Gene transfer from organelles to the nucleus: Frequent and in big chunks

William Martin*

Institute of Botany III, Universität Düsseldorf, Universitätstrasse 1, D-40225 Düsseldorf, Germany

Chloroplasts arose >1.2 billion years ago (1) when a free-living cyanobacterium became an endosymbiont in a eukaryotic host. Since that time, chloroplast genomes have undergone severe reduction, because chloroplast genomes encode between 50 and 200 proteins, whereas cyanobacterial genomes encode several thousand. Accordingly, endosymbiotic theories have always assumed that the cyanobacterial ancestor of plastids relinquished much of its genetic autonomy: "it is not surprising that chloroplasts lost their ability to live independently long ago," as Mereschkowsky put it in 1905 (2). In today's terms, that means that during the course of evolution, genes must have been transferred from the ancestral chloroplast to the nucleus, where they acquired the proper expression and targeting signals to allow the encoded proteins to be synthesized on cytosolic ribosomes and reimported into the organelle with the help of a transit peptide. This process, a special kind of lateral gene transfer called endosymbiotic gene transfer (3), appears to be very widespread in nature: ≈18% of the nuclear genes in *Arabidopsis* seem to come from cyanobacteria (4), and obvious remnants of the chloroplast DNA have been found in higher plant nuclear chromosomes (5). Evolutionary biologists have long been able to infer endosymbiotic gene transfer from evolutionary sequence comparisons but have not been able to watch it happen in the lab until now. In this issue of PNAS, Stegemann *et al.* (6) report gene transfer from the tobacco chloroplast genome to nuclear chromosomes under laboratory conditions. Their findings, together with other recent developments, open up new chapters in our understanding of organelle–nuclear DNA dynamics and have far-reaching evolutionary implications.

The experimental design used by Stegemann *et al.* (6) was simple and effective. Using a technology called chloroplast transformation (7), they introduced a cassette containing two foreign genes into tobacco chloroplast DNA. The first one encoded spectinomycin resistance (*aad*) under the control of a chloroplast-specific promoter; the second one encoded kanamycin resistance (*npt*) under the control of a nuclear-specific promoter. They took advantage of the fact that whole tobacco plants can be regenerated from single cells. By subjecting transformed tobacco tissues to several rounds of selection on medium containing spectinomycin, they were able to obtain tobacco plants that were homoplastomic for *aad* and *npt*; that is, all copies of the chloroplast DNA in all plastids in those plants contained the new cassette. By placing small sections of leaves from those *aad*/*npt* homoplastomic lines on kanamycin-containing medium, they initiated selection for strong expression of the *npt* gene under the control of the nuclear-specific promoter. That was the key step, because on kanamycin medium, only such tobacco cells will survive whose nuclear DNA has incorporated a segment of the genetically modified chloroplast DNA containing the new *npt* gene.

With that simple scheme, they obtained 12 independent kanamycin-resistant regenerant plants, and using some rough calculations, they estimate that ≈1 in every 5,000,000 tobacco leaf cells that they assayed contained a highly expressed, independently transferred, chloroplast-derived *npt* gene in the nucleus. But was the *npt* gene in those 12 plants in the nucleus? Simple genetics says yes: the resistance gene behaved as a normal Mendelian dominant marker. Using their resistant plants as pollen donors in crosses to wild-type tobacco, they obtained a 1:1 ratio of kanamycin-resistant to kanamycin-sensitive progeny. Because chloroplast DNA is not transmitted through the pollen in tobacco (8), this ratio means but one thing: the *npt* gene, which they had originally inserted into chloroplast DNA, had found its way via a natural mechanism from the chloroplast to the nucleus and was being expressed there.

On an evolutionary or environmental scale, 1 in 5,000,000 cells is a whopping number; to some it will be unbelievable (9) and by no means will it make everybody happy. Some biotechnologists are adamant that foreign genes introduced into the chloroplast can be sequestered there and thus will not escape via pollen (introgress) from cultivated fields into wild species like nuclear genes can (9). So for many the first question will be: Are these surprising findings reproducible? The answer is yes (10). In independent work, Huang *et al.* (11) reported chloroplast-to-nucleus gene transfer also by using the *npt* gene. They transformed a different region of the chloroplast genome, placed a typical nuclear intron in the *npt* gene, and used a different approach to look for kanamycin resistance (11). Whereas Stegemann *et al.* (6) took the low road, assaying leaf tissue on a few convenient Petri dishes and finding 12 independent transfers among some 60,000,000 vegetative cells, Huang *et al.* (11) took the high road, assaying seeds from pollen outcrosses of homoplastomic plants and finding 16 independent transfers among 250,000 male gametes so tested. Of course, it is one thing to get a transferred gene expressed when it brings along its own nuclear promoter, as in the present findings (6, 11), but quite another to acquire a good promoter through recombination (9, 10). But with two independent laboratories reporting massive rates of chloroplast DNA escape to the nucleus in laboratory regimens, it is time to consider the mechanistic and evolutionary implications of such findings.

In what physical form are these chloroplast genes making their way to the nucleus? In principle, there are three simple possibilities: as bulk chloroplast DNA, mRNA, or cDNA (possibly virus-mediated). Stegemann *et al.* (6) checked the pollen-outcrossed progeny that possessed only the expressed nuclear copy of the *npt* gene to see whether the *aad* gene was still physically linked to *npt*. It was, suggesting that a contiguous piece of bulk chloroplast DNA had escaped from the plastid and had recombined into a nuclear chromosome. Huang *et al.* (11) made the same observation, and because they had furthermore inserted a nuclear-specific intron GT–AG into their *npt* gene, the involvement of an mRNA or cDNA intermediate in their 16 organelle-to-nucleus transfer events can reasonably be excluded. That raises the question of whether bulk DNA recombination is also involved in chloroplast-to-nucleus gene transfer events in nature.

See companion article on page 8828.

*E-mail: w.martin@uni-duesseldorf.de.

www.pnas.org/cgi/doi/10.1073/pnas.1633606100

A substantial and growing body of evidence indicates this to be the case. The *Arabidopsis* genome revealed 17 different tRNA- and intron-containing insertions of recently integrated chloroplast DNA in nuclear chromosomes, indicating that recombination between nuclear DNA and escaped chloroplast DNA is at work (12). However, only 11 kb total of recently integrated chloroplast DNA was found in that genome, fueling arguments that transfers of this type may play only a minor role in nature (9). But the rice genome tells a different story: chromosome 10 alone contains a recent 33-kb insertion of chloroplast DNA in addition to a 131-kb insertion representing nearly the entire plastid genome (13). Such transfers can hardly involve cDNA.

Is bulk DNA integration specific to chloroplasts, or do mitochondria also occasionally donate a copy of their genome to the nucleus? *Arabidopsis* chromosome 2 contains a complete copy of the 367-kb mitochondrial genome near the centromere (14, 15), rice chromosome 10 contains 57 fragments of mitochondrial DNA ranging from 80 to 2,500 bp in size (13), the yeast genome contains 34 small fragments of mitochondrial DNA (16), and even the human genome contains a hefty spattering of 59 fragments of mitochondrial DNA >2 kb in length, totaling 280 kb, including one 14.6-kb nearly complete copy of the human mitochondrial genome (17). This means that bulk DNA transfer from organelles to the nucleus of the type that Stegemann *et al.* (6) found is no fluke and no experimental artifact. It is real, it is very widespread, and in rare cases it can even result in mutations that cause human disease (18).

So how does intact chloroplast DNA physically get to the nucleus? Is it extruded from intact chloroplasts, is it pulled, is it packaged and delivered, is there a specific machinery to do this, or is it a chance byproduct of occasional chloroplast lysis, which releases genome copies to the cytosol that become available substrates for nuclear recombination? New results from Chris Howe's laboratory provide incisive clues, suggesting the latter to be true (19). If the mechanism of transfer involves chloroplast lysis, then a firm prediction follows: organisms that only possess one chloroplast per cell should be unable to undergo transfer of the type observed in tobacco, which possesses ≈100 chloroplasts per cell. This is because plant cells that lose their only chloroplast will not survive. Lister *et al.* (19) made similar constructs to those used in the tobacco experiments (6, 11), but they transformed the green algae *Chlamydomonas reinhardtii*. *C. reinhardtii* has one (and only one) chloroplast per cell. They examined ≈13 billion homoplastomic transformants looking for a single stable nuclear transfer, and found none at all. Of course, that negative result could be due to any number of things, so they checked to see whether nature had already done the experiment for them. They compared the *Chlamydomonas* chloroplast genome sequence to the complete nuclear genome sequence data available for *Chlamydomonas* and found nothing, no large transferred chloroplast DNA fragments of the type typical of higher plant genomes, no small fragments, nothing. Lister *et al.*'s (19) findings suggest that there may indeed be a causal connection between the number of plastids that can be killed for donating DNA and the frequency of organelle-to-nucleus transfers.

Taken together, the present data suggest that chloroplast DNA reaches the nucleus as intact pieces of the chloroplast genome or, in the case of rice chromosome 10, as a complete chloroplast DNA molecule that undergoes some form of recombination with nuclear DNA. The sequences of a few border regions where the *npt* gene had been integrated revealed no sequence attributes that would provide clear hints as to what type of recombination had taken place (11). However, one further observation seems noteworthy. Huang *et al.* (11) assayed gametes (the products of meiosis) and found transfers at a frequency that is 2–3 orders of magnitude greater than the frequency found by Stegemann *et al.* (6), who assayed somatic cells (the products of mitosis). This points to a need to clarify whether meiotic recombination or the degradation of plastids during pollination might have a positive effect on organelle-to-nucleus gene transfer frequency.

An organelle-to-nucleus gene transfer mechanism supposing cDNA intermediates still finds indirect support from phylogenetic surveys among higher plants (20), because mitochondrial gene copies can be found in the higher plant nucleus that lack the introns sometimes found in the corresponding mitochondrion-encoded copies. However, mitochondrial introns are often mobile, and their mobility can mimic the involvement of cDNA mechanisms even under a bulk transfer process (21). Genome sequence data, and now laboratory work (6, 11), have brought forth abundant evidence for frequent and direct recombination of organelle chromosomes into nuclear chromosomes; if cDNAs or mRNAs play an important role in evolutionary gene transfer, evidence for that should be seen in genome sequences.

Given that DNA is being transferred at dramatic rates from organelles to the nucleus, one has to wonder why there is any DNA left in organelles at all. John F. Allen (22) has put forward what many consider to be the only thoroughly convincing explanation for the stubborn persistence of organelle genomes. His argument is simple: organelles need to be in control of the expression of genes encoding components of their electron transport chain so that they can synthesize those components as they are needed to maintain redox balance, thus avoiding the production of reactive oxygen species, which are exceedingly toxic. The still widely regarded view that particularly hydrophobic proteins cannot be imported by organelles, and hence must be encoded in organelle genomes, is distinctly at odds with the findings that the vast majority of membrane-integral proteins in chloroplasts and mitochondria (substrate transport proteins), in addition to the most hydrophobic protein in chloroplasts (the chlorophyll-binding protein), are all nuclear encoded (22).

Thus, organelles are not fundamentally unable to import hydrophobic proteins. But let us imagine a single chloroplast (among many) that requires a bit more photosystem I core protein, for example, to reestablish a balanced electron flow through its thylakoid membrane and to combat oxidative stress. If it has to ask the nucleus to supply more photosystem I core protein (which is always encoded in plastids), the nucleus will do so, but it will thereby supply more photosystem I to all of the other plastids, who will in turn immediately demand less photosystem I and more photosystem II, and so forth. Gene regulation and redox balance throughout that cell will go berserk. Neither will the plastid that had a problem in the first place be properly serviced, nor will the others who were doing fine until the first one complained. By virtue of this compelling need for gene regulation maintaining redox balance at the level of individual organelles, so argues Allen (22), organelles must be able to regulate and express particular kinds of genes themselves. Of course, that problem would be neatly solved were there only one organelle per cell, but as Lister *et al.* (19) point out, organisms with only one organelle per cell may have a fundamental problem donating their remaining organelle genes to the nucleus.

Projected into the depths of evolutionary time and into the breadth of biodiversity, organelle-to-nucleus transfer rates of the magnitude observed by Stegemann *et al.* (6) have consequences. First, they indicate that ≈1 in 5,000,000 somatic cells in the average tobacco

plant has a sizeable and freshly incorporated piece of chloroplast DNA somewhere in its genome. Those numbers are comparable to or exceed nucleotide mutation rates such that organelle-to-nucleus transfer might be pounding away at the genome, creating deleterious alleles just like mutation does (6, 18). Furthermore, Huang *et al.*'s (11) result indicates that 1 of every 16,000 tobacco plants carries a fresh chunk of chloroplast DNA in the nucleus that it acquired just one generation ago. Thus, although all plants in an average Virginia tobacco field may look very similar, they may harbor some differences with regard to what chloroplast DNA they have in their nuclei. Furthermore, these transfer rates are understimates (6, 11) because in both cases only (*i*) highly expressing transfers and (*ii*) the resistance-conferring segement of the chloroplast DNA were scored.

A constant trickle of organelle DNA into the nucleus might also be an underappreciated source of biological novelty (6). This can be illustrated if we briefly turn back the clock a billion and a half years or so and imagine the first plant cell just at the moment it obtained the cyanobacterial endosymbiont that would someday become the chloroplast. In contrast to modern chloroplasts, that first endosymbiont would not have possessed a ready-made protein import apparatus in its inner and outer membrane that would have allowed it to recognize and cleave transit peptides (23). But it surely would have been capable of donating a full cyanobacterial genome's worth of genes to the chromosomes of its host, provided that an endsymbiont lysed once in a while, which is not outrageously improbable. In the absence of the symbiont's ability to import the products of transferred genes, that initial stage of endosymbiosis would have been a genetic one-way street in the direction of the host. The result would have been competition between intruding genes and preexisting ones for the coding of cytosolic proteins (24). As long as the functions encoded by transferred genes were not worse than preexisting cytosolic functions (for example, glycolytic enzymes), the odds would work in favor of the intruding genes replacing their preexisting homologues, because if the first transfer does not work, maybe the second one will, or maybe the 25,000th, and so forth. As long as there is an expendable number of endosymbionts in the cell, their DNA is an inexhaustible source of genetic starting material for the host. Only when a protein import apparatus becomes invented can the endosymbiont start relinquishing genes to its host (organelle genome reduction), rather than just donating.

Doolittle (25) rightly called the situation just described a gene transfer ratchet, because if an endosymbiont, either chloroplast or mitochondrion, lyses, genes are transferred to the host, but if the host lyses, there are no progeny. However, he also suggested that in addition to organelles, food bacteria (many eukaryotes prey on prokaryotes) could also become a rich source of new nuclear genes: "you are what you eat" (25). Indeed, outright lateral gene transfer from bacterial donors to eukaryotes has become a popular explanatory principle (26). Even the human genome was initially claimed to contain hundreds of such bacterial acquisitions (27), but closer inspection by experienced eyes revealed that those claims for rampant lateral transfer to humans were premature (28, 29). Given that lateral gene transfer (LGT) claims are usually founded in gene phylogenies (26), and given that gene phylogenies have severe limitations (30), it seems prudent to ask: Just how good is the evidence for rampant LGT from food (or other) bacteria to eukaryotic nuclei as opposed to the evidence for transfers from organelles?

That question is relevant because there is now direct experimental evidence for DNA transfer rates from chloroplasts (6, 11), and pioneering work in yeast provided similar experimentally determined rates for transfer from mitochondria (31). Thus, it seems that the onus is on LGT advocates to show that they can measure rates of transfer from food (or other) bacteria to eukaryotic nuclei that compare with the rates now demonstrable in the laboratory for organelles. It also seems that if LGT is as prevalent as is often claimed, eukaryotic genomes should harbor just as much clear-cut evidence for recent acquisitions of prokaryotic chromosome segments as they harbor hard evidence for recent DNA transfer from organelles.

In time, we will become accustomed to seeing a freshly transferred copy of chloroplast DNA or mitochondrial DNA somewhere in nuclear chromosomes with every new genome published. Perhaps we will be even more surprised if recent organelle donations are missing. But until a freshly transferred chunk of a foreign prokaryotic chromosome turns up in a eukaryotic genome sequence, some of us will continue to look to organelles first as the most likely donors of surprising nuclear sequences (32). After all, the once highly acclaimed dinosaur bone DNA turned out ultimately to have stemmed from the human mitochondrion (33). That lesson should remind us that once DNA is transferred from organelles to the nucleus, it can take on many a surprising guise.

1. Butterfield, N. J. (2000) *Paleobiology* **263,** 386–404.
2. Mereschkowsky, C. (1905) *Biol. Centralbl.* **25,** 593–604.
3. Brown, J. R. (2003) *Nat. Rev. Genet.* **4,** 121–132.
4. Leister, D. (2003) *Trends Genet.* **19,** 47–56.
5. Ayliffe, M. A. & Timmis, J. N. (1992) *Theor. Appl. Genet.* **85,** 229–238.
6. Stegemann, S., Hartmann, S., Ruf, S. & Bock, R. (2003) *Proc. Natl. Acad. Sci. USA* **100,** 8828–8833.
7. Maliga, P. (2002) *Curr. Opin. Plant Biol.* **5,** 164–172.
8. Bogorad, L. (2000) *Trends Biotechnol.* **18,** 257–263.
9. Daniell, H. & Parkinson, C. L. (2003) *Nat. Biotechnol.* **21,** 374–375.
10. Huang, C. Y., Ayliffe, M. A. & Timmis, J. N. (2003) *Nat. Biotechnol.* **21,** 489–490.
11. Huang, C. Y., Ayliffe, M. A. & Timmis, J. N. (2003) *Nature* **422,** 72–76.
12. The *Arabidopsis* Genome Initiative (2000) *Nature* **408,** 796–815.
13. The Rice Chromosome 10 Sequencing Consortium (2003) *Science* **300,** 1566–1569.
14. Lin, X., Kaul, S., Rounsley, S., Shea, T. P., Benito, M. I., Town, C. D., Fujii, C. Y., Mason, T., Bowman, C. L., Barnstead, M., *et al.* (1999) *Nature* **402,** 761–768.
15. Stupar, R. M., Lilly, J. W., Town, C. D., Cheng, Z., Kaul, S., Buell, C. R. & Jiang, J. (2001) *Proc. Natl. Acad. Sci. USA* **98,** 5099–5103.
16. Ricchetti, M., Fairhead, C. & Dujon, B. (1999) *Nature* **402,** 96–100.
17. Mourier, T., Hansen, A. J., Willerslev, E. & Arctander, P. (2001) *Mol. Biol. Evol.* **18,** 1833–1837.
18. Turner, C., Killoran, C., Thomas, N. S., Rosenberg, M., Chuzhanova, N. A., Johnston, J., Kemel, Y., Cooper, D. N. & Biesecker, L. G. (2003) *Hum. Genet.* **112,** 303–309.
19. Lister, D. L., Bateman, J. M., Purton, S. & Howe, C. J. (2003) *Gene*, in press.
20. Adams, K. L., Daley, D. O., Qiu, Y. L., Whelan, J. & Palmer, J. D. (2000) *Nature* **408,** 354–357.
21. Henze, K. & Martin, W. (2001) *Trends Genet.* **17,** 383–387.
22. Allen, J. F. (2003) *Philos. Trans. R. Soc. London B* **358,** 19–37.
23. Bölter, B., Soll, J., Schulz, A., Hinnah, S. & Wagner, R. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 15831–15836.
24. Bogorad, L. (1975) *Science* **188,** 891–898.
25. Doolittle, W. F. (1998) *Trends Genet.* **14,** 307–311.
26. Gogarten, J. P. (2003) *Curr. Biol.* **13,** R53–R54.
27. International Human Genome Sequencing Consortium (2001) *Nature* **409,** 860–921.
28. Salzberg, S. L., White, O., Peterson, J. & Eisen, J. A. (2001) *Science* **292,** 1903–1906.
29. Stanhope, M. J., Lupas, A., Italia, M. J., Koretke, K. K., Volker, C. & Brown, J. R. (2001) *Nature* **411,** 940–944.
30. Penny, D., McComish, B. J., Charleston, M. A. & Hendy, M. D. (2001) *J. Mol. Evol.* **53,** 711–723.
31. Thorsness, P. E. & Fox, T. D. (1990) *Nature* **346,** 376–379.
32. Martin, W. (1999) *BioEssays* **21,** 99–104.
33. Zischler, H., Höss, M., Handt, O., von Haeseler, A., van der Kuyl, A. C., Goudschmidt, J. & Pääbo, S. (1995) *Science* **268,** 1192–1193.