Review TRENDS in Genetics Vol.21 No.12 December 2005

Origin, evolution and genetic effects of nuclear insertions of organelle DNA

Dario Leister

Department Biologie I, Botanik, Ludwig-Maximilians-Universität, Menzingerstr. 67, D-80638 München, Germany

In eukaryotes, nuclear genomes are subject to an influx of DNA from mitochondria and plastids. The nuclear insertion of organellar sequences can occur during the illegitimate repair of double-stranded breaks. After integration, nuclear organelle DNA is modified by point mutations, and by deletions. Insertion of organelle DNA into nuclear genes is not rare and can potentially have harmful effects. In humans, some insertions of nuclear mitochondrial DNA are associated with heritable diseases. It remains to be determined whether nuclear organelle DNA can contribute beneficially to gene evolution.

Introduction

Eukaryotic cells contain up to three genomes: the nuclear DNA (nDNA; see Glossary), and the genomes of mitochondria and plastids. The latter organelles are descended from prokaryotic endosymbionts. Interorganellar DNA transfer redistributes genetic material between these organelles and, of the six possible types of DNA transfer, four have been observed (Box 1). Promiscuous DNA sequences originating from nucleus-to-mitochondrion and plastid-to-mitochondrion transfer have been identified, but only plastid-derived mitochondrial tRNA genes are functional. However, the transfer of DNA from mitochondrion or plastid to the nucleus has contributed significantly to the evolution and function of eukaryotic genomes. DNA transfer to the nucleus is ubiquitous and ongoing, and during the early phase of organelle evolution it resulted in a massive relocation of organellar genes [1,2]. In many eukaryotes, including animals, the transfer of functional genes is now rare or has ceased altogether [3], but transfer of certain mitochondrial and plastid genes has occurred frequently in flowering plants in evolutionarily recent times [4,5].

Almost all present-day nuclear transfers of mitochondrial (mtDNA) and plastid (ptDNA) DNA – collectively referred to here as organelle DNA (orgDNA) – give rise to noncoding sequences, the so-called NUMTs (nuclear mtDNA) and NUPTs (nuclear ptDNA) – here collectively termed nuclear organelle DNA (norgDNA). The systematic, genome-wide analysis of norgDNA in human, *Arabidopsis* and rice, as well as experiments designed to trace DNA transfer under laboratory conditions, have provided insights into the mode of origin and divergence of norgDNA, and its impact on gene and genome evolution. In this review, I summarize current knowledge and recent progress in the field of organelle-to-nucleus transfer of DNA, emphasizing the underlying cellular and genetic mechanisms and their evolutionary consequences.

Corresponding author: Leister, D. (leister@lrz.uni-muenchen.de).
Available online 10 October 2005

Genomic organization of norgDNAs

In almost all eukaryotes studied there exist nDNA sequences that are homologous to mtDNA or ptDNA [6–9]. Although in the African malaria mosquito and in the honeybee no NUMTs have been detected so far, NUMTs vary in copy number from few copies in the genomes of *Caenorhabditis*, *Drosophila* and dog, to more than 500 in rice and *Arabidopsis* [6,8]. NUPTs are rare in *Chlamydomonas* and *Plasmodium*, but frequent in *Arabidopsis* and rice [7]. The average size of norgDNA ranges between 42 bp (*Plasmodium* NUPTs) and 647 bp (*Neurospora* NUMTs), thus varying by more than one order of magnitude [6,7]. The largest norgDNA loci known are found in the domestic cat [10] and in flowering plants [11–13] (Figure 1a). The fraction of the nuclear genome represented by norgDNA is usually less than 0.1% (e.g. 0.016% in humans [14]) but it can be several times higher in flowering plants [6,7]. There is no

Glossary

Autophagy: a process of intracellular degradation in which cytoplasmic components, including organelles, are directed to the lysosome/vacuole by a membrane-mediated process.
Double-stranded break (DSB): a type of DNA damage that differs from other DNA lesions in that both strands of the double helix are damaged, preventing use of the complementary DNA strand as template for repair. DSBs are particularly genotoxic and potent inducers of chromosomal aberrations, including chromosomal translocations. DSBs can be repaired by nonhomologous end-joining.
mtDNA: mitochondrial DNA – DNA present in the mitochondria.
nDNA: nuclear DNA – DNA occurring in the nucleus.
Nonhomologous end-joining (NHEJ): a type of DSB repair, also called 'illegitimate' repair, that uses little or no homology to couple DNA ends. NHEJ includes the recognition of DSBs, processing of the ends to yield short overhangs that allow for searching and annealing of homologous patches, and end-to-end ligation. In this process, linear DNA fragments of extranuclear origin can be utilized to join the break sides.
norgDNA: the fraction of nuclear DNA that derives from the insertion of organelle DNA.
NUMT: nuclear mtDNA – the fraction of nuclear DNA that derives from mitochondria.
NUPT: nuclear ptDNA – the fraction of nuclear DNA that derives from plastids.
orgDNA: organelle DNA – DNA that occurs in or derives from cell organelles.
Pallister–Hall syndrome (PHS): an extremely rare genetic disorder that can be associated with malformation and/or decreased function of the hypothalamus, the presence of extra fingers and/or toes and other symptoms. Cases in which a positive family history has not been found are thought to represent new mutations (sporadic cases).
ptDNA: plastid DNA – DNA present in chloroplasts and other plastids.

 0168-9525/$ - see front matter doi:10.1016/j.tig.2005.09.004

Box 1. DNA flow between different genetic compartments

Six types of DNA transfer are conceivable between the three DNA-containing organelles: nucleus, plastid and mitochondrion. In ptDNAs, no sequence of nuclear or mitochondrial origin has yet been detected, indicating that nucleus-to-plastid or mitochondrion-to-plastid transfer occurs extremely rarely or not at all. During the early phase of organelle evolution, organelle-to-nucleus DNA transfer (designated in Figure I as 'a') resulted in a massive relocation of functional genes to the nucleus: in yeast, as many as 75% of all nuclear genes could derive from protomitochondria [62], whereas ~4500 genes in the nucleus of *Arabidopsis* are of plastid descent [63]. Cases of present-day organelle-to-nucleus DNA transfer, revealed by the presence of NUMTs and NUPTs, are known in most species studied so far. Among the few eukaryotic organisms in which norgDNA has not been detected are the malaria mosquito (*Anopheles gambiae*) and the honeybee (*Apis mellifera*). Mitochondrial chromosomes contain segments homologous to chloroplast sequences, as well as sequences of nuclear origin, providing indirect evidence for plastid-to-mitochondrion and nucleus-to-mitochondrion transfer of DNA (Figure I: 'b' and 'c'). Thus, a few percent of the mtDNA of flowering plants derives from ptDNA, whereas retrotransposons seem to be the major source of nucleus-derived mtDNA. Interestingly, although plastid-to-mitochondrion and nucleus-to-mitochondrion DNA transfer have been detected in almost all plant mitochondrial chromosomes sequenced so far [64,65], there is no evidence for the incorporation of nDNA into the mitochondrial genome of maize [66].

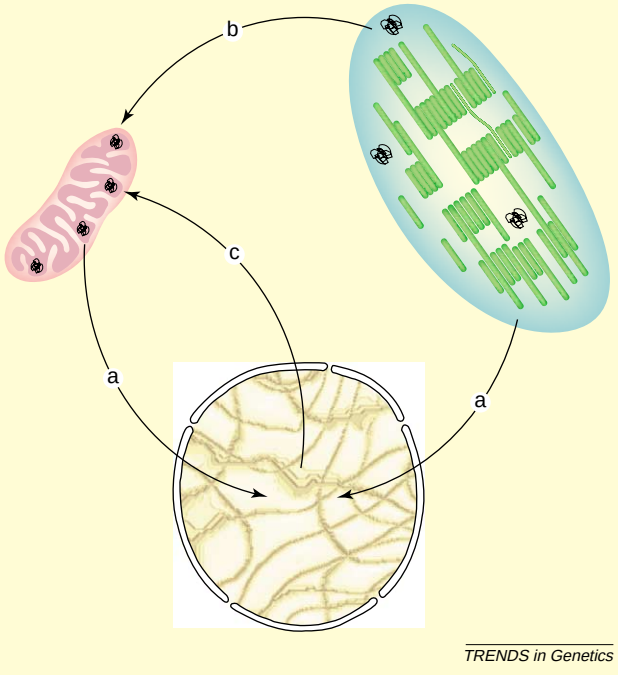

Figure I. Schematic overview of known types of intercompartment DNA transfer. **(a)** Organelle-to-nucleus; **(b)** chloroplast-to-mitochondrion; **(c)** nucleus-to-mitochondrion.

obvious correlation between the abundance of norgDNA and the size of the nuclear or organellar genomes, or of the gene density in the nuclear genome [6]. This, together with the wide variation in the NUPT:NUMT ratio in different plant species [7], indicates that the number and/or stability of organelles in the germ line, or species-specific mechanisms that control the accumulation/loss of nDNA, might be responsible for the interspecific diversity in NUMT and NUPT accumulation.

The genomic organization of norgDNAs has been analysed in greatest detail in yeast, humans, *Arabidopsis* and rice. In the first drafts of the human genome sequence, 200–350 NUMTs were identified [9,14–17], the variation in the number being caused by differences in the search algorithms and thresholds used. Human NUMT loci have a size of up to 14 654 bp [15,16]; they are evenly distributed within and among chromosomes [14,17], and up to 9857 bp of the 16 571 bp contained in human mtDNA can be found at a single locus [16]. A subset of human NUMTs are highly rearranged, comprising sequences derived from different regions of the organellar chromosome that have undergone inversions, deletions and duplications [14,16,18] (Figure 1b). In a more advanced draft of the human genome sequence, 211 and 247 NUMTs, respectively, were identified in two studies [19,20]. Mishmar *et al.* found that more than half of the human NUMT repertoire localizes to repetitive sequences, in particular to mobile elements, and that recent NUMT insertions preferentially occurred in AT-rich isochores [20]. Those NUMTs, which are relatively recent and specific to human, are predominantly inserted in genes [19].

In *Homo sapiens*, phylogenetic analyses have suggested that mtDNA integrates continuously (without a change in integration rate) into the nuclear genome [14,15,17,21], but multiple post-insertional duplications of ancient NUMTs were also reported [16]. Analyses of the largest human NUMTs attempted to quantify the relative contributions of *de novo* nuclear insertions of mtDNA and duplications of pre-existing NUMTs. Between 32% [21] and 85% [17] of the human NUMTs tested were proposed to be a result of *de novo* insertion events, and NUMT duplication might have been discontinuous (with a peak in duplication rate) during primate evolution [21]. Continued transfer of mtDNA to the human nuclear genome was inferred from comparisons of the NUMT repertoires in human and chimpanzee, which revealed the relatively recent acquisition of 27 NUMTs by *H. sapiens* [19].

Analysis of the complete chromosomal sequences of the flowering plants *Arabidopsis* and rice showed that their genomes are particularly rich in norgDNA [6,7]. They also contain the largest insertions of orgDNA reported so far, namely the 620-kb NUMT in *A. thaliana* [12,13] and the 131-kb NUPT in rice [11] (Figure 1a). In both species, norgDNA is frequently nonrandomly organized, forming clusters of insertions that are physically linked to varying degrees ('loose clusters' and 'tight clusters' [7]). Sequence similarity between NUPTs/NUMTs and pt/mtDNA correlates with the size of the integrant. This implies that the primary insertions were large, but decayed over evolutionary time into smaller fragments with diverging sequences, via tight and loose clusters as intermediates [7]. The insertion of non-orgDNA, including transposable elements, into NUPTs and NUMTs contributed significantly to this fragmentation process [22]. Indeed, most of the large rice NUPTs seem to decay within one million years [23]. A more detailed study of the composition of tight clusters showed that, in *Arabidopsis* and rice, two main types of nuclear insertions coexist. One is characterized by long DNA stretches that are colinear with orgDNA, whereas the other consists of mosaics of orgDNA, often derived from both plastids and mitochondria [22]. In some

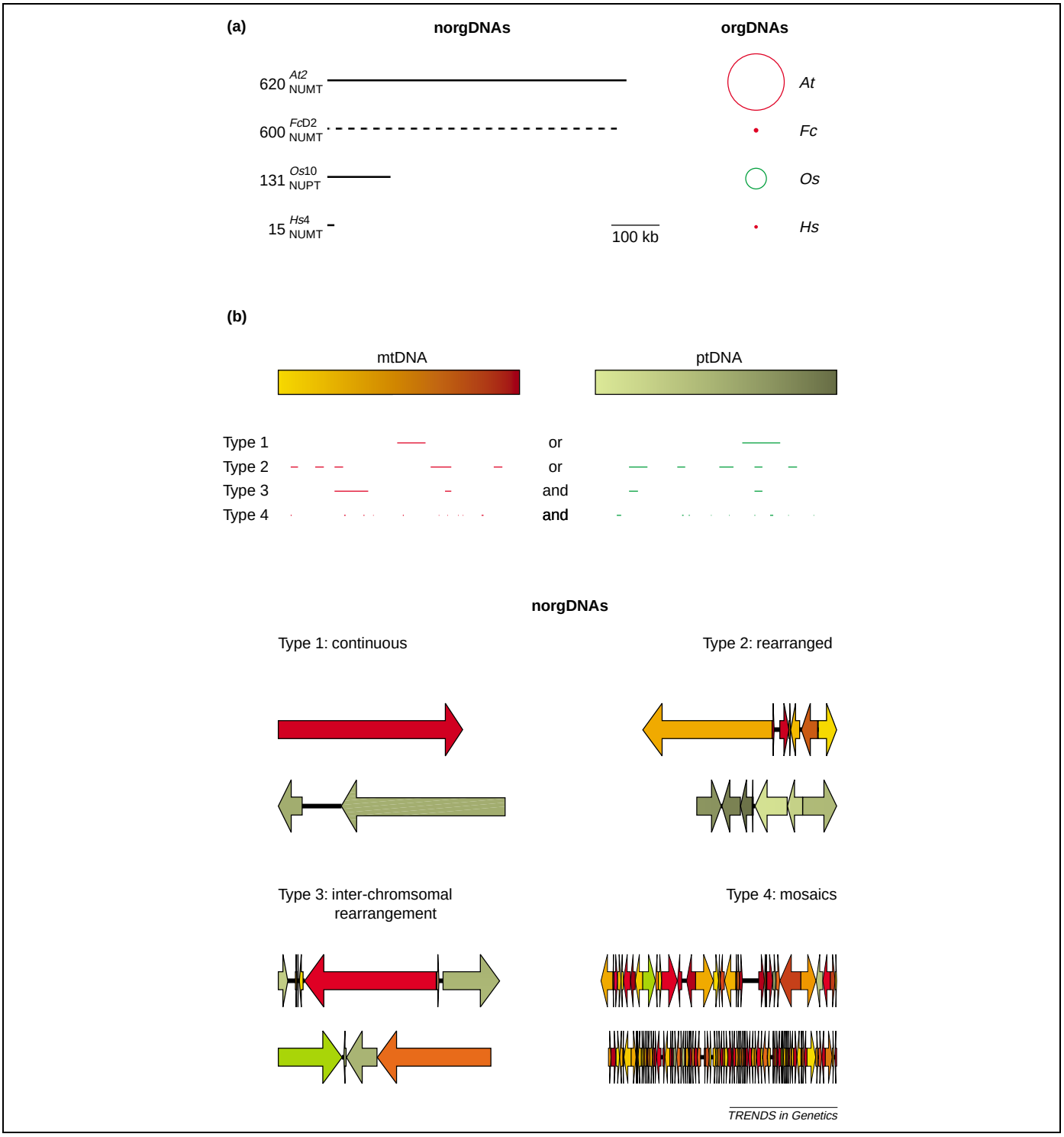

Figure 1. Organization of norgDNAs. **(a)** The largest known norgDNA loci (left) and the respective orgDNAs (right) are displayed as lines and circles, respectively. The norgDNAs are designated according to their size, origin and chromosomal location; e.g. '620^{At2}_{NUMT}' refers to the 620-kb NUMT on chromosome 2 of *A. thaliana*. *Fc* refers to *Felis catus* (domestic cat), *Os* to *O. sativa* (rice) and *Hs* to *H. sapiens*. The orgDNAs are drawn to scale and their colouring indicates their origin (red, mtDNA; green, ptDNA). Actually, the 600-kb cat NUMT is thought to represent a 76-repeat tandem array of a 7.9-kb mtDNA fragment [10]. **(b)** In the upper panel, schematic drawings of mitochondrial and plastid chromosomes (mtDNA and ptDNA) and their representation in the four types of norgDNA (indicated by red or green lines below their regions of origin) are shown. Organelle chromosomes and lines representing norgDNAs are not drawn to scale. In the lower panel, colouring indicates the positions of the homologous sequence in the organelle chromosome (reddish: mtDNA/NUMT, greenish: ptDNA/NUPT). The orientation of norgDNAs relative to orgDNA is indicated by arrowheads (left, reverse orientation). Two main types of organization of norgDNA have been observed in animals and plants: continuous stretches of nDNA colinear with mtDNA or ptDNA (type 1), or rearranged norgDNAs derived from different regions of one organelle chromosome with random orientation (type 2). Because of the presence of both NUMTs and NUPTs in plant nuclear genomes, rearrangements involving different chromosomes can also be detected. The type 3 norgDNAs are complex loci containing both NUMTs and NUPTs, whereas type 4 refers to loci containing large numbers of short NUPTs and NUMTs ('mosaics'). A set of norgDNAs exists that are interspersed with nDNA of non-organelle descent (indicated by black lines) – these are DNA sequences that in the majority of cases resemble transposable elements and disrupt continuous norgDNA sequences. These insertions of non-organelle DNA occurred after establishment of norgDNAs, contributing to locus expansion and to the decay of large continuous norgDNAs into smaller fragments.

cases, the mosaic-type of norgDNA was found to represent up to ~100 distinct fragments, which were, on average, smaller than 100 bp. In the two model plants, ~25% of norgDNA is located within genes. However, this value must be considered in light of the fact that exons and introns together make up 40–50% of the genomes of *Arabidopsis* and rice [7]. In rice, large NUPTs preferentially localize to the pericentromeric regions of the chromosomes; such insertions here should be less deleterious than integrations in other chromosomal regions [23].

Mechanisms of nuclear insertion of orgDNA

Early studies suggested that mtDNA is transferred to the nucleus via an RNA intermediate [24], but this notion has since been revised in light of the results of experimental and bioinformatics studies carried out in yeast and other eukaryotes. These studies showed that (i) any segment of an organellar genome can be transferred to the nucleus, and (ii) that large norgDNAs exist that span several genes or even entire organellar chromosomes [6,7,14,15,22,25,26]. Moreover, analysis of human NUMTs has provided no evidence for splicing or polyadenylation of organellar nucleic acids before insertion [14], indicating that migration of mtDNA sequences to the nucleus is predominantly DNA mediated.

Escape of orgDNA and its uptake into the nucleus has been experimentally demonstrated in yeast [27] and tobacco [28,29]. In yeast, mtDNA can integrate into the nuclear genome by nonhomologous end-joining (NHEJ) repair (illegitimate repair) of double-stranded breaks (DSBs) [30,31]. DSBs are regarded as the most potentially deleterious form of DNA damage that can be induced *in vivo* by exogenous and endogenous sources [32–35] (Figure 2). Repair of DSBs by NHEJ requires little or no sequence homology (0 to 4 bp, 'micro-identities') between the termini, enabling the noncomplementary ends of DSBs and orgDNA to be 'pasted' to one another [30,31,33]. In tobacco, the spectrum of sequences inserted at chromosomal breaks is broader than that seen in yeast [36], whereas no insertions could be detected in *Arabidopsis* [37], implying that DSB repair mechanisms can show major differences even among closely related eukaryotes [34]. Similar patterns of terminal microidentities, however, have been observed for NUMT insertions in yeast and humans [18,19], indicating that DSB-repair-mediated insertion of orgDNA by NHEJ might be a phenomenon common to all eukaryotes. Moreover, the analysis of a reciprocal constitutional translocation in humans, assumed to derive from NHEJ repair of gametogenesis-associated DSBs, indicated that

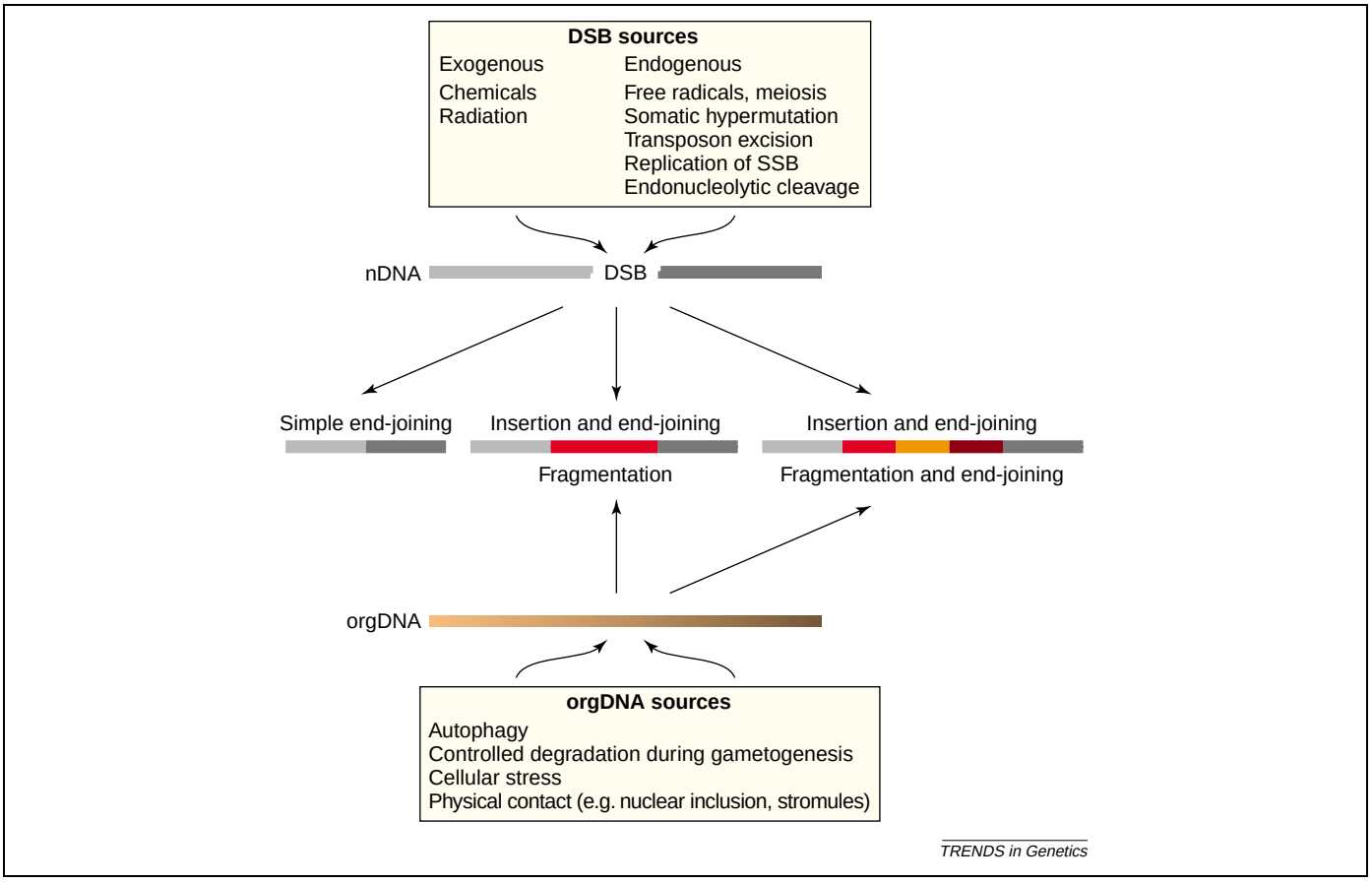

Figure 2. Model of norgDNA generation. DSBs are induced by exogenous and endogenous sources, whereas multiple mechanisms operate in the release of orgDNA. DSB repair by NHEJ can result in different outcomes: simple end-joining of the processed ends of nDNA; the insertion of extrachromosomal DNA with little or no homology to the terminal ends of the DSB; or the insertion of complex sequences to patch the break. Also the complex sequences are thought to derive from an end-joining mechanism. According to this model, any increase in the frequency of formation of DSBs should influence the rate of norgDNA generation. In this context it is interesting to note that the *de novo* 72-bp mtDNA insertion associated with a sporadic case of Pallister–Hall syndrome was temporally and geographically associated with high-level radioactive contamination following the Chernobyl accident [39].

Box 2. Interorganelle associations and developmental processes affecting organelle integrity

Several mechanisms, as well as physical interactions between organelles, have been proposed to facilitate transfer of DNA between mitochondrion (mt), plastid (pt) and nucleus (n). During autophagy, gametogenesis or cellular stress, organelles are degraded and their DNAs are released (Figure Ia). The co-insertion of sequences of mitochondrial and plastid origin into the nuclear genome, resulting in complex norgDNA loci (see Figures 1 and 2), implies the concomitant release of ptDNA and mtDNA. In many eukaryotes, including humans and flowering plants, orgDNA is maternally inherited. Therefore, organelle-to-nucleus transfer of DNA is thought to occur preferentially when a programmed degeneration of organelles takes place: in flowering plants during pollen development and in mammals when sperm mtDNA is released from degenerating mitochondria shortly after penetration of the egg by the sperm cell. Whereas *Arabidopsis* and rice show maternal inheritance of orgDNA, other species exhibit an independent control of the inheritance of mitochondria and plastids [67], indicating that the transfer of ptDNA and mtDNA to the nucleus might also occur at different rates. Fusion of mitochondria [68,69] should have no relevance for organelle-to-nucleus transfer of DNA (Figure Ib), and has been associated with the controversially discussed horizontal transfer of mitochondrial genes [55]. Direct physical association of the nucleus with mitochondria or chloroplasts [70,71], as well as the nuclear inclusion of mitochondria [72,73], might also contribute to DNA exchange (Figure Ic). Plastids can form tubular extensions of their membranes, termed stromules (for stroma-filled tubules), which can associate with other organelles [74], providing the possibility for DNA transfer from plastids to other genetic compartments (Figure Id).

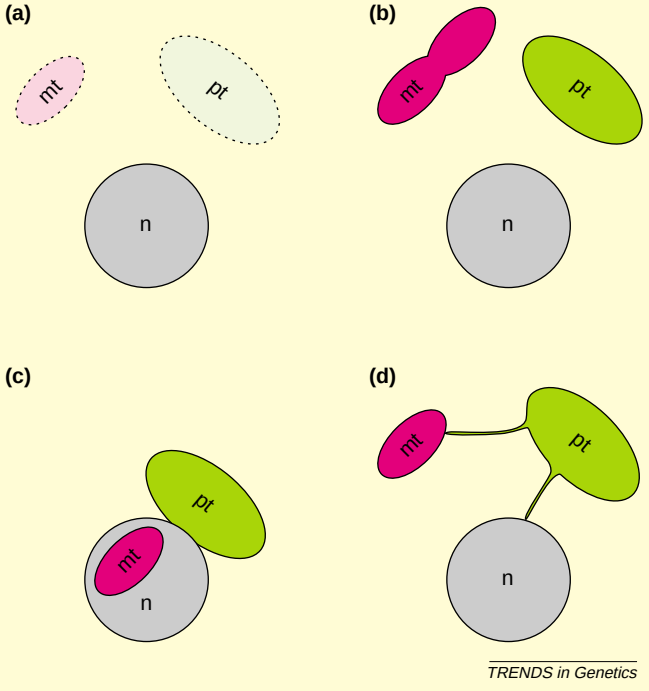

Figure I. Cellular mechanisms that have been proposed to contribute to DNA transfer. **(a)** Lysis of organelles during autophagy, gametogenesis or fertilization. **(b)** Fusion of mitochondria. **(c)** Nuclear inclusion of mitochondria or nuclear attachment of plastids. **(d)** Stromules connecting plastids with mitochondria and/or the nucleus.

a *de novo* insertion of mtDNA had been captured at the breakpoint junction [38]. The preferential insertion of recent human NUMTs into genes [19], as well as the finding that several heritable human diseases are associated with NUMT insertions [38–40], is compatible with the proposal that transcription *per se* also promotes the formation of DSBs [41]. Random end-joining of linear orgDNA fragments, or of T-DNA (transfer DNA) sequences during *Agrobacterium*-mediated T-DNA transformation of plants, can apparently take place before or during insertion into the nuclear genome [18,22,30,31,34,42], and is also likely to involve the NHEJ mechanism.

This leads to the question of how DNA escapes from organelles. Few experimental data are available on mechanisms of organelle-to-nucleus DNA transfer. Several possibilities have been proposed, including the release of DNA during the disruptions of organelle membranes that might occur during autophagy, organelle fusion or division, cell stress, and illegitimate transport of DNA via the nuclear import machinery [27,43,44]. The controlled degradation of organelles during gametogenesis [14,28] and physical interactions between nucleus and organelles have also been considered (Box 2). Because the incidence of complex loci containing DNA from both organelles in nDNA is high [22], it can be concluded that the concomitant release of ptDNA and mtDNA is not a rare event, probably taking place under conditions that affect both organelles, such as cell stress or gametogenesis-associated organellar degradation [2,7]. In the case of nuclear insertions of complete and almost intact copies of mitochondrial or plastid chromosomes, the transfer might be facilitated by physical contact between the donor organelle and the nucleus.

Use and misuse of norDNAs in phylogenetic studies

When orgDNA has been incorporated into the nuclear genome, it is exposed to the evolutionary influences that act on this compartment (Figure 3). In species where the mutation rate in the nucleus is much lower than in mitochondria or plastids, norgDNAs represent molecular fossils. This is the case in humans and other animals, and has been exploited to trace ancestral states of mtDNAs and improve mtDNA-based phylogenies by providing suitable outgroups [9,45]. Moreover, norgDNA insertion polymorphisms, as a subclass of insertion-deletion polymorphisms, are valuable markers for population and evolutionary studies [19,46,47].

Spontaneous mutations of nDNA are poorly characterized, because their low rate precludes direct experimental studies. In plants, where – with few exceptions [48] – nucleotide substitutions occur less frequently in orgDNA than in the nuclear genome [49–51], norgDNAs serve as probes for the types of mutation that occur in nDNA. In two independent studies, Huang *et al.* [46] and Noutsos *et al.* [22] analysed large and recent norgDNA inserts in *Arabidopsis* and rice to study the mutational modification of norgDNA in the nucleus. A predominance of C→T and G→A transition mutations was found, which was associated with the gradual 'acclimatization' of the integrated sequence to the nucleotide composition of the host chromosome [22,46]. The prevalence of C→T (G→A) transitions can be interpreted in the context of the predominantly non-functional nature of norgDNA. This lack of function should favour loss of transcription [52,53] and,

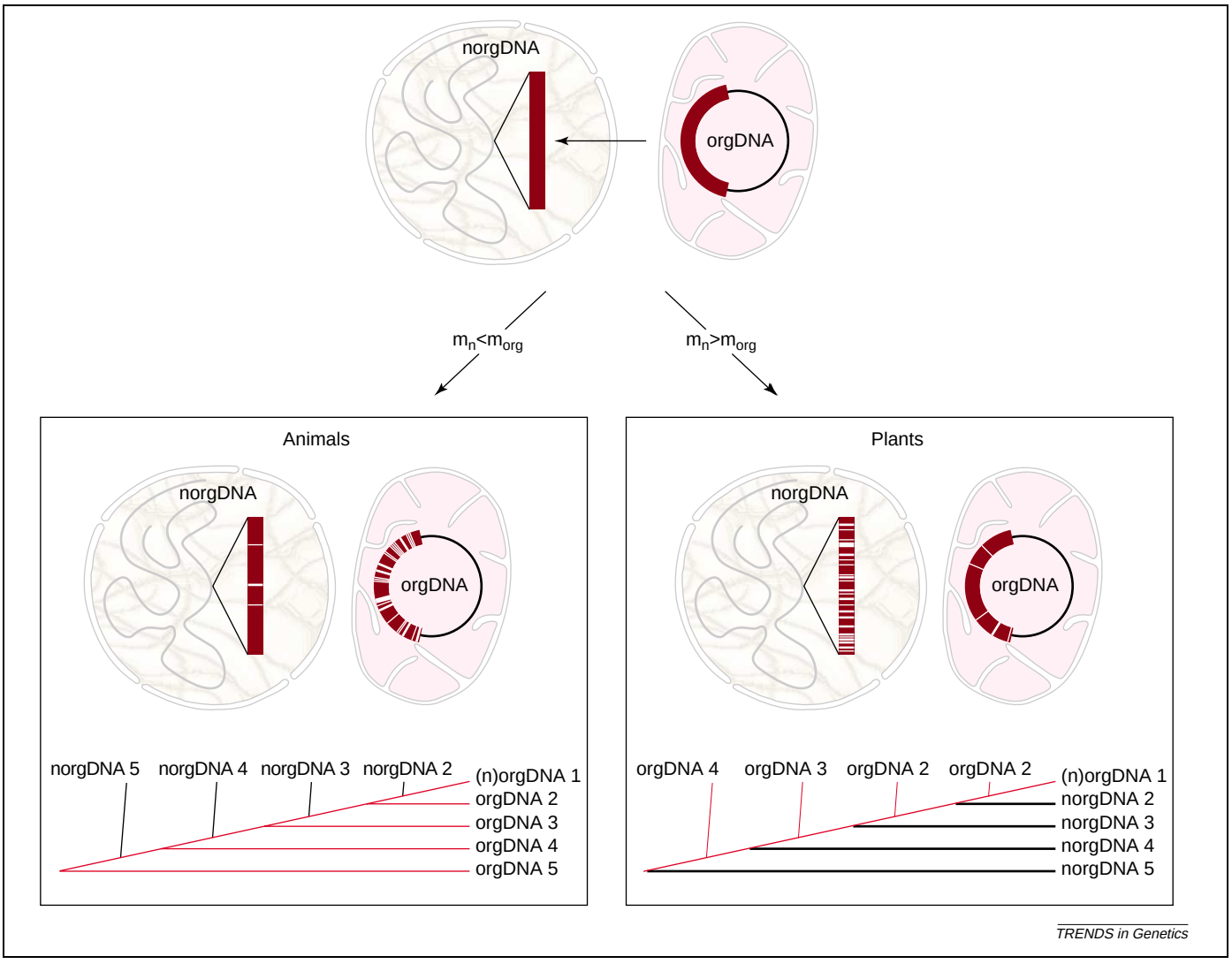

Figure 3. Phylogenies and norgDNA. Upon insertion in the nucleus, norgDNA and its organellar counterpart are identical in sequence (see cladograms: orgDNA1= norgDNA1= (n)orgDNA1), but are exposed to different mutational environments. Depending on the ratio of nuclear and organellar mutation rates, two types of sequence evolution are possible. When the nuclear substitution rate (μn) is lower than the substitution rate of the organelle (μorg), as in animals, norgDNA evolves slower than its organellar equivalent. In this case, extant NUMTs represent molecular fossils, helpful in tracing ancient states of organelle genomes and providing suitable outgroups for mtDNA phylogenies. In several plant species, the opposite case is observed: norgDNA evolves faster than orgDNA, which makes it possible to study the pattern of nuclear mutations in relatively recent NUMTs or NUPTs. In the figure, red segments represent norgDNA and corresponding orgDNA sequences, whereas white lines symbolize mutations. In the cladograms at the bottom, red lines indicate the evolution of orgDNA and black ones the evolution of norgDNA, whereas the length of lines indicates the degree of sequence divergence. Note that although the rate of nucleotide substitutions is relatively low in plant organelles, rearrangements of noncoding sequences occur relatively frequently. In this respect, plant norgDNA sequences can serve as molecular fossils for orgDNA rearrangements.

concomitantly, hypermethylation of cytosine residues. The G–T mismatch created by spontaneous deamination of 5-methylcytosine can be restored to the G–C pair, but repair should create an A–T pair with the same probability [54], resulting in C→T transitions (and G→A on the opposite strand). Deletion of DNA from large norgDNA loci in rice and *Arabidopsis* is more than compensated for by insertions of non-orgDNA, leading to a local expansion at the insertion [22]. Deletions tend to occur between perfect repeats, indicating that they originate by replication slippage [22,46]. Regions flanked by direct repeats are, however, rarely duplicated, indicating that they are more efficiently removed than generated [22].

In contrast to their low point mutation rate, plant mitochondrial genomes frequently undergo rearrangement, expanding and contracting over relatively short evolutionary times [50]. Thus, plant NUMTs should exist that reflect ancient arrangements of mtDNA. Indeed, the 620-kb NUMT on chromosome 2 of *A. thaliana* contains two regions that are homologous to mtDNA from *Brassica napus* but are not found in present-day *Arabidopsis* mtDNA. Hence, they might represent regions that were lost from the mitochondrial chromosome of *Arabidopsis* only after generation of the NUMT [46]. Similarly, a 92-kb rice NUMT contains a region homologous to mtDNA of maize that is no longer present in rice mtDNA [22]. Finally, alternative configurations of mitochondrial and plastid chromosomes exist as a result of recombination across repeat regions. Insertions of orgDNA into the nucleus have led to the fixation of a set of alternative configurations in plant norgDNAs [22,46].

It has been repeatedly stated that norgDNA can be co-amplified accidentally during polymerase chain reaction

(PCR) experiments designed to obtain mtDNA sequences, leading to incorrect phylogenetic reconstructions [9,15,45,55]. Such artifacts have resulted in the mistaken identification of human NUMTs as ancient mtDNA from dinosaurs or monkeys, or in the misinterpretation of NUMTs as indications of pathogenic mtDNA mutations [9]. Recently, Martin [55] and Thalmann *et al.* [56,57] cautioned against the use of PCR-amplified sequences in the construction of mtDNA phylogenies. In humans and great apes, the variety of NUMT sequences that show great similarity to authentic mtDNA renders any standard PCR analysis of mtDNA difficult – and in gorilla perhaps even impossible [56,57]. At worst, the erroneous assumption that NUMT sequences represent genuine mtDNA could even suggest the occurrence of horizontal gene transfer [55]. Therefore, it has been proposed that explicit measures – such as amplification strategies relying upon the circularity of the mtDNA molecule, or the use of extracts enriched for mtDNA as templates for PCR, or specific hybridization and cloning of the noncoding regions flanking the amplified sequences – need to be taken to authenticate mtDNA sequences in newly studied taxa or when any irregularities are noted [9,55–57].

Roles of norgDNA in gene and genome evolution

The rate of the continued colonization of nuclei by orgDNA has been analysed in several species. In yeast, the probability of migration of episomal DNA from mitochondria to the nucleus of any cell is 2×10^{-5} [27], whereas in male gametes of tobacco the chloroplast-to-nucleus DNA transfer frequency is at least 6.4×10^{-5} [28]. In *Homo sapiens*, phylogenetic [17] and cross-species [19] analyses resulted in remarkably similar estimates of the rate of NUMT integration in the germ line, namely $5.1–5.6\times10^{-6}$ per germ cell per generation. As a consequence, in *H. sapiens* and tobacco, on average, any two haploid genomes should differ in at least two loci, caused by the presence or absence of orgDNA. Although it has occasionally been proposed that orgDNA might act as a nuclear mutagen [58–60], until recently the nuclear insertion of orgDNA was considered to be essentially harmless [21,45]. Recent human NUMTs, however, were shown to insert preferentially into genes, most probably resulting in modifications of their exon and intron patterns [19]. This potentially harmful mutagenic effect of orgDNA is supported by several associations of NUMTs with inherited disease in humans. Two cases of familial disease associated with mtDNA insertions are known: a germline insertion of a 41-bp mtDNA sequence into the breakpoint junction of a chromosomal translocation [38], and a case of a severe inherited bleeding disorder where a 251-bp NUMT has introduced a novel splicing site into the gene for coagulation factor VII (*F7*) [40]. A *de novo* 72-bp mtDNA insertion into the *GLI3* gene for a zinc-finger transcription factor created a premature stop codon and was responsible for a sporadic case of Pallister–Hall syndrome [39]. In addition, one of the human genes targeted by a recent intronic NUMT insertion is *MADH2*, a tumour suppressor gene that is often mutated in colorectal carcinomas [19]. Because around 25% of NUPTs and NUMTs are found within exons or introns in flowering plants [7], it can be concluded that nuclear insertions of orgDNA represent a considerable challenge to the functional integrity of their nuclear genomes.

Outlook

The ongoing sequencing of eukaryotic genomes will extend the inventory of known norgDNAs. Whereas the identification of recently inserted and fixed norgDNAs is straightforward, new insertions present only in few individuals, as well as old insertions with short and/or diverged sequences, are more difficult to recognize (Box 3). In general, a common procedure should be applied for identification of norgDNA in different species, allowing recognition of complex and rearranged norgDNA loci as single insertion events, as well as facilitating interspecific comparisons of norgDNA repertoires. The abundance of NUMTs in some species is a serious threat to the establishment of mtDNA-based phylogenies. Therefore, the PCR-based amplification of authentic mtDNA

Box 3. Outstanding questions

(i) Are norgDNAs more than 'only' mutagens?

In some cases, NUMT sequences fused to nDNA sequences are transcribed [18,75], and expressed sequence tags (ESTs) exist that contain both NUPT and non-plastid-derived nDNA [76]. These chimeric genes can be interpreted as the tip of the iceberg, and raise a central question: has the bombardment of the nuclear genome by orgDNA only been neutral or harmful for gene evolution, or can we also recognize beneficial effects? The point here is to understand if norgDNAs can influence nuclear processes such as replication (e.g. as autonomous replicating sequences (ARSs) or transcription (e.g. as promoter sequences) [30], or if they can even rebuild genes and their products by providing new exon modules. If they can, this would constitute, in addition to the ancient transfer of entire prokaryotic genes to the nucleus, a further contribution that organelles make to the evolution of nuclear genomes.

(ii) What is the impact of norgDNA insertions in somatic cells?

Little is known about the frequency and physiological implications of orgDNA insertions in the nuclei of somatic cells. Especially in *H. sapiens*, it would be interesting to study possible implications of NUMT insertions for somatic mutation and oncogenic transformation.

(iii) How can new norgDNA insertions be systematically identified and how frequent are they?

In genome sequencing projects, recent norgDNA insertions present only in a few individuals or in small family groups are difficult to detect. Whereas it is relatively straightforward to screen individuals for the absence of norgDNAs that are present in the reference genome sequence (see, e.g., Ref. [19]), the reverse approach – to identify recent norgDNAs not represented in the reference genome – is challenging. Future studies on norgDNAs could be oriented to the detection of these ongoing insertions, their frequency, and their impact on cell function.

(iv) What is the reason for the marked interspecific diversity in norgDNA accumulation?

No clear explanation exists for the interspecific diversity of NUPTs and NUMTs in copy number and length distribution. For *Chlamydomonas reinhardtii* and *Plasmodium falciparum*, the low number of norgDNAs can be attributed to the presence of only one orgDNA-donor organelle per cell [6,7]. In addition, the efficiency of nuclear import of orgDNA might differ between species. Another possible explanation relates to interspecific differences in the efficiency of integration of orgDNA into the nuclear genome; thus, the repair of DSBs in tobacco is associated with a broad spectrum of insertions of filler sequences whereas in *Arabidopsis* no insertions were observed [37]. The difference in DSB repair between *Arabidopsis* and tobaccoo, in turn, has been associated with a more efficient exonucleolytic degradation of DNA in *Arabidopsis* [34].

sequences used for phylogenetic analyses must be standardized. Although the amplification of overlapping fragments by long-range PCR necessarily generates genuine mtDNA sequences in the case of small mitochondrial chromosomes [56], a similar procedure has yet to be developed for large mitochondrial genomes [55].

Plant species are most informative with respect to organelle-to-nucleus DNA transfer as a result of the presence of two distinct types of organelle and of orgDNA. Whereas previous analyses of norgDNAs considered only species with uniparental (maternal) inheritance of orgDNA, the analysis of angiosperm species with biparental inheritance or with independent inheritance of mitochondria and plastids should enable us to elucidate the role in organelle-to-nucleus DNA transfer of the programmed degeneration of plastids and mitochondria during gametogenesis. Two species of choice for such analyses are *Medicago truncatula* and *Musella lasiocarpa*: whereas *Medicago* inherits plastids (and ptDNA) biparentally, *Musella* shows biparental inheritance of mtDNA [61].

Finally, future studies need to clarify whether nuclear insertion of orgDNA by NHEJ-DSB repair is a common mechanism, whether differences in DSB repair account for the interspecific diversity in norgDNA abundance, and whether – and to what extent – norgDNA contributes beneficially to gene evolution (Box 3).

Acknowledgements

Thanks to Paul Hardy, Anja Schneider, Maarten Koornneef and Franceso Salamini for critical reading of the manuscript. Work in the author's laboratory is supported by the Deutsche Forschungsgemeinschaft (LE 1265/1, /3, /8, /9 and /10; GK 306) and the European Community (HPRN-CT-2002–00248 [PSI-CO]).

References

1 Leister, D. and Schneider, A. (2004) Evolutionary contribution of plastid genes to plant nuclear genomes and its effects on composition of the proteomes of all cellular compartments. In *Organelle, Genomes and Eukaryote Phylogeny: An Evolutionary Synthesis in the Age of Genomics* (The Systematics Association Special Volume Series 68) (Horner, D.S., ed.), pp. 237–255, CRC Press
2 Timmis, J.N. *et al*. (2004) Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nat. Rev. Genet.* 5, 123–135
3 Boore, J.L. (1999) Animal mitochondrial genomes. *Nucleic Acids Res.* 27, 1767–1780
4 Adams, K.L. *et al*. (2001) Multiple losses and transfers to the nucleus of two mitochondrial succinate dehydrogenase genes during angiosperm evolution. *Genetics* 158, 1289–1300
5 Adams, K.L. *et al*. (2000) Repeated, recent and diverse transfers of a mitochondrial gene to the nucleus in flowering plants. *Nature* 408, 354–357
6 Richly, E. and Leister, D. (2004) NUMTs in sequenced eukaryotic genomes. *Mol. Biol. Evol.* 21, 1081–1084
7 Richly, E. and Leister, D. (2004) NUPTs in sequenced eukaryotes and their genomic organization in relation to NUMTs. *Mol. Biol. Evol.* 21, 1972–1980
8 Pereira, S.L. and Baker, A.J. (2004) Low number of mitochondrial pseudogenes in the chicken (*Gallus gallus*) nuclear genome: implications for molecular inference of population history and phylogenetics. *BMC Evol. Biol.* 4, 17
9 Bensasson, D. *et al*. (2001) Mitochondrial pseudogenes: evolution's misplaced witnesses. *Trends Ecol. Evol.* 16, 314–321
10 Lopez, J.V. *et al*. (1994) Numt, a recent transfer and tandem amplification of mitochondrial DNA to the nuclear genome of the domestic cat. *J. Mol. Evol.* 39, 174–190
11 The Rice Chromosome 10 Sequencing Consortium. (2003) In-depth view of structure, activity, and evolution of rice chromosome 10. *Science* 300, 1566–1569
12 Stupar, R.M. *et al*. (2001) Complex mtDNA constitutes an approximate 620-kb insertion on *Arabidopsis thaliana* chromosome 2: implication of potential sequencing errors caused by large-unit repeats. *Proc. Natl. Acad. Sci. U. S. A.* 98, 5099–5103
13 Lin, X. *et al*. (1999) Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*. *Nature* 402, 761–768
14 Woischnik, M. and Moraes, C.T. (2002) Pattern of organization of human mitochondrial pseudogenes in the nuclear genome. *Genome Res.* 12, 885–893
15 Mourier, T. *et al*. (2001) The Human Genome Project reveals a continuous transfer of large mitochondrial fragments to the nucleus. *Mol. Biol. Evol.* 18, 1833–1837
16 Tourmen, Y. *et al*. (2002) Structure and chromosomal distribution of human mitochondrial pseudogenes. *Genomics* 80, 71–77
17 Bensasson, D. *et al*. (2003) Rates of DNA duplication and mitochondrial DNA insertion in the human genome. *J. Mol. Evol.* 57, 343–354
18 Blanchard, J.L. and Schmidt, G.W. (1996) Mitochondrial DNA migration events in yeast and humans: integration by a common end-joining mechanism and alternative perspectives on nucleotide substitution patterns. *Mol. Biol. Evol.* 13, 537–548
19 Ricchetti, M. *et al*. (2004) Continued colonization of the human genome by mitochondrial DNA. *PLoS Biol.* 2, E273
20 Mishmar, D. *et al*. (2004) Mitochondrial DNA-like sequences in the nucleus (NUMTs): insights into our African origins and the mechanism of foreign DNA integration. *Hum. Mutat.* 23, 125–133
21 Hazkani-Covo, E. *et al*. (2003) 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
22 Noutsos, C. *et al*. (2005) Generation and evolutionary fate of insertions of organelle DNA in the nuclear genomes of flowering plants. *Genome Res.* 15, 616–628
23 Matsuo, M. *et al*. (2005) The rice nuclear genome continuously integrates, shuffles, and eliminates the chloroplast genome to cause chloroplast-nuclear DNA flux. *Plant Cell* 17, 665–675
24 Nugent, J.M. and Palmer, J.D. (1991) RNA-mediated transfer of the gene *coxII* from the mitochondrion to the nucleus during flowering plant evolution. *Cell* 66, 473–481
25 Shafer, K.S. *et al*. (1999) Mechanisms of mitochondrial DNA escape to the nucleus in the yeast *Saccharomyces cerevisiae*. *Curr. Genet.* 36, 183–194
26 Huang, C.Y. *et al*. (2004) Simple and complex nuclear loci created by newly transferred chloroplast DNA in tobacco. *Proc. Natl. Acad. Sci. U. S. A.* 101, 9710–9715
27 Thorsness, P.E. and Fox, T.D. (1990) Escape of DNA from mitochondria to the nucleus in *Saccharomyces cerevisiae*. *Nature* 346, 376–379
28 Huang, C.Y. *et al*. (2003) Direct measurement of the transfer rate of chloroplast DNA into the nucleus. *Nature* 422, 72–76
29 Stegemann, S. *et al*. (2003) High-frequency gene transfer from the chloroplast genome to the nucleus. *Proc. Natl. Acad. Sci. U. S. A.* 100, 8828–8833
30 Yu, X. and Gabriel, A. (1999) Patching broken chromosomes with extranuclear cellular DNA. *Mol. Cell* 4, 873–881
31 Ricchetti, M. *et al*. (1999) Mitochondrial DNA repairs double-strand breaks in yeast chromosomes. *Nature* 402, 96–100
32 Shin, D.S. *et al*. (2004) Structure and function of the double-strand break repair machinery. *DNA Repair (Amst.)* 3, 863–873
33 van Gent, D.C. *et al*. (2001) Chromosomal stability and the DNA double-stranded break connection. *Nat. Rev. Genet.* 2, 196–206
34 Puchta, H. (2005) The repair of double-strand breaks in plants: mechanisms and consequences for genome evolution. *J. Exp. Bot.* 56, 1–14
35 Karran, P. (2000) DNA double strand break repair in mammalian cells. *Curr. Opin. Genet. Dev.* 10, 144–150
36 Salomon, S. and Puchta, H. (1998) Capture of genomic and T-DNA sequences during double-strand break repair in somatic plant cells. *EMBO J.* 17, 6086–6095
37 Kirik, A. *et al*. (2000) Species-specific double-strand break repair and genome evolution in plants. *EMBO J.* 19, 5562–5566

38 Willett-Brozick, J.E. *et al*. (2001) Germ line insertion of mtDNA at the breakpoint junction of a reciprocal constitutional translocation. *Hum. Genet.* 109, 216–223
39 Turner, C. *et al*. (2003) Human genetic disease caused by *de novo* mitochondrial-nuclear DNA transfer. *Hum. Genet.* 112, 303–309
40 Borensztajn, K. *et al*. (2002) Characterization of two novel splice site mutations in human factor VII gene causing severe plasma factor VII deficiency and bleeding diathesis. *Br. J. Haematol.* 117, 168–171
41 Gonzalez-Barrera, S. *et al*. (2002) Transcription and double-strand breaks induce similar mitotic recombination events in *Saccharomyces cerevisiae*. *Genetics* 162, 603–614
42 Blanchard, J.L. and Schmidt, G.W. (1995) Pervasive migration of organellar DNA to the nucleus in plants. *J. Mol. Evol.* 41, 397–406
43 Campbell, C.L. and Thorsness, P.E. (1998) 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
44 Thorsness, P.E. and Weber, E.R. (1996) Escape and migration of nucleic acids between chloroplasts, mitochondria, and the nucleus. *Int. Rev. Cytol.* 165, 207–234
45 Perna, N.T. and Kocher, T.D. (1996) Mitochondrial DNA: molecular fossils in the nucleus. *Curr. Biol.* 6, 128–129
46 Huang, C.Y. *et al*. (2005) Mutational decay and age of chloroplast and mitochondrial genomes transferred recently to angiosperm nuclear chromosomes. *Plant Physiol.* 138, 1723–1733
47 Thomas, R. *et al*. (1996) Novel mitochondrial DNA insertion polymorphism and its usefulness for human population studies. *Hum. Biol.* 68, 847–854
48 Cho, Y. *et al*. (2004) From the Cover: Mitochondrial substitution rates are extraordinarily elevated and variable in a genus of flowering plants. *Proc. Natl. Acad. Sci. U. S. A.* 101, 17741–17746
49 Wolfe, K.H. *et al*. (1987) Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proc. Natl. Acad. Sci. U. S. A.* 84, 9054–9058
50 Palmer, J.D. and Herbon, L.A. (1988) Plant mitochondrial DNA evolves rapidly in structure, but slowly in sequence. *J. Mol. Evol.* 28, 87–97
51 Gaut, B.S. (1998) Molecular clocks and nucleotide substitution rates in higher plants. *Evol. Biol.* 30, 93–120
52 Paszkowski, J. and Whitham, S.A. (2001) Gene silencing and DNA methylation processes. *Curr. Opin. Plant Biol.* 4, 123–129
53 Bender, J. (2004) Chromatin-based silencing mechanisms. *Curr. Opin. Plant Biol.* 7, 521–526
54 Holliday, R. and Grigg, G.W. (1993) DNA methylation and mutation. *Mutat. Res.* 285, 61–67
55 Martin, W. (2005) Molecular evolution: Lateral gene transfer and other possibilities. *Heredity* 94, 565–566
56 Thalmann, O. *et al*. (2004) Unreliable mtDNA data due to nuclear insertions: a cautionary tale from analysis of humans and other great apes. *Mol. Ecol.* 13, 321–335
57 Thalmann, O. *et al*. (2005) Nuclear insertions help and hinder inference of the evolutionary history of gorilla mtDNA. *Mol. Ecol.* 14, 179–188
58 Hadler, H.I. (1989) Comment: mitochondrial genes and cancer. *FEBS Lett.* 256, 230–232
59 Richter, C. (1988) Do mitochondrial DNA fragments promote cancer and aging? *FEBS Lett.* 241, 1–5
60 Shay, J.W. and Werbin, H. (1992) New evidence for the insertion of mitochondrial DNA into the human genome: significance for cancer and aging. *Mutat. Res.* 275, 227–235
61 Zhang, Q. *et al*. (2003) Examination of the cytoplasmic DNA in male reproductive cells to determine the potential for cytoplasmic inheritance in 295 angiosperm species. *Plant Cell Physiol.* 44, 941–951
62 Esser, C. *et al*. (2004) A genome phylogeny for mitochondria among α-proteobacteria and a predominantly eubacterial ancestry of yeast nuclear genes. *Mol. Biol. Evol.* 21, 1643–1660
63 Martin, W. *et al*. (2002) Evolutionary analysis of *Arabidopsis*, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. *Proc. Natl. Acad. Sci. U. S. A.* 99, 12246–12251
64 Unseld, M. *et al*. (1997) The mitochondrial genome of *Arabidopsis thaliana* contains 57 genes in 366,924 nucleotides. *Nat. Genet.* 15, 57–61
65 Notsu, Y. *et al*. (2002) The complete sequence of the rice (*Oryza sativa* L.) mitochondrial genome: frequent DNA sequence acquisition and loss during the evolution of flowering plants. *Mol. Genet. Genomics* 268, 434–445
66 Clifton, S.W. *et al*. (2004) Sequence and comparative analysis of the maize NB mitochondrial genome. *Plant Physiol.* 136, 3486–3503
67 Sodmergen *et al*. (2002) Reduction in amounts of mitochondrial DNA in the sperm cells as a mechanism for maternal inheritance in *Hordeum vulgare*. *Planta* 216, 235–244
68 Arimura, S. *et al*. (2004) Frequent fusion and fission of plant mitochondria with unequal nucleoid distribution. *Proc. Natl. Acad. Sci. U. S. A.* 101, 7805–7808
69 Westermann, B. (2003) Mitochondrial membrane fusion. *Biochim. Biophys. Acta* 1641, 195–202
70 Mota, M. (1963) Electron microscope study of relationship between nucleus and mitochondria in *Chlorophytum capense* (L.) Kuntze. *Cytologia (Tokyo)* 28, 409–416
71 Ehara, T. *et al*. (1990) Interactions between the nucleus and cytoplasmic organelles during the cell cycle of *Euglena gracilis* in synchronized cultures. IV. An aggregate form of chloroplasts in association with the nucleus appearing prior to chloroplast division. *Exp. Cell Res.* 190, 104–112
72 Yu, H.S. and Russell, S.D. (1994) Occurrence of mitochondria in the nuclei of tobacco sperm cells. *Plant Cell* 6, 1477–1484
73 Jensen, H. *et al*. (1976) Ultrastructure of mitochondria-containing nuclei in human myocardial cells. *Virchows Arch. B Cell Pathol.* 21, 1–12
74 Kwok, E.Y. and Hanson, M.R. (2004) Stromules and the dynamic nature of plastid morphology. *J. Microsc.* 214, 124–137
75 Maiti, A.K. *et al*. (2001) Isolation, *in silico* characterization and chromosomal localization of a group of cDNAs from ciliated epithelial cells after *in vitro* ciliogenesis. *Genome Biol.* 2, DOI:10.1186/gb-2001-2-7-research (http://genomebiology.com/2001/2/7/research/0026)
76 Shahmuradov, I.A. *et al*. (2003) Abundance of plastid DNA insertions in nuclear genomes of rice and *Arabidopsis*. *Plant Mol. Biol.* 52, 923–934