transit-peptide-bearing copies in the host cell's chromosomes (Fig. 1b) should have been the prevalent mechanism – a single lysed organelle every few million years would more than suffice.

And looking back one step further into the very earliest stages of organelle evolution, there must have been a time when even the protein-import apparatus specific for these organelles[28,29] had not yet evolved. During that early phase of evolution, transit peptides were absent. Hence, genes that were transferred from endosymbionts to the chromosomes of their host would have been expressed only in the cytosol, and fixed only on the basis of the contribution of those cytosolic gene products to the overall fitness of the cell. In that way, biochemical pathways once germane to endosymbionts could have been transferred to the host's cytosol through the simple relocation of the corresponding genes[30]. In general agreement with that reasoning, recent estimates suggest that about 600 *Arabidopsis* nuclear genes of cyanobacterial (plastid) origin encode cytosolic proteins[31]. There is still much to learn about gene transfer from organelles, both in terms of mechanisms and terms of how it has shaped the contours of eukaryotic genomes – even the genomes of eukaryotes that have lost their organelles[32]!

Acknowledgements
We thank Koh-ichi Kadowaki (Tsukuba, Japan) for kindly sharing data before publication and Volker Knoop (Ulm, Germany) for discussions. Work in the authors's lab is funded by the DFG and by SFB-TR/1.

References

1 Jacobs, H.T. *et al.* (1983) Mitochondrial DNA sequences in the nuclear genome of *Strongylocentrotus purpuratus*. *J. Mol. Biol.* 165, 609–632
2 Farrely, F. and Butow, R.A. (1983) Rearranged mitochondrial genes in the yeast nuclear genome. *Nature* 301, 296–301
3 Nugent, J.M. and Palmer, J.D. (1991) RNA-mediated transfer of the gene coxII from the mitochondrion to the nucleus during flowering plant evolution. *Cell* 66, 473–481
4 Brennicke, A. *et al.* (1993) The mitochondrial genome on its way to the nucleus: different stages of gene transfer in higher plants. *FEBS Lett.* 325, 140–145
5 Knoop, V. *et al.* (1995) The gene for ribosomal protein S10 is present in mitochondria of pea and potato but is absent from those of *Arabidopsis* and *Oenothera*. *Curr. Genet.* 27, 559–564
6 Adams, K.L. *et al.* (2000) Repeated ancient and diverse transfers of a mitochondrial gene to the nucleus in flowering plants. *Nature* 408, 354–357
7 Steinhauser, S. *et al.* (1999) Plant mitochondria RNA editing. *J. Mol. Evol.* 48, 303–312
8 Geigé, P. and Brennicke, A. (1999) RNA editing in *Arabidopsis* mitochondria effect 441 C to U changes in ORFs. *Proc. Natl. Acad. Sci. U. S. A.* 96, 15324–15329
9 Unseld, M. *et al.* (1997) The mitochondrial genome of *Arabidopsis thaliana* contains 57 genes in 366924 nucleotides. *Nat. Genet.* 15, 57–61
10 Malek, O. *et al.* (1997) Evolution of trans-splicing plant mitochondrial introns in pre-Permian times. *Proc. Natl. Acad. Sci. U. S. A.* 94, 553–558
11 Sharp, P.A. (1994) Split genes and RNA splicing. *Cell* 77, 805–815
12 Lin, X.Y. *et al.* (1999) Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*. *Nature* 402, 761–768
13 Stupar, R.M. *et al.* (2001) Complex mtDNA constitutes an approximate 620-kb insertion on *Arabidopsis thaliana* chromosome 2: Implication of potential sequencing errors caused by large-unit repeats. *Proc. Natl. Acad. Sci. U. S. A.* 98, 5099–5103
14 Berg, O.G. and Kurland, C.G. (2000) Why mitochondrial genes are most often found in nuclei. *Mol. Biol. Evol.* 17, 951–961
15 Blanchard, J.L. and Lynch, M. (2000) Organellar genes – why do they end up in the nucleus? *Trends Genet.* 16, 315–320
16 Li, W-H. and Graur, D. (1999) *Fundamentals of Molecular Evolution*, Sinauer
17 Kadowaki, K-I. *et al.* (1996) Targeting presequence acquisition after mitochondrial gene transfer to the nucleus occurs by duplication of existing target signals. *EMBO J.* 15, 6652–6661
18 Long, M. *et al.* (1996) Exon shuffling and the origin of the mitochondrial targeting function in plant cytochrome *c*1 precursor. *Proc. Natl. Acad. Sci. U. S. A.* 93, 7727–7731
19 Kubo, N. *et al.* (1999) A single nuclear transcript encoding mitochondrial RPS14 and SDHB of rice is processed by alternative splicing: common use of the same mitochondrial targeting signal for different proteins. *Proc. Natl. Acad. Sci. U. S. A.* 96, 9207–9211
20 Figueroa, P. *et al.* (1999) Transfer of *rps14* from the mitochondrion to the nucleus in maize implied integration within a gene encoding the iron–sulphur subunit of succinate dehydrogenase and expression by alternative splicing. *Plant J.* 18, 601–609
21 Burger, G. *et al.* (1996) Genes encoding the same three subunits of respiratory complex II are present in the mitochondrial DNA of two phylogenetically distant eukaryotes. *Proc. Natl. Acad. Sci. U. S. A.* 93, 2328–2332
22 Kubo, N. *et al.* Mitochondrial sequence migrated downstream to a nuclear *V-ATPase B* gene is transcribed but non-functional. (in press)
23 The *Arabidopsis* Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408, 796–815
24 Rujan, T. and Martin, W. (2001) How many genes in *Arabidopsis* come from cyanobacteria? An estimate from 386 protein phylogenies. *Trends Genet.* 17, 113–120
25 Lang, B.F. *et al.* (1997) An ancestral mitochondrial DNA resembling a eubacterial genome in miniature. *Nature* 387, 493–497
26 Martin, W. *et al.* (1998) Gene transfer to the nucleus and the evolution of chloroplasts. *Nature* 393, 162–165
27 Cho, Y. *et al.* (1998) Explosive invasion of plant mitochondria by a group I intron. *Proc. Natl. Acad. Sci. U. S. A.* 95, 14244–14249
28 Heins, L. *et al.* (1998) The protein translocation apparatus of chloroplast envelopes. *Trends Plant Sci.* 3, 56–61
29 Schatz, G. and Dobberstein, B. (1996) Common principles of protein translocation across membranes. *Science* 271, 1519–1526
30 Martin, W. and Müller, M. (1998) The hydrogen hypothesis for the first eukaryote. *Nature* 392, 37–41
31 Abdallah, F. *et al.* (2000) A prediction of the size and evolutionary origin of the proteome of chloroplasts of *Arabidopsis*. *Trends Plant Sci.* 5, 141–142
32 Embley, T.M. and Hirt, R.P. (1998) Early branching eukaryotes? *Curr. Opin. Genet. Dev.* 8, 624–629