ENDOSYMBIOTIC GENE TRANSFER: ORGANELLE GENOMES FORGE EUKARYOTIC CHROMOSOMES

Jeremy N. Timmis, Michael A. Ayliffe‡, Chun Y. Huang* and William Martin§*

Genome sequences reveal that a deluge of DNA from organelles has constantly been bombarding the nucleus since the origin of organelles. Recent experiments have shown that DNA is transferred from organelles to the nucleus at frequencies that were previously unimaginable. Endosymbiotic gene transfer is a ubiquitous, continuing and natural process that pervades nuclear DNA dynamics. This relentless influx of organelle DNA has abolished organelle autonomy and increased nuclear complexity.

CYANOBACTERIA
The group of pigmented, photosynthetic bacteria that contains the endosymbiont ancestors of chloroplasts.

α-PROTEOBACTERIA
A subgroup of gram-negative bacteria, often called the purple bacteria, that are thought to be the endosymbiont ancestors of mitochondria.

**School of Molecular and Biomedical Science, The University of Adelaide, South Australia 5005, Australia.*
‡CSIRO Plant Industry, GPO Box 1600, Australian Capital Territory 2601, Australia.
§Institute of Botany III, University of Düsseldorf, Düsseldorf 40225, Germany.
Correspondence to J.N.T.
e-mail: jeremy.timmis@adelaide.edu.au
doi:10.1038/nrg1271

Mitochondria and plastids were once free-living prokaryotes. They have retained the bulk of their prokaryotic biochemistry but harbour only a remnant of the eubacterial genome that their respective ancestors possessed. Over time, chloroplast (the photosynthetic plastid) and mitochondrial genomes have shrunk by orders of magnitude from the size of fully-fledged eubacterial genomes to approximately the size of plasmids. Concurrently, eukaryotic nuclear genomes have been the recipients of mitochondrial (mt) and chloroplast (cp) DNA donations and have expanded, often to enormous size and complexity. Signs of such chromosome rebuilding through endosymbiotic gene transfer are unmistakable in sequenced genomes.

Studies make it clear that, during the roughly two billion years since eukaryotes arose, many genes have relocated from the ancestral organellar genomes to the nucleus. Many of these genes have become functionally competent nuclear copies that now drive the biogenesis of mitochondria and chloroplasts, but some others have evolved to control further essential cellular processes. As remodelled nuclear copies of organelle genes usurped the functions of those located in the organelle, biochemical pathways were transferred wholesale from the organelles to the cytosol and the mitochondrial and plastid genomes were reduced in size.

More than 20 years have elapsed since it first became apparent that mitochondrial- and chloroplast-DNA sequences are also present in the nuclear genomes of most eukaryotic species. Genome sequencing projects have now uncovered abundant organelle-to-nucleus transfers. Rates of organelle DNA transfer to the nucleus are now measurable in the laboratory, and the results of the first studies show that it occurs at staggeringly high frequencies. Controversy surrounds the impact of endosymbiotic gene transfer on eukaryote genome evolution, on transgenic crop technology and on natural variation within species. Here, we review comparisons of nuclear, organellar, CYANOBACTERIAL and α-PROTEOBACTERIAL genomes that address endosymbiotic gene transfer. We discuss direct observations of organelle-to-nucleus gene transfer in the laboratory, their evolutionary implications and their consequences for organelle-based transgene containment strategies in genetically modified (GM) crops.

Organelle genomes — prokaryotic remnants

That eukaryotic organelles contain genes with a non-Mendelian mode of inheritance was inferred at the beginning of the 1900s (REF. 1) (BOX 1), as was an endosymbiotic origin for organelles[2]. However, it was not until the 1970s that the notion that organelles originated from endosymbiotic prokaryotes gained some acceptance[3] and, later still, that sequence comparisons unequivocally

DISOMIC
The condition in which there are two sets of similar (homologous) chromosomes, such that there are two alleles for each gene locus. These homologous chromosomes pair at meiosis and their segregation and transmission results in Mendelian inheritance.

HAPLOID
The condition in which there is only a single chromosome, or set of chromosomes, such that all loci are represented by only a single allele.

CYTOPLASMIC ORGANELLES
Here, confined to mean mitochondria and plastids.

PROMISCUOUS DNA
DNA that is present in more than one genetic compartment of the eukaryotic cell.

identified proteobacteria and cyanobacteria as ancestors of mitochondria and chloroplasts, respectively[4] (BOX 2). To account for the observation that many proteins that are encoded by the nuclear genome are essential to chloroplasts and mitochondria[5,6], it was suggested that genes had been relocated from the ancestral organelles to the nucleus during evolution[7].

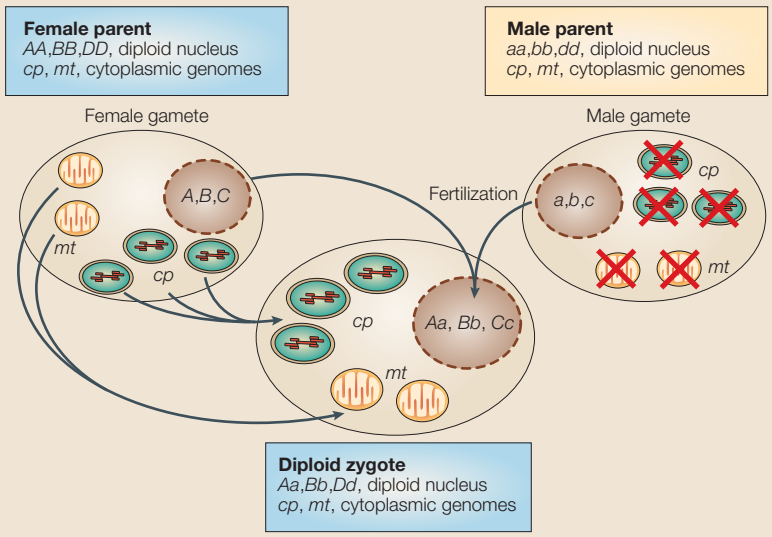
This hypothesis has proved to be robust, although the details of the process are more complex than initially predicted. Cytoplasmic organelles contain a minuscule set of genes compared with the nuclear genome. Both chloroplasts and mitochondria generally contain multiple circular haploid genomes that are present as monomers and multimers. The protein-coding capacity of organelle genomes varies markedly across eukaryotic lineages. Sequenced plastid genomes encode from 20 to 200 proteins[8,9] and mitochondrial genomes encode anything from 3 to 67 proteins[10,11] (see TABLE 1). However, some unusual mitochondrial genomes are composed of many linear chromosomes with one gene each[12] and there are even cases in which the mitochondrial genome is missing altogether[13]. Similarly, circular DNA molecules with one gene each have been identified in some plastids[14] and occasionally the entire plastid has apparently disappeared[15,16]. In addition to their genomic diversity, organelles themselves are biochemically diverse. Some mitochondria consume oxygen, some produce hydrogen (hydrogenosomes)[13], some make ATP with or without oxygen[17] and some extremely reduced types (mitosomes) make no ATP at all, but make iron-sulphur clusters for the cell instead[18]. Plastids are an even more diverse assemblage of organelles[16].

The closest cousins of mitochondrial and chloroplast genomes are free-living α-proteobacteria and cyanobacteria respectively, but which lineages among those groups gave rise to present-day organelles remains unresolved[10,19,20]. The modern α-proteobacterium *Mesorhizobium loti* harbours 7 Mb of DNA that encodes more than 6,700 proteins and its relative *Bradyrhizobium japonicum* contains a 9.1 Mb genome and more than 8,300 proteins. The cyanobacterium *Nostoc* PCC 7,120 has a 6.4 Mb genome, which encodes approximately 5,400 proteins, whereas that of *Nostoc punctiforme* is >9 Mb, coding for more than 7,200 proteins. Comparing these genome sizes with those of organelles puts the magnitude of organelle genome reduction into perspective (TABLE 1).

Although microbial parasites can also have highly reduced genomes, genome reduction in parasites (loss of genes and functions) is fundamentally different from genome reduction in organelles (loss of genes, but not functions), a distinction that is too seldom underscored (BOX 3).

Box 1 | Inheritance of cytoplasmic and nuclear genes

The nuclear genomes of most higher eukaryotic organisms are diploid and are characterized by DISOMIC inheritance and sexual reproduction. So, nuclear genes come in allelic pairs that are often subtly different from each other. Gametes that result from meiosis are HAPLOID and carry only a single allele (in this example, alleles *A*, *B* and *C* from the female parent and *a*, *b* and *c* from the male parent). The zygote that results from fertilization inherits one nuclear allele of each gene from each parent (that is, at the three example loci, it is *Aa*, *Bb* and *Cc*). By contrast, the CYTOPLASMIC ORGANELLES characteristically contain multiple, homogeneous genomes that are usually inherited from one parent only (in this example, and most commonly, the female parent). In tobacco and many other plants, the mitochondrial and chloroplast genomes are specifically degraded before fertilization (red crosses). There are many exceptions to this common inheritance pattern of genes in mitochondria and chloroplasts (for a review, see REF. 119).



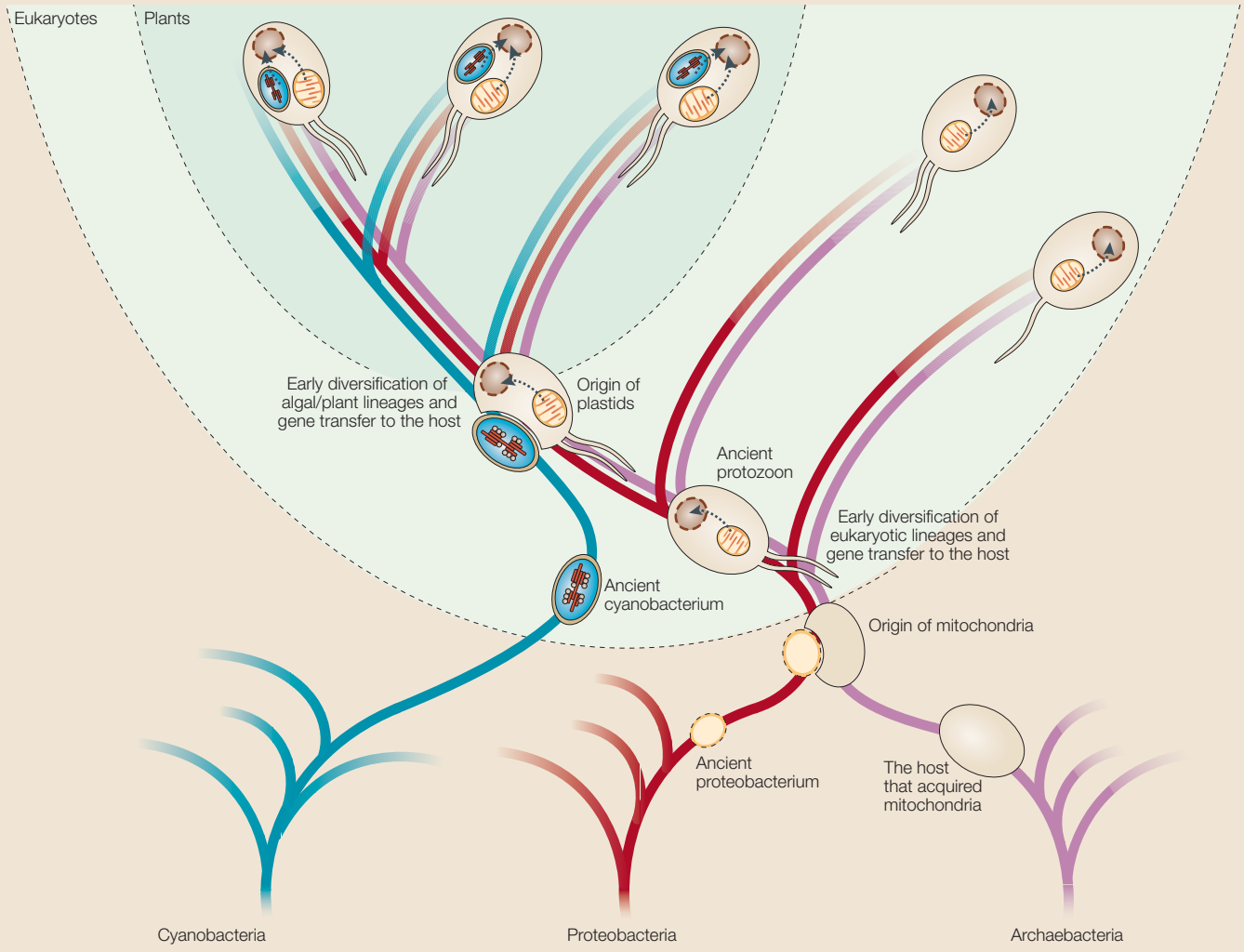
Comparative analyses

Today, we know that genes have been transferred to the nucleus from the ancestral genomes of organelles but key questions are, at present, how many and what kinds of genes were transferred, how did the transfer occur and how often did it occur. Two main categories of comparative analyses address these questions. First, studies that identify copies of genes that are still present in an organelle genome but that are also in other compartments of the same cell indicate the process of DNA movement and identify evolutionarily recent transfer events. Second, studies that analyse nuclear genomes, organelle genomes and the genomes of candidate prokaryotic ancestors to identify genes in the nucleus that are no longer present in the organelle show more ancient transfer events.

Evolutionarily recent transfers shown by genome comparisons. Initial evidence that DNA could move among cell compartments came when fragments of cpDNA were found in the maize mitochondrial genome[21]. Reports of mtDNA sequences[22,23] and chloroplast sequences[24] in nuclear DNA followed, and the term 'PROMISCUOUS DNA' was coined by J. Ellis[25] to connote DNA mobility among the genetic compartments of eukaryotic cells (FIG. 1). These findings were important not only because they provided evidence that gene transfer among cell compartments could occur, but also because they indicated that this might be a continuing process given how recently (in evolutionary terms) the DNA must have been transferred. Subsequently, complete copies of mtDNA were discovered in cat genomes and

Box 2 | **Endosymbiotic evolution and the tree of genomes**

Intracellular endosymbionts that originally descended from free-living prokaryotes have been important in the evolution of eukaryotes by giving rise to two cytoplasmic organelles. Mitochondria arose from α-proteobacteria and chloroplasts arose from cyanobacteria. Both organelles have made substantial contributions to the complement of genes that are found in eukaryotic nuclei today. The figure shows a schematic diagram of the evolution of eukaryotes, highlighting the incorporation of mitochondria and chloroplasts into the eukaryotic lineage through endosymbiosis and the subsequent co-evolution of the nuclear and organelle genomes. The host that acquired plastids probably possessed two flagella[113]. The nature of the host cell that acquired the mitochondrion (lower right) is fiercely debated among cell evolutionists. The host is generally accepted by most to have an affinity to ARCHAEBACTERIA but beyond that, biologists cannot agree as to the nature of its intracellular organization (prokaryotic, eukaryotic or intermediate), its age, its biochemical lifestyle or how many and what kind of genes it possessed[120]. The host is usually assumed to have been unicellular and to have lacked mitochondria.



ARCHAEBACTERIA
An ancient group of organisms that have ribosomes and cell membranes that distinguish them from eubacteria. They sometimes show environmentally extreme ecology.

NUMT
An acronym to describe nuclear integrants of mitochondrial DNA.

the term 'NUMTS' was coined to designate these nuclear stretches of mtDNA[26]. Numts have been found in the nuclear genomes of grasshoppers[27], primates[28,29] and shrimps[30], and are often mistaken for *bona fide* mtDNA[31,32].

Eukaryotic genome sequences have more fully exposed the scale of integrated mitochondrial and cpDNA in the nuclear genome. Fragments of organelle DNA are becoming recognized as a normal attribute of nearly all eukaryotic chromosomes. For example, the yeast genome contains tracts with 80–100% similarity to mtDNA that range in size from 22 to 230 base pairs (bp) integrated at 34 sites[33]. This range of sequence divergence indicates recurrent transfer events, from ancient to contemporary. The human genome has at least 296 different numts of between 106 bp and 14,654 bp (90% of the mitochondrial genome) that cover the entire mtDNA circle[34]. Other studies tallied 612 mtDNA insertions in the human genome[35], a greater number because different sequence conservation criteria for identifying numts are used in different studies[36]. Older numts are more abundant in the human genome than recent INTEGRANTS, indicating that mtDNA can be amplified once inserted[36,37] and many are organized as tandem repeats[35]. Barely detectable numts are present in *Plasmodium*[32], but highly conserved numts have now

been identified in genome data for *Rattus norvegicus*, *Caenorhabditis elegans*, *Drosophila melanogaster*, *Schizosachharomyces pombe*, *Ciona intestinalis*, *Neurospora crassa* and *Fugu rubripes*[38].

Plant genomes provide even more abundant evidence of endosymbiotic DNA transfer. The nuclear genome of *Arabidopsis* contains a large (~620 kb) insert of mtDNA on chromosome 2 (REFS 39,40), as well as an extra ~7 kb that is distributed among 13 small integrants and 17 insertions of cpDNA ('NUPTS': the plastid counterparts of numts), totalling 11 kb (REF. 41). In the rice genome, chromosome 10 alone

Table 1 | **Sizes and coding content of some organelle and prokaryote genomes**

Genome	Length [kbp]	Number of protein-coding genes	GenBank accession number
Algae			
cp *Porphyra purpurea*	191	200	PPU38804
cp *Cyanidium caldarium*	165	197	AF022186
cp *Guillardia theta*	122	148	AF041468
cp *Cyanophora paradoxa*	136	136	CPU30821
cp *Odontella sinensis*	120	124	OSCHLPLXX
cp *Euglena gracilis*	143	58	CLEGCGA
Land plants			
cp *Marchantia polymorpha*	121	84	CHMPXX
cp *Chlorella vulgaris*	151	78	AB001684
cp *Nicotiana tabacum*	156	76	CHNTXX
cp *Oryza sativa*	134	76	X15901
cp *Zea mays*	140	76	ZMA86563
cp *Pinus thunbergii*	120	69	PINCPTRPG
Non-phosynthetic plastids			
cp *Toxoplasma gondii*	35	26	U87145
cp *Eimeria tenella*	35	28	AY217738
cp *Epifagus* virginiana	70	21	EPFCPCG
Cyanobacteria			
Synechocystis sp.	3573	3168	AB001339
Prochlorococcus marinus	1660	1884	NC_005071
Nostoc PCC 7120	6413	5368	AP003602
Nostoc punctiforme	~9000	~7400	http://www/jgi/doe/gov
Plants and algae			
mt *Pylaiella littoralis*	59	52	NC_003055
mt *Marchantia polymorpha*	187	41	MPOMTCG
mt *Laminaria digitata*	38	39	AJ344328
mt *Cyanidioschyzon merolae*	32	34	NC_000887
mt *Arabidopsis thaliana*	367	31	MIATGENA
mt *Chondrus crispus*	26	25	MTCCGNME
mt *Scenedesmus obliquus*	43	20	NC_002254
Various protists and fungi			
mt *Reclinomonas americana*	69	67	NC_001823
mt *Malawimonas jakobiformis*	47	49	AF295546
mt *Naegleria gruberi*	50	46	NC_002573
mt *Rhodomonas salina*	48	44	NC_002572
mt *Dictyostelium discoideum*	56	40	NC_000895
mt *Phytophthora infestans*	38	40	NC_002387
mt *Acanthamoeba castellanii*	42	36	U12386
mt *Cafeteria roenbergensis*	43	34	NC_000946
mt *Monosiga brevicollis*	77	32	AF538053
mt *Physarum polycephalum*	63	20	AB027295
mt *Harpochytrium sp*	24	14	AY182006
mt *Candida albicans*	40	13	NC_002653
mt *Cryptococcus neoformans*	25	12	NC_004336
mt *Plasmodium falciparum*	6	3	NC_001677
Anaerobic mitochondria			
mt Hydrogenosomes*	0	0	
α-proteobacteria			
Caulobacter crescentus	4017	3767	AE006573
Mesorhizobium loti	7596	7281	BA000012
Bradyrhizobium japonicum	~9100	~8300	BA000040
Yeast			
(nuclear)	13,469	6,327	http://www.ebi.ac.uk

An excellent, up-to-date list of sequenced organelle genomes is available at http://megasun.bch.umontreal.ca/ogmp/projects/other/all_list.html. Prokaryote data was gratefully received from http://dna-res.kazusa.or.jp and http://www.jgi.doe.gov/JGI_microbial/html. *Hydrogenosomes are anaerobic forms of mitochondria that usually lack a genome. cp, chloroplast genome; mt, mitochondrial genome.

INTEGRANT
Here, used to describe nuclear tracts of DNA that resemble plastid DNA or mitochondrial DNA.

NUPT
An acronym to describe nuclear integrants of plastid DNA.

Box 3 | Genome reduction in organelles and parasites

In addition to organelles, microbial parasites can also have highly reduced genomes. Examples are *Rickettsia prowazeckii* with ~830 protein-coding genes, *Mycoplasma genitalium* with ~470 and the parasitic eukaryote *Encephalitozoon cuniculi* with ~2,000 genes (half as many as *Escherichia coli*). However, the mechanism of genome reduction in parasites differs fundamentally from that in organelles. Whereas parasites simply lose the genes that they no longer need[70], organelles do require the products of many of the genes that they relinquish to the chromosomes of their host. So, organelle genome reduction is not simply an extension of parasite genome reduction. The nature of the two processes — reduction through specialization to a nutrient-rich intracellular environment in the case of parasites versus reduction through export of essential genes to the host's genetic apparatus with import of thousands of essential proteins from the cytosol in the case of organelles — could not differ more.

contains 28 cpDNA fragments >80 bp long, including two very large insertions (33 kb (REF. 42) and ~131 kb (REF. 43)). Similarly, the draft sequence of rice chromosome 1 shows many plastid insertions[44]. MtDNA insertions are also plentiful in the rice genome: chromosome 10 has 57 such segments that range from 80 to 2,552 bp (REF. 43). Whether larger nuclear genomes generally harbour more promiscuous DNA than smaller ones, such as *Arabidopsis*, remains to be seen.

Transferred organelle DNA segments and sometimes even complete organelle genomes are more or less ubiquitous as integrated constituents of eukaryotic genomes, as early studies had indicated[45,46], but the evolutionary consequences of such transfers have yet to be fully explored.

Recurrent transfers and convergent gene losses. The deletion and functional replacement of mitochondrial genes by nuclear copies has effectively stopped in higher animals[47] in which mitochondria encode 12 to 13 proteins, but the process is still actively continuing in higher plants, which have larger numbers of mitochondrially-encoded proteins. Thorough studies among flowering plants have uncovered many cases of transfer that result in expressed genes[48,49]. For example, the mitochondrial *rps10* gene has been independently transferred to the nucleus many times[49,50]. Most mitochondrial transfer and activation events seem to involve recombination into pre-existing promoter and/or TRANSIT PEPTIDE-coding regions[8,48–57].

Similar to *rps10*, the chloroplast translation initiation factor 1 gene (*infA*) also shows striking evidence of mobility[58]. Nuclear relocations of this chloroplast gene were accompanied by MUTATIONAL DECAY and/or deletion of the corresponding chloroplast sequence. However, unlike *rps10*, characterization of transplanted nuclear *infA* genes indicated the appearance of *de novo* transit peptides rather than the parasitization of existing nuclear genes. Genes such as *infA* underscore earlier findings from genome analyses that parallel losses in independent lineages are regular occurrences in chloroplast genome evolution[59]. Plotting the process of chloroplast genome reduction over time (gene losses from cpDNA) onto a chloroplast genome phylogeny[19] shows that parallel losses in independent lineages outnumber unique losses that are shared by descendant lineages by a ratio of more than10 to 1. Therefore, the similarity in gene content among contemporary plastid genomes is the result of immensely convergent evolution (FIG. 2).

Ancient transfers that are shown by genome comparisons. The recognition of functional gene relocations, like the identification of numts and nupts, relied on sequence similarity between nuclear and organelle genomes. The discovery of genes that were transferred to the nucleus, but are no longer present in most organelle genomes, requires a different approach. To detect such transfers, searches of nuclear genomes with sequences that are present in the organelle genome of at least one species have been undertaken. The PROTIST *Reclinomonas americana* has the mitochondrial genome with the largest known gene content, and homologues of nearly all these genes have been found in the nucleus of other eukaryotes[60]. Similarly, most genes that are contained in the larger chloroplast genomes can be found as transferred homologues among plant and algal nuclear genomes[19,59]. Comparative studies have also shown gene transfer during secondary symbiosis — the origin of plastids from eukaryotic algae instead of from cyanobacteria — in the evolution of unicellular eukaryotes[61–63].

A search for nuclear genes in the *Arabidopsis* genome that branch with cyanobacterial homologues in PHYLOGENETIC trees (or that have homologues in cyanobacteria only) showed that approximately 1,700 of the 9,368 *Arabidopsis* proteins that are sufficiently conserved in sequence to allow phylogenetic analysis, come from cyanobacteria, indicating that roughly 18% of the protein-coding genes in *Arabidopsis* are acquisitions from the plastid[19]. Many of these acquired genes clearly control cellular systems other than chloroplast biogenesis and many are targeted to the cytosol or the secretory pathway[64]. Among eukaryotes, 630 nuclear-encoded proteins were identified that originated from mitochondria[65], of which <30% were predicted to be targeted to mitochondria in yeast and human. So, the proteins encoded by many nuclear genes that are derived from organelle DNA ultimately take on new functions in new compartments.

Targeting and retargeting of proteins that are encoded by transferred genes. Case studies[8,66–68] and genome-wide analyses[64,69] show that the relationship between organelle gene donations and organelle protein imports is complex and difficult to predict. This contrasts with the older idea that proteins were always targeted to the cell compartment from which the genes that encoded them originated (the PRODUCT SPECIFICITY COROLLARY[7] (see REF. 66 for a discussion) to endosymbiotic theory). In *Arabidopsis*, fewer than half the proteins that are identifiable as acquisitions from cyanobacteria are predicted to be targeted to chloroplasts. Many are targeted to the cytosol, the secretory pathway or the mitochondrion. Conversely, a similar proportion of proteins that are targeted to the plastid do not seem to be acquired from cyanobacteria[70,71]. Clearly, the products of nuclear genes that originated from endosymbionts are free to explore

TRANSIT PEPTIDE
A peptide sequence, often at the N-terminus of a precursor protein, that directs a gene product to its specific cellular destination.

MUTATIONAL DECAY
The process that describes the random changes that might occur in a DNA sequence in the absence of selection pressure.

PROTIST
A single-celled eukaryote.

PHYLOGENETICS
Reconstruction of the evolutionary relationships between sequences using any of a variety of inference methods.

PRODUCT SPECIFICITY COROLLARY
The situation in which the product of a gene that is donated by a cytoplasmic organelle to the nucleus is expected to be returned to that organelle.

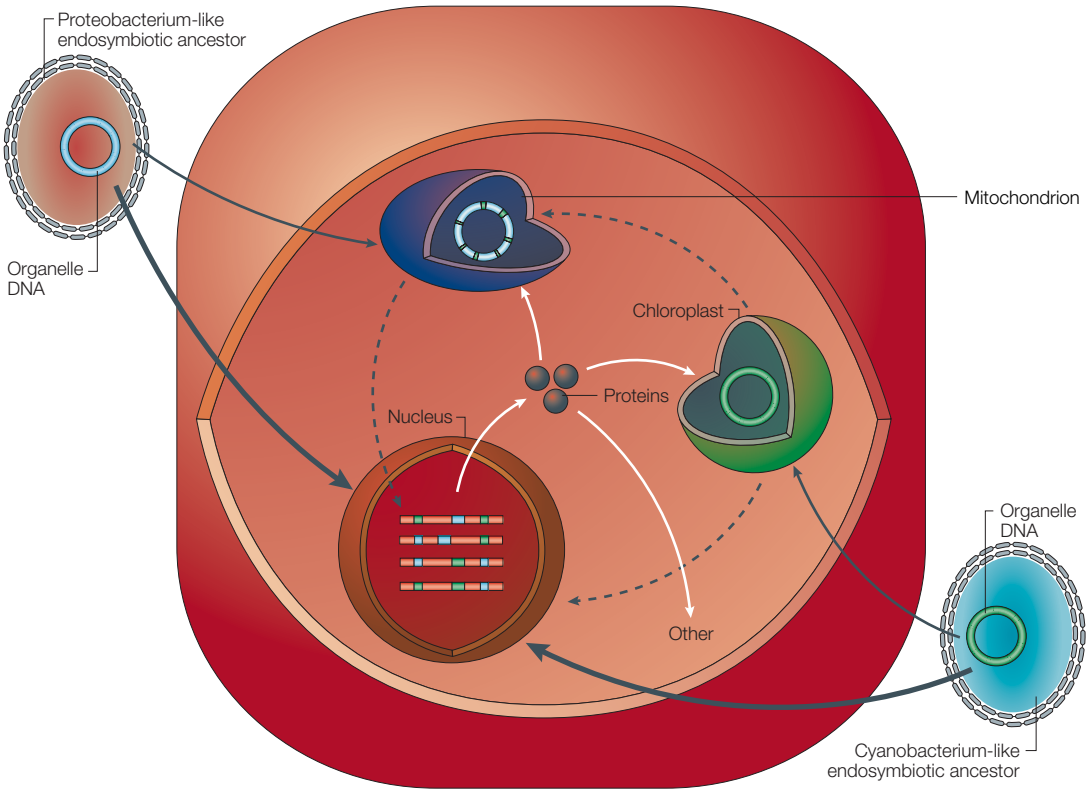


Figure 1 | **Organellar DNA mobility and the genetic control of biogenesis of mitochondria and chloroplasts.** The eukaryotic mitochondrion is derived from a proteobacterial endosymbiotic ancestor but most of the genes that were originally present in this ancestor's genome have been transferred to the nucleus (thick black arrow), with only a small number being retained in the organelle (blue circle). Similarly, most of the genes from the cyanobacterial endosymbiont ancestor of the chloroplast were also transferred to the nucleus (thick black arrow). So, as a result, cytoplasmic organelles are heavily dependent on nuclear genes and import more than 90% of their proteins from the cytoplasm (white arrows). The dotted arrows indicate how DNA of mitochondrial (blue) and chloroplast (green) origin is still being transferred to the nucleus. Chloroplast and nuclear sequences are also found in the mitochondrial genome but little or no promiscuous DNA is located in the chloroplast.

EPISOME
A unit of genetic material that is composed of a series of genes that sometimes has an independent existence in a host cell and at other times is integrated into a chromosome of the cell, replicating itself along with the chromosome.

BIOLISTIC TRANSFORMATION
A commonly used transformation method in which metal beads are coated with gene contructs and shot into cells.

LEAF EXPLANTS
Small sterile sections of leaf or other plant tissue from which whole plants might sometimes be regenerated.

UNIPARENTAL INHERITANCE
The mode of inheritance that generally characterizes the genes of cytoplasmic organelles in which only one of the two sexual partners contributes to the offspring.

new patterns of compartmentalization in the cell[11]. Moreover, gene donations from organelles often lead to functional replacement of pre-existing and functionally equivalent host genes, a process known as endosymbiotic gene replacement[69].

The number of proteins that are predicted to be imported into mitochondria varies markedly across eukaryotic groups, ranging from ~150 proteins in the parasitic fungus *Encephalitozoon cuniculi* to ~4,000 proteins in humans. Only ~50 proteins were common to the mitochondria of all non-parasitic eukaryotes[72]. Similarly, the number of nuclear-encoded proteins that are predicted to be targeted to chloroplasts differs by a factor of two between rice and *Arabidopsis*[73]. Such predictions still have clear limitations but are improving with the accumulation of more direct experimental data for localization[73,74].

For biochemical pathways that are present in both the original host and its endosymbionts, competition can ensue[8,66]. In some cases, the pathway of the symbiont can predominate[18,75] but hybrid pathways can develop from both host and endosymbiont sources[66,76,77]. Organelle division is a prime example of lineage-specific mixing and matching of endosymbiotically inherited functions with newly evolved, eukaryote-specific biochemistry[78,79].

In summary, retargeting of proteins among organelles and, most notably, the cytosol is a highly dynamic and influential process in eukaryotic evolution. When genes are donated from organelles to the nucleus, there is no homing device that automatically re-routes the protein product back to the donor organelle. Rather, chance, natural selection and lineage diversification seem to govern the intracellular targeting fate of genes that organelles donate to the chromosomes of their host. In this sense, gene donations from organelles are important starting material for the evolution of new genes that are specific to the eukaryotic lineage.

Laboratory estimates of transfer frequencies

Comparative genome analyses show us that gene transfers have occurred at different times in the past, and indicate that the process is continuing. The challenge has been to get direct empirical estimates of the frequency at which DNA is being transferred among cellular compartments.

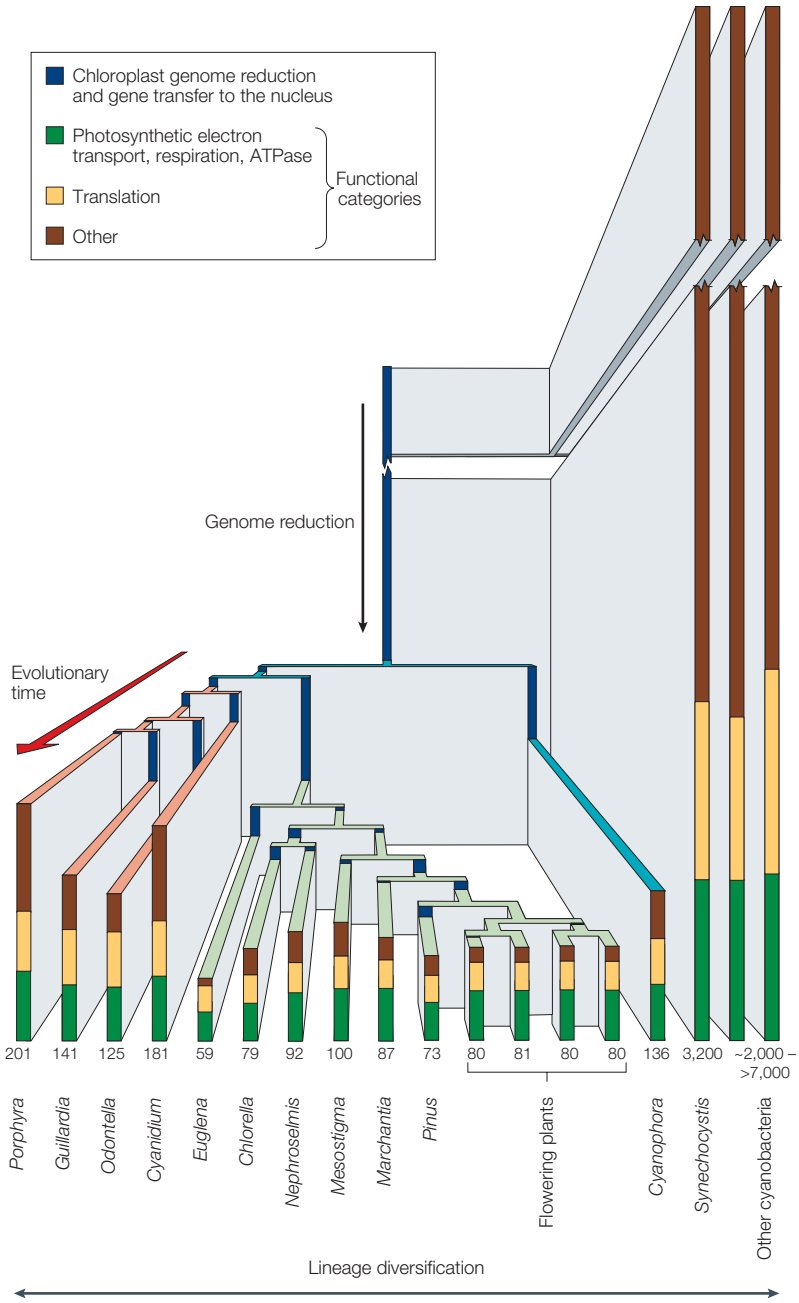


Figure 2 | **Reduction of the chloroplast genome over time.** We know that plastids originated more than 1.2 billion years ago, because fossil red algae of that age have been found[121]. The ancestor of plastids was a free-living cyanobacterium and therefore must have possessed several thousand genes as did its contemporaries. Subsequent to the invention of a plastid protein import apparatus (a prerequisite for relocating genes that encode proteins required by the organelle to the nucleus), plastids relinquished most of their genes to the genome of their host cell. This gene relocation process occurred massively at the onset of endosymbiosis and continued in parallel during algal diversification, yet the same core set of genes (for photosynthesis and translation) has been retained in all lineages. The size of the bars shown indicates the genome sizes of chloroplasts from a diversity of plant lineages, from red algae (*Porphyra*) to angiosperms (flowering plants) and *Cyanophora* (belonging to the most ancient lineage of photosynthetic eukaryotes), and their free-living cyanobacterial relatives (cyanobacteria). The reduction in chloroplast genome size has been mapped onto a phylogenetic tree of the relationships among these genomes. Numbers at the end of branches indicate the number of genes that are present in the respective genome. These genes are divided into three functional categories that are represented by the three different colours making up the bars. Data from REF. 19.

Mitochondrion-to-nucleus transfers. In yeast, a recombinant plasmid, which was introduced into a genome-lacking mitochondrion, was shown to relocate to the nucleus as an EPISOME (that is, not recombined into nuclear DNA) at a frequency of 2×10^{-5} per cell per generation[80]. A lower frequency (5×10^{-6} per cell per generation) of episomal relocation was observed when the plasmid was integrated into the mitochondrial chromosome[81]. In these experiments, the released DNA was episomal, indicating that release of DNA from the yeast mitochondrion is frequent, but integration might be rare in yeast nuclei because of their characteristically high level of reliance on homologous recombination for DNA incorporation. Newer work indicates that mtDNA escape in yeast occurs through an intracellular mechanism that depends on the composition of the growth medium and the genetic state of the mitochondrial genome, and is independent of an RNA intermediate[82].

Chloroplast-to-nucleus transfers in higher plants. Only more recently has it been possible to quantify the process of chloroplast-to-nucleus DNA transfer. To determine the frequency of plastid DNA transfer and integrative recombination into the higher plant nuclear genome, the plastome of tobacco was transformed with a neomycin phosphotransferase gene (*neoSTLS2*) that was tailored for expression only in the nuclear genome[83] (BOX 4). In 16 out of ~250,000 seedlings, the *neoSTLS2* marker had been integrated into a nuclear chromosome, each time in a different location, which equates to a chloroplast-to-nucleus DNA transfer frequency of one in 16,000 gametes tested. The diversity of insertion locations indicates that the marker might be transposed during meiotic or postmeiotic events during male gamete formation because the extreme alternative explanation for these integrations — a single transfer event that is subsequently amplified by somatic cell division — would lead to the same integration site being found in all plants with chloroplast integrants[83]. In agreement with the DNA integrations induced by BIOLISTIC TRANSFORMATION, these transfers show no particular preference for recombination sites in either the nuclear or plastid genomes.

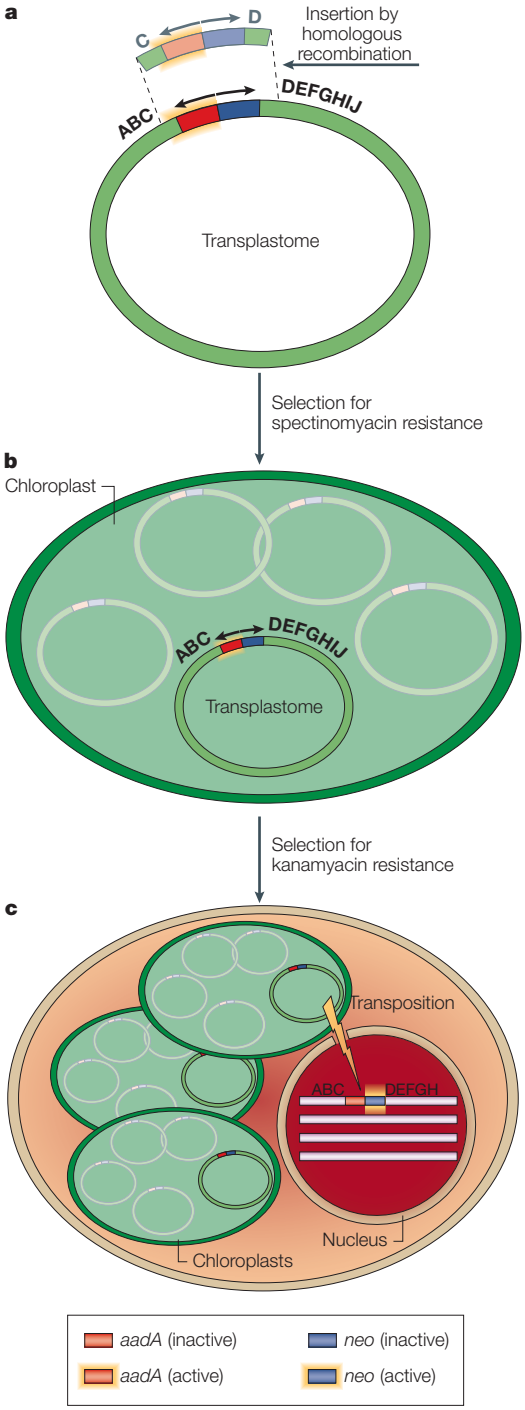
Using a similar experimental strategy with a transgene in a different plastomic location, the frequency of chloroplast-to-nucleus transposition was estimated in tobacco somatic cells[84]. Leaf tissue from transplastomic tobacco that contained an intron-less *neo* gene was cultured on medium that contained high concentrations of kanamycin (100–400 mg/L). Twelve highly resistant plants were regenerated, 11 of which showed Mendelian inheritance of the antibiotic-resistant phenotype. After a courageous approximation of the number of regenerateable cells that were present in LEAF EXPLANTS, a chloroplast-to-nucleus transposition frequency of one event in ~five million somatic cells was estimated[84]. Taken at face value, the frequency in somatic cells is ~300 times lower than that in male gametes of the same species[83]. The programmed degeneration of plastids that occurs during pollen-grain development — the process that underpins UNIPARENTAL INHERITANCE of plastid genes (FIG. 1) — might explain this difference. After the chloroplast genomes are

Box 4 | **Design of experiments that showed DNA transfer from organelles to nucleus in real time**

A construct that consists of chloroplast sequences (*C* and *D*) that flank two selectable marker genes is inserted into the chloroplast genome through homologous recombination, thereby transforming the native plastome into a TRANSPLASTOME (**a**). In the experiments of Huang *et al.*[83], the flanking chloroplast sequences were in the inverted repeats of the tobacco plastome. One of the selectable genes (*aadA* in the case illustrated) is designed for exclusive expression in the chloroplast and incorporation of this marker confers spectinomycin resistance. The other gene, a neomycin phosphotransferase gene *neoSTLS2* (that encodes NPTII and incorporates a nuclear intron; here *neo*), is tailored, by virtue of a nuclear-specific promoter and the presence of a nuclear intron in the reading frame, for expression only when it is transposed to the nucleus. Continuous selection of growing leaf cells on spectinomycin medium allows transformed plastomes to be selected and eventually the transplastome entirely replaces the native chloroplast genome, such that all copies of the chloroplast genome contain the two selectable marker genes (**b**).

Selection of cells or progeny seedlings on kanamycin medium allows the detection of the rare cases in which the *neo* gene has changed its location, such that strong expression is promoted from the nuclear environment (**c**). The progeny of self-fertilized transplastomic plants were not screened directly[83]. Rather, to eliminate low-level expression of *neoSTLS2* from the chloroplast genome, transplastomic plants were crossed with wild-type female plants such that, because of strict maternal inheritance of tobacco plastids (BOX 1), progeny that contained only wild-type chloroplasts were produced. Therefore, chloroplast-to-nucleus transposition must have occurred at some stage during the life cycle of the male parent of the seedlings that were screened on kanamycin plates.

The observation that 1 in 16,000 male tobacco gametes contained a newly integrated segment of chloroplast DNA (REF. 83) was unpredictably high, but it must be an underestimate of the true chloroplast-to-nucleus transposition frequency. In this experiment, the detection strategy enabled the identification only of those events that resulted in an entire, expressed *neoSTLS2* gene in the nucleus. Other regions of the tobacco plastome that integrated in the nucleus without this selectable marker necessarily remained undetected. A similar strategy was used by Stegmann *et al.*[84] and by Lister *et al.*[85].



degraded, fragmented DNA could become more available for transfer. In yeast, an enhanced rate of mtDNA escape to the nucleus has been linked to increased degradation of abnormal mitochondria[85]. Similarly, in animals — which have a sequestered germline — the transfer of mitochondrial sequences to the nucleus might be most likely during sperm mitochondrial degeneration following fertilization[37]. Assuming that a proportion of the somatic events[84] occurs early and enters the germline, we would expect, given a sufficient sample size, to find whole flowering branches that clonally transmit *de novo* nupts to pollen nuclei.

Chloroplast-to-nucleus transfers in algae. Chloroplast-to-nucleus transfer has also been investigated in the unicellular alga, *Chlamydomonas reinhardtii*[71]. Some 13

TRANSPLASTOME
The condition of a plastid genome after the insertion of non-native genes.

billion haploid cells were screened, but no chloroplast-to-nucleus DNA transfer was detected. This finding implies a transfer of at least six orders of magnitude lower than in higher plants, which is consistent with the paucity of integrated cpDNA in the nuclear genome of *C. reinhardtii*[71]. The presence of only one chloroplast in this alga might explain this difference: if chloroplast lysis and subsequent genome degradation is required for endosymbiotic gene transfer (see above), then organisms with only one essential chloroplast would not survive. However, specific degradation of cpDNA in *C. reinhardtii* does occur during zygote formation in the MT⁻ STRAIN at fertilization so it would be interesting to determine whether gene transfer can occur after mating.

Mechanisms, constraints and consequences

The above findings provide a glimpse of how often DNA is transferred among cellular compartments — but what has been learned about the mechanisms and constraints that govern the process?

As in pre-genomic sequencing studies[45,46], most numts and nupts had >95% nucleotide identity to the homologous organelle genes, probably because most of these sequences are rapidly eliminated or perhaps because regions with lower homology were not sought. The lack of divergence in most known numts and nupts indicates that they must have been transferred to the nucleus recently. Also, in all these examples, there has been no evidence of preferential transfer of a particular region or type of organelle sequence — for example, coding versus non-coding regions, introns in the case of the nuclear copy of mtDNA in *Arabidopsis*, structural RNA genes or promoters.

How is organelle DNA transferred? How do organelle genes physically get into the nucleus to recombine with chromosomal DNA? There are two opposing and hotly debated views on this topic that are based on different observations and that can be termed 'bulk DNA' and 'cDNA intermediates'.

The 'bulk DNA' view, based on early comparative studies[53], yeast genetics[53] and genome sequence comparisons[52], argues that recombination between escaped organelle DNA molecules and nuclear chromosomes is the mechanism of gene transfer followed by further recombination[52]. It is founded on evidence from experimental transfer in yeast[53] and on the observation that in comparisons between organelle DNA and nuclear DNA of the same species (numts and nupts), intergenic spacers and other non-coding regions of organelle DNA are found in nuclear-transferred copies as often as are coding sequences[26,34,39–43]. When whole organelle DNA molecules that are >100 kb long are found recombined into eukaryotic chromosomes[39–43] that contain the organelle introns, the tRNAs and hundreds of kb of organelle non-coding regions[39,40], bulk DNA transfer seems likely[52].

The view of 'cDNA intermediates' holds that cDNAs of organelle mRNAs are the vehicle of gene transfer to the nucleus[48–51]. It is based on taxon sampling studies in flowering plants, in which genes that are present in the mtDNA of some species are sought among many other species to monitor loss of the gene in the mtDNA, accompanied by the occurrence of a functional copy in the nuclear DNA (reviewed in REFS 48,49). As mitochondrial protein-coding genes of flowering plants often have introns and RNA EDITING[48], and because the nuclear copies of mtDNA so detected lack the organelle-specific introns and the edited sites, the inference has been that cDNA intermediates of edited and spliced mRNAs are directly involved in the physical transfer process[48–51].

However, there are alternative interpretations for observations that underpin the cDNA intermediate view[52]. When cDNA intermediates of spliced and edited higher-plant mitochondrial transcripts occur, they should arise in the mitochondrion and so be more likely to recombine with mtDNA — thereby erasing edited sites and introns in the mitochondrial gene — than with nuclear DNA[52]. In this way, cDNA-mediated dynamics of intron loss and edited sites in higher-plant mtDNA can mimic potential cDNA intermediates in sequence comparisons, even if bulk mtDNA only were physically being transferred for recombination in the nucleus[52]. Adding to this is the complex nature of flowering plant mtDNA genomes, which are a heterogeneous mixture of DNA molecules that are smaller than the 'master copy'[48].

Although the possibility that cDNA intermediates might be involved in the transfer of genes from mitochondria to the nucleus in flowering plants cannot be excluded, as is often suggested[48–51,68], there are alternative interpretations of the same data[52]. On a broader scale, evidence that implicates cDNA intermediates in mtDNA transfers in eukaryotic groups other than flowering plants is so far lacking (yeast mutants lacking mitochondrial RNA polymerase transfer mtDNA efficiently[53]), as is evidence that implicates cDNA in the transfer of chloroplast genes. The view of bulk DNA transfer finds support from genome comparisons[33–44] and from experimental organelle-to-nucleus studies in yeast and higher plants[80–85]. In the early evolution of mitochondria and chloroplasts, in which the brunt of endosymbiotic gene transfer occurred[10,11,19] (FIG. 2), cDNA intermediates would have been unnecessary in our view, because editing and introns in free-living α-proteobacteria and cyanobacteria are extremely rare at best.

Where does transferred DNA integrate? There is no evidence from genomes that organelle DNA is integrated into preferred chromosomal regions or sequence contexts. In human nuclear DNA, 98% of all numts are integrated into sequences that are not annotated as potential genes and the remaining 2% are in introns[37]. Integration of mtDNA has been implicated in human somatic mutations leading to neoplasms[86], but only one case of a numt causing heritable mutation has been reported[87]. The time and place of this insertion coincided with the Chernobyl nuclear reactor accident[87], indicating that radiation-induced mtDNA fragmentation (or recombination involving pre-existing numts) might have been involved.

MT⁻ STRAIN
One of the two mating types (the other is mt⁺) of *Chlamydomonas reinhardtii*; one of each is required to form a zygote.

RNA EDITING
Changes in the RNA sequence after transcription is completed. Examples include modification of C to U or of A to I by deamination, or insertion and/or deletion of particular bases.

Evidence has been found for a clustering of genes that are involved in mitochondrial nucleic-acid processing in *Arabidopsis*[88], but it is unlikely that the clustering reflects a relic of ancestral mitochondrial genome organization. From the standpoint of transcription, evidence has been found for important groups of nuclear-encoded genes that respond to the physiological state of the plastid in *Arabidopsis*, but it is unclear whether this reflects a relic of ancestral plastid gene regulation[89].

Why do some genes remain in organelles? Given the continual ingress of organelle DNA into the nucleus, why should there be any genes left in organelles? The main, and hotly debated, theories on this issue fall largely into two groups: 'hydrophobicity'[49] and 'redox control'[90]. The former view holds that hydrophobic proteins are poorly imported by organelles, and so must be encoded in organelle genomes[49]. It accounts for organelle encoding of some membrane-integral proteins in chloroplasts and mitochondria, but not for all (for example, light-harvesting proteins or mitochondrial importers). The 'redox control' view holds that individual organelles need to control the expression of genes that encode components of their electron-transport chain so that they can be synthesized when they are needed to maintain redox balance, thereby avoiding the production of highly toxic reactive oxygen species[90]. Both sides can draw on good experimental evidence in their favour[91–93], adding to the controversy.

However, many mitochondria have a modified genetic code, extensive RNA editing or both. In such cases, the mitochondrial genes are locked in place[93], as are genes for such proteins that have evolved organelle-specific assembly or translation mechanisms[94]. The types of gene that plastids (FIG. 2) and mitochondria of diverse lineages most tenaciously retain fit well with the redox-control hypothesis[90], as do the findings that hydrogenosomes (anaerobic mitochondria) lack electron transport and a genome, despite possessing many hydrophobic proteins in their membranes[13]. Clearly, many factors can influence the evolutionary trends of organelle genome persistence[49,54,90–96], and continued debate on this important issue is assured.

Transfers and natural variation. The observation that 1 in 16,000 male gametes in tobacco contains a brand new plastid DNA integrant raises the question as to why plant nuclear genomes are not overflowing with such sequences. A process of sequence elimination must counterbalance insertion. Genome sequencing has shown that although organelle sequences that are present in the nuclear genome vary in age, most are very similar to their organelle counterpart (often having >95% sequence identity). If promiscuous sequence elimination were a slow process, we would expect that most of these sequences would be highly divergent. These data indicate a high turnover of organelle sequences in the nucleus that is similar to that observed for other non-essential, intergenic sequences[97], but with recombination events occasionally leading to selection fixation as functional genes[98].

How important is the continuing contribution of organellar DNA to genetic variation in the nuclear genome? Unlike transposable elements, nuclear-encoded organelle sequences cannot undergo autonomous transposition, which probably explains their modest colonization of the nuclear genome compared with that achieved by *bona fide* transposons. Nonetheless, if the transfer rates that were calculated recently are typical, one in every few thousand plants we see in nature possesses a fresh piece of cpDNA in its nucleus that it acquired only one generation ago[83], a frequency that compares with the nuclear mutation rate[84,98], in which case nearly every plant would have its own individual nupt content.

GM crops and gene transfer

In addition to its pivotal role in eukaryotic genome evolution, gene transfer from organelles to the nucleus is also relevant in the debate that concerns the containment of GM crops. The chloroplast genetic compartment is the focus of one important strategy for transgene containment in GM crops[99]. The high estimates for chloroplast-to-nucleus transfer rates[83,84] cast doubt on claims that chloroplast transgenes can be reliably contained in most potential crop plants through strict maternal inheritance. This has evoked a heated debate[100,101] on the issue of whether crop biotechnology strategies that focus on cpDNA should recognize a measurable potential for escape of transgenes in pollen nuclei.

Of course, the experimental strategies that are applied to detect high-frequency transfer used genes that carry their own nuclear promoter so that any integrant that carries the whole gene can readily be detected. By contrast, genes that are tailored for expression in the chloroplast would need to acquire a nuclear promoter through recombination, analogous to the PROMOTER-TRAPPING technique. The functional transfer and expression of an organelle-specific transgene should therefore be orders of magnitude less frequent than the primary event of DNA relocation. The introduction of organelle-specific introns into chloroplast transgenes might further aid the prevention of their functional transfer to the nucleus. The chloroplast-specific marker genes that are used to select for plastome transformation were nearly always transferred to the nucleus along with the selectable marker but not expressed from that new location[83,84]. Therefore, it should be possible to equip genes that are designed for high activity in the chloroplast with tight control of expression that would essentially preclude any nuclear function. Nuclear-specific suicide cassettes that are introduced adjacent to or within cpDNA transgenes might also aid containment strategies.

Endosymbiotic gene transfer: bigger questions

Is endosymbiotic gene transfer a quirk of evolution that affects only the tips in the tree of life, or is it the mechanism that forged eukaryotes out of prokaryotes? The answer to this question hinges on the issue of how many genes eukaryotes acquired from their mitochondrial and chloroplast endosymbionts[102,103] and what

PROMOTER TRAP
A genetic engineering technique that involves randomly inserting into the genome constructs that encode an easily detectable marker, such as GFP, but contain no promoter sequences. Marker expression is only detected when the construct lands near an endogenous genomic promoter.

kinds of genes those free-living prokaryotes possessed. Today, thinking on the relatedness of prokaryotes and eukaryotes is still dominated by the 'universal' tree of rRNA, which would have us believe that eukaryotes should possess mainly, if not exclusively, genes that are archaebacterial in origin[104]. As we gather more data, this view is looking increasingly insecure[85]. For example, Rivera *et al.*[105] found that about two-thirds of the genes in the yeast genome that had identifiable prokaryotic homologues are more similar to eubacterial homologues than to archaebacterial homologues. Current views on how so many genes of eubacterial origin came to reside in the eukaryotic nuclear genome fall into a range, the extremes of which can be labelled as 'mitochondria' and 'lateral gene transfer' (LGT). One extreme ('mitochondria'), holds the view that all these eubacterial-like genes ultimately stem from the ancestral mitochondrial genome. At the other extreme is the view that the overall gene contribution from mitochondria is small and that various eukaryotic lineages have acquired eubacterial genes either through lineage-specific lateral transfers or from a mysterious symbiont in the ancient past.

The main difficulty with the 'mitochondria' view is that most eubacterial-like genes in the eukaryotic genome do not branch specifically with α-proteobacterial homologues in phylogenetic trees. To account for this, the 'mitochondria' view requires a few tenable corollary assumptions. First, gene phylogenies are imperfect (that is, gene trees produce erroneous branches for mathematical reasons)[19,106–109]. Second, sampling is incomplete (that is, more α-proteobacterial genomes will provide a fuller picture)[20]. Third, LGT among prokaryotes in the ~two billion years since the origin of mitochondria has mixed up the chromosomes of free-living prokaryotes, thereby confounding today's trees[110]. Fourth, when genes are transferred from organelles to the nucleus, they undergo a phase of evolution during which they acquire some odd mutations before they become fully functional[8], which can alter their position in gene trees[76].

The main difficulty with the LGT view is the lack of direct evidence from eukaryotic genomes in its favour. Initial analysis of the human genome sequence indicated that LGT from free-living prokaryotes to eukaryotes is widespread. However, a broader sampling of eukaryotic lineages showed that the initial evidence was far from conclusive[69] — a salutary reminder that our sample of genomes (specifically among eukaryotes, cyanobacteria and α-proteobacteria) is still extremely small. Nonetheless, there do seem to be some lineage-specific acquisitions in eukaryotes, as recent findings attest[103,110]. However, most of the evidence that favours LGT over eukaryotes is based on conflicting gene trees[63], whereas theory and practice indicate that conflicting trees are to be expected even without LGT (REFS 106–109).

The argument that mysterious symbionts from the ancient past donated genes to eukaryotes has recently taken a blow. Mystery symbionts helped explain how 'primitive' eukaryotes that were thought to lack mitochondria, such as the paradigmatic *Giardia intestinalis*, could have obtained their eubacterial genes[111,112]. However, new data show that *Giardia* is not primitive[108,113], and that it has mitochondria[18] too, so a possible mitochondrial origin of eubacterial-like genes applies to the *Giardia* genome as well.

On balance, the evidence from the increasing number of eukaryotic genome sequences might tend to favour the 'mitochondria' end of this range of views. Specifically, there is abundant evidence for integrated fragments of organelle DNA but no reports of an integrated bacterial chromosome segment have emerged from genome sequences[98], not even *Drosophila*, which harbours symbiotic α-proteobacteria[20] (a possible transfer in a different insect[114] aside). So, organelle-to-nucleus transfer is widespread, continuing, real and abundant, but outright prokaryote-to-nucleus transfer seems to be rare, as a simple calculation illustrates. The sum of ~800 individual insertions of organelle DNA in yeast, human and *Arabidopsis* genomes, as well as chromosome 1 and 10 of rice, totals ~1.2 Mb of DNA, but there is no reported evidence at all for integrated chromosome fragments from free-living prokaryotes in those genomes. Accordingly, the contribution of LGT from prokaryotes to eukaryotic DNA in recent evolutionary history can be inferred from the small sample, which is, at most, 1/800th that of the contribution from organelles, and, at most, 1/12,000th that of organelles in terms of length of integrated DNA fragments that are more than 100 bp long. In terms of its directly observable impact on those eukaryotic genomes today, endosymbiotic gene transfer tends to outpace LGT.

Looking into the depths of eukaryotic evolutionary history when organelle genomes were still as big as their prokaryotic ancestors, and when the host genome lacked everything that it later acquired from organelles, the downpour of DNA from organelles must have decisively shaped the eukaryotic genome. After all, at the time when ancestral mitochondria first took up residence in their host, there were neither transit peptides nor was there a protein import machinery[115] to target the protein products of transferred genes back to mitochondria. Accordingly, early transfers from the primitive mitochondrion (a fully-fledged eubacterium) would have enriched the archaebacterial-derived chromosomes of the host with a whole genome's worth of eubacterial genes, over and over again. However, expressed products could have been targeted only to the host's cytosol and plasma membrane and not to the organelle[116]. Only after protein import into mitochondria had evolved (and later, independently for chloroplasts[117]), could the process of organelle genome reduction begin. In its youth, endosymbiotic gene transfer was a powerful and chimaera-generating mechanism of natural variation that is truly unique to the eukaryotic lineage. Indeed, a look into prokaryotic chromosomes shows that they possess nearly all the attributes of eukaryotic chromosomes[118]; what is unique in eukaryotes is that there is more than one genetic compartment[1,2].

Conclusions

Only 20 years have passed since gene transfers between eukaryotic genetic compartments became known. Individual case studies and genome sequences have marshalled overwhelming evidence for its continuous workings over evolutionary time. However, with organelle transformation now on the go, endosymbiotic gene transfer has changed from a compelling logical inference into a directly observable process, opening the door to new progress in determining mechanisms and possibly even evolutionary experimentation. The future prospects seem bright for understanding the deeper evolutionary importance of endosymbiotic gene transfer and its role in shaping the compartmentalized chromosomes and heterogeneous biochemical organization of organelle-bearing (eukaryotic) cells.

1. Baur, E. Das Wesen und die Erblichkeitsverhältnisse der 'Varietates albomarginatae hort' von *Pelargonium zonale*. *Z. Vererbungsl.* **1**, 330–351 (1909).
2. Mereschkowsky, C. Über Natur und Ursprung der Chromatophoren im Pflanzenreiche. *Biol. Centralbl.* **25**, 593–604 (1905). [English translation *Eur. J. Phycol.* **34**, 287–295, 1999].
The starting point of endosymbiotic theory. Outlines the reasoning that we still use today to explain the origin of plastids from cyanobacteria, a paper that was many decades ahead of its time.
3. Margulis, L. *Origin of Eukaryotic Cells* 349 (Yale Univ. Press, New Haven, 1970).
The rediscovery of endosymbiotic theory after Wilson's 1928 condemnation of symbiosis as an evolutionary mechanism (see also reference 2).
4. Gray, M. W. & Doolittle, W. F. Has the endosymbiont hypothesis been proven? *Microbiol. Rev.* **46**, 1–42 (1982).
5. Bogorad, L. Evolution of organelles and eukaryotic genomes. *Science* **188**, 891–898 (1975).
6. Ellis, R. J. Chloroplast proteins: synthesis, transport and assembly. *Ann. Rev. Pl. Physiol.* **32**, 111–137 (1981).
7. Weeden, N. F. Genetic and biochemical implications of the endosymbiotic origin of the chloroplast. *J. Mol. Evol.* **17**, 133–139 (1981).
8. Martin, W. & Herrmann, R. G. Gene transfer from organelles to the nucleus: how much, what happens, and why? *Plant Physiol.* **118**, 9–17 (1998).
9. Simpson, C. L. & Stern, D. B. The treasure trove of algal chloroplast genomes. Surprises in architecture and gene content, and their functional implications. *Plant Physiol.* **129**, 957–966 (2002).
10. Gray, M. W., Burger, G. & Lang, B. F. Mitochondrial evolution. *Science* **283**, 1476–1481 (1999).
11. Lang, B. F., Gray, M. W. & Burger, G. Mitochondrial genome evolution and the origin of eukaryotes. *Annu. Rev. Genet.* **33**, 351–397 (1999).
12. Burger, G., Forget, L., Zhu, Y., Gray, M. W. & Lang, B. F. Unique mitochondrial genome architecture in unicellular relatives of animals. *Proc. Natl Acad. Sci. USA* **100**, 892–897 (2003).
13. Embley, T. M. *et al.* Hydrogenosomes, mitochondria and early eukaryotic evolution. *IUBMB Life* **55**, 387–395 (2003).
An incisive and up-to-date review that covers the biology and evolutionary significance of hydrogenosomes, the mitochondria that endosymbiotic theory nearly forgot.
14. Zhang, Z., Green, B. R. & Cavalier-Smith, T. Single gene circles in dinoflagellate chloroplast genomes. *Nature* **400**, 155–159 (1999).
15. Hannaert, V. *et al.* Plant-like traits associated with metabolism of *Trypanosoma* parasites. *Proc. Natl Acad. Sci. USA* **100**, 1067–1071 (2003).
16. Cavalier-Smith T. Membrane heredity and early chloroplast evolution. *Trends Plant Sci.* **5**, 174–182 (2000).
17. Tielens, A. G., Rotte, C., van Hellemond, J. J. & Martin, W. Mitochondria as we don't know them. *Trends Biochem. Sci.* **27**, 564–572 (2002).
18. Tovar, J. *et al.* Mitochondrial remnant organelles of *Giardia* function in iron-sulphur protein maturation. *Nature* **426**, 172–176 (2003).
19. Martin, W. *et al.* Evolutionary analysis of *Arabidopsis*, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. *Proc. Natl Acad. Sci. USA* **99**, 12246–12251 (2002).
Shows that approximately 18% of the nuclear genes in *Arabidopsis* come from the ancestral plastid genome.
20. Wu, M. *et al.* The genome sequence and evolution of the reproductive parasite *Wolbachia pipientis* wMel: a streamlined α-proteobacterium massively infected with mobile genetic elements. *PLoS Biology* (in the press) .
Incisive evolutionary insights into endosymbiont genome biology with a genome phylogeny for mitochondrial origin.
21. Stern, D. B. & Lonsdale, D. M. Mitochondrial and chloroplast genomes of maize have a 12-kilobase DNA sequence in common. *Nature* **299**, 698–702 (1982).
The paper that initiated progress; showed that DNA was able to migrate between the genetic compartments of eukaryotes.
22. Jacobs, H. T. *et al.* Mitochondrial DNA sequences in the nuclear genome of *Strongylocentrotus purpuratus*. *J. Mol. Biol.* **165**, 609–632 (1983).
23. Farrely, F. & Butow, R. A. Rearranged mitochondrial genes in the yeast nuclear genome. *Nature* **301**, 296–301 (1983).
24. Timmis, J. N. & Scott, N. S. Spinach nuclear and chloroplast DNAs have homologous sequences. *Nature* **305**, 65–67 (1983).
25. Ellis, R. J. Promiscuous DNA — chloroplast genes inside plant mitochondria. *Nature* **299**, 678–679 (1982).
26. Lopez, J. V., Yuhki, N., Masuda, R., Modi, W. & O'Brien, S. J. Numt, a recent transfer and tandem amplification of mitochondrial DNA to the nuclear genome of the domestic cat. *J. Mol. Evol.* **39**, 174–190 (1994).
27. Bensasson, D., Zhang, D. X. & Hewitt, G. M. Frequent assimilation of mitochondrial DNA by grasshopper nuclear genomes. *Mol. Biol. Evol.* **17**, 406–415 (2000).
28. Mundy, N. I., Pissinatti, A. & Woodruff, D. S. Multiple nuclear insertions of mitochondrial cytochrome *b* sequences in callitrichine primates. *Mol. Biol. Evol.* **17**, 1075–1080 (2000).
29. Lu, X. M., Fu, Y. X. & Zhang, Y. P. Evolution of mitochondrial cytochrome *b* pseudogene in genus *Nycticebus*. *Mol. Biol. Evol.* **19**, 2337–2341 (2002).
30. Williams, S. T. & Knowlton, N. Mitochondrial pseudogenes are pervasive and often insidious in the snapping shrimp genus *Alpheus*. *Mol. Biol. Evol.* **18**, 1484–1493 (2001).
31. Olson, L. E. & Yoder, A. D. Using secondary structure to identify ribosomal numts: cautionary examples from the human genome. *Mol. Biol. Evol.* **19**, 93–100 (2002).
Together with reference 32, points out that numts are often mistaken for genuine mitochondrial DNA sequences.
32. Bensasson, D., Zhang, D., Hartl, D. L. & Hewitt, G. M. Mitochondrial pseudogenes: evolution's misplaced witnesses. *Trends Ecol. Evol.* **16**, 314–321 (2001).
33. Ricchetti, M., Fairhead, C. & Dujon, B. Mitochondrial DNA repairs double strand breaks in yeast chromosomes. *Nature* **402**, 96–100 (1999).
The initial genome-wide survey for numts; indicates a role for recombination in numt integration.
34. Mourier, T., Hansen, A. J., Willerslev, E. & Arctander, P. The human genome project reveals a continuous transfer of large mitochondrial fragments to the nucleus. *Mol. Biol. Evol.* **18**, 1833–1837 (2001).
35. Tourmen, Y. *et al.* Structure and chromosomal distribution of human mitochondrial pseudogenes. *Genomics* **80**, 71–77 (2002).
36. Hazkani-Covo, E., Sorek, R. & Graur, D. Evolutionary dynamics of large numts in the human genome: rarity of independent insertions and abundance of post-insertion duplications. *J. Mol. Evol.* **56**, 169–174 (2003).
A careful and detailed inspection of numt duplication dynamics during human and primate genome evolution.
37. Woischnik, M. & Moraes, C. T. Pattern of organisation of human mitochondrial pseudogenes in the nuclear genome. *Genome Res.* **12**, 885–893 (2002).
38. Richly, E. & Leister, D. Numts in sequenced eukaryotic genomes. *Mol. Biol. Evol.* (in the press).
39. Lin, X. Y. *et al.* Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*. *Nature* **402**, 761–768 (1999).
Reports a nearly complete, ~270-kb copy of the 367-kb *Arabidopsis* mtDNA near the centromere.
40. Stupar, R. M. *et al.* 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. USA* **98**, 5099–5103 (2001).
Shows that the 270-kb copy in reference 39 is really the complete 367-kb circle.
41. The *Arabidopsis* Genome Initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**, 796–815 (2000).
42. Yuan, Q. *et al.* Genome sequencing of 239-kb region of rice chromosome 10L reveals a high frequency of gene duplication and a large chloroplast DNA insertion. *Mol. Genet. Genom.* **267**, 713–720 (2002).
First hints from genome sequences of large cpDNA chunks (>30 kb) that are integrated in nuclear chromosomes.
43. The Rice Chromosome 10 Sequencing Consortium. In-depth view of structure, activity, and evolution of rice chromosome 10. *Science* **300**, 1566–1569 (2003).
Reports a nearly complete cpDNA genome chunk (130 kb) that is integrated in the nuclear chromosome.
44. Shahmuradov, I. A., Akbarova, Y. Y., Solovyev, V. V. & Aliyev, J. A. Abundance of plastid DNA insertions in nuclear genomes of rice and *Arabidopsis*. *Plant Mol. Biol.* **52**, 923–934 (2003).
45. Ayliffe, M. A. & Timmis, J. N. Tobacco nuclear DNA contains long tracts of homology to chloroplast DNA. *Theor. Appl. Genet.* **85**, 229–238 (1992).
46. Ayliffe, M. A. & Timmis, J. N. Plastid DNA sequence homologies in the tobacco nuclear genome. *Mol. Gen. Genet.* **236**, 105–112 (1992).
47. Boore, J. L. Animal mitochondrial genomes. *Nucleic Acids Res.* **27**, 1767–1780 (1999).
48. Brennicke, A., Grohmann, L., Hiesel, R., Knoop, V. & Schuster, W. The mitochondrial genome on its way to the nucleus: different stages of gene transfer in higher plants. *FEBS Lett.* **325**, 140–145 (1993).
49. Adams, K. L. & Palmer, J. D. Evolution of mitochondrial gene content: gene loss and transfer to the nucleus. *Mol. Phylog. Evol.* **29**, 380–395 (2003).
A lucid review of flowering plant mitochondrial gene migration to the nucleus.
50. Adams, K. L., Daley, D. O., Qiu, Y. L., Whelan, J. & Palmer, J. D. Repeated, recent and diverse transfers of a mitochondrial gene to the nucleus in flowering plants. *Nature* **408**, 354–357 (2000).
51. Adams, K. L, Qiu Y.-L., Stoutemyer, M. & Palmer J. D. Punctuated evolution of mitochondrial gene content: high and variable rates of mitochondrial gene loss and transfer to the nucleus during angiosperm evolution. *Proc. Natl Acad. Sci. USA* **99**, 9905–9912 (2002).
A broad survey of mitochondrial genome reduction that is accompanied by nuclear integration events in flowering plant evolution.
52. Henze, K. & Martin, W. How are mitochondrial genes transferred to the nucleus? *Trends Genet.* **17**, 383–387 (2001).
53. Thorsness, P. E. & Weber, E. R. Escape and migration of nucleic acids between chloroplasts, mitochondria, and the nucleus. *Int. Rev. Cytol.* **165**, 207–234 (1996).
An excellent review that covers the mechanics of gene transfer from organelles to the nucleus in the pre-genome era.
54. Blanchard, J. L. & Lynch, M. Organellar genes — why do they end up in the nucleus? *Trends Genet.* **16**, 315–320 (2000).
55. Figueroa, P., Gomez, I., Holuigue, L., Araya, A. & Jordana, X. 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 (1999).
Fortuitous recombination in the establishment of active gene transfers.
56. Kubo, N., Harada, K., Hirai, A. & Kadowaki, K. A single nuclear transcript encoding mitochondrial RPS14 and SSDHB of rice is processed by alternative splicing: common use of the same mitochondrial targeting signal for different proteins. *Proc. Natl Acad. Sci. USA* **96**, 9207–9211 (1999).
The establishment of active gene transfers can involve the recruitment of pre-existing transit peptide regions.

57. Long, M., de Souza, S. J., Rosenberg, C. & Gilbert, W. Exon shuffling and the origin of the mitochondrial targeting function in plant cytochrome c1 precursor. *Proc. Natl Acad. Sci. USA* **93**, 7727–7731 (1996).
Gene transfers can involve conversion of a pre-existing nuclear coding region for a cytosolic enzyme into a transit peptide.
58. Millen, R. S. *et al.* Many parallel losses of *infA* from chloroplast DNA during angiosperm evolution with multiple independent transfers to the nucleus. *Plant Cell* **13**, 645–658 (2001).
59. Martin, W. *et al.* Gene transfer to the nucleus and the evolution of chloroplasts. *Nature* **393**, 162–165 (1998).
60. Lang, B. F. *et al.* An ancestral mitochondrial DNA resembling a eubacterial genome in miniature. *Nature* **387**, 493–497 (1997).
The still unsurpassed mitchondrial genome in terms of gene content and streamlined organization.
61. Henze, K. *et al.* A nuclear gene of eubacterial origin in *Euglena gracilis* reflects cryptic endosymbioses during protist evolution. *Proc. Natl Acad. Sci. USA* **92**, 9122–9126 (1995).
62. Stibitz, T. B., Keeling, P. J., & Bhattacharya, D. Symbiotic origin of a novel actin gene in the cryptophyte *Pyrenomonas helgolandii*. *Mol. Biol. Evol.* **17**, 1731–1738 (2000).
63. Archibald, J. M., Rogers, M. B., Toop, M., Isheda, K. & Keeling, P. J. Lateral gene transfer and the evolution of plastid-targeted proteins in the secondary plastid-containing alga *Bigelowiella natans*. *Proc. Natl Acad. Sci. USA* **100**, 7678–7683 (2003).
64. Leister, D. Chloroplast research in the genomics age. *Trends Genet.* **19**, 47–56 (2003).
65. Gabaldón, T. & Huynen, M. A. Reconstruction of the proto-mitochondrial metabolism. *Science* **301**, 609 (2003).
66. Martin, W. & Schnarrenberger, C. 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 (1997).
67. Gallois, J. L. *et al.* The *Arabidopsis* chloroplast ribosomal protein L21 is encoded by a nuclear gene of mitochondrial origin. *Gene* **274**, 179–185 (2001).
68. Adams, K. L. *et al.* Genes for two mitochondrial ribosomal proteins in flowering plants are derived from their chloroplast or cytosolic counterparts. *Plant Cell* **14**, 931–943 (2002).
69. Brown, J. R. Ancient horizontal gene transfer. *Nature Rev. Genet*. **4**, 121–132 (2003).
An incisive review of gene movement across genomes, including the role of endosymbiotic transfers.
70. Gil, R. *et al.* The genome sequence of *Blochmannia floridanus*: comparative analysis of reduced genomes. *Proc. Natl Acad. Sci. USA* **100**, 9388–9393 (2003).
Underscores how reductive evolution in endosymbiotic bacteria leads to massive gene losses through biochemical parasitism of the host.
71. Lister, D. L., Bateman, J. M., Purton, S. & Howe, C. J. DNA transfer from chloroplast to nucleus is much rarer in *Chlamydomonas* than in tobacco. *Gene* **316**, 33–38 (2003).
72. Richly, E., Chinnery, P. F. & Leister, D. Evolutionary diversification of mitochondrial proteomes: implications for human disease. *Trends Genet.* **19**, 356–362 (2003).
73. Richly, E. & Leister, D. An improved prediction of chloroplast proteins reveals diversities and commonalities in the chloroplast proteomes of *Arabidopsis* and rice. *Gene* (in the press).
74. Huh, W.-K. *et al.* Global analysis of protein localization in budding yeast. *Nature* **425**, 686–691 (2003).
75. Lange, B. M., Rujan, T., Martin, W. & Croteau, R. Isoprenoid biosynthesis: the evolution of two ancient and distinct pathways across genomes. *Proc. Natl Acad. Sci. USA* **97**, 13172–13177 (2000).
76. Schnarrenberger, C. & Martin, W. Evolution of the enzymes of the citric acid cycle and the glyoxylate cycle of higher plants: a case study of endosymbiotic gene transfer. *Eur. J. Biochem.* **269**, 868–883 (2002).
77. Aravind, L., Anantharaman, V. & Iyer, L. M. Evolutionary connections between bacterial and eukaryotic signaling systems: a genomic perspective. *Curr. Opin. Microbiol.* **6**, 490–497 (2003).
78. Osteryoung, K. W. & Nunnari, J. The division of endosymbiotic organelles. *Science* **302**, 1698–1704 (2003).
79. McFadden, G. I. & Ralph, S. A. Dynamin: the endosymbiosis ring of power? *Proc. Natl Acad. Sci. USA* **100**, 3557–3559 (2003).
80. Thorsness, P. E. & Fox, T. D. Escape of DNA from the mitochondria to the nucleus in the yeast, *Saccharomyces cerevisiae*. *Nature* **346**, 376–379 (1990).
The initial experimental measurement of the frequency of transfer of DNA between genetic compartments.
81. Thorsness, P. E. & Fox, T. D. Nuclear mutations in *Saccharomyces cerevisiae* that affect the escape of DNA from mitochondria to the nucleus. *Genetics* **134**, 21–28 (1993).
82. Shafer, K. S., Hanekamp, T., White, K. H. & Thorsness, P. E. Mechanisms of mitochondrial DNA escape to the nucleus in the yeast *Saccharomyces cerevisiae*. *Curr. Genet*. **36**, 183–194 (1999).
83. Huang, C. Y., Ayliffe, M. A. & Timmis, J. N. Direct measurement of the transfer rate of chloroplast DNA into the nucleus. *Nature* **422**, 72–76 (2003).
The initial experimental measurement of the frequency of integrative transfer of DNA between chloroplast and nucleus.
84. Stegemann, S., Hartmann, S., Ruf, S. & Bock, R. High-frequency gene transfer from the chloroplast genome to the nucleus. *Proc. Natl Acad. Sci. USA* **100**, 8828–8833 (2003).
85. Campbell, C. L. & Thorsness, P. E. Escape of mitochondrial DNA to the nucleus in yme1 yeast is mediated by vacuolar-dependent turnover of abnormal mitochondrial compartments. *J. Cell Sci.* **111**, 2455–2464 (1998).
86. Shay, J. W. & Werbin, H. New evidence for the insertion of mitochondrial DNA into the human genome: significance for cancer and aging. *Mutat. Res.* **275**, 227–235 (1992).
87. Turner, C. *et al.* Human genetic disease caused by *de novo* mitochondrial-nuclear DNA transfer. *Hum. Genet.* **112**, 303–309 (2003).
88. Elo, A. *et al.* Nuclear genes that encode mitochondrial proteins for DNA and RNA metabolism are clustered in the *Arabidopsis genome*. *Plant Cell* **15**, 1619–1631 (2003).
89. Richly, E. *et al.* Covariations in the nuclear chloroplast transcriptome reveal a regulatory master-switch. *EMBO Rep.* **4**, 491–498 (2003).
90. Allen, J. F. The function of genomes in bioenergetic organelles. *Phil. Trans. Roy. Soc. B* **358**, 19–38 (2003).
A comprehensive treatment of competing views on the issue of why some genes remain within organelle genomes.
91. Pérez-Martínez, X. *et al.* Subunit II of cytochrome *c* oxidase in chlamydomonad algae is a heterodimer encoded by two independent nuclear genes. *J. Biol. Chem.* **276**, 11302–11309 (2001)
92. Daley, D. O. *et al.* Intracellular gene transfer: reduced hydrophobicity facilitates gene transfer for subunit 2 of cytochrome c oxidase. *Proc. Natl Acad. Sci. USA* **99**, 10510–15015 (2002).
93. Pfannschmidt, T., Nilsson, A., & Allen, J. F. Photosynthetic control of chloroplast gene expression. *Nature* **397**, 625–628 (1999).
94. Naithani S., Saracco S. A., Butler C. A. & Fox T. D. Interactions among COX1, COX2, and COX3 mRNA-specific translational activator proteins on the inner surface of the mitochondrial inner membrane of *Saccharomyces cerevisiae*. *Mol. Biol. Cell* **14**, 324–333 (2003).
95. Doolittle, W. F. You are what you eat: a gene transfer ratchet could account for bacterial genes in eukaryotic nuclear genomes. *Trends Genet.* **14**, 307–311 (1998).
96. Race, H. L., Herrmann, R. G. & Martin, W. Why have organelles retained genomes? *Trends Genet.* **15**, 364–370 (1999).
97. Devos, K. M., Brown, J. K. M. & Bennetzen J. L. Genome size reduction through illegitimate recombination counteracts genome expansion in *Arabidopsis*. *Genome Res.* **12**, 1075–1079 (2002).
98. Martin, W. Gene transfers from organelles to the nucleus: frequent and in big chunks. *Proc. Natl Acad. Sci. USA* **100**, 8612–8614 (2003).
99. Maliga, P. Engineering the plastid genome of higher plants. *Curr. Opin. Plant Biol.* **5**, 164–172 (2002).
100. Daniell, H. & Parkinson, C. L. Jumping genes and containment. *Nature Biotechnol.* **21**, 374–375 (2003).
A challenge to the experimental data for gene transfer from chloroplasts by proponents of plastid transgene technology, rebutted head-to-head in reference 101.
101. Huang, C. Y., Ayliffe, M. A. & Timmis, J. N. Organelle evolution meets biotechnology. *Nature Biotechnol.* **21**, 489–490 (2003).
102. Martin, W. & Russell, M. On the origins of cells: a hypothesis for the evolutionary transitions from abiotic geochemistry to chemoautotrophic prokaryotes, and from prokaryotes to nucleated cells. *Phil. Trans. Roy. Soc. Lond. B* **358**, 59–85 (2003).
103. Doolittle, W. F. *et al.* How big is the iceberg of which organellar genes in nuclear genomes are but the tip? *Philos. Trans. Roy. Soc. Lond. B* **358**, 39–57 (2003).
104. Woese, C., Kandler, O. & Wheelis, M. L. Towards a natural system of organisms: proposal for the domains archaea, bacteria and eucarya. *Proc. Natl Acad. Sci. USA* **87**, 4576–4579 (1990).
The current higher-level taxonomic model, with eukaryotes as sisters to archaebacteria.
105. Rivera, M. C., Jain, R., Moore, J. E. & Lake, J. A. Genomic evidence for two functionally distinct gene classes. *Proc. Natl Acad. Sci. USA* **95**, 6239–6244 (1998).
A landmark paper that uncovers more eubacterial genes than archaebacterial genes in the yeast genome.
106. Penny, D., Foulds, L. R. & Hendy, M. D. Testing the theory of evolution by comparing phylogenetic trees constructed from five different protein sequences. *Nature* **297**, 197–200 (1982).
107. Cummings, M. P., Otto, S. P. & Wakeley, J. Sampling properties of DNA-sequence data in phylogenetic analysis. *Mol. Biol. Evol.* **12**, 814–822 (1995).
108. Embley, T. M. & Hirt, R. P. Early branching eukaryotes? *Curr. Opin. Genet. Dev.* **8**, 655–661 (1998).
109. Rokas, A., Williams, B. L., King, N. & Carroll, S. B. Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature* **425**, 798–804 (2003).
110. Gogarten P. J., Doolittle W. F. & Lawrence J. G. Prokaryotic evolution in light of lateral gene transfer. *Mol. Biol. Evol.* **19**, 2226–2238 (2002).
111. Hedges, S. B. *et al.* A genomic timescale for the origin of eukaryotes. *BMC Evol. Biol.* **1**, 4 (2001).
112. Hartman, H. & Fedorov, A. The origin of the eukaryotic cell: a genomic investigation. *Proc. Natl Acad. Sci. USA* **99**, 1420–1425 (2002).
113. Stechmann, A. & Cavalier-Smith, T. Rooting the eukaryote tree by using a derived gene fusion. *Science* **297**, 89–91 (2002).
A milestone relating to the issue of which eukaryotes might be the most ancient.
114. Kondo, N., Nikoh, N., Ijichi, N., Shimada, M. & Fukatsu, T. Genome fragment of *Wolbachia* endosymbiont transferred to X chromosome of host insect. *Proc. Natl Acad. Sci. USA* **99**, 14280–14285 (2002).
115. Neupert, W. Protein import into mitochondria. *Ann. Rev. Biochem.* **66**, 683–717 (1997).
116. Martin, W. & Müller, M. The hydrogen hypothesis for the first eukaryote. *Nature* **392**, 37–41 (1998).
117. Soll, J. Protein import into chloroplasts. *Curr. Opin. Plant. Biol.* **5**, 529–535 (2002).
118. Bendich, A. J. & Drlica, K. Prokaryotic and eukaryotic chromosomes: what's the difference? *BioEssays* **22**, 481–486 (2000).
An enlightening survey of chromosome attributes in prokaryotes and eukaryotes.
119. Birky, C. W. The inheritance of genes in mitochondria and chloroplasts: laws, mechanisms and models. *Annu. Rev. Genet.* **35**, 125–148 (2001).
120. Martin, W., Hoffmeister, M., Rotte, C. & Henze, K. An overview of endosymbiotic models for the origins of eukaryotes, their ATP-producing organelles (mitochondria and hydrogenosomes), and their heterotrophic lifestyle. *Biol. Chem.* **382**, 1521–1539 (2001).
121. Butterfield, N. J. *Bangiomorpha pubescens* n. gen., n. sp.: implications for the evolution of sex, multicellularity, and the Mesoproterozoic/ Neoproterozoic radiation of eukaryotes. *Paleobiology* **26**, 386–404 (2000).
A fossil red algae of 1. 2 billion years of age anchors plant evolution in the Precambrian age.

Acknowledgements

We thank M. Limpert for help in preparing the manuscript, the Australian Research Council, the Australian–German Joint Research Cooperation Scheme and the Deutsche Forschungsgemeinschaft for financial support, and D. Leister for valuable discussions and permission to modify published figures. Countless individual report on numts, nupts and eukaryotic genes that were acquired from organelles are available; we apologize to all for having to focus on selected and more recent work.

Competing interests statement

The authors declare that they have no competing financial interests.

Online links

DATABASES

The following terms in this article are linked online to:

TAIR: http://www.arabidopsis.org
rps10

FURTHER INFORMATION

CyanoBase: http://www.kazusa.or.jp/cyano/cyano.html
DOE Joint Genome Institute:
http://www.jgi.doe.gov/JGI_microbial/html
The Institute for Genomic Research: http://www.tigr.org/tdb
The Organelle Genome Megasequencing Program:
http://megasun.bch.umontreal.ca/ogmpproj.html
Access to this interactive links box is free online.