Repeats Of Unusual Size in Plant Mitochondrial Genomes: Identification, Incidence and Evolution

Emily L. Wynn[1] and Alan C. Christensen[2]

School of Biological Sciences

University of Nebraska

Lincoln, Nebraska 68588-0666, USA

[1] ORCID 0000-0001-6902-9301

[2] ORCID 0000-0002-1125-3172

1 **Running title: Repeats in plant mitochondria**

2

3 **Key words: Plant mitochondrial genomes, Repeated sequence, Genome**

4 **rearrangement, Organelle genome evolution**

5

Author for Correspondence: Alan C. Christensen, School of Biological Sciences, University of Nebraska – Lincoln, Lincoln, NE, USA, 68588-0666, Phone 402-472-0681, FAX 402-472-8722, email achristensen2@unl.edu

Abstract

Plant mitochondrial genomes have excessive size relative to coding capacity, a low mutation rate in genes and a high rearrangement rate. They also have non-tandem repeats in two size groups: a few large repeats which cause isomerization of the genome by recombination, and numerous repeats longer than 50bp, often found in exactly two copies per genome. It appears that repeats in the size range from several hundred to a few thousand base pair are underrepresented. The repeats are not well-conserved between species, and are infrequently annotated in mitochondrial sequence assemblies. Because they are much larger than expected by chance we call them Repeats Of Unusual Size (ROUS). The repeats consist of two functional classes, those that are involved in genome isomerization through frequent crossing over, and those for which crossovers are rare unless there are mutations in DNA repair genes, or the rate of double-strand breakage is increased. We systematically described and compared these repeats, which are important clues to mechanisms of DNA maintenance in mitochondria. .We developed a tool to find non-tandem repeats and analyzed the complete mitochondrial sequences from 135 plant species. We observed an interesting difference between taxa: the repeats are larger and more frequent in the vascular plants. Analysis of closely related species also shows that plant mitochondrial genomes evolve in dramatic bursts of breakage and rejoining, complete with DNA sequence gain and loss, and the repeats are included in these events. We suggest an adaptive explanation for the existence of the repeats and their evolution.

Introduction

It has long been known that plant mitochondrial genomes are much larger than those of animals (Ward, B. L. *et al.* 1981) and include significant amounts of non-coding DNA (Schuster, W. and A. Brennicke 1994). These genomes also often have repeats of several kb, leading to multiple isomeric forms of the genome (Folkerts, O. and M. R. Hanson 1989; Klein, M. *et al.* 1994; Palmer, J. D. and L. A. Herbon 1988; Palmer, J. D. and C. R. Shields 1984; Siculella, L. *et al.* 2001; Sloan, D. B. *et al.* 2010; Stern, D. B. and J. D. Palmer 1986). Plant mitochondrial genomes have very low mutation rates, but paradoxically have such high rearrangement rates that there is virtually no conservation of synteny (Drouin, G. *et al.* 2008; Palmer, J. D. and L. A. Herbon 1988; Richardson, A. O. *et al.* 2013; Wolfe, K. *et al.* 1987).

In addition to the large, frequently recombining repeats, there are often other repeated sequences in the size range of 1kb and lower (Arrieta-Montiel, M. P. *et al.* 2009; Forner, J. *et al.* 2005). Ectopic recombination between these homeologous repeats has been shown to increase when double-strand breakage is increased, or in plants mutant for DNA maintenance genes (Abdelnoor, R. V. *et al.* 2003; Shedge, V. *et al.* 2007; Wallet, C. *et al.* 2015). Understanding the repeats is critical to fully understanding the mechanisms of DNA maintenance and evolution in plant mitochondria, yet they have never been systematically identified and analyzed. In addition to being infrequently and inconsistently annotated and described in mitochondrial genome sequences, repeats are often described as long, short and intermediate-length (Arrieta-Montiel, M. P. *et al.* 2009; Davila, J. I. *et al.* 2011; Miller-

Messmer, M. *et al.* 2012). The repeats are sometimes thought to be distributed into two size classes (one of up to several hundred bp and another of several kb), but this is derived from early studies of Arabidopsis and a few other species in which repeats were described and annotated.

The most likely hypothesis that explains the peculiar characteristics of plant mitochondrial genomes is that double-strand break repair (DSBR) is abundantly used in plant mitochondria, perhaps to the exclusion of nucleotide excision and mismatch repair pathways (Christensen, A. C. 2014; Christensen, A. C. 2018). Double-strand break repair is very accurate when the repair is template-based, accounting for the low mutation rate in genes, but the nonhomologous end-joining or break-induced-replication pathways can account for the creation of repeats and chimeric genes, expansions, and loss of synteny through rearrangements.

The lack of a coherent nomenclature and the inconsistent reporting and annotation of repeated sequences leads to a number of questions. What is the best way to discover and characterize them? Is the size distribution really bimodal in angiosperms? Are there repeats in the mitochondria of other groups of green plants? How do they differ between groups? And can they be followed through evolutionary lineages like genes? Are the repeats themselves somehow adaptive, or are they a side-effect of DSBR that is neutral or nearly neutral? The availability in recent years of complete mitochondrial genome sequences across a wide variety of taxa of green plants allows us to begin addressing these questions. We describe a

computational strategy for finding non-tandem repeats within genomes. Using this tool we describe the phylogenetic distribution of repeats in both size classes, examine their evolution in a family of closely related angiosperms, and propose an hypothesis for the evolutionary significance of the repeats and the DSBR processes that produce them.

Materials and Methods

Sequence data and manipulation

DNA sequences were downloaded as FASTA format files from GenBank (https://www.ncbi.nlm.nih.gov/genbank/). BLAST searches (Altschul, S. F. *et al.* 1990) were done using version 2.7.1 on a Linux-based machine. In addition to the sequences shown in Table 1, mitochondrial genomes from several Brassica species were used to compare close relatives. These sequences are as follows: *Brassica carinata*; JF920287, *Brassica rapa*; JF920285, *Brassica oleracea fujiwase*; AP012988, *Brassica napus polima*; FR715249, *Brassica juncea*; JF920288. Alignments were done using the clustalW implementation in the VectorNTI 11.5 software package (ThermoFisher).

Repeat Analysis

Custom Python scripts are in Supplementary Materials. The script ROUSFinder.py (Supplemental File S1) uses blastn to perform a pairwise ungapped comparison of a sequence with itself, both strands separately, using a word size of 50, E value of 10,000, reward for a match +1, penalty for a mismatch -9, percent identity cutoff

99%. The script then concatenates the two output files and the full length identity is deleted. Alignments are then sorted and compared to identify and remove duplicate repeats, and an output file of the repeats in fasta format is created. This output file is then used as a query with the genome as subject to locate every copy of that repeat, create a table, and a table of binned sizes. The output can also be formatted for GenBank annotation. MultipleRepeats.py (Supplemental File S2) automates running ROUSFinder.py on every sequence within a directory.

Data Availability

The authors state that all data necessary for confirming the conclusions presented in this article are represented fully within the article, including python scripts in Supplemental Material and accession numbers of DNA sequences shown in Table 1. Supplemental Material available at FigShare.

Results

Repeats in plant mitochondrial genomes

The existence of large non-tandem repeats in plant mitochondrial genomes is well known by now, but they have not been systematically identified and analyzed. Prior studies used variations of BLAST (Altschul, S. F. *et al.* 1990) to find repeats (Alverson, A. J. *et al.* 2011a; Alverson, A. J. *et al.* 2010; Alverson, A. J. *et al.* 2011b; Liu, Y. *et al.* 2014) or REPuter (Hecht, J. *et al.* 2011; Kurtz, S. and C. Schleiermacher 1999). Other available software packages specifically identify tandem repeats, or repeats matching known repetitive sequences. Due to the ready availability of

BLAST and the flexibility of its use, and because most prior work used it, we wrote and used a Python script called ROUSFinder.py that uses BLAST to identify non-tandem repeats within mitochondrial genomes. The parameters for identification of a sequence repeat were relatively stringent and included a blastn word size of 50, a percent identify cutoff of 99% and match/mismatch scores of +1/-9. Any choice of parameters will necessarily identify some false positives and false negatives. These parameters were chosen in order to find duplicate copies of sequence that were either recently created or recently corrected by gene conversion. A duplication longer than 100 bases that has several mismatches in the center of the repeat unit will be identified as two different repeats in this way. However, the mismatches in the center are indicative of either two independent events producing the two parts of the repeat, or mutation and drift that have escaped gene conversion. Because we are concerned with the recombination behavior of the repeats we therefore choose to call these two different repeats. To analyze and identify repeats in a single sequence for further study or annotation would require additional manual curation of the output.

The species we used represent a significant subset of the complete mitochondrial genome sequences from green plants in GenBank and are shown in Table 1. Sequences available on GenBank are not a random sample across taxa (food crops are very over-represented, for example), so to reduce sampling bias somewhat we used only one species per genus. Incomplete sequences or sequences with gaps are not handled well by BLAST without further curation, so these were not used. Species with multiple distinct chromosomes were also not used because of the additional

layer of complexity from inter- and intra-chromosomal repeats. The full output is in Supplemental Table S1. The repeats seen in plant mitochondrial genomes are much larger than those found in random sequence (data not shown), suggesting that they arise from specific biological processes and are not stochastic. For this reason we call them “Repeats Of Unusual Size” or ROUS (Christensen, A. C. 2018).

BLAST is an excellent tool for identification of repeated sequences. Our script automates the task of identifying repeats in both direct and inverted orientations, removes the full-length match, and provides the output in a convenient format that can be used for annotation or in spreadsheets for further analysis.

Phylogenetic clustering

The distribution of repeat sizes forms distinct clusters between the phylogenetic groups (see Figure 1). Because there are different numbers of species in each group, and some species have an order of magnitude more total repeats than others, we represent the data as the fraction of species within that group that have at least one repeat within a given size range. The complete output is in Supplemental Table S1. There are several complete mitochondrial genomes from chlorophytes and bryophytes to compare to angiosperms. Within the chlorophytes, repeats of greater than 200bp are rare. The exceptions are the prasinophytes (discussed below) and a few interesting cases. *Chlamydomonas reinhardtii* has a 532 bp inverted repeat at the termini of its linear chromosome. *Dunaliella salina*, *Kirchneriella aperta* and *Polytoma uvella* have novel structures at a small number of loci that consist of

overlapping and nested repeats and palindromes (Smith, D. R. *et al.* 2010). The function of these structures is unknown, but they are unusual and not common in the chlorophytes. The prasinophyte group resembles the rest of the chlorophytes in having no ROUS greater than 200bp but many of them include two copies of a single large repeat between 9.5 and 14.4 kb. This is similar to many chloroplast genomes and it is possible that this structure is involved in replication (Bendich, A. J. 2004). The bryophytes generally resemble the chlorophytes; there are no ROUS greater than 200bp.

In contrast to the chlorophytes and bryophytes, the Marchantiophyta (liverworts) and Anthocerotophyta (hornworts) have ROUS greater than 200bp in size, but none bigger than 867bp. Few taxa within the liverworts and hornworts have been sequenced, so this group is underrepresented, but the members available to date are consistent. The other lineages of streptophytic green algae (referred to as charophytes in GenBank) resemble the chlorophytes albeit with a slightly higher upper limit. In this group the largest repeat is found in *Chlorokybus atmophyticus* and is 291bp.

The ferns and lycophytes are strikingly different from the previous groups. Unfortunately the number of species sequenced is low. They have large numbers of repeats and the repeat sizes range well above 200bp, up to 10 kb. Some members of these groups, such as Huperzia, are similar to the bryophytes, but others are large and have significant repeat content (Guo, W. *et al.* 2017). These groups are also very

underrepresented among available mitochondrial sequences (in part due to the complexity caused by the repetitive nature of the genomes (Grewe, F. *et al.* 2009)), but the patterns are noticeably different from the nonvascular plants described above.

The angiosperms are represented very well in the sequence databases. Only one member of this group does not have any ROUS above 200bp (*Medicago truncatula*). A small number of angiosperms, scattered among plant families, lack repeats larger than 1 kbp, and approximately half include repeats larger than 9 kbp. *Silene conica*, a species with multiple large chromosomes not included in our dataset has a nearly 75kb sequence found in both chromosomes 11 and 12 (Sloan, D. B. *et al.* 2012). Gymnosperms are also underrepresented, but appear to be similar to the other vascular plants. Interestingly, the gymnosperms *Ginkgo biloba* and *Welwitschia mirabilis* resemble angiosperms, while *Cycas taitungensis* is more similar to ferns. The *C. taitungensis* mitochondrion has numerous ROUS, including many that are tandemly repeated. Five percent of this genome consists of the mobile Bpu element, a remarkable level of repetitiveness (Chaw, S. M. *et al.* 2008).

It is only in the vascular plants that the number and size of repeated sequences in mitochondrial genomes has been expanded. The vascular plants generally only have genomes a few times larger than the bryophytes, liverworts and hornworts, but the repeats are expanded well beyond proportionality to size. In addition, random sequences of comparable length do not have any repeats of the sizes discussed here

(data not shown). Some taxa, such as the Geraniaceae, Plantago, and Silene include species with significantly expanded genomes (Park, S. *et al.* 2015; Parkinson, C. L. *et al.* 2005; Sloan, D. B. *et al.* 2012). These species are outliers in the magnitude of the genome sizes and number of repeats, but the underlying processes are likely to be the same. The overall picture is that there was a significant change in mitochondrial DNA maintenance mechanisms roughly coincident with the origin of the vascular plants.

Repeat sizes and frequency in angiosperms.

Large repeats of several kilobases have been identified in several species and shown to be recombinationally active, isomerizing angiosperm mitochondrial genomes (Folkerts, O. and M. R. Hanson 1989; Klein, M. *et al.* 1994; Palmer, J. D. and L. A. Herbon 1988; Palmer, J. D. and C. R. Shields 1984; Siculella, L. *et al.* 2001; Sloan, D. B. 2013; Stern, D. B. and J. D. Palmer 1986). A few species have been reported to lack such structures (Palmer, J. D. 1988). The first comprehensive catalog of repeated sequences shorter than 1000 base pairs was done in *Arabidopsis thaliana*, and they were shown to be recombinationally active in some mutant backgrounds, but not generally in wild type (Arrieta-Montiel, M. P. *et al.* 2009; Davila, J. I. *et al.* 2011; Miller-Messmer, M. *et al.* 2012; Shedge, V. *et al.* 2007). Is the spectrum of repeat sizes in Arabidopsis, and its bimodality, typical for angiosperms? Figure 2 illustrates the presence of repeats in the size range of 50bp to over 10,000 bp in 58 angiosperms. The overall pattern is that there is a multimodal distribution of sizes that is often bimodal. Gaps in the distribution are indicated in yellow in Figure 2,

however, the size cutoffs of the gap are somewhat variable. Most species have a paucity of repeats between 600 and 10,000bp. Twelve of the 58 species have no repeats larger than 600bp, leaving open the question of if they isomerize through recombination or not. All of the other species have a large repeat of somewhere between 800bp and 65kbp. The total length of repeats in a species does not correlate with genome size (linear regression $r^2 = 0.08$, data not shown), additional evidence that these are not produced by stochastic processes, and suggesting that they occur and change faster than speciation does.

Alignment of repeats within the Brassicales

Understanding the evolution of the repeated sequences requires analysis of homologous repeats in related species. Of the species with sequenced mitochondrial genomes, the Brassicales order of plants has a number of such species. Within the *Brassica* genus there are three diploid species: *Brassica rapa*, *Brassica nigra* and *Brassica oleracea*, and three allotetraploid species (Cheng, F. *et al.* 2017). The diploid nuclear genomes are called the A, B and C genomes, respectively. Based on both nuclear and mitochondrial sequences it appears that *Brassica carinata* has the *B. nigra* and *B. oleracea* nuclear genomes (BBCC) and the *B. nigra* mitochondrial genome, while *Brassica juncea* has the *B. nigra* and *B. rapa* nuclear genomes (BBCC) and the *B. rapa* mitochondrial genome. *Brassica napus* is of two subspecies, *polima* and *napus*. Both have the *B. oleracea* and *B. rapa* nuclear genomes (AACC), but *B. napus polima* appears to have the *B. rapa* mitochondrial genome and *B. napus napus* has the *B. oleracea* mitochondrial genome (Chang, S. *et al.* 2011; Franzke, A. *et al.*

2011; Grewe, F. *et al.* 2014). Thus it appears that the hybridization event between *B. oleracea* and *B. rapa* occurred at least twice, with each species being the maternal parent. In the analysis below we use the *B. napus polima* mitochondrial genome. We compared these Brassica species to *Raphanus sativus* and *Sinapis arvensis*. These species are the closest relatives of the Brassicas with complete mitochondrial genome sequences (Grewe, F. *et al.* 2014). Several of these species were mapped prior to genomic sequencing, and repeated sequences and isomerization of the genomes was observed (Palmer, J. D. 1988; Palmer, J. D. and L. A. Herbon 1986).

All eight of these species include one pair of long repeats, ranging in length from 1.9 to 9.7kb. However, these species fall into two groups. Each group has a large repeat that is found as a single copy in the other group. The B. nigra group consists of *R. sativus*, *S. arvensis*, *B. nigra* and *B. carinata*; the B. rapa group consists of *B. rapa*, *B. oleracea*, *B. napus* and *B. juncea*. A tree showing the relationships of these species, but without measures of distance, is shown in Figure 3. Part A shows the long repeat and neighboring sequences from the B. nigra group and the homologous single-copy sequences from the B. rapa group. Part B compares the long repeat from the B. rapa group to the unique homologous region from the B. nigra group.

The number of changes in coding DNA between plant species is generally low, making mutation rate estimates difficult, and these low rates may also be affected by sequencing errors (Sloan, D. B. *et al.* 2018). Grewe et al examined the synonymous substitution rates in genes of Brassicales mitochondrial genomes (Grewe, F. *et al.*

2014) and found them to be very low, consistent with most land plants. However, the presence of repeats allows mutations in non-coding DNA to be examined qualitatively. The long repeats of *R. sativus, S. arvensis*, *B. nigra* and *B. carinata* differ by large block substitutions and insertion/deletions (alignments are shown in Supplemental Figure S1). Where two copies are present in a species there are very few difference between copies, and they are generally near the boundaries of the repeats. Although significant differences can arise during speciation events, both copies of a repeat within a species remain identical. This supports the hypothesis that copies of repeated DNA are maintained as identical sequence by frequent recombination and gene conversion.

The long repeat of *B. nigra* and *B. carinata* underwent massive change in the lineage leading to the other four Brassica species. The first 1.6kb and the last 1.7kb of the repeat of *B. nigra* are conserved in the *B. rapa* group, and the *ccmB* gene still flanks the repeat on one side. However, the last 1.7kb are inverted and separated from the first 1.6kb by 3.3kb of a sequence of unknown origin. An additional difference is seen in *B. oleracea* wherein *rps7* now flanks the repeat rather than *ccmB*. Other major changes appear to have occurred in the time since *B. nigra* diverged from the ancestor of *B. oleracea* and *B. rapa*; a comparison of the complete mitochondrial genomes of *B. rapa* and *B. nigra* reveal at least 13 segments of DNA that have been rearranged. No major rearrangements have occurred between *B. nigra* and *B. carinata*, nor between *B. rapa*, *B. juncea* and *B. napus polima*. *B. oleracea* differs from *B. rapa* by approximately six rearrangement events.

At the same time that the *B. nigra* long repeat was being dramatically altered in the lineage leading to *B. rapa* and *B. oleracea*, a new long repeat appeared, which includes the coding sequence of the *cox2* gene. This new long repeat is maintained throughout this group of four species, and the flanking genes are also conserved (alignments are shown in Supplemental Figure S2). The *cox2* gene is single copy in *R. sativus*, *S. arvensis*, *B. nigra* and *B. carinata*, and is in a nearly syntenic arrangement with neighboring genes.

Discussion

The availability of complete mitochondrial genome sequences from many taxa of green plant allows us to compare the repeat structures across taxa. Although Large Repeats and ROUS have been known for some time, their functions (if any) and evolution are largely mysterious. It has been suggested that their existence and maintenance are outgrowths of double-strand break repair events such as nonhomologous end-joining (NHEJ), break-induced replication (BIR) and gene conversion (Christensen, A. C. 2018). We describe here a Python script that uses BLAST (Altschul, S. F. *et al.* 1990) to find non-tandem repeats within genomes, and use it to analyze plant mitochondrial DNA. Comparison of repeats between closely related species showed that repeat differences between species were largely due to rearrangements and block substitutions or insertions, which could be due to NHEJ and BIR, while the two copies of the repeat were identical within a species, suggesting continuing repair by gene conversion or homologous recombination.

The phylogenetic distribution of complex repeated structures in mitochondria appear to be common to the vascular plants and significantly different from the more primitive non-vascular taxa. This suggests that the common ancestor of lycophytes, ferns and seed plants adopted a new mechanism or strategy of mitochondrial genome replication and repair that led to a proliferation of repeats and increases in size. Complete sequences of more species, particularly in the lycophytes and ferns, is necessary to add clarity but the ancestor of vascular plants evidently made a transition to increased use of double-strand break repair in their mitochondria, leading to the genomic gymnastics seen today in plants.

The analysis of repeats in the *Brassica* species suggests that mitochondrial genomes can remain relatively static for long periods of time, but can also diverge very rapidly resembling punctuated equilibrium (Gould, S. J. and N. Eldredge 1977) that includes major rearrangements, sequence loss, and gain of sequences of unknown origin. The mechanisms and frequency are unknown, but it suggests that a lineage can experience a burst of genome recombination, breakage and rejoining, dramatically rearranging and altering the mitochondrial genome, as if it had been shattered and rebuilt. These events occur on a time scale that is faster than that of speciation, leading to the high levels of divergence, and a lack of strong correlations with the phylogeny.

Qualitative differences have been described between the repeats shorter and longer than about 1kb (Arrieta-Montiel, M. P. *et al.* 2009; Klein, M. *et al.* 1994; Mower, J. P. *et al.* 2012). The clustering within phylogenetic groups (Figure 2) and the trend towards bimodality (Figure 3) suggest differences between the large repeats and the smaller ones. We suggest that the term "large repeats" be reserved for those ROUS that are involved in genome isomerization. A working definition could be those ROUS larger than 1000bp, but functional analysis may reveal different size cutoffs in different species. Functional analysis can be done by analyzing clones big enough to include the repeats (Klein, M. *et al.* 1994), by long read sequencing (Shearman, J. R. *et al.* 2016) or Southern blotting. Functional analysis of the large repeats is an important step in understanding genome structure and evolution (Guo, W. *et al.* 2016; Guo, W. *et al.* 2017; Sloan, D. B. 2013) and may reveal different size cutoffs between species, which would reveal important differences in the replication and repair machinery and dynamics.

We doubt that there is an adaptive advantage to large size and abundant rearrangements in the genomes of plant mitochondria. We suggest that these are merely correlated traits accompanying the adaptive advantage of a greatly increased reliance on double-strand break repair. DNA repair is critically important because damage is more likely in mitochondria than the nucleus due to the changes in pH and redox potential, and the presence of reactive oxygen species. The strategy followed by animals is to minimize mutational targets by reducing genome size (Lynch, M. *et al.* 2006; Smith, D. R. 2016). However, with multiple copies of

mitochondrial DNA in each cell, an alternative trajectory is to increase the use of template DNA in repair. The template-based accuracy of double-strand break repair is accompanied by the creation of chimeras, rearrangements and duplications when templates are not identical or cannot be found by the repair enzymes. Dramatic expansions, rearrangements and losses, accompanied by low substitution rates in genes is characteristic of flowering plant mitochondria. Selection on gene function maintains the genes, while the expansions and rearrangements must be nearly neutral. Once mitochondria evolved very efficient double-strand break repair, and a mechanism for inducing double-strand breaks at the sites of many types of damage, more primitive mechanisms, such as nucleotide excision repair can and have been lost (Gualberto, J. M. *et al.* 2014; Gualberto, J. M. and K. J. Newton 2017) without obvious evolutionary cost.

The adaptive value of increased and efficient double-strand break repair is probably to avoid mutations in the essential genes of mitochondria, and is possible because of the abundance of double-stranded template molecules in each cell. However this mechanism of repair has an additional correlated trait. There are bacterial species, such as *Deinococcus radiodurans*, that excel at double-strand break repair and can rebuild even significantly fragmented genomes (Krisko, A. and M. Radman 2013). This species is notoriously resistant to ionizing radiation, but the adaptive value of the trait is thought to be desiccation resistance, because dehydration also produces double-strand breaks (Mattimore, V. and J. R. Battista 1996). Radiation resistant bacteria in unrelated phylogenetic groups show more genome rearrangements and

loss of synteny than their radiation sensitive relatives (Repar, J. *et al.* 2017), suggesting that abundant double-strand break repair is the cause of both the resistance to significant double-strand breakage and the loss of synteny. An interesting possibility is that very efficient double-strand break repair in plant mitochondria also confers desiccation resistance as a correlated trait. Because mitochondria are metabolically active immediately upon imbibition of seeds, DNA damage must be repaired very efficiently and rapidly (Paszkiewicz, G. *et al.* 2017). Efficient repair of desiccation-mediated damage in all cellular compartments is a prerequisite to being able to produce seeds or spores for reproduction. It is possible that the DNA repair strategy of plant mitochondria was one of several factors (including desiccation resistance of the nuclear and plastid genomes, presumably by distinct mechanisms) that are beneficial to vascular plants. The evidence of the repeats suggests that the transition to double-strand break repair in mitochondria occurred at approximately the same time as the transition to vascularity in plants, and it may have been one of several traits that enabled their success. In addition, once the life cycles of land plants included periods of desiccation in spores and seeds, double-strand breakage would have increased, accompanied by increases in rearrangements, expansions, and chimeras. The mechanisms of double-strand break repair continue to be important for understanding the evolution of plant mitochondrial genomes.

Acknowledgements

We are grateful to Jeff Mower and Brandi Sigmon for many helpful comments on the manuscript, and Alex Kozik (UC Davis) for beta testing. ACC is grateful to Meric Lieberman and Isabelle Henry (U.C. Davis) for introducing him to Python scripting. ELW thanks Maya Khasin for support and encouragement. This work was supported in part by a grant from the National Science Foundation (MCB-1413152).

Supplemental Materials

Supplemental File S1. Python script ROUSFinder.py

Supplemental File S2. Python script MultipleRepeats.py

Supplemental Table S1. Repeat sizes of all species used in this study. Bins include repeats larger than the size in the header, up to the next bin size.

Supplemental Figure S1. Alignment of the long repeats from *Raphanus sativus*, *Sinapis arvensis*, *Brassica nigra* and *Brassica carinata* with the homologous sequences from *Brassica rapa*, *Brassica juncea*, *Brassica napus polima* and *Brassica oleracea*. Part a is interleaved, and part b is in sequential fasta format.

Supplemental Figure S2. Alignment of the long repeats from *Brassica rapa*, *Brassica juncea*, *Brassica napus polima* and *Brassica oleracea*, with the homologous sequences from *Raphanus sativus*, *Sinapis arvensis*, *Brassica nigra* and *Brassica carinata*. Part a is interleaved, and part b is in sequential fasta format.

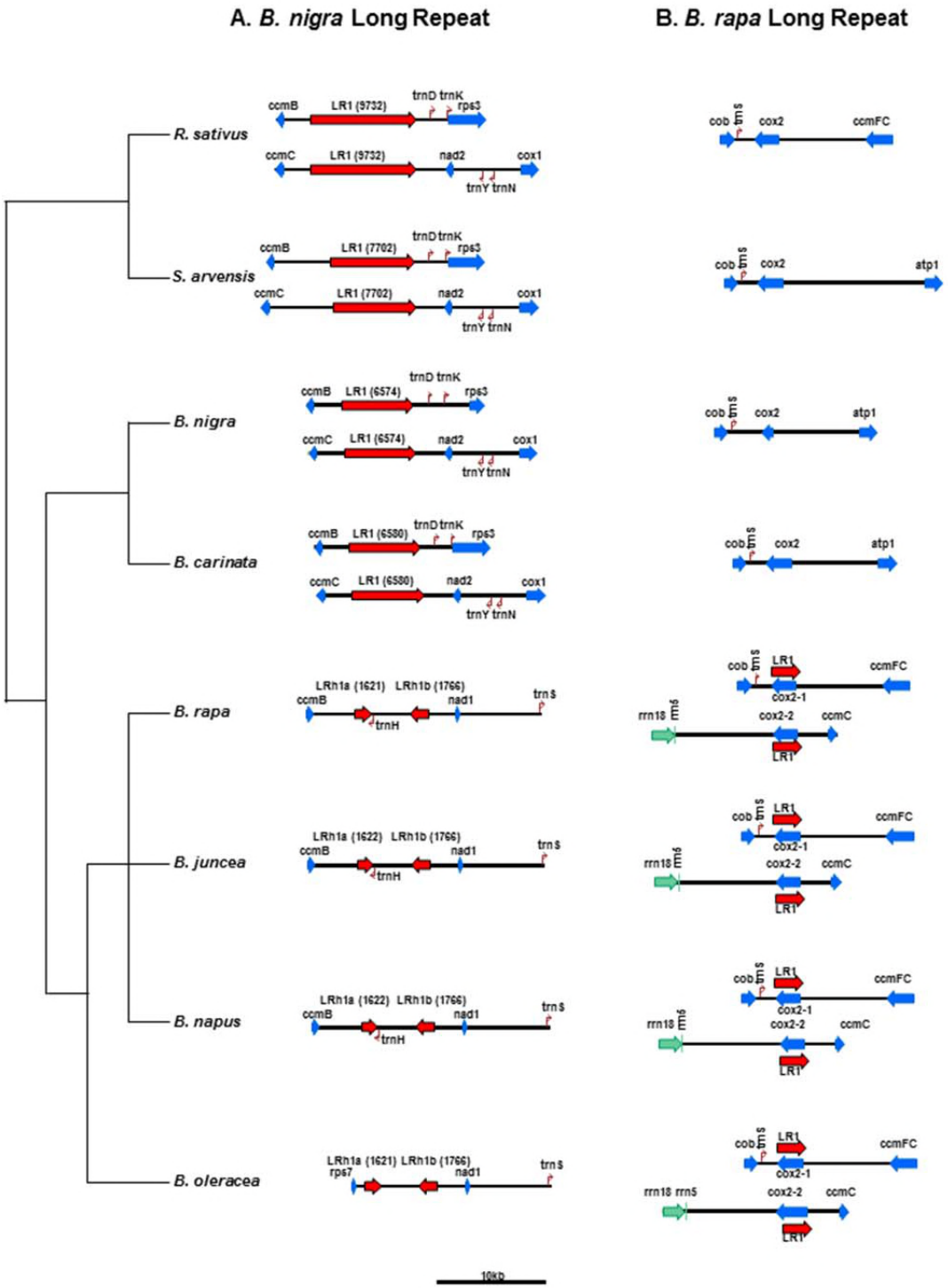
Figure 1. Size distributions of repeats in groups of species. The number of species represented in each group is shown. Headings indicate the bins of repeat sizes and the numbers indicate the fraction of species in that group that have at least one repeat of that size. Heat map color coding is blue for the highest value and yellow for zero.

	#sp	50-199	200-499	500-999	1000-2499	2500-4999	5000-7499	7500-9999	≥ 10000
Chlorophyta	26	0.88	0.19	0.08	0.00	0.00	0.00	0.00	0.00
Prasinophytes	8	0.63	0.00	0.00	0.00	0.00	0.00	0.50	0.25
Anthocerotophyta	2	1.00	0.50	1.00	0.00	0.00	0.00	0.00	0.00
Marchantiophyta	4	1.00	1.00	0.50	0.00	0.00	0.00	0.00	0.00
Bryophytes	23	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Charophyta	8	0.88	0.25	0.00	0.00	0.00	0.00	0.00	0.00
Fern	2	1.00	1.00	1.00	1.00	1.00	0.00	1.00	0.00
Lycophyte	2	1.00	1.00	1.00	1.00	1.00	0.00	0.50	0.00
Gymnosperm	3	1.00	1.00	0.33	0.67	0.33	0.33	0.00	0.00
Angiosperm	58	1.00	0.97	0.43	0.43	0.43	0.36	0.26	0.41

Figure 2. Distribution of repeat sizes among angiosperms. Species are sorted by increasing genome size. The number of repeats of each size class is shown. Blue shading indicates a number greater than zero. Yellow indicates a group of contiguous size ranges that contains zero repeats.

	Genome size	50-99	100-199	200-299	300-399	400-499	500-599	600-699	700-799	800-899	900-999	1000-1499	1500-1999	2000-2499	2500-2999	3000-3499	3500-3999	4000-4499	4500-4999	5000-5499	5500-5999	6000-6499	6500-6999	7000-7499	7500-7999	8000-8499	8500-8999	9000-9499	9500-9999	≥ 10000
Brassica nigra	232407	70	8	5	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0
Sinapis arvensis	240024	59	23	10	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0
Eruca vesicaria	247696	84	35	5	7	0	0	0	0	0	0	0	2	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	2
Silene latifolia	253413	8	4	0	0	0	0	0	0	0	0	35	2	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Schrenkiella parvula	255773	39	22	4	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Raphanus sativus	258426	50	11	7	4	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0
Ipomoea nil	265768	174	43	4	0	0	0	2	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Medicago truncatula	271618	40	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Diplostephium hartwegii	277718	52	10	8	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Daucus carota subsp. sativus	281132	55	21	2	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	5
Helianthus annuus	301004	42	17	7	0	2	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
Allium cepa	316363	67	8	0	0	2	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Spinacia oleracea	329613	181	9	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0
Lagerstroemia indica	333948	1152	21	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Ajuga reptans	352069	18	4	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Ziziphus jujuba	365190	41	12	4	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Arabidopsis thaliana	367808	91	34	14	6	2	4	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	2	0	0	0	0	0	0	0
Citrullus lanatus	379236	62	36	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0
Lotus japonicus	380861	60	33	4	7	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	2
Glycine max	402558	79	127	11	3	0	0	0	0	0	0	0	9	0	7	2	4	0	2	0	3	2	4	0	0	0	0	0	0	0
Batis maritima	403930	41	23	2	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Vigna angularis	404466	27	50	6	4	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Cannabis sativa	415602	76	15	4	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Millettia pinnata	425718	54	9	2	0	0	0	3	0	0	0	0	0	2	0	0	2	0	0	0	0	2	0	0	0	0	0	0	0	2
Nicotiana tabacum	430597	83	34	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	5
Butomus umbellatus	450826	54	28	4	4	0	0	0	0	0	3	0	3	2	5	0	0	0	2	0	0	2	0	0	0	0	0	0	0	0
Triticum aestivum	452528	87	29	6	3	0	0	0	0	0	0	2	3	4	0	0	2	3	0	2	0	4	0	2	0	0	0	0	0	0
Sorghum bicolor	468628	75	17	7	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	4
Aegilops speltoides	476091	64	24	4	3	3	0	0	0	0	0	0	2	0	0	4	0	0	4	4	3	0	0	2	0	0	2	0	0	0
Carica papaya	476890	46	17	4	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2
Oryza sativa Indica	491515	152	17	3	0	2	0	0	0	0	0	3	2	0	0	0	0	3	0	0	0	2	0	2	0	0	2	0	2	6
Castilleja paramensis	495499	33	13	0	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0
Salvia miltiorrhiza	499236	77	18	4	2	0	0	2	0	0	0	0	0	2	0	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0
Beta vulgaris	501020	94	18	3	13	3	0	0	0	3	0	3	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	4
Hyoscyamus niger	501401	1435	138	5	10	5	2	0	0	0	0	0	0	2	2	0	0	0	0	0	0	0	0	2	0	2	0	0	2	2
Ricinus communis	502773	204	4	4	0	0	0	0	2	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0
Boea hygrometrica	510519	41	8	4	0	0	0	0	0	2	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Capsicum annuum	511530	91	21	0	2	0	0	0	0	0	0	3	2	0	0	0	0	0	0	0	0	0	0	2	2	0	0	0	0	2
Hordeum vulgare	525599	76	20	5	7	4	2	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	2	0	2	4
Mimulus guttatus	525671	41	12	9	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	2	2
Heuchera parviflora	542954	69	34	7	4	0	3	0	0	0	0	3	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	2
Rhazya stricta	548608	53	9	5	0	0	3	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4
Liriodendron tulipifera	553721	190	35	6	0	4	0	4	0	2	0	2	2	0	0	2	0	0	0	2	0	0	0	0	0	0	0	0	0	2
Viscum album	565432	286	34	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Zea mays strain NB	569630	77	15	6	0	2	2	2	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	6
Hibiscus cannabinus	569915	108	32	6	5	5	2	0	0	2	0	0	4	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0
Betula pendula	581505	444	44	2	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Vicia faba	588000	58	55	2	6	5	0	3	0	0	0	0	5	3	2	2	2	0	0	0	2	0	0	2	2	0	0	2	0	11
Salix suchowensis	644437	54	12	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	2	0	0
Gossypium hirsutum	668584	81	70	38	25	11	2	2	0	0	0	2	0	0	0	0	2	0	2	0	0	2	0	2	0	0	2	0	0	4
Lolium perenne	678580	101	33	4	2	12	8	8	4	2	0	14	6	4	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	6
Cocos nucifera	678653	110	38	19	0	0	0	0	0	0	0	4	2	4	2	0	0	2	0	0	0	0	0	3	3	0	0	0	0	4
Tripsacum dactyloides	704100	8356	1261	7	5	0	0	0	0	0	0	0	2	0	3	0	0	0	0	0	0	0	0	0	0	0	2	0	0	4
Phoenix dactylifera	715001	64	18	0	4	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Geranium maderense	737091	305	55	17	6	2	0	0	0	2	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Vitis vinifera	773279	219	35	6	4	2	2	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Populus tremula	783442	78	22	4	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Corchorus capsularis	1999602	665	125	8	2	0	2	2	0	0	0	0	0	2	2	0	0	0	0	0	0	0	0	2	0	0	0	0	0	4

Figure 3. Long repeats in the Brassicales. A phylogenetic tree is shown at left, derived from Grewe *et al* (Grewe, F. *et al.* 2014). In part A are the regions surrounding the long repeat in *R. sativus*, *S. arvensis*, *B. nigra* and *B. carinata*. The homologus sequence from *B. rapa*, *B. napus*, *B. juncea* and *B. oleracea* is also shown. Part A shows the regions surrounding the long repeat in *B. rapa*, *B. napus*, *B. juncea* and *B. oleracea*, and the homologous region in the prior four species. Branch lengths in the tree are not to scale. The sequences are depicted at the scale shown in the figure.



1

2

1 Table 1. List of species and mitochondrial DNA accession numbers

Genus species	Group	Subgroup	Accession Number
Auxenochlorella protothecoides	Chlorophytes	Chlorophyta	KC843974.1
Botryococcus braunii	Chlorophytes	Chlorophyta	LT545992.1
Bracteacoccus aerius	Chlorophytes	Chlorophyta	KJ806265.1
Chlamydomonas eustigma	Chlorophytes	Chlorophyta	BEGY01000520.1
Chlamydomonas reinhardtii	Chlorophytes	Chlorophyta	EU306617.1
Chlorella sp. ArM0029B	Chlorophytes	Chlorophyta	KF554428.1
Dunaliella salina strain GN	Chlorophytes	Chlorophyta	KX641169.1
Eudorina sp.	Chlorophytes	Chlorophyta	KY442294.1
Gloeotilopsis planctonica	Chlorophytes	Chlorophyta	KX306823.1
Gloeotilopsis sarcinoidea	Chlorophytes	Chlorophyta	KX306822.1
Hariotina sp. MMOGRB0030F	Chlorophytes	Chlorophyta	KU145405.1
Kirchneriella aperta	Chlorophytes	Chlorophyta	NC_024759.1
Lobosphaera incisa	Chlorophytes	Chlorophyta	KP902678.1
Microspora stagnorum	Chlorophytes	Chlorophyta	KF060942.1
Mychonastes homosphaera	Chlorophytes	Chlorophyta	NC_024760.1
Neochloris aquatica	Chlorophytes	Chlorophyta	NC_024761.1
Oltmannsiellopsis viridis	Chlorophytes	Chlorophyta	DQ365900.1
Ostreococcus tauri	Chlorophytes	Prasinophytes	CR954200.2
Ourococcus multisporus	Chlorophytes	Chlorophyta	NC_024762.1
Polytoma uvella	Chlorophytes	Chlorophyta	NC_026572.1
Prototheca wickerhamii	Chlorophytes	Chlorophyta	U02970.1
Pseudendoclonium akinetum	Chlorophytes	Chlorophyta	AY359242.1
Tetradesmus obliquus	Chlorophytes	Chlorophyta	CM007918.1
Trebouxiophyceae sp. MX-AZ01	Chlorophytes	Chlorophyta	JX315601.1
Ulva flexuosa	Chlorophytes	Chlorophyta	KX455878.1
Ulva linza	Chlorophytes	Chlorophyta	NC_029701.1
Nephroselmis olivacea	Chlorophytes	Nephroselmidophyceae	AF110138.1
Bathycoccus prasinos	Chlorophytes	Prasinophytes	NC_023273.1
Cymbomonas tetramitiformis	Chlorophytes	Prasinophytes	KX013548.1
Micromonas sp. RCC299	Chlorophytes	Prasinophytes	FJ859351.1
Monomastix sp. OKE-1	Chlorophytes	Prasinophytes	KF060939.1
Prasinoderma coloniale	Chlorophytes	Prasinophytes	KF387569.1
Pycnococcus provasolii	Chlorophytes	Prasinophytes	GQ497137.1
Pyramimonas parkeae	Chlorophytes	Prasinophytes	KX013547.1
Megaceros aenigmaticus	Anthocerotophyta	Anthocerotophyta	EU660574.1
Phaeoceros laevis	Anthocerotophyta	Anthocerotophyta	GQ376531.1
Aneura pinguis	Marchantiophyta	Marchantiophyta	KP728938.1
Marchantia polymorpha	Marchantiophyta	Marchantiophyta	NC_001660.1
Pleurozia purpurea	Marchantiophyta	Marchantiophyta	NC_013444.1
Treubia lacunosa	Marchantiophyta	Marchantiophyta	JF973315.1
Anomodon attenuatus	Bryophytes	Bryophytes	JX402749.1
Atrichum angustatum	Bryophytes	Bryophytes	KC784956.1
Bartramia pomiformis	Bryophytes	Bryophytes	KC784955.1
Brachythecium rivulare	Bryophytes	Bryophytes	KR732319.1
Bucklandiella orthotrichacea	Bryophytes	Bryophytes	KP742835.1
Buxbaumia aphylla	Bryophytes	Bryophytes	KC784954.1
Climacium americanum	Bryophytes	Bryophytes	KC784950.1
Codriophorus aciculare	Bryophytes	Bryophytes	KP453983.1
Funaria hygrometrica	Bryophytes	Bryophytes	KC784959.1
Hypnum imponens	Bryophytes	Bryophytes	KC784951.1
Nyholmiella gymnostoma	Bryophytes	Bryophytes	KX578030.1
Orthotrichum callistomum	Bryophytes	Bryophytes	KX578029.1
Oxystegus tenuirostris	Bryophytes	Bryophytes	KT326816.1
Physcomitrella patens	Bryophytes	Bryophytes	NC_007945.1
Ptychomnion cygnisetum	Bryophytes	Bryophytes	KC784949.1
Sanionia uncinata	Bryophytes	Bryophytes	KP984757.1

Sphagnum palustre	Bryophytes	Bryophytes	KC784957.1
Stoneobryum bunyaense	Bryophytes	Bryophytes	KX578031.1
Syntrichia filaris	Bryophytes	Bryophytes	KP984758.1
Tetraphis pellucida	Bryophytes	Bryophytes	KC784953.1
Tetraplodon fuegianus	Bryophytes	Bryophytes	KT373818.1
Ulota phyllantha	Bryophytes	Bryophytes	KX578033.1
Zygodon viridissimus	Bryophytes	Bryophytes	KX711975.1
Chaetosphaeridium globosum	Streptophyta	Charophyta	AF494279.1
Chara vulgaris	Streptophyta	Charophyta	AY267353.1
Chlorokybus atmophyticus	Streptophyta	Charophyta	NC_009630.1
Closterium baillyanum	Streptophyta	Charophyta	NC_022860.1
Entransia fimbriata	Streptophyta	Charophyta	KF060941.1
Klebsormidium flaccidum	Streptophyta	Charophyta	KP165386.1
Nitella hyalina	Streptophyta	Charophyta	JF810595.1
Roya obtusa	Streptophyta	Charophyta	KF060943.1
Ophioglossum californicum	Tracheophyta	Fern	NC_030900.1
Psilotum nudum	Tracheophyta	Fern	NC_030952.1
Huperzia squarrosa	Tracheophyta	Lycophyte	NC_017755.1
Selaginella moellendorffii	Tracheophyta	Lycophyte	GL377739.1
Ginkgo biloba	Tracheophyta	Gymnosperm	KM672373.1
Welwitschia mirabilis	Tracheophyta	Gymnosperm	NC_029130.1
Cycas taitugensis	Tracheophyta	Gymnosperm	AP009381.1
Aegilops speltoides	Tracheophyta	Angiosperm	AP013107.1
Ajuga reptans	Tracheophyta	Angiosperm	KF709392.1
Allium cepa	Tracheophyta	Angiosperm	KU318712.1
Arabidopsis thaliana	Tracheophyta	Angiosperm	BK010421
Batis maritima	Tracheophyta	Angiosperm	KJ820684.1
Beta vulgaris	Tracheophyta	Angiosperm	BA000024.1
Betula pendula	Tracheophyta	Angiosperm	LT855379.1
Boea hygrometrica	Tracheophyta	Angiosperm	JN107812.1
Brassica nigra	Tracheophyta	Angiosperm	KP030753.1
Butomus umbellatus	Tracheophyta	Angiosperm	KC208619.1
Cannabis sativa	Tracheophyta	Angiosperm	KU310670.1
Capsicum annuum	Tracheophyta	Angiosperm	KJ865410.1
Carica papaya	Tracheophyta	Angiosperm	EU431224.1
Castilleja paramensis	Tracheophyta	Angiosperm	KT959112.1
Citrullus lanatus	Tracheophyta	Angiosperm	GQ856147.1
Cocos nucifera	Tracheophyta	Angiosperm	KX028885.1
Corchorus capsularis	Tracheophyta	Angiosperm	KT894204.1
Daucus carota subsp. sativus	Tracheophyta	Angiosperm	JQ248574.1
Diplostephium hartwegii	Tracheophyta	Angiosperm	KX063855.1
Eruca vesicaria	Tracheophyta	Angiosperm	KF442616.1
Geranium maderense	Tracheophyta	Angiosperm	NC_027000.1
Glycine max	Tracheophyta	Angiosperm	JX463295.1
Gossypium hirsutum	Tracheophyta	Angiosperm	NC_027406.1
Helianthus annuus	Tracheophyta	Angiosperm	CM007908.1
Heuchera parviflora	Tracheophyta	Angiosperm	KR559021.1
Hibiscus cannabinus	Tracheophyta	Angiosperm	MF163174.1
Hordeum vulgare	Tracheophyta	Angiosperm	AP017300.1
Hyoscyamus niger	Tracheophyta	Angiosperm	KM207685.1
Ipomoea nil	Tracheophyta	Angiosperm	AP017303.1
Lagerstroemia indica	Tracheophyta	Angiosperm	KX641464.1
Liriodendron tulipifera	Tracheophyta	Angiosperm	NC_021152.1
Lolium perenne	Tracheophyta	Angiosperm	JX999996.1
Lotus japonicus	Tracheophyta	Angiosperm	JN872551.2
Medicago truncatula	Tracheophyta	Angiosperm	KT971339.1
Millettia pinnata	Tracheophyta	Angiosperm	JN872550.1
Mimulus guttatus	Tracheophyta	Angiosperm	JN098455.1
Nicotiana tabacum	Tracheophyta	Angiosperm	BA000042.1

Oryza sativa Indica	Tracheophyta	Angiosperm	NC_007886.1
Phoenix dactylifera	Tracheophyta	Angiosperm	JN375330.1
Populus tremula	Tracheophyta	Angiosperm	KT337313.1
Raphanus sativus	Tracheophyta	Angiosperm	NC_018551.1
Rhazya stricta	Tracheophyta	Angiosperm	KJ485850.1
Ricinus communis	Tracheophyta	Angiosperm	HQ874649.1
Salix suchowensis	Tracheophyta	Angiosperm	KU056812.1
Salvia miltiorrhiza	Tracheophyta	Angiosperm	KF177345.1
Schrenkiella parvula	Tracheophyta	Angiosperm	KT988071.2
Silene latifolia	Tracheophyta	Angiosperm	HM562727.1
Sinapis arvensis	Tracheophyta	Angiosperm	KM851044.1
Sorghum bicolor	Tracheophyta	Angiosperm	DQ984518.1
Spinacia oleracea	Tracheophyta	Angiosperm	KY768855.1
Tripsacum dactyloides	Tracheophyta	Angiosperm	NC_008362.1
Triticum aestivum	Tracheophyta	Angiosperm	AP008982.1
Vicia faba	Tracheophyta	Angiosperm	KC189947.1
Vigna angularis	Tracheophyta	Angiosperm	AP012599.1
Viscum album	Tracheophyta	Angiosperm	NC_029039.1
Vitis vinifera complete	Tracheophyta	Angiosperm	FM179380.1
Zea mays strain NB	Tracheophyta	Angiosperm	AY506529.1
Ziziphus jujuba	Tracheophyta	Angiosperm	KU187967.1

1
2

Literature Cited

Abdelnoor, R. V., R. Yule, A. Elo, A. C. Christensen, G. Meyer-Gauen *et al.*, 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.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers and D. J. Lipman, 1990 Basic local alignment search tool. J Mol Biol 215**:** 403-410.

Alverson, A. J., D. W. Rice, S. Dickinson, K. Barry and J. D. Palmer, 2011a Origins and recombination of the bacterial-sized multichromosomal mitochondrial genome of cucumber. Plant Cell 23**:** 2499-2513.

Alverson, A. J., X. Wei, D. W. Rice, D. B. Stern, K. Barry *et al.*, 2010 Insights into the evolution of mitochondrial genome size from complete sequences of Citrullus lanatus and Cucurbita pepo (Cucurbitaceae). Mol Biol Evol 27**:** 1436-1448.

Alverson, A. J., S. Zhuo, D. W. Rice, D. B. Sloan and J. D. Palmer, 2011b The mitochondrial genome of the legume Vigna radiata and the analysis of recombination across short mitochondrial repeats. PLoS One 6**:** e16404.

Arrieta-Montiel, M. P., V. Shedge, J. Davila, A. C. Christensen and S. A. Mackenzie, 2009 Diversity of the Arabidopsis Mitochondrial Genome Occurs via Nuclear-Controlled Recombination Activity. Genetics 183**:** 1261-1268.

Bendich, A. J., 2004 Circular Chloroplast Chromosomes: The Grand Illusion. The Plant Cell 16**:** 1661-1666.

Chang, S., T. Yang, T. Du, Y. Huang, J. Chen *et al.*, 2011 Mitochondrial genome sequencing helps show the evolutionary mechanism of mitochondrial genome formation in Brassica. BMC Genomics 12**:** 497.

Chaw, S. M., A. C. Shih, D. Wang, Y. W. Wu, S. M. Liu *et al.*, 2008 The mitochondrial genome of the gymnosperm Cycas taitungensis contains a novel family of short interspersed elements, Bpu sequences, and abundant RNA editing sites. Mol Biol Evol 25**:** 603-615.

Cheng, F., J. Liang, C. Cai, X. Cai, J. Wu *et al.*, 2017 Genome sequencing supports a multi-vertex model for Brassiceae species. Curr Opin Plant Biol 36**:** 79-87.

Christensen, A. C., 2014 Genes and Junk in Plant Mitochondria—Repair Mechanisms and Selection. Genome Biol Evol 6**:** 1448-1453.

Christensen, A. C., 2018 Mitochondrial DNA Repair and Genome Evolution, pp. 11-31 in *Annual Plant Reviews, 2nd Edition, Plant Mitochondria*, edited by D. C. Logan. Wiley-Blackwell, New York, NY, USA.

Davila, J. I., M. P. Arrieta-Montiel, Y. Wamboldt, J. Cao, J. Hagmann *et al.*, 2011 Double-strand break repair processes drive evolution of the mitochondrial genome in Arabidopsis. BMC Biol 9**:** 64.

Drouin, G., H. Daoud and J. Xia, 2008 Relative rates of synonymous substitutions in the mitochondrial, chloroplast and nuclear genomes of seed plants. Mol Phylogenet Evol 49**:** 827-831.

Folkerts, O., and M. R. Hanson, 1989 Three copies of a single recombination repeat occur on the 443 kb master circle of the Petunia hybrida 3704 mitochondrial genome. Nucleic Acids Res 17**:** 7345-7357.

Forner, J., B. Weber, C. Wietholter, R. C. Meyer and S. Binder, 2005 Distant sequences determine 5' end formation of cox3 transcripts in Arabidopsis thaliana ecotype C24. Nucleic Acids Res. 33**:** 4673-4682.

Franzke, A., M. A. Lysak, I. A. Al-Shehbaz, M. A. Koch and K. Mummenhoff, 2011 Cabbage family affairs: the evolutionary history of Brassicaceae. Trends Plant Sci 16**:** 108-116.

Gould, S. J., and N. Eldredge, 1977 Punctuated Equilibria: The Tempo and Mode of Evolution Reconsidered. Paleobiology 3**:** 115-151.

Grewe, F., P. P. Edger, I. Keren, L. Sultan, J. C. Pires *et al.*, 2014 Comparative analysis of 11 Brassicales mitochondrial genomes and the mitochondrial transcriptome of Brassica oleracea. Mitochondrion.

Grewe, F., P. Viehoever, B. Weisshaar and V. Knoop, 2009 A trans-splicing group I intron and tRNA-hyperediting in the mitochondrial genome of the lycophyte Isoetes engelmannii. Nucleic Acids Res 37**:** 5093-5104.

Gualberto, J. M., D. Mileshina, C. Wallet, A. K. Niazi, F. Weber-Lotfi *et al.*, 2014 The plant mitochondrial genome: dynamics and maintenance. Biochimie 100**:** 107-120.

Gualberto, J. M., and K. J. Newton, 2017 Plant Mitochondrial Genomes: Dynamics and Mechanisms of Mutation. Annu Rev Plant Biol.

Guo, W., F. Grewe, W. Fan, G. J. Young, V. Knoop *et al.*, 2016 Ginkgo and Welwitschia Mitogenomes Reveal Extreme Contrasts in Gymnosperm Mitochondrial Evolution. Mol Biol Evol 33**:** 1448-1460.

Guo, W., A. Zhu, W. Fan and J. P. Mower, 2017 Complete mitochondrial genomes from the ferns Ophioglossum californicum and Psilotum nudum are highly repetitive with the largest organellar introns. New Phytol 213**:** 391-403.

Hecht, J., F. Grewe and V. Knoop, 2011 Extreme RNA editing in coding islands and abundant microsatellites in repeat sequences of Selaginella moellendorffii mitochondria: the root of frequent plant mtDNA recombination in early tracheophytes. Genome Biol Evol 3**:** 344-358.

Klein, M., U. Eckert-Ossenkopp, I. Schmiedeberg, P. Brandt, M. Unseld *et al.*, 1994 Physical mapping of the mitochondrial genome of *Arabidopsis thaliana* by cosmid and YAC clones. Plant J. 6**:** 447-455.

Krisko, A., and M. Radman, 2013 Biology of extreme radiation resistance: the way of *Deinococcus radiodurans*. Cold Spring Harb Perspect Biol 5**:** a012765.

Kurtz, S., and C. Schleiermacher, 1999 REPuter: fast computation of maximal repeats in complete genomes. Bioinformatics 15**:** 426-427.

Liu, Y., R. Medina and B. Goffinet, 2014 350 my of mitochondrial genome stasis in mosses, an early land plant lineage. Mol Biol Evol 31**:** 2586-2591.

Lynch, M., B. Koskella and S. Schaack, 2006 Mutation Pressure and the Evolution of Organelle Genomic Architecture. Science 311**:** 1727-1730.

Mattimore, V., and J. R. Battista, 1996 Radioresistance of Deinococcus radiodurans: functions necessary to survive ionizing radiation are also necessary to survive prolonged desiccation. J Bacteriol 178**:** 633-637.

Miller-Messmer, M., K. Kuhn, M. Bichara, M. Le Ret, P. Imbault *et al.*, 2012 RecA-dependent DNA repair results in increased heteroplasmy of the Arabidopsis mitochondrial genome. Plant Physiol 159**:** 211-226.

Mower, J. P., A. L. Case, E. R. Floro and J. H. Willis, 2012 Evidence against equimolarity of large repeat arrangements and a predominant master circle

structure of the mitochondrial genome from a monkeyflower (Mimulus guttatus) lineage with cryptic CMS. Genome Biol Evol 4**:** 670-686.

Palmer, J. D., 1988 Intraspecific variation and multicircularity in Brassica mitochondrial DNAs. Genetics 118**:** 341-351.

Palmer, J. D., and L. A. Herbon, 1986 Tricircular mitochondrial genomes of Brassica and Raphanus: reversal of repeat configurations by inversion. Nucleic Acids Res 14**:** 9755-9764.

Palmer, J. D., and L. A. Herbon, 1988 Plant mitochondrial DNA evolves rapidly in structure, but slowly in sequence. J Mol Evol 28**:** 87-97.

Palmer, J. D., and C. R. Shields, 1984 Tripartite structure of the Brassica campestris mitochondrial genome. Nature 307**:** 437.

Park, S., F. Grewe, A. Zhu, T. A. Ruhlman, J. Sabir *et al.*, 2015 Dynamic evolution of Geranium mitochondrial genomes through multiple horizontal and intracellular gene transfers. New Phytol 208**:** 570-583.

Parkinson, C. L., J. P. Mower, Y. L. Qiu, A. J. Shirk, K. Song *et al.*, 2005 Multiple major increases and decreases in mitochondrial substitution rates in the plant family Geraniaceae. BMC Evol Biol 5**:** 73.

Paszkiewicz, G., J. M. Gualberto, A. Benamar, D. Macherel and D. C. Logan, 2017 Arabidopsis Seed Mitochondria Are Bioenergetically Active Immediately upon Imbibition and Specialize via Biogenesis in Preparation for Autotrophic Growth. Plant Cell 29**:** 109-128.

Repar, J., F. Supek, T. Klanjscek, T. Warnecke, K. Zahradka *et al.*, 2017 Elevated Rate of Genome Rearrangements in Radiation-Resistant Bacteria. Genetics 205**:** 1677-1689.

Richardson, A. O., D. W. Rice, G. J. Young, A. J. Alverson and J. D. Palmer, 2013 The "fossilized" mitochondrial genome of *Liriodendron tulipifera*: ancestral gene content and order, ancestral editing sites, and extraordinarily low mutation rate. BMC Biol 11**:** 29.

Schuster, W., and A. Brennicke, 1994 The Plant Mitochondrial Genome: Physical Structure, Information Content, RNA Editing, and Gene Migration to the Nucleus. Annual Review of Plant Physiology and Plant Molecular Biology 45**:** 61-78.

Shearman, J. R., C. Sonthirod, C. Naktang, W. Pootakham, T. Yoocha *et al.*, 2016 The two chromosomes of the mitochondrial genome of a sugarcane cultivar: assembly and recombination analysis using long PacBio reads. Sci Rep 6**:** 31533.

Shedge, V., M. Arrieta-Montiel, A. C. Christensen and S. A. Mackenzie, 2007 Plant mitochondrial recombination surveillance requires unusual RecA and MutS homologs. Plant Cell 19**:** 1251-1264.

Siculella, L., F. Damiano, M. R. Cortese, E. Dassisti, G. Rainaldi *et al.*, 2001 Gene content and organization of the oat mitochondrial genome. Theoretical and Applied Genetics 103**:** 359-365.

Sloan, D. B., 2013 One ring to rule them all? Genome sequencing provides new insights into the 'master circle' model of plant mitochondrial DNA structure. New Phytol 200: 978-985.

Sloan, D. B., A. J. Alverson, J. P. Chuckalovcak, M. Wu, D. E. McCauley *et al.*, 2012 Rapid evolution of enormous, multichromosomal genomes in flowering plant mitochondria with exceptionally high mutation rates. PLoS Biol 10: e1001241.

Sloan, D. B., A. J. Alverson, H. Storchova, J. D. Palmer and D. R. Taylor, 2010 Extensive loss of translational genes in the structurally dynamic mitochondrial genome of the angiosperm Silene latifolia. BMC Evol Biol 10: 274.

Sloan, D. B., Z. Wu and J. Sharbrough, 2018 Correction of Persistent Errors in Arabidopsis Reference Mitochondrial Genomes. The Plant Cell 30: 525-527.

Smith, D. R., 2016 The mutational hazard hypothesis of organelle genome evolution: 10 years on. Mol Ecol 25: 3769-3775.

Smith, D. R., R. W. Lee, J. C. Cushman, J. K. Magnuson, D. Tran *et al.*, 2010 The Dunaliella salina organelle genomes: large sequences, inflated with intronic and intergenic DNA. BMC Plant Biol 10: 83.

Stern, D. B., and J. D. Palmer, 1986 Tripartite mitochondrial genome of spinach: physical structure, mitochondrial gene mapping, and locations of transposed chloroplast DNA sequences. Nucleic Acids Res 14: 5651-5666.

Wallet, C., M. Le Ret, M. Bergdoll, M. Bichara, A. Dietrich *et al.*, 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.

Ward, B. L., R. S. Anderson and A. J. Bendich, 1981 The mitochondrial genome is large and variable in a family of plants (cucurbitaceae). Cell 25: 793-803.

Wolfe, K., W. Li and P. Sharp, 1987 Rates of nucleotide substitution vary greatly among plant mitochondrial,chloroplast and nuclear DNAs. Proc. Natl. Acad. Sci. U S A 84: 9054-9058.