Review

The plant mitochondrial genome: Dynamics and maintenance

José M. Gualberto, Daria Mileshina, Clémentine Wallet, Adnan Khan Niazi, Frédérique Weber-Lotfi, André Dietrich

Institut de Biologie Moléculaire des Plantes, CNRS and Université de Strasbourg, 12 rue du Général Zimmer, 67084 Strasbourg, France

A R T I C L E   I N F O

Article history:
Received 22 July 2013
Accepted 17 September 2013
Available online 26 September 2013

Keywords:
Mitochondria
mtDNA
Replication
Recombination
Repair

A B S T R A C T

Plant mitochondria have a complex and peculiar genetic system. They have the largest genomes, as compared to organelles from other eukaryotic organisms. These can expand tremendously in some species, reaching the megabase range. Nevertheless, whichever the size, the gene content remains modest and restricted to a few polypeptides required for the biogenesis of the oxidative phosphorylation chain complexes, ribosomal proteins, transfer RNAs and ribosomal RNAs. The presence of autonomous plasmids of essentially unknown function further enhances the level of complexity. The physical organization of the plant mitochondrial DNA includes a set of sub-genomic forms resulting from homologous recombination between repeats, with a mixture of linear, circular and branched structures. This material is compacted into membrane-bound nucleoids, which are the inheritance units but also the centers of mitochondrial genome dynamics, generation of mitotypes and is involved in the evolution of the mitochondrial DNA. In line with, or as a consequence of its complex physical organization, replication of the plant mitochondrial DNA is likely to occur through multiple mechanisms, potentially involving recombination processes. We give here a synthetic view of these aspects.

© 2013 Elsevier Masson SAS. All rights reserved.

1. Introduction

Mitochondria are key players in plant development, fitness and reproduction. Their contribution to energy production, metabolism and cell homeostasis relies on the performance of their own genetic system that makes them semi-autonomous. As in other organisms, the mitochondrial genome in plants encodes a series of essential polypeptides that build up the complexes of the oxidative phosphorylation chain, together with nuclear-encoded subunits. But plant mitochondrial DNAs (mtDNAs) have remarkable features that distinguish them from their animal and fungal counterparts. In particular, higher plants harbor large mtDNAs that are highly variable in size and structural organization. On top of that, in many plant species, mitochondria contain various forms of plasmids that replicate independently from the main chromosome. In most plant species, the mtDNA gene sequences evolve very slowly, as compared to animal mtDNA sequences, and point mutations are rare. It is believed that this is because plant mitochondria contain an active DNA recombination system that allows copy correction of mutations. Indeed, numerous studies have shown that plant mitochondrial genomes undergo extensive and high frequency homologous recombination (HR). Such processes make plant mtDNA prone to rearrangements. When not lethal, mtDNA mutations/rearrangements can generate severe phenotypes or cause cytoplasmic male sterility (CMS). Hence the requirement for efficient DNA repair and maintenance pathways in a context where oxidative pressure, replication defaults or environmental hazard generate base modifications and strand breaks. Finally, mitochondrial genome dynamics generates heteroplasmy, with alternative mtDNA configurations coexisting with the main mtDNA. Segregation of alternative mitotypes significantly contributes to the rapid evolution of the plant mtDNA structure. The present review addresses these issues, with special emphasis on the specificities of the plant mitochondrial genetic system.

2. Plant mitochondrial DNA structure and organization

2.1. Genome size and content

The structure of higher plant mitochondrial genomes has a number of unique features. Whereas most animals possess circular mtDNAs of 15–17 kb in size, the mitochondrial genomes of plants...
Recombination of double-strand breaks can invade homologous double-stranded DNA, forming a D-loop and leading to the establishment of replication forks.

Complex array of linear molecules. Branched forms of the mtDNA could be intermediates of recombination-dependent processes: 3 single-stranded sequences resulting from the recombination of double-strand breaks can invade homologous double-stranded DNA, forming a D-loop and leading to the establishment of replication forks.

Fig. 1. Circular and linear models of the mtDNA structure in plants. The dynamics of the mtDNA of plants allow several models. a) A circular genome model representing the different organisations of the mtDNA of A. thaliana ecotype C24. The two pairs of large, frequently recombining repeats (A and B) and their orientations are represented by blue and red arrows. The different parts of the genome are colored in a range of gray scale. Intramolecular recombination events are represented with dashed lines, and in orange, recombination leading to subgenomic molecules. Dotted line shows possible intermolecular recombination. b) The mtDNA has been observed as predominantly constituted by a complex array of linear molecules. Branched forms of the mtDNA could be intermediates of recombination-dependent processes: 3 single-stranded sequences resulting from the recombination of double-strand breaks can invade homologous double-stranded DNA, forming a D-loop and leading to the establishment of replication forks.

are much larger and differ greatly in size, even between very close species or within species [1,2]. They commonly range between 200 and 750 kb in angiosperms [1], but with a tremendous further extension in some lineages. As an example, proliferation of dispersed repeats, expansion of existing introns and acquisition of sequences (including sequences of nuclear, plastid, viral and bacterial origin) account for the expansion of the cucumber (Cucumis sativus) mitochondrial genome into a tri-chromosome structure with components of 1556, 84, and 45 kb in size [3]. Sequencing of the mtDNA from two Silene species with exceptionally high mutation rate revealed enormous mitochondrial genomes of 6.7 (Silene noctiflora) and 11.3 (Silene conica) megabase-pairs resulting from massive proliferation of non-coding content [4]. Moreover, these genomes are distributed into a large number of circular chromosomes: 59 for S. noctiflora and over 128 for S. conica. These chromosomes range themselves in size from 44 to 192 kb. As a comparison, the more typical, slowly evolving mitochondrial genome of Silene latifolia accounts for a single chromosome of 253 kb [4]. Remarkably, plant mitochondrial genomes are large but their ploidy appears to be low. Whereas thousands of mtDNA copies can be present in a mammalian cell, much lower mtDNA levels were detected in plants. Preuten et al. [5] analyzed the ploidy of several individual mitochondrial genes in various samples of Arabidopsis thaliana, Nicotiana tabacum and Hordeum vulgare, using real time quantitative PCR. The copy numbers of the investigated genes both differed from each other and varied greatly between organs or during development, with the highest estimates around 280 copies per cell for the atp1 gene in mature leaves. This was actually less than the mean number of mitochondria per cell, which was around 450 in protoplasts derived from mature leaves. Thus, individual mitochondria in plants may contain only part of the genome or potentially no DNA [5].

Despite their large sizes, land plant mtDNAs do not contain a significantly greater number of genes than mitochondrial genomes of other lineages. The known genes usually range between 50 and 60, while multiple cis- or trans-spliced introns and large intergenic regions complete the genomes. Protein genes in plant mtDNAs encode subunits of the oxidative phosphorylation chain complexes but also proteins involved in the biogenesis of these complexes, as well as several ribosomal proteins. First to be entirely sequenced, the 367 kb A. thaliana mtDNA (ecotype C24) codes for 32 proteins, 3 ribosomal RNAs (5S, 18S and 26S rRNAs) and 22 tRNAs [6], while the 16.5 kb human mtDNA codes for 13 proteins, 2 rRNAs (12S and 16S) and 22 tRNAs [7]. With a total of 57 genes versus 37, this makes a 1.5 fold difference in gene content for a 22 fold difference in size. Identified genes represent a minor part of the plant mtDNA content (10% in A. thaliana), whereas the major part (60% in A. thaliana) has no recognizable origin and function [6]. Introns and duplications expand (8% and 7% in A. thaliana, respectively), as well as integrated nuclear and plastid sequences (5% as a whole in A. thaliana). Further putative open reading frames of significant length can be detected, which also add up to 10% of the genome in A. thaliana [6].

Differences in gene content among plant species are also not linked to genome size. The large cucumber mtDNA carries just 4 more protein genes than the A. thaliana genome within 1.3 megabase of extra sequence [3]. On the other hand, it can be noted that the cucumber mitochondrial genome has integrated 7 more tRNA genes of plastid origin, whereas 2 tRNA genes of authentic mitochondrial origin present in other species are missing. Similarly, the 253 kb mtDNA of S. latifolia and the 11.3 megabase mtDNA of S. conica carry the same reduced set of 25 protein genes and, while the S. latifolia mtDNA contains 9 tRNA genes, that of S. conica has only 2 [4]. The intergenic sequences with no recognizable origin or function thus represent 91.5% (10.3 megabase) of the S. conica mitochondrial genome. With the emergence of second-generation sequencing technologies, the number of completed plant mitochondrial genome sequences rapidly increased in the last few years,

2.2. Physical organization

The physical structure of plant mitochondrial genomes is usually represented by a circle of double-stranded DNA, called “master chromosome”, housing the complete set of mitochondrial genes. However, such a structure is derived from mapping and sequence assembly and has not been directly observed. Thus, although it might exist, the master circle must be rare. On the other hand, the 620 kb mitochondrial insertion characterized in A. thaliana nuclear chromosome 2 [8] seems to derive from a concatemer of mitochondrial genomes, suggesting that such concatamers may exist in vivo. Plant mtDNAs usually contain repeated sequences dispersed throughout the genome, which can be in direct or inverted orientation. They can pair up and recombine, redistributing sequences and generating sub-genomic circular DNA molecules (Fig. 1a).

Indeed, numerous studies have shown that plant mitochondrial genomes undergo extensive and high frequency homologous recombination. Especially, recombination between a few large repeats (several kb in size) is frequent and reciprocal, leading to multiple, interconverting configurations of the genome. These recombinationally active long repeats are not homologous between species, but the process is conserved. In A. thaliana, two large repeats of 6.5 and 4.2 kb were shown to be involved in recombination [9]. In the N. tabacum mtDNA, there are three pairs of large repeated sequences, with lengths of 18 kb, 6.9 kb and 4.7 kb [10]. But the number of recombinogenic repeats can be much larger, as for example in wheat where there are 10 pairs of large repeats [11]. As a consequence, the mtDNA of plants cannot be physically represented as a unique, organized structure that could be used to determine phylogenetic relationships between species. Furthermore, besides sub-genomic circles, early observations highlighted linear forms [12]. In the case of the maize (Zea mays) CMS-S cytoplasm, it was proposed that the mitochondrial genome would exist mainly as multiple linear molecules. These would result from recombination of the mitochondrial chromosome with a linear plasmid (see Section 4 below) specific to CMS-S. The sequencing data seemed to be consistent with this premise [2]. Also in early studies, the morphology of carefully isolated mtDNA was consistent with a predominance of linear molecules [13]. Altogether, it appears that the entity of a plant mitochondrial genome is an array of inter-convertible linear and circular DNA molecules that most likely reflect recombination intermediates between repeated sequences, yielding a mixture of various subgenomic DNA molecules with different structures and contents (Fig. 1). The concept of a circular master chromosome remains because the question of how multipartite and branched DNA molecules are properly transmitted to the next generation is unsolved. Finally, a significant portion of the mtDNA (2–8%) was reported to be in a single-stranded form in Chenopodium album suspension culture or whole plants [14].

Homologous recombination between the numerous intermediate size repeats (ISRs, usually ranging from 50 to 600 bp in size) also present in the genome is generally infrequent but, when occurring, leads to further complex rearrangements. It has been proposed that most of the recombination processes involving ISRs result from Break-Induced Replication (BIR) pathways [15]. But other pathways, such as Single-Strand Annealing (SSA), might also be involved [16]. These low frequency recombination activities are non-allélic, and the process is asymmetrical, resulting in accumulation of only one of the expected reciprocal recombination products and leading to duplication or deletion of genomic sequences. In addition, illegitimate recombination processes involving very short sequence homologies of a few nucleotides (microhomologies) have also been described [17], which can lead to gene chimeras. Both low frequency recombination between ISRs and error-prone repair processes mediated by microhomologies (see below) yield low-copy number alternative configurations (or “mitotypes”) of the mitochondrial genome that contribute to the natural heteroplasmy of the plant mtDNA. Finally, additional processes might be at work, depending on the species. In particular, the large size of the mitochondrial genomes in the genera Cucumis and Silene have been linked with the accumulation of short (30–53 bp) repeated sequences [3,4,18].

Contrary to a prevailing idea, the mtDNA is not naked but is packed into nucleoprotein particles called nucleoids [19]. In these membrane-anchored particles, the DNA is associated with a number of factors involved in its compaction and metabolism. Nucleoids are considered as the heritable units of mtDNA and their structure influences mtDNA segregation and transcriptional regulation [20]. The actual set of nucleoid-associated proteins is still discussed and seems to vary between the different organisms, suggesting that the composition of mitochondrial nucleoproteins has been reinvented several times along with mtDNA evolution [19]. Mitochondrial nucleoids contain a set of core proteins involved in DNA maintenance and expression, as well as peripheral factors that are components of signaling pathways [21]. Plant mitochondrial nucleoids were characterized as dynamic entities, with chromatin-like or fibril-like structures, depending on the developmental stage or tissue [22,23]. Chromatin-like structures were associated with a vesicle component.

Whereas a nucleoid-enriched proteome has been established for chloroplasts [24], information on the composition of their mitochondrial counterparts remains scarce. Plant nucleoids appeared to share a qualitatively similar phospholipid composition with the whole organelle and to contain polypeptides of both inner and outer membrane origin [22]. The association of the identified proteins with isolated nucleoids was a function of the detergent concentration in the isolation procedure, so that the structural or functional significance has to be further evaluated. Little is known about the proteins responsible for nucleoid organization in plants [25]. The presence of nucleosome-like structures was suggested, depending on the developmental stage [22,26]. It was reported that histone H3 and histone deacetylase might be targeted to mitochondria in cauliflower [27,28]. However, such a view was contradicted by studies in tobacco cells [29]. Actin was persistently recovered in plant nucleoids, where it seemed to associate with the mtDNA [22]. Import of actin to the inside of mitochondria was indeed reported and its potential function in the organelles was discussed [30]. It would bind to the mtDNA in complexes also containing porin and the ADP/ATP carrier. Plant mitochondrial nucleoids likely have a heterogeneous genomic organization. Open circles, supercoils, complex forms, and linear molecules with interspersed sigma-shaped structures and/or loops represented about 70% of the mtDNA molecules present in mung bean nucleoids [23]. As in fungi or mammals, mitochondrial nucleoids in plants are not only compaction and segregation units but they support the entire DNA metabolism. Isolated mung bean nucleoids were reported to run DNA synthesis and were capable of directing regulated RNA synthesis [22]. Membrane-association of the plant mitochondrial base excision repair (BER) components (see Section 6.2) also supports the assumption that mtDNA maintenance and repair pathways take place in the membrane-bound nucleoids [31,32].

3. Mitochondrial DNA replication

Replication of plant mitochondrial genomes is far from being understood. Several factors involved in organellar genome
replication have been identified by sequence homology and from proteomic data. Among those, plant organellar DNA polymerases were identified by their similarities to the known mitochondrial DNA polymerase of animals and yeast. In *A. thaliana*, there are two organellar DNA polymerases, Pol1A and Pol1B, which are dual-targeted to both mitochondria and plastids [33–35] and which are apparently redundant in their functions [36,37]. Individually, each of them is dispensable, because the mutants show no visible phenotypes, apart from a small reduction in mtDNA and chloroplast DNA (cpDNA) copy numbers. Conversely, the double mutant is not viable, showing that the two polymerases are redundant for organellar genome replication [37]. However, the situation might not be the same in other species. While single mutants of each of the *A. thaliana* organellar DNA polymerases have a very mild effect on mtDNA and cpDNA copy numbers, in *Z. mays*, deficiency in a plastid-localized DNA polymerase results in albino plants because of a drastic reduction in cpDNA copy number [38]. This suggests that in maize the organellar DNA polymerases are not redundant and might be specific for each type of organelle.

Additional mtDNA replication factors have been identified, which are, in most cases, also dual-targeted to chloroplasts and mitochondria. A helicase similar to the Twinkle helicase of animal mitochondria is present in all plant genomes sequenced, and is targeted to both types of organelles in *A. thaliana*. Like its T7 phage homolog, it combines helicase and primase activities in a single protein [35,39]. Genes coding for organellar Type I and Type II topoisomerases have also been identified [24,35,40] and there are homologs of bacterial single-stranded DNA-binding (SSB) proteins in plant mitochondria [41].

Regarding the mechanisms of replication, they are still open to debate, mainly because the large size and multipartite complexity of the plant mtDNA render this process difficult to analyze. In a search for replication origins as they had been described in metazoan mitochondrial genomes, two putative fragments (oriA and oriB) were identified from the genome of *Petunia hybrida*. These fragments were able to initiate DNA synthesis in an *in vitro* DNA synthesizing system derived from purified *P. hybrida* mitochondria [42]. DNA synthesis was performed with both the A and B origins in the same plasmid, in complementary strands, initiated first in the A-origin and proceeded in the direction of the B-origin, after which replication was also initiated in the B-origin. Based on these observations, the authors proposed a model of mtDNA replication in plants structurally comparable to the one of mammalian mtDNA heavy and light strands [43]. Similarly, a search for sequences of *Papaver somniferum* (Opium poppy) mtDNA able to promote autonomous plasmid replication in yeast revealed a 917 bp fragment with high A/T content [44]. However, these experiments strongly relied on the premises that most of the plant mtDNA exists as genome-size circles replicating according to a theta model of circular DNA replication. But, as commented on in Section 2.2, while data arising from restriction mapping can imply circular mtDNA genomes, direct observation of plant mtDNA failed to support that assumption. Early experiments aiming to characterize the topology of the mtDNA of plants mainly revealed linear and small circular molecules, and the inability to purify full-size circular genomes was attributed to artifactual shearing of large molecules (reviewed in Ref. [12]). Structural analysis of plant mtDNA using pulsed-field gel electrophoresis and moving pictures of DNA migration in agarose revealed that most of the genome is contained in large linear molecules and complex branched molecular structures larger in size than the predicted genome size [13,45]. Analysis of the mtDNA of the liverwort *Marchantia polymorpha* by digestion with restriction enzymes with one recognition site per genome and pulsed-field gel electrophoresis showed that the genome is constituted of linear and head-to-tail concatemers [46]. These concatemeric molecules could be generated both by rolling circle replication and by recombination-mediated replication, as in phase T4. The latter model could also account for the complex branched mtDNA molecules (Fig. 1b).

Rolling-circle and recombination-dependent replication have also been inferred from electron microscopy observation of the *C. album* mitochondrial genome and plasmids (see Section 4 below). In these studies, the mtDNA preparations contained mostly linear molecules of variable size, but also subgenomic circular DNA molecules, circular molecules with tails (sigma-like structures), as well as more complex rosette-like and catenate-like molecules [47]. The distribution of single-stranded mtDNA in the sigma structures and the detection of entirely single-stranded molecules indicated a rolling-circle type of replication of subgenomic circles [48], while analysis of replicating mtDNA in cell cultures also revealed characteristic recombination intermediates compatible with a phase T4-like mechanism of recombination-mediated replication [49]. Dramatic changes in the absolute and relative amounts of the different replicative structures occurred during cell growth, unveiling the plasticity of the mtDNA structure. Taken together, the pulsed-field gel electrophoresis, movies of migrating DNA and electron microscope observations suggest the coexistence of a recombination-dependent mode of replication initiation with a rolling-circle mode of replication. In the yeast *Candida albicans*, a comprehensive topological analysis of the mtDNA also supported a model of recombination-driven replication initiation [50], and inverted repeats in the genome seem to be predominantly involved in replication initiation via homologous recombination. As in *C. albicans*, replication of plant mtDNA might initiate by recombination involving the large repeats that are usually present in plant genomes.

However, the present models of recombination-dependent replication do not satisfactorily explain how the complex populations of mtDNA molecules are transmitted in a near homoplasmic fashion from cell to cell and from one generation to another. At present, it cannot be excluded that replication involves different processes and factors according to the plant tissue and the developmental stage. These questions have to be addressed in order to understand how the mtDNA is faithfully and completely transmitted, and how its copy number is regulated, probably in coordination with the replication of the plastid and nuclear genomes.

### 4. Mitochondrial plasmids

In addition to a large and dynamic main genome, mitochondria of many higher plant species contain a variety of smaller DNA molecules whose size can range from 0.7 to over 20 kb [51–54]. These can be regarded as extrachromosomal replcions or plasmids that can be autonomously replicated in mitochondria because they are usually present at a high stoichiometry relative to the main genome [54]. Multimeric forms of these mitochondrial plasmids have been observed that likely result from homologous recombination processes [55]. The pattern of mitochondrial plasmids is species-specific and can be different among varieties of the same plant species [56]. In Silene species, plasmids further contribute to the complexity of mitochondrial genetics [57]. Although remnants of putative integration into the mtDNA were sometimes reported [58], most plant mitochondrial plasmid DNAs have essentially no sequence similarity with the main chromosome and cannot be considered as subgenomic forms. In that respect, they resemble mitochondrial plasmids found in several fungus species, such as *Neurospora crassa* or *Neurospora intermedia*. Their presence or absence has no correlation with species peculiarities but their amplification changes along the development. Additional plasmid
variants can arise from recombination events [59]. The nuclear context participates in the control of mitochondrial plasmid appearance and copy number [59].

Mitochondrial plasmids in plants were studied more extensively in the 1980’s, when they were expected to provide vectors for mitochondrial transformation. The field subsequently became less active. They can in turn be classified into two types, circular and linear plasmids. Circular plasmids have been found in a diverse group of plant species. Their size is often in the range of 1-2.5 kb, although it can extend to 9 kb or more [51-53]. They can share short regions of homology or contain sequences homologous to the nuclear genome (reviewed in Ref. [54]). In early studies, replication of the mttp2 (1.7 kb) and mttp3 (1.5 kb) circular mitochondrial plasmids of Vicia faba was reported to start at a specific origin, close to sequences that can fold into hairpin structures, and to proceed in the same, single direction around the molecules [60]. An asymmetric rolling circle replication mechanism with unique features was documented for the C. album mp1 circular DNA plasmid (1.3 kb) [61]. This plasmid would contain a putative replication origin sharing homology with double-stranded rolling circle origin (dsO) or transfer origin (oriT) nicking motifs of bacterial plasmids [62]. Typical rolling circle intermediates and concatemers were characterized [63]. Sigma-like molecules with long double-stranded tails were identified and their structure suggested the occurrence of strand switching during rolling circle replication, as for bacterial replicons [62]. A T-rich regions containing tandem sequences that resemble the yeast ARS (autonomously replicating sequence)-type replication origin were highlighted in the Z. mays 1.4 and 1.9 kb circular plasmids [64,65]. Within their limited size, plant circular mitochondrial plasmids mostly carry open reading frames of modest length, although some reach 1 kb. Nevertheless, these orfs are transcribed, as established in a number of species [64-69]. The function of such transcripts is not understood. Notably, differences were reported in the set of circular plasmids and transcripts derived thereof between male fertile and CMS lines of Beta vulgaris [70].

Linear mitochondrial plasmids have been characterized in Beta, Brassica, Daucus, Sorghum and Zea species, but they seem to be less frequent than circular forms [54]. Linear plasmids are generally larger and commonly range from 2 to 12 kb [54]. They usually contain terminal inverted repeats (TIRs) and their 5’-termini were reported to carry specific covalently-bound proteins (reviewed in Ref. [54]). Such “invertron” features are shared with different classes of fungal mitochondrial plasmids, virus or phage DNAs and transposable elements [71]. The terminal proteins are presumably involved in the initiation of DNA replication at both 5’ termini. Five linear plant mitochondrial plasmids have been studied more closely: the Z. mays S1 [72], S2 [73] and 2.3 kb [74] plasmids, the B. vulgaris 10.4 kb plasmid (GenBank accession Y10854), and the Brassica napus 11.6 kb plasmid [75]. The terminal inverted repeats at their ends range from 170 bp (Z. mays 2.3 kb plasmid) to 407 bp (B. vulgaris 10.4 kb plasmid). TIR sequences are unrelated, except for the S1 and S2 plasmids, which share identical terminal repeats. Apart from the Z. mays 2.3 kb plasmid, linear mitochondrial plasmids contain several open reading frames (up to 6 in the 11.6 kb B. napus plasmid) that are larger than those present in circular plasmids. Several of these orfs are considered to encode DNA or RNA polymerases, while the others have no recognizable function. Like the nuclear-encoded RNA polymerases that transcribe the main mitochondrial genome, linear plasmid-encoded RNA polymerases are of phage-type. But the two classes of enzymes are different. Plasmid-encoded DNA polymerases have similarities with group B DNA polymerases characteristic of phages and viruses. The presence of DNA and RNA polymerase genes in linear mitochondrial plasmids supports the idea that these plasmids are autonomously replicated and transcribed. It has been proposed that gene conversion involving the TIRs (considered as palindromic telomerases) can shape the architecture of mitochondrial linear genetic elements, including plasmids [76]. Transcription and translation of putative genes carried by linear plasmids have been detected, especially of those coding for RNA or DNA polymerases likely to be required for expression and replication [77,78].

Expression of orfs with unknown function was established in particular for the 11.6 kb linear plasmid of B. napus, demonstrating that these are functional genes [75]. Similarly, the protein encoded by orf1 of the Z. mays 2.3 kb plasmid could be detected [79]. This polypeptide is similar to a domain of the putative RNA polymerase encoded by the Z. mays S2 linear plasmid, but it was hypothesized that it is rather the protein that covalently binds to the 5’ ends of the 2.3 kb plasmid [79]. Interestingly, transcription appeared to start inside or just at the end of the TIR sequence in both the 11.6 kb B. napus plasmid and the 2.3 kb Z. mays plasmid [75,79]. The latter carries the only functional tRNA[^1] gene of Z. mays mitochondria and is therefore the only plant mitochondrial episome with a clear function, explaining its ubiquitous presence in Z. mays lines [74]. The presence of the S1 and S2 plasmids of Z. mays and their exchanges with the main mtDNA were reported to be involved in CMS or non-chromosomal stripe (NCS) phenotypes, although a strict correlation remained difficult to establish (discussed in Ref. [54]). Conversely, mtDNA rearrangements involving short repeats shared with the S2 plasmid were reported to be responsible for both CMS-S reversion to fertility and NCS mutation [80,81]. However, CMS-S reversion is not necessarily accompanied by a loss of the S plasmids [82]. Similarly, association of the 11.6 kb linear plasmid with CMS in Brassica was found not to be exclusive [83]. In Lolium perenne, a linear plasmid-like element carrying the characteristic DNA and RNA polymerase orfs was found to be integrated into the main mtDNA in a CMS line, while in fertile lines it is present as a low-copy number extrachromosomal replicon [84].

Mitochondrial plasmids are transmitted uniparentally from the maternal plant to the progeny, together with the main genome. However, deviations from a strict maternal inheritance were observed for the Beta maritima 10.4 kb linear plasmid [85]. The possibility of paternal inheritance through the pollen was established and extensively studied for the 11.6 kb Brassica linear plasmid [86,87].

The origin of mitochondrial plasmids is mostly unknown but their presence in many plant species indicates that their acquisition is not a rare event. Short similarities were highlighted between several circular plasmids from different species, raising the idea of a common ancestor [88]. Mitochondrial plasmids often contain sequences that have nuclear counterparts, which in turn suggests intercompartment exchange. Conversely, the invertron structure of linear plasmids [71] rather points out to a common ancestral invertron of viral origin [54]. However, the orf sequences and the structural organization differ among plant mitochondrial linear plasmids, in agreement with independent origins. Finally, it has been speculated that horizontal DNA transfer, in particular from fungi, might have played a role in the acquisition of linear plasmids by plant mitochondria [54].

### 5. Recombination control

#### 5.1. Recombination factors

At the core of the organelar homologous recombination (HR) activities, there are eubacterial-type RecA proteins inherited from the corresponding prokaryotic endosymbiont ancestors. RecA catalyzes DNA strand-exchange during homologous recombination. Initially, RecA polymerizes on single-stranded DNA to form a long
presynaptic DNA–protein filament. In the presence of homologous double-stranded DNA, RecA catalyzes strand exchange, forming a D-loop that is a common intermediate in homologous recombination pathways. The process can be modulated by many factors. These include SSB proteins, which have very high affinity for single-stranded DNA that they sequester from RecA, and recombination mediators, which can displace SSBs and promote DNA annealing. After strand invasion, several outcomes are possible, according to the recombination pathway that is engaged (reviewed in Ref. [89]) and the available recombination factors. The processes that are active in plant mitochondria are still not identified.

In *A. thaliana* and all flowering plants, there are two mitochondrial RecA-like proteins, RECA2 and RECA3 (Table 1). While RECA3 is strictly mitochondrial, RECA2 was described as dual-targeted [90], but loss of RECA2 only affects mitochondrial functions [16]. By heterologous expression in *Escherichia coli*, it was shown that RECA2 and RECA3 are functional RecAs, with overlapping but also specific activities [16]. Accordingly, the *A. thaliana* reca2 and reca3 mutants displayed different phenotypes: reca2 plants are unable to develop past the seedling stage, while reca3 plants are phenotypically normal [16,90].

The essential functions of RECA2 are still unknown, but they might be related to C-terminal sequences that it shares with most known RecA proteins and that are not present in RECA3. The reca2 and reca3 mutants have normal mtDNA copy numbers, indicating that RecA functions are not essential for replication. But both reca2 and reca3 mutants display molecular phenotypes of increased recombination between ISRs [16]. The RecA-independent homologous recombination activity is exacerbated in reca2 mutants, where accumulation of crossover products is an order of magnitude higher than in reca3 mutants. As discussed below, RECA2 and RECA3, as well as additional genes, are apparently involved in surveillance pathways that restrict mtDNA recombination. The reshuffling of the mitochondrial genome by recombination is probably the cause of the observed reca2 lethal phenotypes.

As RECA2 and RECA3, additional core elements of the plant mtDNA maintenance machinery were inherited from the bacterial ancestors of plant organelles. That is the case of the eubacterial-type single-stranded DNA-binding proteins SSB. In bacteria, SSB is an abundant essential protein that cooperatively binds single-stranded DNA, forming protein filaments that saturate long stretches of the DNA. Binding of SSB protects the exposed single-stranded DNA from nucleases, but also melts secondary structures that inhibit elongation of RecA–single-stranded DNA filaments. However, the high-affinity binding of SSB also limits RecA assembly onto single-stranded DNA [91]. To overcome recombination inhibition by SSB, recombination mediators are required. Mediators can be other single-stranded DNA-binding proteins that promote the displacement of SSB and the assembly of RecA [92]. In *A. thaliana* and in most flowering plants, there are two SSB genes, SSB1 and SSB2 (Table 1) [33,41]. Both proteins are targeted to mitochondria ([41] and our unpublished results). SSB1 was shown to have similar properties as *E. coli* SSB: it specifically binds single-stranded DNA but not double-stranded DNA and stimulates RecA-dependent strand invasion [41]. Essential functions for SSB2 were confirmed in *A. thaliana*, where ssb2 mutants are unable to develop past the seedling stage (our unpublished results).

In addition to eubacterial-type SSBs, plant mitochondria have specific single-stranded DNA-binding proteins. The precise molecular functions of these proteins are not well characterized, but they might have multiple and possibly redundant roles in mtDNA replication, recombination and repair. Among them there are the organellar single-stranded DNA-binding proteins OSB (Table 1), which in *A. thaliana* form a small family of four proteins targeted to either mitochondria or plastids, or dual-targeted to both organelles [93]. OSB proteins are characterized by the presence of a central domain similar to the OB-fold domain followed by one, two or three C-terminal PDF domains. The PDF domain is a well-conserved domain of 50 amino acids only found in OSB proteins. It is a Predicted (SUBA3, http://suba.plantenergy.uwa.edu.au).

**Table 1** Nuclear-encoded proteins involved or suspected to be involved in organellar DNA replication, recombination and repair in *A. thaliana*.

<table>
<thead>
<tr>
<th>Category</th>
<th>Name</th>
<th>AGI</th>
<th>Targeting</th>
<th>Function</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA polymerase</td>
<td>Pol1A</td>
<td>At1g50840</td>
<td>Mt + Cp</td>
<td>Replication, Repair</td>
<td>[34–37,144]</td>
</tr>
<tr>
<td></td>
<td>Pol1B</td>
<td>At3g20540</td>
<td>Mt + Cp</td>
<td>Replication, Repair</td>
<td>[35,39,145]</td>
</tr>
<tr>
<td>Topoisomerase</td>
<td>Twinkle</td>
<td>At1g30680</td>
<td>Mt + Cp</td>
<td>Replication</td>
<td>[40]</td>
</tr>
<tr>
<td></td>
<td>GyrIIA</td>
<td>At3g10690</td>
<td>Mt + Cp</td>
<td>Replication</td>
<td>[40]</td>
</tr>
<tr>
<td></td>
<td>GyrII1B</td>
<td>At3g10270</td>
<td>Cp</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>GyrII2B</td>
<td>At5g04130</td>
<td>Mt</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recombinase</td>
<td>RECA2</td>
<td>At2g19490</td>
<td>Mt + Cp</td>
<td>Replication surveillance</td>
<td>[96]</td>
</tr>
<tr>
<td></td>
<td>RECA3</td>
<td>At3g10140</td>
<td>Mt</td>
<td>Recombination surveillance, Repair</td>
<td>[96,147]</td>
</tr>
<tr>
<td>ssDNA-binding/recombination mediator</td>
<td>SSB1</td>
<td>At4g11060</td>
<td>Mt</td>
<td>Replication, Repair</td>
<td>[33,41]</td>
</tr>
<tr>
<td></td>
<td>SSB2</td>
<td>At3g18580</td>
<td>Mt</td>
<td>Replication</td>
<td>[93]</td>
</tr>
<tr>
<td></td>
<td>OSB1</td>
<td>At3g18580</td>
<td>Mt</td>
<td>Recombination surveillance</td>
<td>[93]</td>
</tr>
<tr>
<td></td>
<td>OSB2</td>
<td>At4g20010</td>
<td>Cp</td>
<td>n.d.</td>
<td>[93]</td>
</tr>
<tr>
<td></td>
<td>OSB3</td>
<td>At5g44785</td>
<td>Mt + Cp</td>
<td>Recombination surveillance</td>
<td>[93]</td>
</tr>
<tr>
<td></td>
<td>OSB4</td>
<td>At1g31010</td>
<td>Mt</td>
<td>Recombination surveillance</td>
<td>Unpublished</td>
</tr>
<tr>
<td></td>
<td>WHY1</td>
<td>At1g14410</td>
<td>Cp</td>
<td>Repair</td>
<td>[95,146]</td>
</tr>
<tr>
<td></td>
<td>WHY2</td>
<td>At1g17260</td>
<td>Mt</td>
<td></td>
<td>[95,146]</td>
</tr>
<tr>
<td></td>
<td>WHY3</td>
<td>At2g07740</td>
<td>Cp</td>
<td></td>
<td>[95,146]</td>
</tr>
<tr>
<td></td>
<td>ODB1</td>
<td>At1g17130</td>
<td>Mt + Nuc?</td>
<td>Repair</td>
<td>[96,97,147]</td>
</tr>
<tr>
<td></td>
<td>ODB2</td>
<td>At5g7870</td>
<td>Cp + Nuc?</td>
<td></td>
<td>[96,97]</td>
</tr>
<tr>
<td>MutS</td>
<td>MSH1</td>
<td>At3g24320</td>
<td>Mt + Cp</td>
<td>Recombination surveillance</td>
<td>[106]</td>
</tr>
<tr>
<td>DNA glycosylase</td>
<td>UNG</td>
<td>At2g18630</td>
<td>Mt</td>
<td>Repair</td>
<td>[31,148]</td>
</tr>
<tr>
<td></td>
<td>FPG</td>
<td>At1g52500</td>
<td>Nuc⁺</td>
<td></td>
<td>[123,124]</td>
</tr>
<tr>
<td></td>
<td>OGG1</td>
<td>At1g21710</td>
<td>Nuc, Cp⁺</td>
<td></td>
<td>[121,122]</td>
</tr>
<tr>
<td>AP endonuclease</td>
<td>ARP</td>
<td>At2g14160</td>
<td>Cp</td>
<td>Repair</td>
<td>[118]</td>
</tr>
<tr>
<td></td>
<td>APE1L</td>
<td>At3g48425</td>
<td>Nuc⁺ + Cp</td>
<td></td>
<td>[127]</td>
</tr>
<tr>
<td></td>
<td>APE2</td>
<td>At4g36050</td>
<td>Nuc⁺ + Mt</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DNA ligase</td>
<td>LG1</td>
<td>At1g08130</td>
<td>Mt + Nuc</td>
<td>Replication, Repair</td>
<td>[130]</td>
</tr>
</tbody>
</table>

molecular phenotype of the mutants for the mitochondrion-targeted OSB1 and OSB4 proteins is the same as that of the recA2 and recA3 mutants: mutant plants accumulate aberrant mtDNA sequences resulting from ectopic recombination between certain ISRs ([93] and our unpublished results). This suggests that OSB proteins are also involved in the surveillance of mtDNA recombination. The specific expression of OSB1 in gametophytic cells further suggests that recombination surveillance by OSB proteins is required to prevent transmission of aberrant mtDNA configurations generated by ectopic recombination ([93]).

Whirly (WHY) proteins are further single-stranded DNA-binding proteins present in plant organelles (reviewed in Refs. [89,94]). In plastids, WHY proteins have multiple DNA- and RNA-related roles [89,94]. In A. thaliana mitochondria there is a single WHY protein (WHY2) (Table 1), whose roles are not yet completely understood. Over-expression of WHY2 affected the transcription of several mitochondrial genets but this could be a non-specific effect due to masking of the mtDNA by WHY2. Rather, like chloroplast WHY1 and WHY3, mitochondrial WHY2 seems to be involved in the surveillance of illegitimate recombination, preventing error-prone pathways of double-strand break (DSB) repair through sequence microhomologies [17,95]. It is possible that WHY proteins function as recombination mediators, promoting RecA-dependent homologous recombination repair pathways to the detriment of alternative illegitimate recombination processes.

Proteins with sequence and structural similarities to eukaryotic RAD52, a recombination mediator, are present in plants [96–97]. The two RAD52-like genes identified in the A. thaliana genome were reported to generate four orfs through differential splicing, with the respective isoforms being distributed between the nucleus, mitochondria and chloroplasts [96]. However, a specific antibody only detected the mitochondrial and chloroplast isoforms [97]. Like WHY2, the RAD52-like mitochondrial protein ODB1 (organellar DNA-binding) is possibly an additional recombination mediator (Table 1). ODB1 co-purified with WHY2 from cauliflower mitochondrial extracts and both proteins were shown to co-immunoprecipitate [97]. ODB1 has a high affinity for single-stranded DNA, with no apparent sequence specificity, but is also able to bind double-stranded DNA, although with much less affinity. On the other hand, ODB1 is able to promote annealing of complementary sequences [97]. In this respect, ODB1 seems to be a functional homolog of yeast mitochondrial Mgm1p101p. Also distantly related to RAD52, the latter is involved in mtDNA maintenance and has recombination mediator properties similar to those of RAD52 [98–100]. In agreement with this assumption, A. thaliana knockout mutants of ODB1 are impaired in homologous recombination-dependent repair of mtDNA breaks generated by genotoxic treatment [97].

An important component of the mitochondrial recombination machinery is the MSH1 gene, which encodes a MutH-like protein (Table 1). MSH1 has essential functions in the surveillance of mtDNA recombination. The gene was identified in A. thaliana as responsible for the cytoplasmic inherited CHM (chloroplast mutator) phenotype. The msh1 mutants display phenotypes of variegated and distorted leaves that correlated with important rearrangements of the mtDNA by recombination [15,90,101–103]. In additional species where MSH1 was knocked down by RNA silencing, mtDNA recombination was also increased, accompanied by phenotypic defects that included male sterility [104]. In A. thaliana, MSH1 is targeted to both mitochondria and chloroplasts [105] and recently it was shown that it is also active in the maintenance of the plastidial genome, explaining the variegation phenotype of msh1 plants [106]. In bacteria, MutS is a protein that participates in mismatch repair and in the suppression of ectopic recombination. Together with MutL, it recruits the endonuclease MutH to mismatches that arise during replication and homologous recombination [107]. There is no known MutH-like gene in A. thaliana, but plant mitochondrial MSH1 contains a C-terminal GIY-YIG-type homing endonuclease domain. It is therefore possible that MSH1 combines mismatch recognition and endonuclease functions. In recombination reactions, MSH1 probably recognizes heteroduplexes in strand-exchange complexes when there is little sequence homology. MSH1 would then promote their rejection [105,108].

Additional putative components of the plant mitochondrial recombination machinery have been identified by their similarity to known factors from other species, including DNA polymerases. As discussed in Section 3, in A. thaliana there are two organellar DNA polymerases, Pol1A and Pol1B (Table 1), which are targeted to both mitochondria and plastids and have redundant roles in DNA replication [33–35,37]. However, Pol1B, but not Pol1A, apparently has specific functions in DNA repair, because pol1B mutants, and not pol1A mutants, are hypersensitive to genotoxic agents that specifically induce DNA breaks in organellar genomes [37].

Finally, although many factors involved in mtDNA recombination have been identified either by their DNA-binding activities or from sequence comparisons, many additional ones are predicted to be required that remain to be identified. These might not be of eubacterial origin, because corresponding genes could not be found in the A. thaliana genome. They might be of viral origin or might have been recruited from the eukaryotic host. For example, an initial fundamental step in homologous recombination of DSBs is the recognition of free DNA extremities and their resection by 5′−3′ exonucleases, to form long 3′-overhangs that are substrates for RecA and other mediators. DNA helicases should also have important roles in recombination, for instance during recombination, to unwind branched duplex DNA and help to process Holliday-junctions by catalyzing branch migration. Several candidates can be pinpointed, from genes coding for proteins with corresponding functional motifs and predicted organellar targeting sequences. However, the putative mitochondrial functions of these proteins remain to be validated.

5.2. Recombination specificity

Homologous recombination involves the exchange between the invading single strand of the donor DNA and the double-stranded recipient DNA. The length of sequence homology required for efficient strand-exchange is characteristic of the recombinase enzymes that catalyze the reaction, which in plant mitochondria are eubacterial-type RecAs [16,90,109]. The minimal pairing region of bacterial RecA can be as small as 25 bp, but efficiency greatly increases with sizes of sequence homology greater than 200 bp [110]. Repeated sequences are the main sites of intramolecular recombination in plant mtDNAs and repeats of diverse size and number have been identified in the plant mitochondrial genomes so far sequenced [33,111]. Accordingly, the efficiencies of recombination observed in plant mitochondria correlate with the size and homology of the repeats. Large repeats of sizes larger than 1 kb mediate high frequency, reciprocal homologous recombination. This type of recombination involving large repeats is responsible for the well described multipartite structure of plant mitochondrial genomes. On the other hand, ISRs recombine sporadically, and often the products of their recombination are not detectable by conventional map-based sequencing approaches. But such products can be visualized on overexposed Southern blots and by PCR. The infrequent ectopic recombination mediated by these repeats leads to many possible alternative configurations of the mtDNA and has been associated with intraspecific variation of the mtDNA [15,103]. But this type of recombination can also lead to deleterious
changes in gene expression, and is under tight control. Thus, mutants of several genes that apparently modulate mtDNA recombination can have dramatically higher frequencies of ectopic recombination between ISRs, correlating with phenotypes of distorted and variegated leaves. Among these, mutants affected in the expression of MSH1 have been extensively characterized in several plant species [90,106,108]. In *A. thaliana*, in addition to MSH1, other genes have been identified whose mutants have the same molecular phenotypes of increased ectopic recombination across ISRs. These are the genes coding for the two mitochondrion-targeted RecA proteins RECA2 and RECA3 [16] and the plant-specific organelar single-stranded DNA-binding proteins OSB1 and OSB4 ([93] and our unpublished results). The mtDNA rearrangements induced by the loss of any of these surveillance genes concern ISRs ranging from about 50 bp up to 556 bp in *A. thaliana* [15,16,103]. These can be identical in sequence or imperfect repeated sequences, and the efficiency of sequence exchange appears to loosely correlate with repeat size and sequence homology. How far apart the repeats are located in the genome and how they are oriented might also influence the recombination efficiency and the differential accumulation of the corresponding crossover products [15,16].

The similar molecular and visible phenotypes of these mutants suggest that the corresponding genes are involved in recombination surveillance processes that restrict recombination when sequence homology between repeats is below a certain size. In the case of MSH1, like for other MutS factors, when there is little sequence homology the protein might recognize heteroduplexes in strand-exchange complexes, and promote their rejection. The C-terminal GIY-YIG-type homing endonuclease domain of MSH1 could be responsible for cleaving the heteroduplexes and starting over the process of sequence homology scanning [105]. However, different recombination pathways might be affected in other mutants. In the case of the RECA2 and RECA3 mutants, the observed increase in ectopic recombination suggests that it is the loss of RecA-dependent recombination that favors RecA-independent pathways under more relaxed control. These could involve the Single-Strand Annealing recombination pathway (SSA) that can occur when complementary single-stranded DNA sequences are exposed [16]. The OSB1 and OSB4 proteins might also be involved in modulating RecA-dependent recombination, for instance by promoting or suppressing binding of RecA to single-stranded DNA.

In addition to ectopic recombination that is suppressed under normal plant conditions, it was observed that genotoxic treatments also result in increased ectopic recombination across ISRs. This can be explained by the mobilization of recombination factors for the repair of DNA breaks by the BIR pathway. This homologous recombination-dependent repair pathway depends on RecA-like activities and accordingly it was found that genotoxic-induced mtDNA recombination is suppressed in RECA3 mutants [16]. Other genes were also found to be involved in this genotoxic-induced repair process: the organellar DNA polymerases Pol1B [37], the Rad52-like protein ODB1 [97] and the plant-specific single-stranded DNA-binding protein WHY2 [17]. The corresponding mutants are not affected in recombination surveillance under normal growth conditions, but under genotoxic treatment they have reduced recombination, as compared to wild-type plants. Concomitant with the loss of the RecA-dependent repair activity, these mutants also display increased illegitimate recombination involving microhomologies of just a few nucleotides [17,37,97]. Such sequence chimeras are of very low abundance, and can only be detected by PCR-based approaches. But collectively they likely constitute a major source of mtDNA heteroplasmy. On an evolutionary time scale, they are probably responsible for the important changes in the mtDNA structure observed between plant species.

6. Mitochondrial DNA repair

6.1. Recombination-dependent repair

Repair of DSBs is essential for genome stability because they can lead to collapsed replication forks, gene loss and genome instability. The sequence of plant mtDNAs revealed numerous examples of gene chimeras or fusions of unrelated sequences likely emerging from repair by illegitimate recombination, involving none or just a few nucleotides. Several of these gene chimeras have been associated with CMS phenotypes.

In chloroplasts there is extensive evidence for the participation of plastid-targeted RecA in DNA repair [112–114]. In *Physcomitrella patens*, it was shown that disruption of a mitochondrial RecA results
in reduced recovery of the mtDNA following genotoxic damage [115]. In the mitochondria of flowering plants, it was demonstrated that RECA3 has specific functions in mtDNA repair [16]. In wild-type A. thaliana plants, treatments that induce DNA repair activities result in the accumulation of specific ISR-derived crossover products. This induced recombination activity is lost in the recA3 background, showing that RECA3 is part of an HR-dependent repair activity induced by genotoxic stress [16]. The possible roles of RECA2 could not be assessed, given the seedling-lethal phenotype of the recA2 mutant.

The patterns of recombination that are triggered by RECA3-dependent repair activities are specific and distinct from RECA-independent ones that spontaneously occur in dependent repair activities are specific and distinct from RECA-independent ones that spontaneously occur in dependent repair activities. This could be tested by qPCR analysis of plants grown in the presence of genotoxic drugs, such as the antibiotic ciprofloxacin that sterically inhibits the organellar-targeted gyrase, thus provoking DSBs in the organellar genomes but not in the nuclear DNA. The studies showed that RECA3 and the RAD52-like protein ODB1 are involved in homologous recombination-dependent repair of the mtDNA [16,97]. Similar experiments confirmed the specific involvement in repair of the single-stranded DNA-binding protein WHY2 and of the organellar DNA polymerase POL1B, but not of POL1A (our unpublished results).

As a consequence of deficient homologous recombination-dependent mtDNA repair, alternative pathways of illegitimate recombination repair are apparently activated, in particular non-homologous end joining (NHEJ) and microhomology-mediated recombination (MMMR) (Fig. 2). This was shown in why2, pol1B and odb1 mutants. Increased repair by NHEJ and MMMR-dependent pathways can explain the chimeric sequences often observed in plant mitochondrial genomes.

Importantly, mutants deficient in mitochondrial recombination-dependent repair can be hypersensitive to genotoxic drugs that induce DNA breaks [16,37], showing that repair of organellar genome lesions significantly contributes to the survival of plants under stress conditions. It is apparent that many studies dealing with plant responses under stress often overlook the contribution of the organellar genomes to these processes.

6.2. Base excision repair

A major consequence of mtDNA oxidation caused by respiration byproducts or environmental conditions is the formation of abnormal bases. Oxidized bases are preferentially replaced through base excision repair (BER). The BER pathway [116] is initiated by a specific DNA glycosylase that cleaves the N-glycosidic bond between the abnormal base and the deoxyribose, thus generating an abasic (AP) site. Two ways are then possible for the process to proceed, short-patch BER (spBER) and long-patch BER (lpBER). In the case of spBER, if the cleavage has been carried out by a monofunctional DNA-glycosylase (1) or a bifunctional DNA glycosylase (2). UNG, uracil-DNA glycosylase; OGG1, 8-oxoguanine-DNA glycosylase; FPG, formamidopyrimidine-DNA glycosylase; APE2 and APEL1, homologs of AP endonucleases from yeast and bacteria. For enzymes followed by question marks, the involvement is not established.

In vivo assays with mitochondrial fractions from Z. mays seedlings [119] and by in vitro and in organello assays with A. thaliana and Solanum tuberosum mitochondria [31]. The enzyme did not seem to have a lyase activity and was in part membrane-associated. In vivo targeting and in vitro import of GFP fusion proteins showed that the enzyme encoded by the single putative A. thaliana UDG gene is indeed targeted into mitochondria [31]. For 8-oxoguanine-DNA glycosylase (OGG1) and formamidopyrimidine-DNA glycosylase (FPG), which are functional analogs, the situation is less clear. A DNA glycosylase activity recognizing 8-oxoguanine in A. thaliana and S. tuberosum mitochondria was characterized, using in vitro and in organello assays (Boesch P., Paulus F. and Dietrich A., unpublished data). Plants are the only organisms where the presence of both OGG1 and FPG has been reported [120]. There are thus two candidates to account for the 8-oxoguanine-cleaving activity. A. thaliana OGG1 is a bi-functional enzyme that was shown to localize to the nucleus by GFP fusion experiments [121]. Based on bioinformatic analysis, Macovei et al. [122] suggested the presence of a putative mitochondrial targeting sequence in the N-terminal part of the Medicago truncatula, A. thaliana, Populus nigra and Oryza sativa OGG1. However, when considering all relevant bioinformatic tools, prediction of OGG1 localization in plants is either
mitochondrial or plastidic, depending on the program used, with an overall preference for plastids. FPG, also a bi-functional enzyme, was found in *M. truncatula* and *A. thaliana* [122,123]. In both species, FPG contains a nuclear targeting sequence. Remarkably, in *A. thaliana* there is a single copy of the FPG gene that comprises ten exons and codes potentially for seven FPG forms, due to alternative splicing of the mRNA [124]. The different forms of FPG could be specific for different oxidized substrates but it can also be hypothesized that they would be localized in different compartments of the cell. As a whole, one might still speculate that OGG1 or FPG would be relocalized to the organelles when the mtDNA or the cpDNA is oxidized. This was shown to be the case for *Saccharomyces cerevisiae* N-glycosylase1 (Ntg1). Ntg1 possesses both a nuclear and a mitochondrial targeting sequence and is sent to the compartment where the DNA is oxidized [125,126]. Sumoylation, a post-translational modification, seems to be involved in Ntg1 nuclear relocalization [125].

After base excision, an AP endonuclease is necessary for the BER pathway to proceed. An AP endonuclease activity was characterized in *A. thaliana* and *S. tuberosum* mitochondria [31]. *A. thaliana* possesses three genes, named ARP, APE11 and APE2 (Table 1), that encode homologs of AP endonucleases from human, yeast and bacteria. Transient expression of GFP fusions established that the ARP protein is targeted to chloroplasts [118]. APE11 was found in a plastid proteome [127]. Alternative splicing yields two forms of APE2, the longest showing organellar targeting predictions. Also in *B. napus*, the putative AP endonuclease is predicted to be targeted to the organelles. Upon elimination of the damaged nucleotide, a DNA polymerase is needed to fill the gap. As discussed in previous sections, in *A. thaliana*, two organellar γ-type DNA polymerases, POL1A and POL1B, are dual-targeted to chloroplasts and mitochondria (Table 1) [33,34,128] and both function in mtDNA replication [36]. Whether one or both of these enzymes are involved in BER remains to be established. At the final step, a ligase is needed to reconnect the two parts of the DNA strand. Cordoba-Canero et al. [129] showed that DNA ligase 1 is required for both short-patch and long-patch BER in the *A. thaliana* nucleus. In this species, transcription of the DNA ligase 1 gene yields one major and two minor mRNA variants differing in the length of the 5′ leader preceding a common orf. According to in planta targeting of GFP fusions, the longer protein resulting from translation starting at the first in frame AUG codon is imported into mitochondria (Table 1), whereas the other forms are not [130]. The mitochondrial isoform actually would be relocalized to the organelles when the mtDNA or the cpDNA is oxidized [125,126]. As in yeast and in animals (reviewed in Ref. [32]), sharing a mitochondrial targeting sequence and is sent to the compartment where the DNA is oxidized [125,126]. Sumoylation, a post-translational modification, seems to be involved in Ntg1 nuclear relocalization [125].

7. Mitochondrial genome segregation and evolution

An important consequence of mtDNA heteroplasmasy is that, from a heteroplasmic parent, individuals can segregate in which certain substoichiometric mitotypes were amplified and became the predominant mtDNA. This process is called substoichiometric shifting (SSS) [131,132]. SSS can result in the activation or silencing of mitochondrial sequences, including the expression of gene chimeras, thus altering mitochondrial gene expression with potentially deleterious consequences to the plant. But SSS is also responsible for the rapid evolution of the mtDNA structure and to cytoplasm-nucleus conflicts, which can result in CMS [133]. This was clearly revealed by intra-specific sequence comparisons showing that intergenic regions change rapidly through sequence duplications, inversions, deletions and insertions resulting from recombination [2]. Through the production of specific stoichiometric changes, the activity of the MSH1 recombination factor appeared to account for the polymorphisms distinguishing *A. thaliana* ecotypes [103]. Mutation or down regulation of genes coding for key recombination factors like MSH1, OSB1 or RECA3 triggers extensive mitochondrial genome reorganization in *A. thaliana* or crop species, leading to marked phenotypes [90,93,103,104]. In OSB1-mutated lines, subgenomic levels of homologous recombination products accumulate, one of them becoming predominant in subsequent generations [93]. Remarkably, the mtDNA structural changes triggered by recombination factor mutation or down regulation are irreversible after SSS, and it is possible to revert to the wild type nuclear genome by backcrossing, while maintaining the modified mtDNA genome that is maternally inherited [93,104]. In that way, down regulating MSH1 allowed the production of stable, non-transgenic, male sterile lines [104].

In the case of non-allelic recombination involving ISRs, a model based on BIR was proposed that could account for the events leading to stoichiometric changes [15,90]. However, it is difficult to reconcile this model with the accumulation of recombined mtDNA in mutants deficient for RecA [16,90,134], which should be required for the BIR pathway. Most probably, both RecA-dependent BIR and additional RecA-independent homologous recombination processes co-exist in plant mitochondria and contribute to the pool of recombinated mtDNA sequences [16]. But while BIR and additional homologous and illegitimate recombination processes can account for the accumulation of re-shuffled mtDNA sequences in somatic tissues, these pathways cannot explain the maintenance and segregation of very low copy number sequences that cannot be re-created *de novo* by recombination [135]. It is therefore possible that selective replication of alternative mtDNA sequences is also involved in the SSS process.

How these molecules segregate and can be amplified during SSS is not well understood. The segregation of alternative mitotypes likely occurs during gametogenesis, and it is therefore significant that several genes described as involved in the suppression of ectopic recombination involving ISRs are expressed in the female gamete [16,90,93]. This suggests that their activities are important to repress the transmission of recombined mtDNA molecules. As it has been speculated in animals, the segregation of mitotypes could involve a genetic bottleneck during oogenesis that would reduce the copy number of transmitted mtDNA molecules. However, there is no evidence that such a process exists in plants, because the mtDNA copy number in egg cells seems to be equivalent to the one in mesophyll cells, at least in *A. thaliana* [136]. Since the mtDNA copy numbers in plants are relatively small, as compared to animals [5], the existence of a bottleneck is probably not necessary.

Models have also been proposed to account for the evolution of mitochondrial genome structures and explain the combination of low mutation rates and genome expansion. A "mutation burden" hypothesis put forward by Lynch et al. [137] proposes that the fundamental features of genome evolution are largely defined by the relative power of random genetic drift and mutation, i.e. primary non-adaptive forces. Especially, high mutation rates would be a barrier to organelle genome evolution. In this respect, that plant and animal mitochondrial genomes evolved to opposite architectural complexity would be a consequence of a two order of magnitude difference in their mutation rates. Such a model would
stand for plant species with low mtDNA mutation rates and high mitochondrial genome expansion, but it remains difficult to apply to species like *S. noctiflora* or *S. conica*, which show both exceptionally high mutation rates in genes and highly expanded genomes. Christensen proposed a model based on DNA repair and recombination mechanisms [138] that might account both for low mutation rates in genes combined with high expansion, rearrangement and mutation rates in the rest of the mitochondrial genome, and for the conundrum posed by species with huge genomes and high mutation rates in genes [4,139,140]. The model postulates different repair mechanisms in transcribed and non-transcribed regions [138]. Accurate template-directed DSB repair by synthesis-dependent strand annealing (SDSA) with enhanced second strand capture, gene conversion and BER would keep mutation rates low in coding regions. Conversely, BIR, NHEJ and further error-prone repair pathways, such as microhomology-mediated BIR, would explain expansion, rearrangements and rapid divergence in non-coding regions. In species with both high gene mutation rates and dramatic expansion of non-coding regions, lack of some BER or mismatch repair mechanisms would promote the accumulation of damaged and unpaired bases in coding regions, the fixation of mutations, the enhancement of DSB formation and the frequency of BIR [138]. This attractive hypothesis implies that transcription-directed repair processes exist in plant mitochondria, although the mechanisms and factors involved remain to be identified.

8. Conclusion

In many aspects, the genetics of plant mitochondria is more complex than that of most other organisms. The capacity to integrate and/or expand intergenic non-coding sequences in the organellar genome is remarkable, especially versus the strictly compact mammalian mtDNA. One might speculate that gain of plasmids and their import competence are shared with fungal mitochondria [142]. However, mammalian organelles are competent as well [143] and their genome size is stable. Also, many years after their discovery, the significance of mitochondrial plasmids and their cross-talk with the main mitochondrial genome are still a mystery. Not only the sequence content but also the physical organization of the plant mtDNA is a complex mix that either results from or supports a variety of replication and maintenance mechanisms. While rare in mammalian organelles, recombination is a major player in shaping plant mitochondrial genomes, redistributing sequences, generating polymorphism, driving evolution and at the same time preserving the genetic information. In the end, the challenge remains to understand how the high plasticity of the plant mtDNA can combine with gene conservation, intercompartment genome coordination and appropriate functional efficiency of the organelles. In this respect, it can be noted that, apart from reported mutants, mtDNA maps and restriction patterns have been reproduced many times for many species and generations, which indicates that, although dynamic, plant mitochondrial genomes remain stable on an observable time scale.

Acknowledgments

Our projects are funded by the French Centre National de la Recherche Scientifique (CNRS, UPR2357), the Université de Strasbourg (UdS), the Agence Nationale de la Recherche (ANR-06-MRAR-037-02, ANR-09-BLAN-0240-01) and the Ministère de la Recherche et de l’Enseignement Supérieur (Investissements d’Avenir/Laboratoire d’Excellence MitoCross).

References


[145] T.E. Shutt, M.W. Gray, Twinkle, the mitochondrial replicative DNA helicase, is widespread in the eukaryotic radiation and may also be the mitochondrial DNA primase in most eukaryotes, J. Mol. Evol. 62 (2006) 588–599.

