Plant Physiol. (1998) 118: 9–17

Update on Gene Transfer from Organelles to the Nucleus

Gene Transfer from Organelles to the Nucleus: How Much, What Happens, and Why?[1]

William Martin* and Reinhold G. Herrmann

Institut für Genetik, Technische Universität Braunschweig, Spielmannstrasse 7, D-38023 Braunschweig, Germany (W.M.); and Botanisches Institut, Ludwig-Maximillians-Universität, Menzingerstrasse 67, D-81927 Munich, Germany (R.G.H.)

Chloroplasts were once free-living cyanobacteria, mitochondria were once free-living proteobacteria, and both have preserved remnants of eubacterial genomes. But from the functional standpoint, both organelles have retained much more of their eubacterial biochemistry than is reflected in their DNA. The discrepancy between the number of genes that organelles encode and the number of eubacterial proteins that they contain is generally explained by something that we have come to know as "endosymbiotic gene transfer." During evolution, organelles export their genes to the nucleus, but reimport the products with the help of transit peptides and protein-import machinery, so that proteins are retained in organelles, but most of the genes are not. This process, over time, concentrates genetic material in nuclear chromosomes. Because gene-regulatory processes under the control of the nucleus are more complex and interrelated than those under the control of organelles, and because organelles naturally tend to come under the control of nuclear regulatory genes (imagine the opposite!), organelle regulatory processes are likely to have been among the first to be transferred successfully to the nucleus. From the standpoint of genes, this process therefore results in a compartmented, but integrated, eukaryotic genetic system under the regulatory dominance of the nucleus (Herrmann, 1997), rather than genetically semiautonomous organelles. However, from the standpoint of the encoded products of transferred genes, a surprising picture is emerging that could be loosely described as "a funny thing happened on the way back to the organelle."

The prerequisite for endosymbiotic gene transfer is protein-import machinery in the two membranes that surround chloroplasts and mitochondria, which allows these organelles to take up cytosolic precursors, cleave the transit peptides, and release the processed polypeptides into the stroma and matrix, respectively. For an overview of what proteins that machinery consists of, how it works, and how it might have evolved, we recommend the recent overviews by Schatz and Dobberstein (1996) for a general summary, and Heins et al. (1998) for chloroplasts in particular. The first clear-cut examples of endosymbiotic gene transfer became known about 10 years ago (for review, see Gray [1992] for a general overview; Brennicke et al. [1993] for the process of gene transfer).

Here we provide a brief summary of organelle genome reduction and its impact on plant cells, skimming the surface with a few examples of gene transfer to the nucleus from both plastids and mitochondria. From the standpoint of gene product function, we will consider factors that (a) might influence the immediate fate of genes that become transferred to the nucleus, and (b) might help to determine whether such transfer events become genetically fixed. We will also consider the question of why genes tend to be transferred from organelles to the nucleus.

PLASTIDS: HOW MANY GENES AND PROTEINS?

In 1998 we have an unfair advantage relative to those who wondered about the genetic "semiautonomy" of chloroplasts in 1978, because we have a much better overview of the number and types of genes contained in plastid genomes. Several chloroplast genomes have been completely sequenced, quite a few more are now being sequenced, and a cyanobacterial genome has been sequenced, with additional ones in the pipeline. How many proteins are encoded by ctDNA? The answer depends upon which plastid one considers; a summary is given in Table I (see also Martin et al., 1998). The nonphotosynthetic plastids in *Epifagus* (a parasite of beech trees) and *Plasmodium* (a parasite of humans) contain about 20 protein-coding genes. Fully functional higher plant chloroplasts encode about 60 to 80 proteins, the rhodophyte *Porphyra* can boast 200, whereas *Odontella* and *Cyanophora* encode on the order of 120 to 130 proteins. By contrast, the genome of the unicellular cyanobacterium *Synechocystis* encodes about 3168 proteins. Table I reveals that plastid genomes generally encode many more proteins than even the largest mitochondrial genomes studied (Lang et al., 1997; Unseld et al., 1997), but on the whole, they contain only about 1% to 5% as many protein-coding genes as a comparatively small cyanobacterial genome.

How many proteins do plastids contain? In an earlier work, Ellis (1981) suggested that the roughly 200 chloroplast proteins then directly resolvable by two-dimensional

[1] This work was supported by the Deutsche Forschungsgemeinschaft.

* Corresponding author; e-mail w.martin@tu-bs.de; fax 49–531–391–5765.

Table I. *Size comparison of sequenced chloroplast genomes, three large mitochondrial genomes, and a cyanobacterial genome*

Due to length limitations in this forum, accession numbers (in parentheses) rather than references are given for plastids.

Genome	Length in Base Pairs	No. of Protein-Coding Genes	No. of rRNA Clusters
ct *Porphyra* (U38804)	191,028	200	2
ct *Cyanophora* (U30821)	135,599	136	2
ct *Odontella* (Z67753)	119,704	124	2
ct *Marchantia* (X04465)	121,024	84	2
ct *Chlorella* (AB001684)	150,613	78	1
ct *Nicotiana* (S54304)	155,844	76	2
ct *Oryza* (X15901)	134,525	76	2
ct *Zea* (X86563)	140,387	76	2
ct *Pinus* (D17510)	119,707	69	1
ct *Euglena* (Z11874)	143,170	58	3
ct *Plasmodium* (X95275-6)	29,422	23	2
ct *Epifagus* (M81884)	70,028	21	2
mt *Reclinomonas*[a]	69,034	63	1
mt *Marchantia*[b]	186,608	41	1
mt *Arabidopsis*[c]	366,924	31	1
Synechocystis sp.[d]	3,573,470	3168	2

[a] Lang et al. (1997). [b] Oda et al. (1992). [c] Unseld et al. (1997). [d] Kaneko et al. (1996).

electrophoresis could be just the "tip of the iceberg." More recently it was shown that thylakoid membranes alone contain at least 75 major proteins (Herrmann et al., 1991; Pillen et al., 1996). We can roughly estimate the number of plastid proteins, considering the number of identified genes in the *Synechocystis* genome that belong to pathways and the functions known to reside partially, predominantly, or exclusively in plastids. Only about 50% (approximately 1600) of the genes in the *Synechocystis* genome have a known or putative function (Kaneko et al., 1996). Using the table provided by Kaneko et al. (1996), a rough calculation reveals that of those 1600 genes in *Synechocystis*, homologs of about 600 might be expected to exist in plastids. This provides a rough lower boundary for the number of chloroplast proteins. Assuming that the other approximately 1500 genes of as-yet-unassignable function in *Synechocystis* harbor another 400 to 500 probable plastid functions, one can obtain a conservative estimate of about 1000 different proteins that might be contained in a fully functional plastid. However, this estimate assumes that plastids do not do more for the plant cell than cyanobacteria do for themselves, which is probably not true (they probably do more). We estimate that the total number of different proteins, including, for example, isoenzymes and proteins involved in plastid-nucleus gene regulatory circuitry, in various types of plastids may be closer to about 2000. However, estimations from Arabidopsis data suggest that this number could approach 5000 (R. Douce, personal communication). Thus, plastids import the vast majority of their proteins, which are encoded in the nucleus.

How many plant nuclear genes are known that descend from cyanobacterial genomes? To our knowledge, nobody has yet compared all 3168 *Synechocystis* proteins to the plant databases and recorded/reported the results in such a manner that would answer that question. We do know, however, that at least 44 genes found in at least one plastid genome have functional, structurally characterized homologs in the nucleus of at least one higher plant (Martin et al., 1998). Surprisingly, genes tend to undergo multiple parallel losses from ctDNA in independent evolutionary lineages; parallel losses even outnumber phylogenetically unique losses by a ratio of about 4 to 1 (Martin et al., 1998).

REDUNDANT FUNCTIONS AND COMPARTMENTATION

What kinds of genes have been lost from organelle genomes? If we tabulate all of the different proteins of known or assignable (by sequence similarity) function that are encoded in sequenced chloroplast genomes, separate them into the functional categories used by Kaneko et al. (1996), and compare the corresponding numbers of genes per category for plastids (in toto) to the *Synechocystis* genome, we can begin to get a feel for what types of genes the ancestral plastid genome (we assume one cyanobacterial origin of plastids) might have had and what types are left (Fig. 1). Some of the encoded functions are gone altogether in higher plants, for example, the genes for proteins of

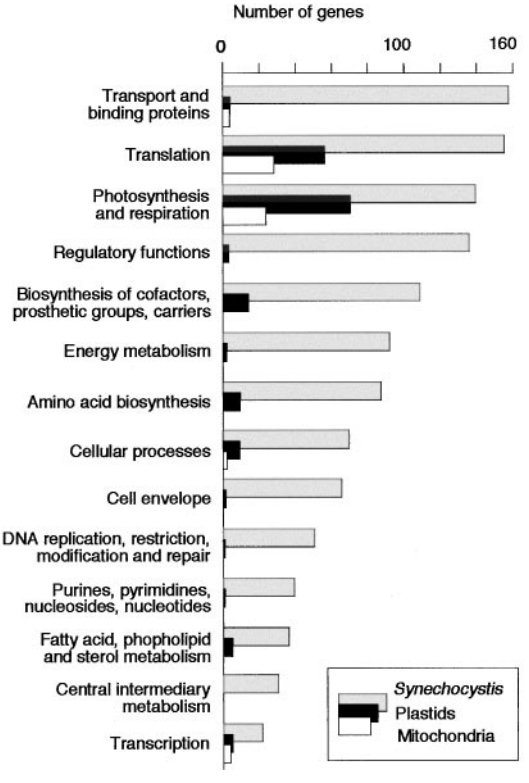

Figure 1. Protein-coding genes found in plastidial, mitochondrial, and a cyanobacterial genome in the functional categories used by Kaneko et al. (1996) (see also http://www.kazusa.or.jp/cyano/cyano.html). For organelles, the sum of all genes found in the genomes listed in Table I was used, whereby a gene found in 10 genomes is counted as 1 gene, not 10. Obviously, the largest plastidial (*Porphyra*) and mitochondrial (*Reclinomonas*) genomes contribute the majority of different genes per category.

phycobilisomes or for synthesis of a eubacterial cell wall. Yet many of the original cyanobacterial functions still exist in plastids, energy metabolism and amino acid biosynthesis for example, but the genes for these proteins are in the nucleus.

Do the products of genes that were transferred to the nucleus always return on a 1-to-1 basis to the organelle that donated the gene? No. Several cases of gene transfer from organelles with evolutionary rerouting of nuclear-encoded gene products are known. This can be illustrated when we consider just a small segment of plant metabolism distributed across the chloroplast and cytosol (Fig. 2). Color coding is used in the figure to summarize the evolutionary history and subcellular compartmentation of several enzymes involved in central carbohydrate metabolism in spinach (for details, see Martin and Schnarrenberger, 1997). Several aspects of the figure are noteworthy.

First, all proteins shown in the figure are encoded in the nucleus except the large subunit of Rubisco, which is plastid encoded. In several algae, chloroplast phosphoglycerate mutase and the subunits of chloroplast pyruvate dehydrogenase are also plastid encoded (Martin et al., 1998). Second, several enzymes in chloroplasts are not cyanobacterial proteins, but rather are proteobacterial proteins encoded by genes of mitochondrial origin (Martin and Schnarrenberger, 1997) and acquired in the nucleus the targeting signals (transit peptide), which redirect them to the chloroplast, where they replaced the function of the preexisting cyanobacterial homolog. Obviously, this is most likely to occur for proteins that were common to the eubacterial antecedents of both mitochondria and chloroplasts, and thus are functionally redundant in eukaryotes through endosymbiosis (Martin and Schnarrenberger, 1997). Third, higher plants possess a largely eubacterial glycolytic pathway in the cytosol. This is an unexpected finding, because the host of mitochondrial origins was, from the standpoint of today's views, either a descendant of the archaebacteria or possibly an archaebacterium outright (Doolittle, 1998). Archaebacteria possess the enzymes of the glycolytic and gluconeogenetic pathways, but for the enzymes shown in the figure, the ancestry of the "host" is not reflected as archaebacterial enzymes in the cytosol. Rather, the archaebacterial enzymes have been replaced by the products of genes that were donated to the nucleus from eubacterial symbionts, chloroplasts, and mitochondria. These genes have not acquired a region encoding a transit peptide (which is simpler than acquiring one) and therefore their products have been left "stranded" in the cytosol (Fig. 2).

There are still quite a few gray areas in the figure where either the higher plant sequences have not been determined or they are known but the gene phylogeny is insufficiently clear (in our view) to make a statement on the origin of the plant nuclear genes. Chloroplastic and cytosolic pyruvate kinases are a good example of sequenced genes with an evolutionary history that is so intriguingly complex (Hattori et al., 1995) that one cannot yet tell where the plant nuclear genes come from. Furthermore, cases are also known in which the compartmentation of individual gene products can change in different lineages over evolutionary time, such that Figure 2, if prepared for *Chlamydomonas* or *Euglena* rather than spinach, would reveal different patterns of origins and compartmentation for the enzymes of the same pathways in those organisms (for an overview, see Martin and Schnarrenberger, 1997).

THE TIMING OF EXPRESSION AND TARGETING

The finding that the products of some genes that were transferred from organelles to the nucleus have remained in the cytosol is both curious and noteworthy. The classical view of endosymbiotic gene transfer, crisply formulated by Weeden (1981), predicts that the products of transferred genes should be targeted specifically to the organelle from which the gene was donated, the product specificity corollary. Under this view, the process of gene transfer would proceed in two stages (Fig. 3, left). First, a copy of the gene would enter the nucleus (by whatever means), but in the same cell (and its descendants) many organelles with many genomes per organelle would still retain the organellar copy, so that a transient state would exist where the gene is potentially active in both compartments. Real examples of such a two-functional copy state are (still) not known, but cases are known where a functional transferred gene exists in the nucleus and a degenerate copy persists in the organelle (Brennicke et al., 1993): in the example of mitochondrial rps19 in Arabidopsis, a defective copy is found in the mitochondrion, and a recently transferred copy with newly acquired domains of ribosomal function is active in the nucleus (Sanchez et al., 1996). For the product of a transferred gene to be reimported, the process of nuclear integration would have to proceed to supply the nuclear-localized organellar gene with the proper organelle-targeting signal, the transit peptide, almost simultaneously with integration. If that occurs, then in the second stage, the organelle copy can become defective and be lost, thereby completing the process of gene transfer. Clearly, the transferred gene has to solve two problems to permit loss of the organellar copy: expression and targeting (Herrmann, 1997; outlined in Fig. 3).

The genetic systems of plastids and the nucleus are quite different. Genome and gene organization in plastids is generally, but not purely, prokaryotic. Transcription occurs with the help of both a plastid-encoded prokaryotic-type RNA polymerase and a nuclear-encoded, single-chain, phage-like, 110-kD RNA polymerase (scRPO) (Hedtke et al., 1997), and different sets of genes are specifically transcribed by these two types of polymerases in plastids (Hajdukiewicz et al., 1997; Kapoor et al., 1997). So when a plastid gene finds its way to the nucleus, by whatever means, it moves from a genetic apparatus that is compact, operon harboring, and intron poor, to one that is inflated, operon splitting, and intron laden (Herrmann, 1997). Such a gene has an immediate problem. It must become selected to avoid becoming a pseudogene ("promiscuous DNA"; see Brennicke, 1993; Herrmann, 1997), and its product must become expressed before selection can eliminate deleterious variants arising from the constant pressure of mutation. If the gene product is to compete with its many plastid-encoded homologs, for selection to work in the

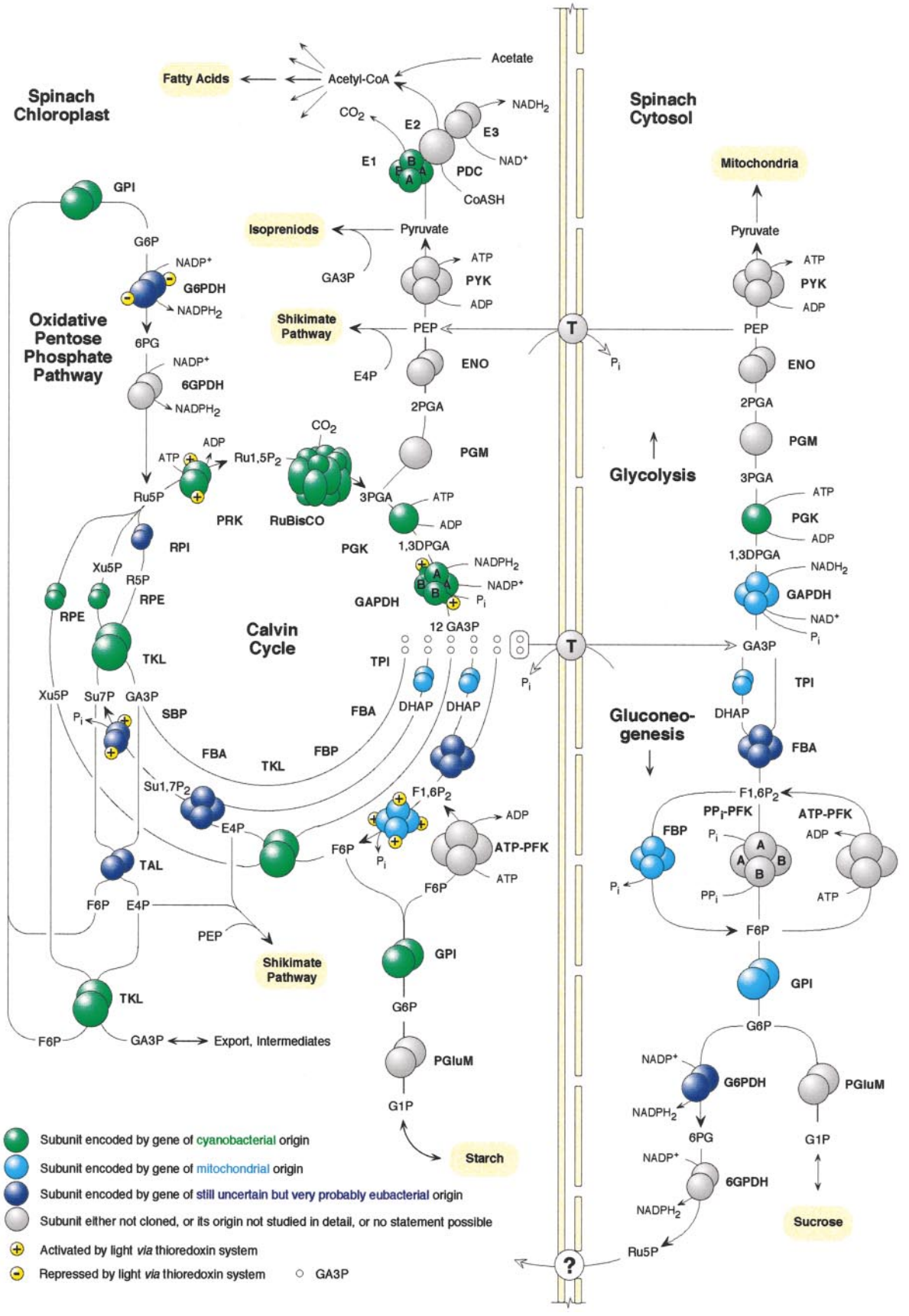

Figure 2. (Legend appears on facing page.)

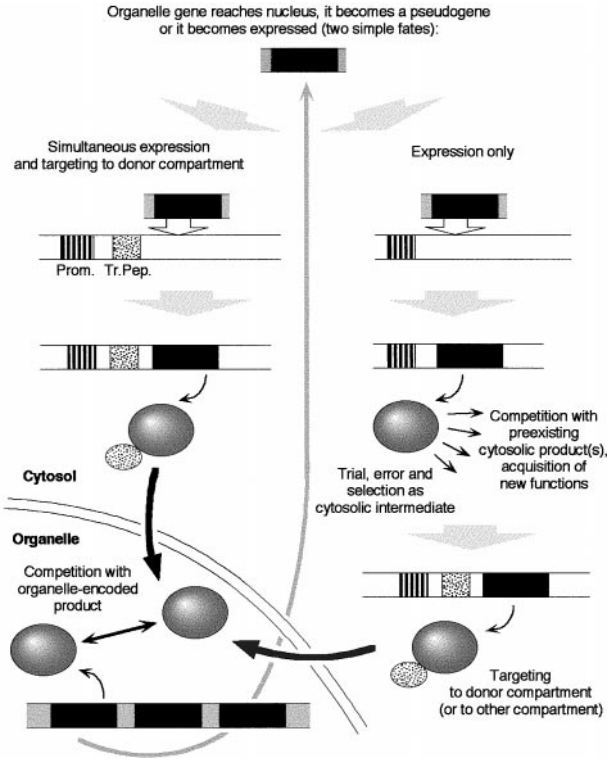

Figure 3. Schematic diagram contrasting the fate of products of transferred genes under two simple scenarios (see text). Prom., Promoter and other regulatory elements necessary for gene expression; Tr.Pep., transit peptide. Gray arrows symbolize processes in time.

organelle, its gene must acquire adequate expression elements and proper targeting signals.

Obviously, targeting signals are useless without expression. Establishment of adequate expression of transferred genes in the nucleus could, in principle, exert a more serious limitation to endosymbiotic gene transfer than the acquisition of the routing signals necessary for directing the gene product to its proper compartment. This view is substantiated by the finding that acquisition of a transit peptide is not as difficult as one might think: 1 in 30 randomly cloned sequences from *Escherichia coli* DNA successfully directed the import of proteins into mitochondria in yeast (Baker and Schatz, 1987). By contrast, stable expression of promoterless constructs into plant nDNA (without selection) occurs at a much lower frequency (Herman et al., 1990), suggesting that, in general, successful stable expression, rather than acquisition of proper targeting signals, might be rate limiting for the integration of chloroplast genes into nuclear chromosomes over geological time. Thus, expression levels sufficient to supply all plastids with product in addition to proper routing are prerequisites for loss of the organellar copy. Stable, sufficient expression may be more difficult to attain than routing signals, suggesting that some genes that are transferred to the nucleus from plastids or mitochondria may undergo a transient phase in which the gene becomes expressed, but without a transit peptide. In that case, the encoded product would be a cytosolic protein until the gene acquires the proper routing signal.

THE CYTOSOL FIRST, THEN THE ORGANELLE?

Let us briefly entertain the notion that cytosolic localization of gene products, which ultimately descend from plastids or mitochondria, such as higher plant cytosolic phosphoglycerate kinase (Martin and Schnarrenberger, 1997), might represent a natural, intermediate stage in the gene-transfer process. Under this view, the first step of endosymbiotic gene transfer would entail successful integration and nuclear expression of a transferred gene prior to the acquisition of a viable organelle-targeting signal (outlined in Fig. 3, right). The product so expressed would wander about the cytosol, possibly interfering with preexisting cytosolic functions, or, in the case of functionally redundant enzymes, competing with them. We consider four simple fates for such transferred genes, depending upon the interactions of their products with cytosolic proteins.

(1) If the donated gene product satisfies the needs of the cell better than the preexisting cytosolic product, as has

Figure 2. (Figure appears on facing page.)
Localization of several enzymes of central carbohydrate metabolism in spinach. Suggested evolutionary origins for the nuclear genes are color coded. Enzymes regulated through the thioredoxin system are indicated. Many enzymes in the figure are allosterically regulated, but no allosteric regulation is indicated here. Enyzme abbreviations are: FBA, Fru-1,6-bisphosphate aldolase; FBP, Fru-1,6-bisphosphatase; GAPDH, glyceraldehyde-3-P dehydrogenase; PGK, 3-phosphoglycerate kinase; PRI, Rib-5-P isomerase; PRK, phosphoribulokinase; RPE, ribulose-5-P 3-epimerase; SBP, sedoheptulose-1,7-bisphosphatase; TKL, transketolase; TPI, triosephosphate isomerase; TAL, transaldolase; GPI, Glc-6-P isomerase; G6PDH, Glc-6-P dehydrogenase; 6GPDH, 6-phosphogluconate dehydrogenase; pGluM, phosphoglucomutase; PGM, phosphoglyceromutase; PFK, phosphofructokinase (pyrophosphate and ATP-dependent); ENO, enolase; PYK, pyrtuvate kinase; PDC, pyruvate dehydrogenase complex (E1, E2, E3 components); and T, translocator. PDC is a multienzyme complex, but only one set of components is drawn here. Note that chloroplast isoenzymes of PGM, ENO, and PYK have not been demonstrated in spinach leaves, but for convenience we have included those enzymes in this figure, since they have been well characterized in the plastids of other higher plants (Plaxton, 1996). Open arrowheads indicate transport rather than conversion. Solid arrowheads indicate physiologically irreversible reactions. For details, see Martin and Schnarrenberger (1997), Johnston et al. (1997; A and B subunits of pyruvate dehydrogenase), Nowitzki et al. (1998; chloroplast and cytosolic GPI), and Wenderoth et al. (1997; chloroplast and cytosolic G6PDH). It seems likely that the nuclear genes for chloroplast and cytosolic FBA (C. Schnarrenberger, personal communication) and G6PDH (A. von Schaewen, personal communication) are of mitochondrial origin, but archaebacterial sequences are still not known for comparison, so the color coding for these enzymes is dark. See also Fischer et al. (1997) and Lange et al. (1998).

been suggested for the origin of the eubacterial glycolytic pathway in the eukaryotic cytosol (for a discussion, see Martin and Schnarrenberger, 1997; Martin and Müller, 1998; Nowitzki et al., 1998), then it can be expected that the gene for the cytosolic protein would be fixed and the speed of fixation would in some way be proportional to the degree of benefit; examples of such cytosolic rerouting are evident in Figure 1. It would only be a matter of time before fortuitous duplication events (Kadowaki et al., 1996) or exon shuffling (Long et al., 1996) gave rise to a copy with proper routing signals, which could eventually compete with the organelle-encoded protein in the organelle to permit loss of the organellar copy.

(2) If the preexisting and intruding products are functionally equivalent, then it becomes a matter of chance as to which one survives. For each "attempted" transfer event, accumulation of mutations could be expected before a successful attempt results in a properly expressed copy in the nucleus encoding a properly routed product.

(3) If the function of the intruding product is more poorly suited to the needs of the cell than the preexisting cytosolic product, or if it has no preexisting counterpart with which to compete, then it will be freed from selection, and will rapidly accumulate mutations. In this case, it is a matter of time before the gene (a) becomes a pseudogene, (b) mutates into something that may be otherwise useful for the cell, or (c) acquires routing signals to get the product into the compartment from which the gene came into a different compartment (see Fig. 1). An interesting recent example in which transferred genes may have mutated to encode something different (but useful) can be found in the chloroplast protein import machinery (Heins and Soll, 1998).

(4) If the intruding product interferes with cytosolic functions in a manner that is detrimental to the cell, nuclear expression will be strongly counterselected. The degree of detriment would behave in a manner proportional to the life span of the transferred gene, and this has been suggested by R.-B. Klösgen (personal communication) as a mechanism that might explain why certain genes tend to remain in organelles, rather than be transferred to the nucleus (Herrmann, 1997). Curiously, one of the early events in eukaryotic programmed cell death (apoptosis) is the export of a mitochodrial protein (Cyt *c*) into the cytosol, a compartment where the protein does not belong (Bossy-Wetzel et al., 1998), indicating that some gene products can indeed be unhealthy when localized in the wrong compartment.

All four of the above possibilities predict that genes, which ultimately gave rise to the properly routed proteins, should have undergone a period of evolution in which the encoded product was freed from selection, or in which new selective pressures were in effect until the product was able to return to the donor organelle. This can be expected to result in a phase of more rapid accumulation of mutations in such genes, and thus in some degree of structural discontinuity in nuclear copies of organelle genes relative to their organelle-encoded counterparts.

Are examples known where integrated nuclear genes of organelle origin pick up new transit peptides from preexisting nuclear sequences? Yes. Kadowaki et al. (1996) reported a very clear-cut case involving nuclear-encoded genes for a mitochondrial ribosomal protein, rps11. There are two genes for rps11 (*RPS11-1* and *RPS11-2*) in the rice nuclear genome that share 92% sequence identity, and a pseudogene for rps11 (still) exists in the mitochondrial genome. Both nuclear copies encode N-terminal transit peptides, and in a rare case we can see by sequence homology where the transit peptides come from. In the case of *RPS11-1*, part of the transit peptide was stolen from the transit peptide in the nuclear gene for mitochondrial atpB, a component of the ATPase. That is, part of the "same" transit peptide is found on two different nuclear genes, *RPS11-1* and *ATPB*. In the case of *RPS11-2*, the region encoding the mature mitochondrial subunit has stolen part of its transit peptide from part of the transit peptide in the rice nuclear gene for mitochondrial Cyt *c* oxidase subunit Vb (*COXVB*). These visible recombination events underscore the role of duplication and recombination in the acquisition of transit peptides (Kadowaki et al., 1996).

Examples are also known where exon shuffling plays a role in the acquisition of transit peptides. A particularly clear case was reported by Long et al. (1996), who found that the transit peptide for mitochondrial Cyt *c* in potato was acquired by exon shuffling between the nuclear gene for Cyt *c* and a gene for a cytosolic protein, glyceraldehyde-3-P dehydrogenase. The 41-amino acid N-terminal transit peptide of the nuclear Cyt *c* gene was stolen from the first three exons of a gene for a glycolytic enzyme, functionally converting a peptide that used to be part of a NAD^+-binding domain into a mitochondrial targeting sequence (Long et al., 1996). Exon shuffling can have an additional effect on the fate of transferred genes, since introns themselves can directly influence gene expression at various levels (Rose and Last, 1997). Thus, whether by recombination in coding sequences or in introns, these examples indicate that transit peptides are indeed not too difficult to acquire once a gene is in the nucleus and functioning.

WHY TRANSFER TO THE NUCLEUS?

Why should genes tend to be transferred from chloroplasts to the nucleus during evolution in the first place? Several possible factors that might favor the transfer of genes to the nucleus were discussed in detail by Allen and Raven (1996), who carefully outlined the importance of redox-associated functions in organelles that might increase the free-radical-induced mutagenic load for genes in organelles, thus favoring their transfer to the nucleus. This is certainly one important factor. Are other factors imaginable?

Could it be that complex gene regulation is only possible in the nucleus? Hardly, because plastid gene expression is regulated, and in a reasonably complex manner (Herrmann, 1997), both at the transcriptional (Hajdukiewicz et al., 1997) and posttranscriptional (Bock and Koop, 1997) levels. Gene regulation in plastids is not simply a miniature prokaryotic system, rather, it is part of an integrated eukaryotic gene regulation system (Herrmann, 1997). This is manifested, for example, in the findings that the genes for σ factors required by the plastid-encoded RNA core polymerase are themselves expressed and regulated by the

nuclear apparatus (Tanaka et al., 1996, 1997), as is the gene for the single-chain RNA core polymerase of plastids (Hedtke et al., 1997). So gene regulation in plastids is subordinate to the nucleus, but that does not directly explain why genes tend to accumulate there, rather than reside in plastids (where they can be regulated as well). If not regulation, what then?

Plastids were once free-living bacteria. When the first plastid entered its host eons ago, it immediately became genetically isolated from its free-living relatives. Upon endosymbiosis, it probably became clonal, asexual. What happens to any organism/population when it is deprived of sex? It cannot recombine out the deleterious mutations that are inevitably going to accumulate in its genome. This phenomenon is known as Muller's ratchet (Muller, 1964). Given sufficient time, Muller's ratchet is thought to ultimately doom asexual populations or species to inescapable extinction. How does Muller's ratchet figure into gene transfer from organelles to the nucleus? When a gene is successfully transferred to the nucleus, it moves from a predominantly asexual to a predominantly sexual genome, restoring recombination, and freeing the gene from the fate of mutational meltdown. In the long term, this factor therefore strongly favors the transfer of genes to the nucleus, a return to recombination.

Is there evidence for the effect of Muller's ratchet in organelle genomes? In general, yes. Muller's ratchet has been shown to effect a rapid accumulation of (probably deleterious) substitutions in tRNA genes of animal mitochondria (Lynch, 1996). Furthermore, proteobacteria that have lived as stably transmitted endosymbionts for many of millions of years, genetically isolated in the body cavity of aphids, also show clear signs of Muller's ratchet in their genomes, manifested as elevated levels of accumulated substitutions in various genes relative to their free-living cousins (Moran, 1996).

In plant mitochondria and chloroplasts, however, the situation is more complicated. In these organelles, the effect of Muller's ratchet is much less pronounced than in animal mitochondria (Lynch, 1997). Perplexingly, the rate of nucleotide substitution in plant organelles is not higher than in the nucleus, as Muller's ratchet would predict, rather, it is lower (Wolfe et al., 1987). This suggests that compensatory factors are at work in plant organelles, which counteract the long-term effects of asexuality. The most obvious of these is genetic recombination between organelles, as is well known in chloroplasts of *Chlamydomonas* (Fischer et al., 1996). Compensation might also be provided by the high polyploidy levels of chloroplasts, which permit recombination between genomes within the same plastid such that deleterious alleles on ctDNA could be sorted out. Plastids do import a nuclear-encoded homolog of cyanobacterial RecA that is functionally involved in recombination and repair in chloroplasts, suggesting that these pathways may be similar in eubacteria and plastids (Cerutti et al., 1995).

Other DNA repair pathways that might help to account for the lower rate of nucleotide substitution, such as nucleotide excision repair, have not yet been characterized from plastids. But one gene involved in this pathway (*mutS*) is encoded in the mitochondrial genome of an animal, where it might influence the mitochondrial substitution rate (Pont-Kingdon et al., 1995). DNA repair in plant organelles may have an influence on their lower nucleotide substitution rate, and by lowering the rate of mutation, longer times would be needed for Muller's ratchet to take effect. Obviously, Muller's ratchet alone does not account in full for the transfer of genes from organelles to the nucleus, but in the early phases of organelle origins, when the majority of organelle genome shrinkage is thought to have occurred, it may have played a prominent role.

WHICH GENES GO FIRST, WHICH GO LAST?

If we were to wait 500 million years and then redo Table I for the same organelle genomes, we would probably find fewer numbers of genes left, and in some cases the number would possibly reach zero. Fortunately, we do not need to wait that long, because in some eukaryotic organelles, hydrogenosomes, the genome has already been assimilated in toto by the nuclear genome. Hydrogenosomes are double-membrane-bounded, ATP-producing organelles of amitochondriate protists; they descend from the same symbiont as mitochondria, but no hydrogenosomes are known (yet) that possess a genome (Martin and Müller, 1998). In hydrogenosomes the process of endosymbiotic gene transfer has gone to completion. From contemporary mitochondrial genomes we can obtain an impression of what types of genes are the last to be lost from these organelles: those for translation and respiration (Fig. 1). In hydrogenosomes no respiration occurs, so there is nothing left to translate, hence proteins of translation, tRNAs and the rRNAs, can be lost as well (Herrmann, 1997).

Conversely, if we were to turn back the clock a billion years or so, we would be able to ask, "Which genes are the first to be lost from organelles?" This is a more difficult question, but from the standpoint of today's data it appears that genes for regulatory functions tend to be fixed in the nucleus more readily than enzymatic or structural functions (for an overview, see Herrmann, 1997). Examples of this are readily visible in several chloroplast multisubunit proteins (in addition to the σ factors for the plastid-encoded RNA polymerase mentioned above). The γ-subunit of the chloroplast ATPase (atpC) possesses a regulatory role for the ATPase complex. The structural subunits atpA, atpB, atpE, atpF, and atpH of the ATPase are encoded in all land plant chloroplast genomes sequenced to date, but the regulatory subunit, atpC, is nuclear encoded in all plants studied to date, and is therefore probably the first gene of this complex to have been transferred (Herrmann, 1997). The same tendency in structure-function distribution can be found within the chloroplast Clp protease: the catalytic subunit (ClpP) is ctDNA encoded in all higher plants studied, whereas the regulatory subunit (ClpC) is nuclear encoded (Martin et al., 1998). Even within Rubisco, a similar hierarchy in distribution of plastid-encoded catalysis (rbcL) and nuclear-encoded regulation (rbcS) can be found. These are possibly manifestations of a general tendency for regulation (genes and proteins) to be concentrated in the nucleus.

In conclusion, many factors figure into endosymbiotic gene transfer. Over time, gene flow within the cell ultimately trickles into the nuclear sink. But once a gene has been transferred, the nuclear sink becomes the source from which the gene product can apparently flow in any direction. If a functionally equivalent gene product becomes routed to an organelle, the organelle gene can eventually be lost. Here we have outlined a few factors that could help to explain why the brunt of organelle genome reduction took place early in evolution, which could also help to explain the slower, but constant flow of genes to the nucleus until today.

ACKNOWLEDGMENTS

We thank Axel Brennicke, Rainer Figge, Ulrich Nowitzki, Antje von Schaewen, and Claus Schnarrenberger for critical discussions.

Received March 30, 1998; accepted May 1, 1998.

LITERATURE CITED

Allen JF, Raven JA (1996) Free-radical-induced mutation versus redox regulation: costs and benefits of genes in organelles. J Mol Evol **42:** 482–492

Baker A, Schatz G (1987) Sequences from a prokaryotic genome or the mouse dihydrofolate reductase gene can restore the import of a truncated precursor protein into yeast mitochondria. Proc Natl Acad Sci USA **84:** 3117–3121

Bock R, Koop H-U (1997) Extraplastidic site-specific factors mediate RNA editing in chloroplasts. EMBO J **16:** 3282–3288

Bossy-Wetzel E, Newmeyer DD, Green DR (1998) Mitochondrial cytochrome *c* release in apoptosis occurs upstream of DEVD-specific caspase activation and independently of mitochondrial transmembrane depolarization. EMBO J **17:** 37–49

Brennicke A, Grohmann L, Hiesel R, Knoop V, Schuster W (1993) The mitochondrial genome on its way to the nucleus: different stages of gene transfer in higher plants. FEBS Lett **325:** 140–145

Cerutti H, Johnson AM, Boynton JE, Gilham NW (1995) Inhibition of chloroplast DNA recombination and repair by dominant negative mutants of *Eschericia coli* RecA. Mol Cell Biol **15:** 3003–3011

Doolittle WF (1998) A paradigm gets shifty. Nature **392:** 15–16

Ellis JR (1981) Chloroplast proteins: synthesis, transport, and assembly. Annu Rev Plant Physiol **32:** 111–137

Fischer K, Kammerer B, Gutensohn M, Arbinger B, Weber A, Häusler RE, Flügge UI (1997) A new class of plastidic phosphate translocators: a putative link between primary and secondary metabolism by the phospho*enol*pyruvate/phosphate antiporter. Plant Cell **9:** 453–462

Fischer N, Stampacchia O, Redding K, Rochaix JD (1996) Selectable marker recycling in the chloroplast. Mol Gen Genet **251:** 373–380

Gray M (1992) The endosymbiont hypothesis revisited. Int Rev Cytol **141:** 233–357

Hajdukiewicz PTJ, Allison LA, Maliga P (1997) The two RNA polymerases encoded by the nuclear and the plastid compartments transcribe distinct groups of genes in tobacco plastids. EMBO J **16:** 4041–4048

Hattori J, Baum BR, Mchugh SG, Blakeley SD, Dennis DT, Miki BL (1995) Pyruvate kinase isozymes: ancient diversity retained in modern plant cells. Biochem Syst Ecol **23:** 773–780

Hedtke B, Börner T, Weihe A (1997) Mitochondrial and chloroplast phage-like RNA polymerases in *Arabidopsis.* Science **277:** 809–811

Heins L, Collison I, Soll J (1998) The protein translocation apparatus of chloroplast envelopes. Trends Plant Sci **3:** 56–61

Heins L, Soll J (1998) Chloroplast biogenesis: mixing the prokaryotic and the eukaryotic? Curr Biol **8:** R215–R217

Herman L, Jacobs A, Van Montagu M, Depicker A (1990) Plant chromosome/marker gene fusion assay for study of normal and truncated T-DNA integration events. Mol Gen Genet **224:** 248–256

Herrmann RG (1997) Eukaryotism, towards a new intrepretation. *In* HEA Schenk, RG Herrmann, KW Jeon, W Schwemmler, eds, Eukaryotism and Symbiosis. Springer, Heidelberg, Germany, pp 73–118

Herrmann RG, Oelmüller R, Bichler J, Schneiderbauer A, Stephuhn J, Wedel N, Tyagi AK, Westhoff P (1991) The thylakoid membrane of higher plants: genes, their expression and interaction. *In* RG Herrmann, BA Larkins, eds, Plant Molecular Biology 2: NATO ASI Series A, Life Sciences, Vol 212. Plenum Press, New York, pp 411–427

Johnston ML, Luethy MH, Miernyk JA, Randall DD (1997) Cloning and molecular analyses of the *Arabidopsis thaliana* plastid pyruvate dehydrogenase subunits. Biochim Biophys Acta **1321:** 200–206

Kadowaki K-I, Kubo N, Ozawa K, Hirai A (1996) Targeting presequence acquisition after mitochondrial gene transfer to the nucleus occurs by duplication of existing target signals. EMBO J **15:** 6652–6661

Kaneko T, Sato S, Kotani H, Tanaka A, Asamizu E, Nakamura Y, Miyajima N, Hirosawa M, Sugiura M, Sasamoto S, and others (1996) Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. DNA Res **3:** 109–136

Kapoor S, Suzuki J-Y, Sugiura M (1997) Identification and functional significance of a new class of non-consensus-type plastid promoters. Plant J **11:** 327–337

Lang BF, Burger G, O'Kelly CJ, Cedergren R, Golding GB, Lemieux C, Sankoff D, Turmel M, Gray MW (1997) An ancestral mitochondrial DNA resembling a eubacterial genome in miniature. Nature **387:** 493–497

Lange BM, Wildung MR, McCaskill D, Croteau R (1998) A family of transketolases that directs isoprenoid biosynthesis via a mevalonate-independent pathway. Proc Natl Acad Sci USA **95:** 2100–2104

Long M, de Souza SJ, Rosenberg C, Gilbert W (1996) Exon shuffling and the origin of the mitochondrial targeting function in plant cytochrome *c*1 precursor. Proc Natl Acad Sci USA **93:** 7727–7731

Lynch M (1996) Mutation accumulation in transfer RNAs: molecular evidence for Muller's ratchet in mitochondrial genomes. Mol Biol Evol **13:** 209–220

Lynch M (1997) Mutation accumulation in nuclear, organelle and prokaryotic transfer RNA genes. Mol Biol Evol **14:** 914–925

Martin W, Müller M (1998) The hydrogen hypothesis for the first eukaryote. Nature **392:** 37–41

Martin W, Schnarrenberger C (1997) The evolution of the Calvin cycle from prokaryotic to eukaryotic chromosomes: a case study of functional redundancy in ancient pathways through endosymbiosis. Curr Genet **32:** 1–18

Martin W, Stoebe B, Goremykin V, Hansmann S, Hasegawa M, Kowallik KV (1998) Gene transfer to the nucleus and the evolution of chloroplasts. Nature **393:** 162–165

Moran NA (1996) Accelerated evolution and Muller's ratchet in endosymbiotic bacteria. Proc Natl Acad Sci USA **93:** 2873–2878

Muller HJ (1964) The relation of recombination to mutational advance. Mut Res **1:** 2–9

Nowitzki U, Flechner A, Kellermann J, Hasegawa M, Schnarrenberger C, Martin W (1998) Eubacterial origin of eukaryotic nuclear genes for chloroplast and cytosolic glucose-6-phosphate isomerase: sampling eubacterial gene diversity in eukaryotic chromosomes through symbiosis. Gene **214:** 205–213

Oda K, Yamato K, Ohta E, Nakamura Y, Takemura M, Nozato N, Akashi K, Kanegae T, Ogura Y, Kohchi T, and others (1992) Gene organization deduced from the complete sequence of

liverwort *Marchantia polymorpha* mitochondrial DNA: a primitive form of plant mitochondrial genome. J Mol Biol **223:** 1–7

Pillen K, Schondelmaier J, Jung C, Herrmann RG (1996) Genetic mapping of genes for twelve nuclear-encoded polypeptides associated with the thylakoid membranes in *Beta vulgaris* L. FEBS Lett **395:** 58–62

Plaxton WC (1996) The organization and regulation of plant glycolysis. Annu Rev Plant Physiol Plant Mol Biol **47:** 185–214

Pont-Kingdon GA, Okada NA, Macfarlane JL, Beagley CT, Wolstenholme DR, Cavalier-Smith T, Clark-Walker GD (1995) A coral mitochondrial *mutS* gene. Nature **375:** 109–111

Rose AB, Last RL (1997) Introns act post-transcriptionally to increase expression of the *Arabidopsis thaliana* tryptophan pathway gene PAT1. Plant J **11:** 455–464

Sanchez H, Fester T, Kloska S, Schroeder W, Schuster W (1996) Transfer of *rps19* to the nucleus involves the gain of an RNP-binding motif which may functionally replace RPS13 in *Arabidopsis* mitochondria. EMBO J **15:** 2138–2149

Schatz G, Dobberstein B (1996) Common principles of protein translocation across membranes. Science **271:** 1519–1526

Tanaka K, Oikawa K, Ohta N, Kuroiwa H, Kuroiwa T, Takahashi H (1996) Nuclear encoding of a chloroplast RNA polymerase sigma subunit in a red alga. Science **272:** 1932–1935

Tanaka K, Tozawa Y, Mochizuki N, Shinozaki K, Nagatani A, Wakasa K, Takahashi H (1997) Characterization of three cDNA species encoding plasmid RNA polymerase sigma factors in *Arabidopsis thaliana*: evidence for the sigma factor heterogeneity in higher plant plastids. FEBS Lett **413:** 309–313

Unseld M, Marienfeld JR, Brandt P, Brennicke A (1997) The mitochondrial genome of *Arabidopsis thaliana* contains 57 genes in 366924 nucleotides. Nat Genet **15:** 57–61

Weeden NF (1981) Genetic and biochemical implications of the endosymbiotic origin of the chloroplast. J Mol Evol **17:** 133–139

Wenderoth I, Scheibe R, von Schaewen A (1997) Identification of the cysteine residues involved in redox modification of plant plastidic glucose-6-phosphate dehydrogenase. J Biol Chem **272:** 26985–26990

Wolfe KH, Li W-H, Sharp P (1987) Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proc Natl Acad Sci USA **84:** 9054–9058