bioRxiv preprint doi: https://doi.org/10.1101/427500; this version posted September 26, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Do Plant Mitochondria Even Need Base Excision Repair?

Emily L. Wynn and Alan C. Christensen*

School of Biological Sciences, University of Nebraska – Lincoln, Lincoln, NE 68588-0666, USA

ORCID IDs: 0000-0001-6902-9301 (E.L.W.), 0000-0002-1125-3172 (A.C.C.)

* To whom correspondence should be addressed, Tel: +1 402 472 0681; FAX +1 402 472 8722; Email: achristensen2@unl.edu

bioRxiv preprint doi: https://doi.org/10.1101/427500; this version posted September 26, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

ABSTRACT

Substitution rates in plant mitochondrial genes are extremely low, indicating strong selective pressure as well as efficient repair. Plant mitochondria possess base excision repair pathways, however, many repair pathways such as nucleotide excision repair and mismatch repair appear to be absent. In the absence of these pathways, many DNA lesions must be repaired by a different mechanism. To test the hypothesis that double-strand break repair (DSBR) is that mechanism, we maintained independent lines of plants deficient in uracil-N-glycosylase (UNG) for 10 generations to determine the repair outcomes when that pathway is missing. In the absence of UNG, there is an increase in in double-strand breaks as assayed by recombination at repeated sequences. Surprisingly, given the single-seed descent bottleneck and the bottleneck of mitochondrial genomes in gametogenesis, no single nucleotide polymorphisms were fixed in any line in generation 10. The pattern of heteroplasmy was also unaltered through 10 generations. These results indicate that double strand break repair is a general system of repair in plant mitochondria, and appears to be so efficient that base excision repair is nearly dispensable. The existence of this general system may explain the seemingly anomalous differences between genes and non-genes in plant mitochondria.

bioRxiv preprint doi: https://doi.org/10.1101/427500; this version posted September 26, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

INTRODUCTION

Plant mitochondrial genomes have very low base substitution rates, while also expanding and rearranging rapidly (1-4). The low nonsynonymous substitution rates in protein coding genes indicates that selective pressure to maintain the genes is high, and the low synonymous substitution rates indicate that the DNA repair mechanisms are very accurate (5,6). However, little is known about the multiple pathways of DNA repair in plant mitochondria. So far, there is no evidence of nucleotide excision repair (NER), nor mismatch repair (MMR) in plant mitochondria (7,8). It has been hypothesized that in plant mitochondria, the types of DNA damage that are usually repaired through NER and MMR are repaired through double-strand break repair (DSBR) (9,10). Plant mitochondria have the nuclear-encoded base excision repair (BER) pathway enzyme Uracil DNA glycosylase (UNG) (7). UNG is an enzyme that can recognize and bind to uracil in DNA and begin the process of base excision repair by enzymatically excising the uracil (U) residue from single stranded or double stranded DNA (11). Uracil can appear in a DNA strand due to the spontaneous deamination of cytosine, or by the misincorporation of dUTP during replication (12). Unrepaired uracil in DNA can lead to G-C to A-T transitions within the genome.

The low substitution rate and the high rearrangement rate of plant mitochondria must be due to both selection and the specific DNA damage repair mechanisms available. These mechanisms must also account for the observations of genome expansion found in land plant mitochondria. Few pathways of repair besides DSBR are known in plant mitochondria, and it is possible that many lesions, including mismatches are repaired by creating double-strand breaks and using a template to repair both strands. Our hypothesis is that DSBR accounts for most of the repair in plant mitochondria, and both error-prone and accurate subtypes of DSBR lead to the observed patterns of genome evolution (13). One way of testing this is to eliminate the pathway of uracil base excision repair and ask if the G-U mispairs that occur by spontaneous deamination are instead repaired by DSBR.

bioRxiv preprint doi: https://doi.org/10.1101/427500; this version posted September 26, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The *Arabidopsis thaliana* mitochondrial genome contains two pairs of very large repeats (4.2 and 6.6kb) that commonly undergo recombination (14-16) producing multiple isoforms of the genome. The mitochondrial genome also contains many smaller repeats between 50 and 600 base pairs (16-19). In wild type plants, these repeats of unusual size (ROUS) recombine at very low rates. However, the ROUS have been shown to recombine at higher rates in several mutant lines defective in plant mitochondrial genome maintenance, such as *msh1* and *reca3* (20-22). In addition, ciprofloxacin (CIP) is a DNA gyrase inhibitor that can cause the formation of double strand breaks (DSBs) in mitochondria (23). Treating plants with CIP has been shown to induce DSBs in *A. thaliana* mitochondria, overwhelming the DSBR homology surveillance system and leading to increased ectopic recombination between ROUS (24). Thus recombination at the ROUS can be used as an indicator of increased DSBR. In this work we show that a loss of uracil base excision repair leads to increased DSBR, supporting our hypothesis.

Numerous proteins known to be involved in the processing of plant mitochondrial DSBs have been characterized. Plants lacking the activity of mitochondrially targeted recA homologs have been shown to be deficient in DSBR (21,25). In addition, it has been hypothesized that the plant MSH1 protein may be involved in binding to DNA lesions and initiating DSBs (9,10). The MSH1 protein contains a mismatch binding domain fused to a GIY-YIG type endonuclease domain which may be able to make DSBs (26,27). Here we provide evidence that in the absence of mitochondrial UNG activity, several genes involved in DSBR, including *MSH1*, are transcriptionally upregulated.

MATERIALS AND METHODS

Plant growth conditions:

Arabidopsis thaliana Columbia-0 (Col-0) seeds were obtained from Lehle Seeds (Round Rock, TX, USA). UNG T-DNA insertion hemizygous lines were obtained from the Arabidopsis Biological Resource Center, line number CS308282. Hemizygous T-DNA lines were self-crossed to obtain homozygous lines

bioRxiv preprint doi: https://doi.org/10.1101/427500; this version posted September 26, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(Genotyping primers: wild-type 5'-TGTCAAAGTCCTGCAATTCTTCTCACA-3' and 5'-TCGTGCCATATCTTGCAGACCACA-3', *ung* 5'-ATAATAACGCTGCGGACATCTACATTTT-3' and 5'-ACTTGGAGAAGGTAAAGCAATTCA-3'). All plants were grown in walk-in growth chambers under a 16:8 light:dark schedule at 22°C. Plants grown on agar were surface sterilized and grown on 1x Murashige and Skoog Basal Medium (MSA) with Gamborg's vitamins (Sigma) with 5µg/mL Nystatin Dihydrate to prevent fungal contamination.

RT-PCR:

RNA was extracted from mature leaves of plants grown in soil during *ung* generation nine and from seedlings grown on MSA with and without 0.5µM CIP(28). Reverse transcription using Bio-Rad iScript was performed and the resulting cDNA was used as a template for qPCR to measure relative transcript amounts. Quantitative RT-PCR data was normalized using *UBQ11* as a housekeeping gene control. Reactions were performed in a Bio-Rad CFX96 thermocycler using 96 well plates and a reaction volume of 20µL/well. SYBRGreen mastermix (Bio-Rad) was used in all reactions. Two biological and three technical replicates were used for each amplification. Primers are listed in Supplementary Table S2. The thermocycling program for all RT-qPCR was a ten-minute denaturing step at 95° followed by 45 cycles of 10s at 95°, 15s at 60°, and 13s at 72°. Following amplification, melt curve analysis was done on all reactions to ensure target specificity. The melt curve program for all RT-qPCR was from 65°-95° at 0.5^{o} increments for 5s each.

ROUS recombination qPCR:

DNA was collected from the mature leaves of Columbia-0 and generation ten *ung* plants, and from Columbia-0 seedlings grown on MSA with and without 0.5µM CIP using the CTAB DNA extraction method (29). qPCR was performed using primers from the flanking sequences of the ROUS. Primers are listed in Supplementary Table S1. Using different combinations of forward and reverse primers, either

bioRxiv preprint doi: https://doi.org/10.1101/427500; this version posted September 26, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

the parental or recombinant forms of the repeat can be selectively amplified (see Figure 1A). The mitochondrially-encoded *COX2* and *RRN18* genes were used as standards for analysis. Reactions were performed in a Bio-Rad CFX96 thermocycler using 96 well plates with a reaction volume of 20µL/well. SYBRGreen mastermix (Bio-Rad) was used in all reactions. Number of biological replicates are listed in figure descriptions, three technical replicates were used for each reaction. The thermocycling program for all ROUS recombination qPCR was a ten-minute denaturing step at 95° followed by 45 cycles of 10s at 95°, 15s at 60°, and a primer specific amount of time at 72° (extension times for each primer pair can be found in Supplementary Table S1). Following amplification, melt curve analysis was done on all reactions to ensure target specificity. The melt curve program for all qPCR was from 65°-95° at 0.5^{o} increments for 5s each.

DNA sequencing:

DNA extraction from frozen leaves of Columbia-0 and generation 10 *ung* was done by a modification of the SPRI magnetic beads method of Rowan *et al* (30,31). Genomic libraries for paired-end sequencing were prepared using a modification of the Nextera protocol (32), modified for smaller volumes following Baym *et al* (33). Following treatment with the Nextera Tn5 transpososome 14 cycles of amplification were done. Libraries were size-selected to be between 400 and 800bp in length using SPRI beads (31). Libraries were sequenced with 150bp paired-end reads on an Illumina HiSeq 4000 by the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley.

Reads were aligned using BWA-MEM v0.7.12-r1039 (34). The reference sequence used for alignment was a file containing the improved Columbia-0 mitochondrial genome (accession BK010421) (35) as well as the TAIR 10 *Arabidopsis thaliana* nuclear chromosomes and chloroplast genome sequences (36). If reads are aligned only to the mitochondrial genome, some reads containing nuclear or chloroplast

bioRxiv preprint doi: https://doi.org/10.1101/427500; this version posted September 26, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

sequences with similarity to mitochondrial sequence are incorrectly aligned to the mitochondria, leading to false SNP calls downstream in the analysis (data not shown). Using Samtools v1.3.1 (37), bam files were sorted for uniquely mapped reads for downstream analysis.

Variants were called using VarDict (38). To minimize the effects of sequencing errors, SNPs called by VarDict were filtered by the stringent quality parameters of Allele Frequency ≥ 0.05, Qmean ≥ 30, MQ ≥ 30, NM ≤ 3, Pmean ≥ 8, Pstd = 1, AltFwdReads ≥ 3, and AltRevReads ≥ 3. The mitochondrial reference genome positions corresponding to *RRN18* and *RRN26* were excluded from analysis because they have similarity to bacterial 16S and 23S ribosomal RNAs, respectively. Sequencing reads from contaminating soil bacteria can be misaligned to these positions and falsely called as low frequency SNPs. No other mitochondrial sequences show such similarity to bacterial genes (data not shown).

RESULTS

Experimental Design

This work was designed as a mutation accumulation study (39). We expected a relatively low mutation rate and chose 23 different *ung* homozygous plants derived from one hemizygous parent. These 23 plants were designated as generation 1 *ung* and were allowed to self-cross. The next generation was derived by single-seed descent from each line, and this was repeated until generation 10 *ung* plants were obtained. Leaf tissue and progeny seeds from each line were kept at each generation.

Increased Double-strand break repair:

If most DNA damage in plant mitochondria is repaired by double-strand break repair (DSBR), supplemented by base excision repair (7), then in the absence of the Uracil-N-glycosylase (UNG) pathway we predict an increase in DSBR. To find evidence of this we assayed ectopic recombination

bioRxiv preprint doi: https://doi.org/10.1101/427500; this version posted September 26, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

between homeologous repeats, which increases when DSBR is increased (21,22,24). We assayed ROUS recombination in *ung* generation ten lines and in wild-type by quantitative PCR (qPCR). Different combinations of primers in the unique sequences flanking the ROUS allow us to determine the relative copy numbers of parental type ROUS and low frequency recombinants (Figure 1A). The mitochondrial genes *COX2* and *RRN18* were used to standardize relative amplification between lines. We and others (18,24) have found that some of the ROUS are well-suited for qPCR analysis and are sensitive indicators of ectopic recombination, increasing in repair-defective mutants and when drugs are used to increase double-strand breaks. We chose to analyze recombination at the three repeats known as Repeats D, L, and B (17). All three show reductions in the parental 2/2 form, while repeat B also shows a reduction in the parental 1/1 form (Figure 1B). This indicates that ectopic recombination is increased, however, this pattern of recombination is different than what is seen in wild-type plants treated with CIP. In the presence of 0.5µM CIP, there is an increase in the recombinant 1/2 and 2/1 forms of the repeats as well (Figure 2).

Substitution Mutations

We also expected that in the absence of UNG there would be an increase in G-C to A-T substitution mutations. To test this prediction, we sequenced 23 independent lines of the *ung* mutant in generation 10 using an Illumina Hi-Seq4000 system. Mitochondrial sequences from these independent *ung* lines were aligned to the Columbia-0 reference genome using BWA-MEM and single nucleotide polymorphisms were identified using VarDict. The average depth of coverage of the mitochondrial genome was 130x and the range for different lines was 50x to 467x.

There were no SNPs that reached fixation to an allele frequency of 1 in any *ung* lines. Mitochondrial genomes are not diploid; each cell can have many copies of the mitochondrial genome. Therefore, it is

bioRxiv preprint doi: https://doi.org/10.1101/427500; this version posted September 26, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

possible that an individual plant could accumulate low frequency mutations in some of the mitochondrial genomes in the cell. VarDict (38) was used to detect heteroplasmic SNPs at allele frequencies as low as 0.05. VarDict's sensitivity in calling low frequency SNPs scales with depth of coverage and quality of the sample, so it is not possible to directly compare heteroplasmic mutation rates in samples with different depths of coverage. However, because the activity of the UNG protein is specific to uracil, the absence of the UNG protein should not have any effect on mutation rates other than G-C to A-T transitions. Because of this, heteroplasmic mutations that are not G-C to A-T transitions can be considered the background rate of heteroplasmic SNP accumulation in plant mitochondria. We therefore compared the numbers of G-C to A-T transitions to all other mutations. If the *ung* mutants are accumulating heteroplasmic SNPs at a faster rate than wild-type, we would expect to see that as an increased number of G-C to A-T transitions compared to other mutation types. In contrast, there is little difference in the ratios of G-C to A-T transitions between *ung* and Col-0 (see Table 1). Because detection of low frequency SNPs depends on read depth, we only analyzed the 7 *ung* samples with an average mitochondrial read depth above 125x for this comparison. In the absence of a functional UNG protein, plant mitochondria do not accumulate cytosine deamination mutations at an increased rate.

Alternative Repair Pathway Genes

Because the *ung* mutants show increased double-strand break repair but not an increase of G-C to A-T transition mutations, we infer that the inevitable appearance of uracil in the DNA is repaired via conversion of a G-U pair to a double-strand break and efficiently repaired by the DSBR pathway. If this is true, genes involved in the DSBR processes of breakage, homology surveillance and strand invasion in mitochondria will be up-regulated in *ung* mutants. To test this hypothesis, we assayed transcript levels of several candidate genes known to be involved in DSBR (8,17,20-22,24,25,40-42) in *ung* lines compared to wild-type using RT-PCR. *MSH1* and *RECA2* were upregulated in *ung* lines (9.68 fold ± 4.64

bioRxiv preprint doi: https://doi.org/10.1101/427500; this version posted September 26, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

and 5.39 fold ± 0.98, respectively – see Figure 3). *RECA3*, *OSB1*, *SSB*, and *RECG* showed no differential expression compared to wild-type.

To investigate the differences in ROUS recombination between *ung* lines and wild-type lines treated with CIP, expression of DSBR genes was compared between wild-type plants with and without CIP treatment. *MSH1* was downregulated in wild-type plants grown in the presence of CIP compared to untreated control plants (0.42 fold ± 0.1 – see Figure 3)

Discussion:

In the mitochondrion as well as in the nucleus and chloroplast, cytosine is subject to deamination to uracil. This could potentially lead to transition mutations, and is dealt with by a specialized base excision repair pathway. The first step in this pathway is hydrolysis of the glycosidic bond by the enzyme Uracil-N-glycosylase (UNG), leaving behind an abasic site (11). An AP endonuclease can then cut the DNA backbone, producing a 3' OH and a 5' dRP. Both DNA polymerases found in *Arabidopsis* mitochondria, POL1A and POL1B, exhibit 5'-dRP lyase activity, allowing them to remove the 5' dRP and polymerize a new nucleotide replacing the uracil (43). In the absence of functional UNG protein, cytosine will still be deaminated in plant mitochondrial genomes, so efficient removal of uracil must be through a different repair mechanism, most likely DSBR (9,10). We have found that in *ung* mutant lines, there is an increase in double-strand break formation as assayed by an increase in ectopic recombination at ROUS, as well as an increase in the expression of genes known to be involved in DSBR, consistent with this hypothesis.

Surprisingly, we have also found that *ung* lines do not accumulate G-C to A-T transition mutations at a higher rate than wild-type. This finding is particularly surprising given the possible bottlenecking of mitochondrial genomes during female gametogenesis, and given the deliberate bottleneck in the experimental design of maintaining the different lines through single-seed descent for 10 generations. This finding supports the hypothesis that plant mitochondria have a very efficient alternative mismatch

bioRxiv preprint doi: https://doi.org/10.1101/427500; this version posted September 26, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

or damage surveillance system that prevents G-C to A-T transitions from becoming fixed in the mitochondrial population, but that the side-effect of increased ectopic recombination may be the selective pressure that maintains the UNG pathway.

The angiosperm MSH1 protein consists of a DNA mismatch binding domain fused to a double-stranded DNA endonuclease domain (26,27). Although mainly characterized for its role in recombination surveillance (22), MSH1 is a good candidate for a protein that may be able to recognize and bind to various DNA lesions and make DSBs near the site of the lesion, thus funneling these types of damage into the DSBR pathway. With many mitochondria and many mitochondrial genomes in each cell there are numerous available templates to accurately repair DSBs through homologous recombination, making this a plausible mechanism of genome maintenance. Here we show that in *ung* lines, *MSH1* is transcriptionally upregulated more than 9-fold compared to wild-type. This further supports the hypothesis that MSH1 initiates repair in plant mitochondria by creating a double-strand break at G-U pairs, and possibly other mismatches and damaged bases.

Several other proteins involved in processing plant mitochondrial DSBs have been characterized. The RECA homologs, RECA2 and RECA3, are homology search and strand invasion proteins (21,22,25,42,44-46). The two mitochondrial RECAs share much sequence similarity, however RECA2 is dual targeted to both the mitochondria and the chloroplast, while RECA3 is found only in the mitochondria (21,22). RECA3 also lacks a C-terminal extension present on RECA2 and most other homologs. This extension has been shown to modulate the ability of RECA proteins to displace competing ssDNA binding proteins in E. coli (47). *Arabidopsis reca2* mutants are seedling lethal and both *reca2* and *reca3* lines show increased ectopic recombination at ROUS (21). Arabidopsis RECA2 has functional properties that RECA3 cannot perform, such as complementing a bacterial recA mutant during the repair of UV-C induced DNA lesions (42). Here we show that in *ung* lines, *RECA2* is transcriptionally upregulated more than 5-fold compared to the wild-type. However, *RECA3* is not upregulated in *ung* lines. Responding to MSH1-initiated DSBs

bioRxiv preprint doi: https://doi.org/10.1101/427500; this version posted September 26, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

may be one of the functions unique to RECA2. The increased expression of *RECA2* in the absence of a functional UNG protein is further evidence that uracil arising in DNA may be repaired through the mitochondrial DSBR pathway.

We also tested the differential expression of other genes known to be involved in processing mitochondrial DSBs. The single stranded binding protein genes *OSB1* and *SSB*, and the helicase *RECG* were not found to be differentially expressed at the transcript level compared to wild-type (data not shown).

The specific patterns of recombination at mitochondrial ROUS are different between wild-type, *ung* mutants, and DSBR mutants. In *msh1* lines, there is an increase in ROUS recombination likely due to relaxed homology surveillance in the absence of the MSH1 protein (22) . In *ung* lines, the mitochondrial recombination machinery is still intact, so any differences in ROUS recombination between *ung* lines and wild-type are solely due to differences in processing uracil in DNA with or without the UNG protein. At many ROUS, when DSBs are chemically induced by ciprofloxacin, parental type repeats are reduced while recombinant type repeats increase in prevalence (24). The *ung* lines show reductions in parental type repeats, but no change in recombinant type repeats. One explanation for this behavior is that the upregulation of *MSH1* in the *ung* lines leads to an increase in DSBs at deaminated cytosine and reduces the accumulation of recombinant ROUS. In contrast, there is a downregulation of *MSH1* in plants exposed to 0.5µM CIP (Figure 3). This downregulation would lead to relaxation of the homology surveillance activity of MSH1, making accumulation of recombinant ROUS more common. Such asymmetric accumulation of recombination products has been documented in plant mitochondrial genomes (17,18,21,22,48).

bioRxiv preprint doi: https://doi.org/10.1101/427500; this version posted September 26, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

To determine the outcomes of genomic uracil in the absence of a functional UNG protein, we sequenced the genomes of several *ung* lines. No fixed mutations of any kind were found in *ung* lines, even after 10 generations of self-crossing. Low frequency heteroplasmic SNPs were found in both wild-type and *ung* lines, but *ung* lines showed no difference in the ratio of G-C to A-T transitions to other mutation types when compared to wild-type.

Clearly the double-strand break repair pathway in plant mitochondria can repair uracil in DNA sufficiently to prevent mutation accumulation, providing evidence that DSBR is the primary mechanism of repair. Why then has the BER pathway been conserved in plant mitochondria where NER and MMR have apparently been lost? DSBR protects the genome efficiently from mutations, but also leads to an increase in ectopic recombination and rearrangements in plant mitochondria. We suggest that for DNA lesions canonically repaired through NER or MMR pathways, the slight increase in ectopic recombination due to repair by DSBR does not incur any fitness cost compared to the strong selection for accurate repair of genes. In Arabidopsis, the most common type of heteroplasmic SNP found in the mitochondrial genome is the G-C to A-T transition; cytosine deamination is therefore a very common type of DNA damage in plant mitochondria. Perhaps the increased ectopic recombination that results from the repair of these very common DNA lesions provides just enough selective pressure to maintain an intact BER pathway in plant mitochondria. The UNG protein is also transported to the plastid (7). UNG might therefore be retained in the mitochondria due to selection in the plastid.

Here we have provided evidence that in the absence of a dedicated BER pathway, plant mitochondrial genomes do not accumulate SNPs at an increased rate. Instead, DNA damage is accurately repaired by double-strand break repair which also causes an increase in ectopic recombination at homeologous repeats. Do plant mitochondria even need base excision repair? The answer appears to be “barely.” It has recently been shown that mice lacking a different mitochondrial BER protein, oxoguanine glycosylase, do not accumulate mitochondrial SNPs (49). Here we show that in plants base-excision

bioRxiv preprint doi: https://doi.org/10.1101/427500; this version posted September 26, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

repair by UNG is similarly unnecessary to prevent mitochondrial mutations. Perhaps a generalized system of DNA repair also exists in mammalian mitochondria similar to the broad capacity of DSBR to repair different lesions in plant mitochondria. Clearly DSBR is efficient and accurate, and the presence of the UNG pathway reduces ectopic recombination slightly. Double strand break repair and recombination are important mechanisms in the evolution of plant mitochondrial genomes, but many key enzymes and steps in the repair pathway are still unknown. Further identification and characterization of these missing steps is sure to provide additional insight into the unique evolutionary dynamics of plant mitochondrial genomes.

Data Availability:

Fastq files generated from Illumina sequencing of *ung* lines and wild-type control are available from the Sequence Read Archive, BioProject ID PRJNA492503.

Funding

This work was supported by a grant from the National Science Foundation to A.C.C. (MCB-1413152), and used the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley, supported by NIH S10 OD018174 Instrumentation Grant.

Conflict of Interest Statement: None declared.

Acknowledgements

We are grateful to Beth Rowan, UC Davis, for advice and protocols for DNA extraction and Illumina sequencing. Thanks to Emma Purfeerst for keeping our lab running and putting up with us during her time here and to Emily Jezewski for help with qPCR. Thanks to Sterling Ericsson and Ana Martinez-Hottovy for discussion and comments on the manuscript. Daniel Schactman helped with the disposal of leaves from generations 2 through 9 of the *ung* lineages.

bioRxiv preprint doi: https://doi.org/10.1101/427500; this version posted September 26, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

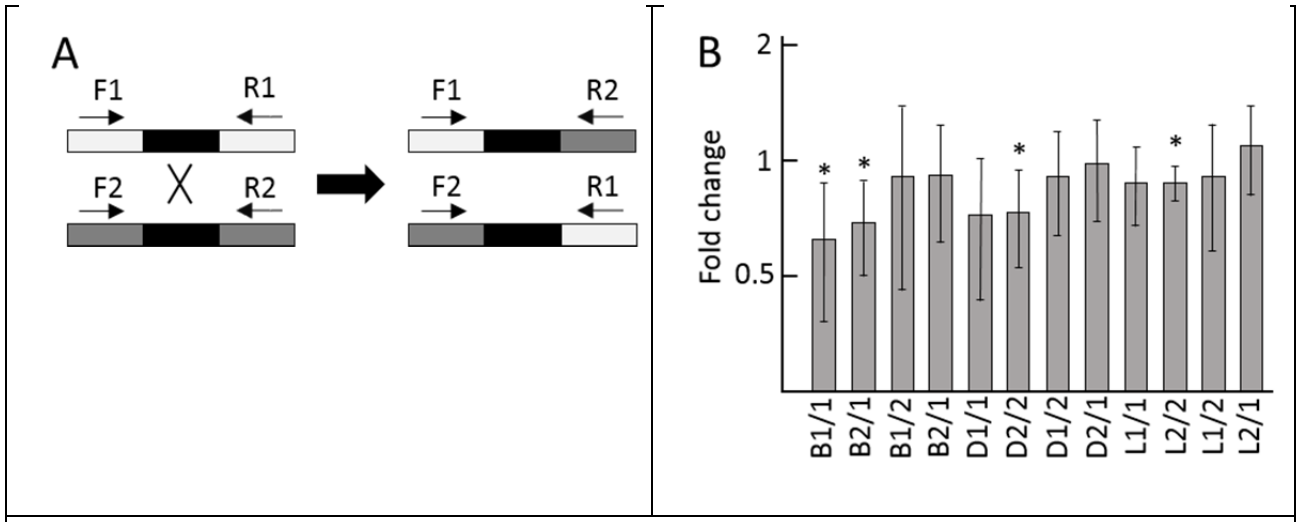

Figure 1: qPCR analysis of ROUS recombination in *ung* lines compared to wild-type. Recombination at ROUS is an indicator of increased double strand breaks in plant mitochondrial genomes. **A)** Primer scheme for detecting parental and recombinant repeats. Using different combinations of primers that anneal to the unique sequence flanking the repeats, either parental type (1/1 and 2/2) or recombinant type (1/2 and 2/1) repeats can be amplified **B)** Fold change of ROUS in *ung* lines relative to wild-type. Error bars are standard deviation of three biological replicates. B1/1, B2/2, D2/2, and L2/2 show significant reduction in copy number (unpaired, 2-tailed Student's t-test, * indicates $p<0.05$)

bioRxiv preprint doi: https://doi.org/10.1101/427500; this version posted September 26, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

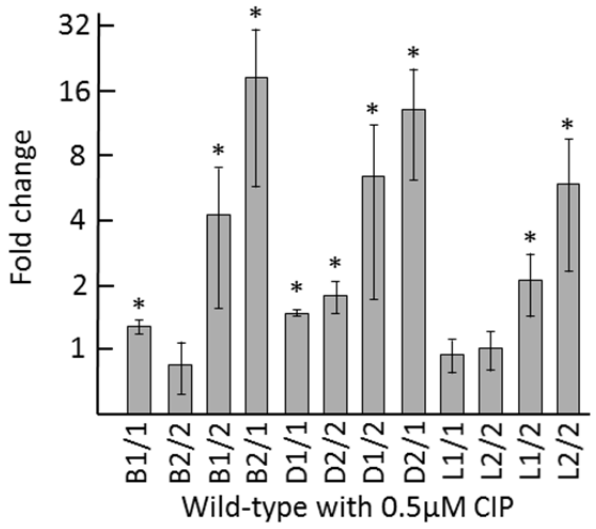

Figure 2: qPCR analysis of ROUS recombination in wild-type plants treated with 0.5µM CIP compared to untreated. Error bars are standard deviation of two biological replicates. B1/1, B1/2, B2/1, D1/1, D2/2, D1/2, D2/2, L1/2, and L2/1 show significant increase in copy number (unpaired, 2-tailed Student's t-test, * indicates $p<0.05$)

bioRxiv preprint doi: https://doi.org/10.1101/427500; this version posted September 26, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

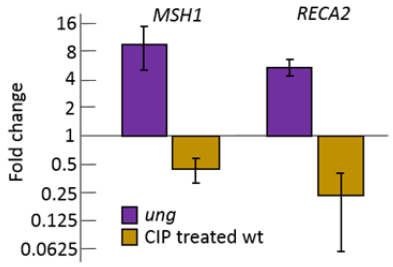

Figure 3: Quantitative RT-PCR assays of MSH1 and *RECA2* in *ung* lines and wild-type lines treated with 0.5µM CIP relative to untreated wild-type. Fold change in transcript level is shown on the Y-axis. Error bars are standard deviation of of two biological replicates. *MSH1* and *RECA2* are transcriptionally upregulated in *ung* lines relative to wild-type (9.68 fold ± 4.64 and 5.39 fold ± 0.98, respectively). *MSH1* and *RECA2* are transcriptionally downregulated in wild-type plants treated with 0.5µM CIP relative to untreated controls (0.42 fold ± 0.1 and 0.22 fold ± 0.16, respectively).

bioRxiv preprint doi: https://doi.org/10.1101/427500; this version posted September 26, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 1: Number of different SNPs in each of 5 *ung* mutant lines. SNP counts are shown for the entire mitochondrial genome, sorted by the type of change.

	GC→AT	**GC→TA**	**GC→CG**	**AT→GC**	**AT→CG**	**AT→TA**	**Total**	**Ratio of GC→AT transitions to total SNPs**
ung115	19	6	3	4	3	3	38	0.50
ung159	20	5	3	5	1	5	39	0.51
ung163	38	10	6	5	3	7	69	0.55
ung176	37	10	4	6	3	6	66	0.56
ung198	31	7	5	12	4	10	69	0.45
ung201	37	7	5	12	3	8	72	0.51
ung203	28	5	3	18	2	4	60	0.47
Col-0	41	8	6	6	3	7	48	0.58

bioRxiv preprint doi: https://doi.org/10.1101/427500; this version posted September 26, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supplemental Table S1: Primers for ROUS recombination assay

Repeat Name	Size	Forward Primer	Reverse Primer	Extension time during qPCR
Repeat L1	248	GTTATGCCATTTTGGGCTTTGCTC	CGCTCGACCGAAGAAATGAGTAAC	16s
Repeat L2	248	CCGGTTGAAAGCTAAGCCGGAG	CCCTCACTGAACCGACTTGAATCTG	16s
Repeat B1	532	AAGTCCTGCTCATATACATAC	AATCCGGCTCGCTCTCGGCTACGTG	30s
Repeat B2	532	TACGGCTTACGAAGTTGATCG	CGACATGAAAAGACGATGTCC	30s
Repeat D1	452	ATAAGAGAGCAAATGATGGTATAGC	GCTGCTTTTAGGAGAGTGATCTG	30s
Repeat D2	452	CAAAGCAAAGGATGGTTCAAC	CGATTGTATCCGACTATTGAGTG	30s
cox2	76	AATAAACGTGATTGACCCAATTCT	TCCGATGAGCAGTCACTCAC	6s
rrn18	62	CCTTGAGCTAGGAGCCTCTTT	CATGCAAGTCGAACGTTGTT	6s

Supplemental Table S2: Primers for RT-PCR. ^ represents exon junctions.

Name	Size	Forward Primer	Reverse Primer
UBQ11	70	TCTCAACCGTGATCAAG^ATGC	TTCCACCTCGAGGGTGATAGT
MSH1	77	AGTTACCAAGCTCCGCACTC	CACGTC^CCTGATGCATTGTG
RECA2	74	GCGAAGAG^ACAACGGGAAGT	GAGCAGTTCTAAACGGCGGA

bioRxiv preprint doi: https://doi.org/10.1101/427500; this version posted September 26, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supplemental Table S3: qPCR primer efficiency

Primer Name	Calibration Curve	R^2	Efficiency	Limit of Detection	No Template Control Results
cox2	y=-3.281x+16.06	0.9996	101.74	32.28	No amplification in 45 cycles
rrn18	y=-3.25x+17.295	0.9987	103.09	30.09	Average Cq = 35.65. Melt curve shows nonspecific amplification
L11	y=-3.681x+20.892	0.9979	86.92	31.5	No amplification in 45 cycles
L22	y=-3.264x+21.288	0.985	102.48	31.82	No amplification in 45 cycles
L12	y=-3.706x+21.916	0.9963	86.14	33.02	No amplification in 45 cycles
L21	y=-3.48x+24.28	0.9838	93.83	30.96	No amplification in 45 cycles
B11	y=-3.76x+20.928	0.9981	84.48	31.87	No amplification in 45 cycles
B22	y=-3.625x+20.355	0.9993	88.74	27.66	No amplification in 45 cycles
B12	y=-4.285x+25.995	0.9985	71.15	34.66	No amplification in 45 cycles
B21	y=-3.445x+23.345	0.9998	95.11	30.26	No amplification in 45 cycles
D11	y=-3.902x+21.65	0.9981	80.42	37.35	No amplification in 45 cycles
D22	y=-4.013x+21.033	0.9975	77.5	33.32	No amplification in 45 cycles
D12	y=-4.045x+27.638	0.9998	76.69	35.76	No amplification in 45 cycles
D21	y=-3.655x+26.272	1	87.76	33.57	No amplification in 45 cycles
ubq11	y=-3.579x+20.534	0.9991	90.29	34.88	No amplification in 45 cycles
msh1	y=-3.308x+33.313	0.9971	100.58	40.26	Average Cq = 42.64. Melt curve shows nonspecific amplification
recA2	y=-3.24x+28.263	0.9999	103.54	34.73	No amplification in 45 cycles

bioRxiv preprint doi: https://doi.org/10.1101/427500; this version posted September 26, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

References

1. Drouin, G., Daoud, H. and Xia, J. (2008) Relative rates of synonymous substitutions in the mitochondrial, chloroplast and nuclear genomes of seed plants. *Mol Phylogenet Evol*, **49**, 827-831.
2. 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.
3. Richardson, A.O., Rice, D.W., Young, G.J., Alverson, A.J. and Palmer, J.D. (2013) The "fossilized" mitochondrial genome of *Liriodendron tulipifera*: ancestral gene content and order, ancestral editing sites, and extraordinarily low mutation rate. *BMC biology*, **11**, 29.
4. Wolfe, K., Li, W. and Sharp, P. (1987) Rates of nucleotide substitution vary greatly among plant mitochondrial,chloroplast and nuclear DNAs. *Proc. Natl. Acad. Sci. U S A*, **84**, 9054-9058.
5. Sloan, D.B. and Taylor, D.R. (2010) Testing for selection on synonymous sites in plant mitochondrial DNA: the role of codon bias and RNA editing. *J Mol Evol*, **70**, 479-491.
6. Wynn, E.L. and Christensen, A.C. (2015) Are Synonymous Substitutions in Flowering Plant Mitochondria Neutral? *J. Molec. Evol*, **81**, 131-135.
7. Boesch, P., Ibrahim, N., Paulus, F., Cosset, A., Tarasenko, V. and Dietrich, A. (2009) Plant mitochondria possess a short-patch base excision DNA repair pathway. *Nucleic Acids Res*, **37**, 5690-5700.
8. Gualberto, J.M. and Newton, K.J. (2017) Plant Mitochondrial Genomes: Dynamics and Mechanisms of Mutation. *Annu Rev Plant Biol*.
9. Christensen, A.C. (2014) Genes and Junk in Plant Mitochondria—Repair Mechanisms and Selection. *Genome biology and evolution*, **6**, 1448-1453.
10. Christensen, A.C. (2018) In Logan, D. C. (ed.), *Annual Plant Reviews, 2nd Edition, Plant Mitochondria*. Wiley-Blackwell, New York, NY, USA, Vol. 50, pp. 11-31.
11. Cordoba-Cañero, D., Dubois, E., Ariza, R.R., Doutriaux, M.P. and Roldan-Arjona, T. (2010) Arabidopsis uracil DNA glycosylase (UNG) is required for base excision repair of uracil and increases plant sensitivity to 5-fluorouracil. *J Biol Chem*, **285**, 7475-7483.
12. Krokan, H.E., Standal, R. and Slupphaug, G. (1997) DNA glycosylases in the base excision repair of DNA. *The Biochemical journal*, **325 (Pt 1)**, 1-16.
13. Christensen, A.C. (2013) Plant mitochondrial genome evolution can be explained by DNA repair mechanisms. *Genome biology and evolution*, **5**, 1079-1086.
14. Klein, M., Eckert-Ossenkopp, U., Schmiedeberg, I., Brandt, P., Unseld, M., Brennicke, A. and Schuster, W. (1994) Physical mapping of the mitochondrial genome of *Arabidopsis thaliana* by cosmid and YAC clones. *Plant J.*, **6**, 447-455.
15. Palmer, J.D. and Shields, C.R. (1984) Tripartite structure of the Brassica campestris mitochondrial genome. *Nature*, **307**, 437.
16. Unseld, M., Marienfeld, J.R., Brandt, P. and Brennicke, A. (1997) The mitochondrial genome of *Arabidopsis thaliana* contains 57 genes in 366,924 nucleotides. *Nat. Genet.*, **15**, 57-61.
17. Arrieta-Montiel, M.P., Shedge, V., Davila, J., Christensen, A.C. and Mackenzie, S.A. (2009) Diversity of the Arabidopsis Mitochondrial Genome Occurs via Nuclear-Controlled Recombination Activity. *Genetics*, **183**, 1261-1268.
18. Davila, J.I., Arrieta-Montiel, M.P., Wamboldt, Y., Cao, J., Hagmann, J., Shedge, V., Xu, Y.Z., Weigel, D. and Mackenzie, S.A. (2011) Double-strand break repair processes drive evolution of the mitochondrial genome in Arabidopsis. *BMC biology*, **9**, 64.
19. Wynn, E.L. and Christensen, A.C. (2018) Repeats Of Unusual Size in Plant Mitochondrial Genomes: Identification, Incidence and Evolution. *bioRxiv*.

bioRxiv preprint doi: https://doi.org/10.1101/427500; this version posted September 26, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

20. Abdelnoor, R.V., Yule, R., Elo, A., Christensen, A.C., Meyer-Gauen, G. and Mackenzie, S.A. (2003) Substoichiometric shifting in the plant mitochondrial genome is influenced by a gene homologous to *MutS*. *Proc. Natl. Acad. Sci. U S A*, **100**, 5968-5973.
21. Miller-Messmer, M., Kuhn, K., Bichara, M., Le Ret, M., Imbault, P. and Gualberto, J.M. (2012) RecA-dependent DNA repair results in increased heteroplasmy of the Arabidopsis mitochondrial genome. *Plant Physiol*, **159**, 211-226.
22. Shedge, V., Arrieta-Montiel, M., Christensen, A.C. and Mackenzie, S.A. (2007) Plant mitochondrial recombination surveillance requires unusual RecA and MutS homologs. *Plant Cell*, **19**, 1251-1264.
23. Evans-Roberts, K.M., Mitchenall, L.A., Wall, M.K., Leroux, J., Mylne, J.S. and Maxwell, A. (2016) DNA Gyrase Is the Target for the Quinolone Drug Ciprofloxacin in Arabidopsis thaliana. *J Biol Chem*, **291**, 3136-3144.
24. Wallet, C., Le Ret, M., Bergdoll, M., Bichara, M., Dietrich, A. and Gualberto, J.M. (2015) The RECG1 DNA Translocase Is a Key Factor in Recombination Surveillance, Repair, and Segregation of the Mitochondrial DNA in Arabidopsis. *Plant Cell*, **27**, 2907-2925.
25. Odahara, M., Inouye, T., Fujita, T., Hasebe, M. and Sekine, Y. (2007) Involvement of mitochondrial-targeted RecA in the repair of mitochondrial DNA in the moss, Physcomitrella patens. *Genes & genetic systems*, **82**, 43-51.
26. Abdelnoor, R.V., Christensen, A.C., Mohammed, S., Munoz-Castillo, B., Moriyama, H. and Mackenzie, S.A. (2006) Mitochondrial genome dynamics in plants and animals: Convergent gene fusions of a *MutS* homolog. *J. Molec. Evol.*, **63**, 165-173.
27. Kleinstiver, B.P., Wolfs, J.M. and Edgell, D.R. (2013) The monomeric GIY-YIG homing endonuclease I-BmoI uses a molecular anchor and a flexible tether to sequentially nick DNA. *Nucleic Acids Res*, **41**, 5413-5427.
28. Onate-Sanchez, L. and Vicente-Carbajosa, J. (2008) DNA-free RNA isolation protocols for Arabidopsis thaliana, including seeds and siliques. *BMC research notes*, **1**, 93.
29. Allen, G.C., Flores-Vergara, M.A., Krasynanski, S., Kumar, S. and Thompson, W.F. (2006) A modified protocol for rapid DNA isolation from plant tissues using cetyltrimethylammonium bromide. *Nature protocols*, **1**, 2320-2325.
30. Rowan, B.A., Patel, V., Weigel, D. and Schneeberger, K. (2015) Rapid and inexpensive whole-genome genotyping-by-sequencing for crossover localization and fine-scale genetic mapping. *G3 (Bethesda, Md.)*, **5**, 385-398.
31. Rowan, B.A., Seymour, D.K., Chae, E., Lundberg, D.S. and Weigel, D. (2017) In White, S. J. and Cantsilieris, S. (eds.), *Genotyping: Methods and Protocols*. Springer New York, New York, NY, pp. 221-242.
32. Caruccio, N. (2011) Preparation of next-generation sequencing libraries using Nextera technology: simultaneous DNA fragmentation and adaptor tagging by in vitro transposition. *Methods Mol Biol*, **733**, 241-255.
33. Baym, M., Kryazhimskiy, S., Lieberman, T.D., Chung, H., Desai, M.M. and Kishony, R. (2015) Inexpensive multiplexed library preparation for megabase-sized genomes. *PLoS One*, **10**, e0128036.
34. Li, H. (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv:1303.3997v2*.
35. Sloan, D.B., Wu, Z. and Sharbrough, J. (2018) Correction of Persistent Errors in Arabidopsis Reference Mitochondrial Genomes. *The Plant Cell*, **30**, 525-527.
36. Berardini, T.Z., Reiser, L., Li, D., Mezheritsky, Y., Muller, R., Strait, E. and Huala, E. (2015) The Arabidopsis information resource: Making and mining the "gold standard" annotated reference plant genome. *Genesis (New York, N.Y. : 2000)*, **53**, 474-485.

bioRxiv preprint doi: https://doi.org/10.1101/427500; this version posted September 26, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

37. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. and Durbin, R. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078-2079.
38. Lai, Z., Markovets, A., Ahdesmaki, M., Chapman, B., Hofmann, O., McEwen, R., Johnson, J., Dougherty, B., Barrett, J.C. and Dry, J.R. (2016) VarDict: a novel and versatile variant caller for next-generation sequencing in cancer research. *Nucleic Acids Res*, **44**, e108.
39. Halligan, D.L. and Keightley, P.D. (2009) Spontaneous Mutation Accumulation Studies in Evolutionary Genetics. *Annual Review of Ecology, Evolution, and Systematics*, **40**, 151-172.
40. Edmondson, A.C., Song, D., Alvarez, L.A., Wall, M.K., Almond, D., McClellan, D.A., Maxwell, A. and Nielsen, B.L. (2005) Characterization of a mitochondrially targeted single-stranded DNA-binding protein in Arabidopsis thaliana. *Mol. Genet. Genomics*, **273**, 115-122.
41. Gualberto, J.M., Mileshina, D., Wallet, C., Niazi, A.K., Weber-Lotfi, F. and Dietrich, A. (2014) The plant mitochondrial genome: dynamics and maintenance. *Biochimie*, **100**, 107-120.
42. Khazi, F.R., Edmondson, A.C. and Nielsen, B.L. (2003) An Arabidopsis homologue of bacterial RecA that complements an E. coli recA deletion is targeted to plant mitochondria. *Mol Gen Genet*, **269**, 454-463.
43. Trasviña-Arenas, C.H., Baruch-Torres, N., Cordoba-Andrade, F.J., Ayala-Garcia, V.M., Garcia-Medel, P.L., Diaz-Quezada, C., Peralta-Castro, A., Ordaz-Ortiz, J.J. and Brieba, L.G. (2018) Identification of a unique insertion in plant organellar DNA polymerases responsible for 5'-dRP lyase and strand-displacement activities: Implications for Base Excision Repair. *DNA Repair (Amst)*, **65**, 1-10.
44. McGrew, D. and Knight, K. (2003) Molecular design and functional organization of the RecA protein. *Critical Reviews in Biochemistry and Molecular Biology*, **38**, 385-432.
45. Rowan, B.A., Oldenburg, D.J. and Bendich, A.J. (2010) RecA maintains the integrity of chloroplast DNA molecules in Arabidopsis. *J Exp Bot*, **61**, 2575-2588.
46. Xu, L. and Marians, K. (2002) A dynamic RecA filament permits DNA polymerase-catalyzed extension of the invading strand in recombination intermediates. *J. Biol. Chem.*, **277**, 14321-14328.
47. Eggler, A.L., Lusetti, S.L. and Cox, M.M. (2003) The C terminus of the Escherichia coli RecA protein modulates the DNA binding competition with single-stranded DNA-binding protein. *J Biol Chem*, **278**, 16389-16396.
48. Sakamoto, W., Tan, S., Murata, M. and Motoyoshi, F. (1997) An unusual mitochondrial atp9-rpl16 cotranscript found in the maternal distorted leaf mutant of *Arabidopsis thaliana*: implication of GUG as an initiation codon in plant mitochondria. *Plant Cell Physiol.*, **38**, 975-979.
49. Kauppila, J.H.K., Bonekamp, N.A., Mourier, A., Isokallio, M.A., Just, A., Kauppila, T.E.S., Stewart, J.B. and Larsson, N.G. (2018) Base-excision repair deficiency alone or combined with increased oxidative stress does not increase mtDNA point mutations in mice. *Nucleic Acids Res*, **46**, 6642-6669.