

12-18-2018

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

Emily L. Wynn

*University of Nebraska-Lincoln*, emilywynn6@gmail.com

Alan C. Christensen

*University of Nebraska - Lincoln*, achristensen2@unl.edu

Follow this and additional works at: <http://digitalcommons.unl.edu/bioscifacpub>



Part of the [Biology Commons](#)

---

Wynn, Emily L. and Christensen, Alan C., "Repeats of Unusual Size in Plant Mitochondrial Genomes: Identification, Incidence and Evolution" (2018). *Faculty Publications in the Biological Sciences*. 697.

<http://digitalcommons.unl.edu/bioscifacpub/697>

This Article is brought to you for free and open access by the Papers in the Biological Sciences at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Faculty Publications in the Biological Sciences by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

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

Emily L. Wynn and Alan C. Christensen<sup>1</sup>

School of Biological Sciences, University of Nebraska, Lincoln, Nebraska 68588-0666

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

**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 abundant non-tandem repeats often including pairs of large repeats which cause isomerization of the genome by recombination, and numerous repeats of up to several hundred base pairs that recombine only when the genome is stressed by DNA damaging agents or mutations in DNA repair pathway genes. Early work on mitochondrial genomes led to the suggestion that repeats in the size range from several hundred to a few thousand base pair are underrepresented. The repeats themselves are not well-conserved between species, and are not always annotated in mitochondrial sequence assemblies. We systematically identified and compared these repeats, which are important clues to mechanisms of DNA maintenance in mitochondria. We developed a tool to find and curate non-tandem repeats larger than 50bp and analyzed the complete mitochondrial sequences from 157 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. We suggest an adaptive explanation for the existence of the repeats and their evolution.

## KEYWORDS

plant  
mitochondrial  
genomes  
repeated  
sequence  
genome  
rearrangement  
organelle  
genome  
evolution

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 non-tandem 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; Shedje, V. *et al.* 2007; Wallet, C. *et al.* 2015). Understanding the repeats is critical to 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 (Alverson, A. J. *et al.* 2011b; Andre, C. *et al.* 1992; Arrieta-Montiel, M. P. *et al.* 2009; Davila, J. I. *et al.* 2011; Folkerts, O. and M. R. Hanson 1989; Sugiyama, Y. *et al.* 2005).

Copyright © 2019 Wynn, Christensen

doi: <https://doi.org/10.1534/g3.118.200948>

Manuscript received July 20, 2018; accepted for publication December 14, 2018; published Early Online December 18, 2018.

This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Supplemental material available at Figshare: <https://doi.org/10.25387/g3.7425680>.

<sup>1</sup>Author for Correspondence: School of Biological Sciences, University of Nebraska – Lincoln, Lincoln, NE 68588-0666, E-mail: [achristensen2@unl.edu](mailto:achristensen2@unl.edu)

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 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? 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 plant mitochondrial 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

Table 1 lists the mitochondrial genome sequences that 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 fujiiwase*; 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 -20. The script then concatenates the two output files and the full length self-identity is deleted. Alignments are then sorted and compared to identify and remove duplicate repeats, and an output file containing each distinct repeat in fasta format is created. The sorting by size allows the script to automate the curation of the repeats by comparing the query start and end coordinates of each identified repeated sequence with the subject start and end coordinates of repeated sequences of the same size. When there are more than two copies of a repeat, BLAST does not report every pairwise hit, so the output file of FASTA-formatted sequences of repeats is then used as a query with the entire 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.

Prior work identifying repeats, especially in *Arabidopsis thaliana* (Arrieta-Montiel, M. P. *et al.* 2009) showed that although BLAST is very useful, it has some characteristics that make it difficult to automate curation of the identified matches of non-tandem DNA sequences. For example, if there are a number of mismatches in a repeat, sometimes BLAST will identify subsets of the repeat sequence, or not give the same alignments when two imperfect repeats are used as queries of the entire genome. When there are three or more copies of a repeat, BLAST will also not identify every possible pairwise alignment, giving a subset instead. When examining the repeats in a single species, these problems can be solved by additional manual curation and inspection of the species. However, for automated curation of sequences from multiple species, some compromises have to be made. The simplest way to curate the repeated sequences is to ensure that the sizes of each repeat in a pair are the same. This ensures that repeats can be matched with each other by examining the coordinates of each copy to find all copies of the repeated sequence. In *Arabidopsis thaliana*, an ungapped blastn search with a match reward of +1 and a mismatch penalty of -18 or lower ensured that different copies of the repeats were the same length, allowing automated curation. We therefore set the penalty parameter to -20 to make the automated curation reliable and fast. In order to more carefully examine the repeats in a single species, the script ROUSFinder2.py (Supplemental File S3) allows the user to set the match reward and mismatch penalty parameters on the command line.

After an initial analysis of sequences available in early 2018, we added additional species to the data in late 2018. These additional species are indicated in Table 1 by an asterisk. These include two hornworts, two liverworts, 3 bryophytes and 14 angiosperms. These new species do not change the patterns or conclusions compared to the earlier analysis, providing additional validation of the use of BLAST and the curation methods described.

### 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: <https://doi.org/10.25387/g3.7425680>.

## 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, and match/mismatch scores of +1/-20. 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. As described in the Methods, they were also chosen to enable automated curation of the repeats that are found in the first

■ Table 1 List of species and mitochondrial DNA accession numbers

Species	Group	Subgroup	Accession
<i>Auxenochlorella protothecoides</i>	Chlorophytes	Chlorophyta	KC843974.1
<i>Botryococcus braunii</i>	Chlorophytes	Chlorophyta	LT545992.1
<i>Bracteacoccus aerius</i>	Chlorophytes	Chlorophyta	KJ806265.1
<i>Chlamydomonas eustigma</i>	Chlorophytes	Chlorophyta	BEGY01000520.1
<i>Chlamydomonas reinhardtii</i>	Chlorophytes	Chlorophyta	EU306617.1
<i>Chlorella</i> sp. ArM0029B	Chlorophytes	Chlorophyta	KF554428.1
<i>Dunaliella salina</i> strain GN	Chlorophytes	Chlorophyta	KX641169.1
<i>Eudorina</i> sp.	Chlorophytes	Chlorophyta	KY442294.1
<i>Gloeotilopsis planctonica</i>	Chlorophytes	Chlorophyta	KX306823.1
<i>Gloeotilopsis sarcinoidea</i>	Chlorophytes	Chlorophyta	KX306822.1
<i>Hariotina</i> sp. MMOGRB0030F	Chlorophytes	Chlorophyta	KU145405.1
<i>Kirchneriella aperta</i>	Chlorophytes	Chlorophyta	NC_024759.1
<i>Lobosphaera incisa</i>	Chlorophytes	Chlorophyta	KP902678.1
<i>Microspora stagnorum</i>	Chlorophytes	Chlorophyta	KF060942.1
<i>Mychonastes homosphaera</i>	Chlorophytes	Chlorophyta	NC_024760.1
<i>Neochloris aquatica</i>	Chlorophytes	Chlorophyta	NC_024761.1
<i>Nephroselmis olivacea</i>	Chlorophytes	Nephroselmidophyceae	AF110138.1
<i>Oltmannsiellopsis viridis</i>	Chlorophytes	Chlorophyta	DQ365900.1
<i>Ourococcus multisporus</i>	Chlorophytes	Chlorophyta	NC_024762.1
<i>Polytoma uvella</i>	Chlorophytes	Chlorophyta	NC_026572.1
<i>Prototheca wickerhamii</i>	Chlorophytes	Chlorophyta	U02970.1
<i>Pseudendoclonium akinetum</i>	Chlorophytes	Chlorophyta	AY359242.1
<i>Tetrademus obliquus</i>	Chlorophytes	Chlorophyta	CM007918.1
<i>Trebouxiophyceae</i> sp. MX-AZ01	Chlorophytes	Chlorophyta	JX315601.1
<i>Ulva flexuosa</i>	Chlorophytes	Chlorophyta	KX455878.1
<i>Ulva linza</i>	Chlorophytes	Chlorophyta	NC_029701.1
<i>Bathycoccus prasinos</i>	Chlorophytes	Prasinophytes	NC_023273.1
<i>Cymbomonas tetramitiformis</i>	Chlorophytes	Prasinophytes	KX013548.1
<i>Micromonas</i> sp. RCC299	Chlorophytes	Prasinophytes	FJ859351.1
<i>Monomastix</i> sp. OKE-1	Chlorophytes	Prasinophytes	KF060939.1
<i>Ostreococcus tauri</i>	Chlorophytes	Prasinophytes	CR954200.2
<i>Prasinoderma coloniale</i>	Chlorophytes	Prasinophytes	KF387569.1
<i>Pycnococcus provasolii</i>	Chlorophytes	Prasinophytes	GQ497137.1
<i>Pyramimonas parkeae</i>	Chlorophytes	Prasinophytes	KX013547.1
<i>Anthoceros angustus*</i>	Anthocerotophyta	Anthocerotophyta	NC_037476.1
<i>Leiosporoceros dussii*</i>	Anthocerotophyta	Anthocerotophyta	NC_039751.1
<i>Megaceros aenigmaticus</i>	Anthocerotophyta	Anthocerotophyta	EU660574.1
<i>Phaeoceros laevis</i>	Anthocerotophyta	Anthocerotophyta	GQ376531.1
<i>Aneura pinguis</i>	Marchantiophyta	Marchantiophyta	KP728938.1
<i>Calypogeia neogaea*</i>	Marchantiophyta	Marchantiophyta	NC_035980.1
<i>Marchantia polymorpha</i>	Marchantiophyta	Marchantiophyta	NC_001660.1
<i>Pleurozia purpurea</i>	Marchantiophyta	Marchantiophyta	NC_013444.1
<i>Treubia lacunosa</i>	Marchantiophyta	Marchantiophyta	JF973315.1
<i>Tritomaria quinqueidentata*</i>	Marchantiophyta	Marchantiophyta	NC_037041.1
<i>Anomodon attenuatus</i>	Bryophytes	Bryophytes	JX402749.1
<i>Atrichum angustatum</i>	Bryophytes	Bryophytes	KC784956.1
<i>Bartramia pomiformis</i>	Bryophytes	Bryophytes	KC784955.1
<i>Brachythecium rivulare</i>	Bryophytes	Bryophytes	KR732319.1
<i>Bucklandiella orthotrichacea</i>	Bryophytes	Bryophytes	KP742835.1
<i>Buxbaumia aphylla</i>	Bryophytes	Bryophytes	KC784954.1
<i>Climacium americanum</i>	Bryophytes	Bryophytes	KC784950.1
<i>Codriophorus aciculare</i>	Bryophytes	Bryophytes	KP453983.1
<i>Funaria hygrometrica</i>	Bryophytes	Bryophytes	KC784959.1
<i>Hypnum imponens</i>	Bryophytes	Bryophytes	KC784951.1
<i>Nyholmiella gymnostoma</i>	Bryophytes	Bryophytes	KX578030.1
<i>Nyholmiella obtusifolia*</i>	Bryophytes	Bryophytes	NC_031767.1
<i>Orthotrichum callistomum</i>	Bryophytes	Bryophytes	KX578029.1
<i>Orthotrichum speciosum*</i>	Bryophytes	Bryophytes	NC_026121.1
<i>Oxystegus tenuirostris</i>	Bryophytes	Bryophytes	KT326816.1
<i>Physcomitrella patens</i>	Bryophytes	Bryophytes	NC_007945.1
<i>Ptychomnion cygnisetum</i>	Bryophytes	Bryophytes	KC784949.1
<i>Racomitrium elongatum*</i>	Bryophytes	Bryophytes	NC_026890.1

(continued)

■ Table 1, continued

Species	Group	Subgroup	Accession
<i>Sanionia uncinata</i>	Bryophytes	Bryophytes	KP984757.1
<i>Sphagnum palustre</i>	Bryophytes	Bryophytes	KC784957.1
<i>Stoneobryum bunyaense</i>	Bryophytes	Bryophytes	KX578031.1
<i>Syntrichia filaris</i>	Bryophytes	Bryophytes	KP984758.1
<i>Tetraphis pellucida</i>	Bryophytes	Bryophytes	KC784953.1
<i>Tetraplodon fuegianus</i>	Bryophytes	Bryophytes	KT373818.1
<i>Ulota phyllantha</i>	Bryophytes	Bryophytes	KX578033.1
<i>Zygodon viridissimus</i>	Bryophytes	Bryophytes	KX711975.1
<i>Chaetosphaeridium globosum</i>	Streptophyta	Charophyta	AF494279.1
<i>Chara vulgaris</i>	Streptophyta	Charophyta	AY267353.1
<i>Chlorokybus atmophyticus</i>	Streptophyta	Charophyta	NC_009630.1
<i>Closterium baillyanum</i>	Streptophyta	Charophyta	NC_022860.1
<i>Entransia fimbriata</i>	Streptophyta	Charophyta	KF060941.1
<i>Klebsormidium flaccidum</i>	Streptophyta	Charophyta	KP165386.1
<i>Nitella hyalina</i>	Streptophyta	Charophyta	JF810595.1
<i>Roya obtusa</i>	Streptophyta	Charophyta	KF060943.1
<i>Ophioglossum californicum</i>	Tracheophyta	Fern	NC_030900.1
<i>Psilotum nudum</i>	Tracheophyta	Fern	NC_030952.1
<i>Huperzia squarrosa</i>	Tracheophyta	Lycophyte	NC_017755.1
<i>Selaginella moellendorffii</i>	Tracheophyta	Lycophyte	GL377739.1
<i>Cycas taitungensis</i>	Tracheophyta	Gymnosperm	AP009381.1
<i>Ginkgo biloba</i>	Tracheophyta	Gymnosperm	KM672373.1
<i>Welwitschia mirabilis</i>	Tracheophyta	Gymnosperm	NC_029130.1
<i>Aegilops speltoides</i>	Tracheophyta	Angiosperm	AP013107.1
<i>Ajuga reptans</i>	Tracheophyta	Angiosperm	KF709392.1
<i>Allium cepa</i>	Tracheophyta	Angiosperm	KU318712.1
<i>Ammopiptanthus mongolicus*</i>	Tracheophyta	Angiosperm	NC_039660.1
<i>Arabidopsis thaliana</i>	Tracheophyta	Angiosperm	BK010421
<i>Arabis alpina*</i>	Tracheophyta	Angiosperm	NC_037070.1
<i>Batis maritima</i>	Tracheophyta	Angiosperm	KJ820684.1
<i>Beta vulgaris</i>	Tracheophyta	Angiosperm	BA000024.1
<i>Betula pendula</i>	Tracheophyta	Angiosperm	LT855379.1
<i>Boea hygrometrica</i>	Tracheophyta	Angiosperm	JN107812.1
<i>Bombax ceiba*</i>	Tracheophyta	Angiosperm	NC_038052.1
<i>Brassica nigra</i>	Tracheophyta	Angiosperm	KP030753.1
<i>Bupleurum falcatum*</i>	Tracheophyta	Angiosperm	NC_035962.1
<i>Butomus umbellatus</i>	Tracheophyta	Angiosperm	KC208619.1
<i>Cannabis sativa</i>	Tracheophyta	Angiosperm	KU310670.1
<i>Capsicum annuum</i>	Tracheophyta	Angiosperm	KJ865410.1
<i>Carica papaya</i>	Tracheophyta	Angiosperm	EU431224.1
<i>Castilleja paramensis</i>	Tracheophyta	Angiosperm	KT959112.1
<i>Chrysanthemum boreale*</i>	Tracheophyta	Angiosperm	NC_039757.1
<i>Citrullus lanatus</i>	Tracheophyta	Angiosperm	GQ856147.1
<i>Citrus sinensis*</i>	Tracheophyta	Angiosperm	MG736621.1
<i>Cocos nucifera</i>	Tracheophyta	Angiosperm	KX028885.1
<i>Corchorus capsularis</i>	Tracheophyta	Angiosperm	KT894204.1
<i>Daucus carota subsp. sativus</i>	Tracheophyta	Angiosperm	JQ248574.1
<i>Diplostephium hartwegii</i>	Tracheophyta	Angiosperm	KX063855.1
<i>Eruca vesicaria</i>	Tracheophyta	Angiosperm	KF442616.1
<i>Geranium maderense</i>	Tracheophyta	Angiosperm	NC_027000.1
<i>Glycine max</i>	Tracheophyta	Angiosperm	JX463295.1
<i>Gossypium hirsutum</i>	Tracheophyta	Angiosperm	NC_027406.1
<i>Helianthus annuus</i>	Tracheophyta	Angiosperm	CM007908.1
<i>Heuchera parviflora</i>	Tracheophyta	Angiosperm	KR559021.1
<i>Hibiscus cannabinus</i>	Tracheophyta	Angiosperm	MF163174.1
<i>Hordeum vulgare</i>	Tracheophyta	Angiosperm	AP017300.1
<i>Hyoscyamus niger</i>	Tracheophyta	Angiosperm	KM207685.1
<i>Ipomoea nil</i>	Tracheophyta	Angiosperm	AP017303.1
<i>Lagerstroemia indica</i>	Tracheophyta	Angiosperm	KX641464.1
<i>Leucaena trichandra*</i>	Tracheophyta	Angiosperm	NC_039738.1
<i>Liriodendron tulipifera</i>	Tracheophyta	Angiosperm	NC_021152.1
<i>Lolium perenne</i>	Tracheophyta	Angiosperm	JX999996.1

(continued)

■ Table 1, continued

Species	Group	Subgroup	Accession
<i>Lotus japonicus</i>	Tracheophyta	Angiosperm	JN872551.2
<i>Medicago truncatula</i>	Tracheophyta	Angiosperm	KT971339.1
<i>Milletia pinnata</i>	Tracheophyta	Angiosperm	JN872550.1
<i>Mimulus guttatus</i>	Tracheophyta	Angiosperm	JN098455.1
<i>Nelumbo nucifera</i> *	Tracheophyta	Angiosperm	KR610474.1
<i>Nicotiana tabacum</i>	Tracheophyta	Angiosperm	BA000042.1
<i>Nymphaea colorata</i> *	Tracheophyta	Angiosperm	KY889142.1
<i>Oryza sativa Indica</i>	Tracheophyta	Angiosperm	NC_007886.1
<i>Phoenix dactylifera</i>	Tracheophyta	Angiosperm	JN375330.1
<i>Platycodon grandiflorus</i> *	Tracheophyta	Angiosperm	NC_035958.1
<i>Populus tremula</i>	Tracheophyta	Angiosperm	KT337313.1
<i>Raphanus sativus</i>	Tracheophyta	Angiosperm	NC_018551.1
<i>Rhazya stricta</i>	Tracheophyta	Angiosperm	KJ485850.1
<i>Ricinus communis</i>	Tracheophyta	Angiosperm	HQ874649.1
<i>Rosa chinensis</i> *	Tracheophyta	Angiosperm	CM009589.1
<i>Salix suchowensis</i>	Tracheophyta	Angiosperm	KU056812.1
<i>Salvia miltiorrhiza</i>	Tracheophyta	Angiosperm	KF177345.1
<i>Schrenkiella parvula</i>	Tracheophyta	Angiosperm	KT988071.2
<i>Senna tora</i> *	Tracheophyta	Angiosperm	NC_038053.1
<i>Silene latifolia</i>	Tracheophyta	Angiosperm	HM562727.1
<i>Sinapis arvensis</i>	Tracheophyta	Angiosperm	KM851044.1
<i>Sorghum bicolor</i>	Tracheophyta	Angiosperm	DQ984518.1
<i>Spinacia oleracea</i>	Tracheophyta	Angiosperm	KY768855.1
<i>Styphnolobium japonicum</i> *	Tracheophyta	Angiosperm	NC_039596.1
<i>Tripsacum dactyloides</i>	Tracheophyta	Angiosperm	NC_008362.1
<i>Triticum aestivum</i>	Tracheophyta	Angiosperm	AP008982.1
<i>Utricularia reniformis</i> *	Tracheophyta	Angiosperm	NC_034982.1
<i>Vicia faba</i>	Tracheophyta	Angiosperm	KC189947.1
<i>Vigna angularis</i>	Tracheophyta	Angiosperm	AP012599.1
<i>Viscum album</i>	Tracheophyta	Angiosperm	NC_029039.1
<i>Vitis vinifera complete</i>	Tracheophyta	Angiosperm	FM179380.1
<i>Zea mays strain NB</i>	Tracheophyta	Angiosperm	AY506529.1
<i>Ziziphys jujuba</i>	Tracheophyta	Angiosperm	KU187967.1

iteration of BLAST. A duplication longer than 100 bases that has several mismatches or a gap in the center of the repeat unit will be identified as two different repeats by this script. However, mismatches in the center of one copy of a repeat 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 chose 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 word size parameter of 50 is chosen to make the output more manageable. Reducing the word size identifies numerous smaller repeats, but the smaller the repeats get, the more complex the curation task of distinguishing identical sequences from similar ones of the same size – a task that at this point still needs to be done manually. Reducing the word size does not change the conclusions about any of the potentially recombinogenic repeats that also distinguish the major groups in the plant kingdom. Previous work in *Arabidopsis thaliana* has shown that crossover products between non-tandem repeats of 556bp or smaller in accession Col-0 and 204bp or smaller in accession Ws are undetectable by PCR (Wallet, C. *et al.* 2015), similar to prior results using Southern blots (Arrieta-Montiel, M. P. *et al.* 2009; Davila, J. I. *et al.* 2011; Sakamoto, W. *et al.* 1996)

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 or wildcard characters (such as N, R, Y, etc.) 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.

### Phylogenetic clustering

The distribution of repeat sizes forms distinct clusters between broad 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. 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 non-tandem repeats greater than 200bp but many of them include

two copies of a single very 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 repeats longer than 200bp.

In contrast to the chlorophytes and bryophytes, the Marchantiophyta (liverworts) and Anthocerotophyta (hornworts) have repeats greater than 200bp in size, but none bigger than 1131bp. 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 underrepresented among available mitochondrial sequences, in part due to the complexity caused by the repetitive nature of lycophyte and fern mitochondrial 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 repeats larger than 200bp (*Medicago truncatula*). A small number of angiosperms 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 repeats, 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 mitochondrial genomes a few times larger than the bryophytes, liverworts and hornworts, but the repeats are expanded well beyond proportionality to size. Some taxa, such as the Geraniaceae, Plantago, and *Silene* include species with significantly expanded mitochondrial DNA (Park, S. *et al.* 2015; Parkinson, C. L. *et al.* 2005; Sloan, D. B. *et al.* 2012). These species are outliers in the mitochondrial genome sizes and the number of repeats, but the underlying DNA replication, recombination and repair processes are likely to be the same. There appears to have been a significant change in mitochondrial DNA maintenance mechanisms 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 72 angiosperms, sorted by the class and order of the species. While individual species often have a bimodal distribution of sizes, there is no size range that is universally absent from the distribution. Thirteen of the 72 species have no repeats larger than 600bp, leaving open the question of whether those particular mitochondrial genomes isomerize through recombination. All of the other species have a large repeat of somewhere between 600bp and 65kbp. There is no pattern of repeat size distribution or total size with the phylogenetic group or total mitochondrial genome size, suggesting 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

In order to test the hypothesis that the repeated sequences change rapidly compared to speciation events, leading to the lack of pattern in the Angiosperm orders, we analyzed 6 closely related species in the Brassica genus. 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* has 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* as outgroups. 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 mitochondrial genome isomerization 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 1.9kb to 9.7kb. However, these species show an interesting pattern. *B. nigra*, *B. carinata*, *R. sativus* and *S. arvensis*, hereafter referred to as group A, each have two copies of a 6.5 to 9.7kb repeat that is only present as single copy sequence in the mitochondria of *B. rapa*, *B. oleracea*, *B. napus* and *B. juncea*, hereafter referred to as group B (see Figure 3). The group B species each have two copies of a long repeat 1.9kb long that is present as single-copy sequence in group A. Figure 3 shows these repeated sequences, aligned only to each other and placed onto the known phylogenetic tree of the Brassicales (Grewe, F. *et al.* 2014). The longest repeats are aligned, and the genes flanking them are shown. Part A shows the long repeat and neighboring sequences from the A group and the homologous single-copy sequences from the B group. Part B compares the long repeat from the B group to the single-copy homologous region from the A group.

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 in the A group differ by large block

Group	Subgroup	# of species	50-199	200-499	500-999	1000-2499	2500-4999	5000-7499	7500-9999	≥ 10000
Chlorophytes	Chlorophyta	26	0.88	0.19	0.08	0.00	0.00	0.00	0.00	0.00
Chlorophytes	Prasinophytes	8	0.63	0.00	0.00	0.00	0.00	0.00	0.50	0.25
Anthocerotophyta	Anthocerotophyta	4	1.00	0.50	1.00	0.25	0.00	0.00	0.00	0.00
Marchantiophyta	Marchantiophyta	6	1.00	1.00	0.67	0.00	0.00	0.00	0.00	0.00
Bryophytes	Bryophytes	26	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Streptophyta	Charophyta	8	0.88	0.25	0.00	0.00	0.00	0.00	0.00	0.00
Tracheophyta	Fern	2	1.00	1.00	1.00	1.00	1.00	0.00	1.00	0.00
Tracheophyta	Lycophyte	2	1.00	1.00	1.00	1.00	1.00	0.00	0.50	0.00
Tracheophyta	Gymnosperm	3	1.00	1.00	0.33	0.67	0.33	0.33	0.00	0.00
Tracheophyta	Angiosperm	72	1.00	0.96	0.47	0.49	0.46	0.36	0.25	0.43

**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 one and yellow for zero.

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 B group of Brassica species (see Figure 3). The first 1.6kb and the last 1.7kb of the long repeat in the A group are conserved in the B 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 (Grewe, F. *et al.* 2014).

At the same time that the long repeat of the A group was being dramatically altered in the lineage leading to *B. rapa* and *B. oleracea*, a new long repeat appeared in the B group, 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 (see Figure 3; sequence alignments are shown in Supplemental Figure S2). The *cox2* gene is single copy in the A group and is in a nearly syntenic arrangement with neighboring genes.

## DISCUSSION

The availability of complete mitochondrial genome sequences from many taxa of green plants allows us to compare the abundance and size distribution of non-tandem repeats across taxa. Although such repeats 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 sequences, and use it to analyze plant mitochondrial DNA. In addition, comparison of repeats between closely related

species within the Brassicales 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.

Repeats in mitochondria appear to be more abundant and larger in the vascular plants than in the non-vascular taxa. This suggests that the first vascular common ancestor of lycophytes, ferns, gymnosperms and angiosperms acquired new mechanisms of mitochondrial genome replication and repair that led to a proliferation of repeats and increases in repeat size and mitochondrial genome 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.

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 by rearrangements, sequence loss, and gain of sequences of unknown origin. This pattern resembles punctuated equilibrium (Gould, S. J. and N. Eldredge 1977). 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 high levels of divergence, and loss of synteny.

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). In general the largest repeats within a species have been found to recombine constitutively, leading to isomerization of the genome into multiple major forms. The shortest repeats (less than 50bp) may be involved in homologous recombination events only rarely, while those of intermediate size, generally in the 100s of base pairs, can recombine in response to genome damage or in DNA maintenance mutants, but do not normally do so in unstressed, non-mutant plants, as noted above. The intermediate size repeats have been primarily analyzed in *Arabidopsis thaliana*, and have been found to recombine in abnormal conditions. In plants treated with ciprofloxacin (which induces mitochondrial double-strand breaks), or in mutants of the mitochondrial *recG* homolog, repeats of 452, 249, 204 and 126bp were seen to recombine (Wallet, C. *et al.* 2015). In mutants of *msh1* (which results in high levels of ectopic recombination), there was some recombination seen between repeats as small as 70bp, but none in repeats of 50bp or smaller (Davila, J. I. *et al.* 2011). This suggests a changing spectrum of function and activity correlated with size, which could also vary by species.







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 (Paszkiwicz, 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.

## ACKNOWLEDGMENTS

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).

## 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. USA* 100: 5968–5973. <https://doi.org/10.1073/pnas.1037651100>
- 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. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2)
- 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. <https://doi.org/10.1105/tpc.111.087189>
- 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. <https://doi.org/10.1093/molbev/msq029>
- 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. <https://doi.org/10.1371/journal.pone.0016404>
- Andre, C., A. Levy, and V. Walbot, 1992 Small repeated sequences and the structure of plant mitochondrial genomes. *Trends Genet.* 8: 128–132.
- 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. <https://doi.org/10.1534/genetics.109.108514>
- Bendich, A. J., 2004 Circular Chloroplast Chromosomes: The Grand Illusion. *Plant Cell* 16: 1661–1666. <https://doi.org/10.1105/tpc.160771>
- 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. <https://doi.org/10.1186/1471-2164-12-497>
- 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. <https://doi.org/10.1093/molbev/msn009>
- Cheng, F., J. Liang, C. Cai, X. Cai, J. Wu *et al.*, 2017 Genome sequencing supports a multi-vertex model for Brassicaceae species. *Curr. Opin. Plant Biol.* 36: 79–87. <https://doi.org/10.1016/j.pbi.2017.01.006>
- Christensen, A. C., 2014 Genes and Junk in Plant Mitochondria—Repair Mechanisms and Selection. *Genome Biol. Evol.* 6: 1448–1453. <https://doi.org/10.1093/gbe/evu115>
- Christensen, A. C., 2018 Mitochondrial DNA repair and genome evolution, pp. 11–31 in *Annual Plant Reviews, Plant Mitochondria*, Ed. 2, edited by Logan, D. C., Wiley-Blackwell, New York, NY
- Cole, T. C. H., H. H. Hilger, and P. F. Stevens, 2017 Angiosperm phylogeny poster (APP) – Flowering plant systematics, 2017. *PeerJ Preprints* 5: e2320v4. <https://doi.org/10.7287/peerj.preprints.2320v4>
- 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. <https://doi.org/10.1186/1741-7007-9-64>
- 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. <https://doi.org/10.1016/j.ympev.2008.09.009>
- 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. <https://doi.org/10.1093/nar/17.18.7345>
- 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. <https://doi.org/10.1093/nar/gki774>
- 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. <https://doi.org/10.1016/j.tplants.2010.11.005>
- Gould, S. J., and N. Eldredge, 1977 Punctuated Equilibria: The Tempo and Mode of Evolution Reconsidered. *Paleobiology* 3: 115–151. <https://doi.org/10.1017/S0094837300005224>
- 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* 19: 135–143. <https://doi.org/10.1016/j.mito.2014.05.008>
- Grewe, F., P. Viehoveer, 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. <https://doi.org/10.1093/nar/gkp532>
- Gualberto, J. M., D. Mileschina, C. Wallet, A. K. Niazi, F. Weber-Lotfi *et al.*, 2014 The plant mitochondrial genome: dynamics and maintenance. *Biochimie* 100: 107–120. <https://doi.org/10.1016/j.biochi.2013.09.016>
- Gualberto, J. M., and K. J. Newton, 2017 Plant Mitochondrial Genomes: Dynamics and Mechanisms of Mutation. *Annu. Rev. Plant Biol.* 68: 225–252. <https://doi.org/10.1146/annurev-arplant-043015-112232>
- 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. <https://doi.org/10.1093/molbev/msw024>
- 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. <https://doi.org/10.1111/nph.14135>
- 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. <https://doi.org/10.1093/gbe/evr027>
- Klein, M., U. Eckert-Ossenkopp, I. Schmiedeborg, 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. <https://doi.org/10.1046/j.1365-313X.1994.06030447.x>
- Krisko, A., and M. Radman, 2013 Biology of extreme radiation resistance: the way of *Deinococcus radiodurans*. *Cold Spring Harb. Perspect. Biol.* 5: a012765. <https://doi.org/10.1101/cshperspect.a012765>
- Kurtz, S., and C. Schleiermacher, 1999 REPuter: fast computation of maximal repeats in complete genomes. *Bioinformatics* 15: 426–427. <https://doi.org/10.1093/bioinformatics/15.5.426>
- 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. <https://doi.org/10.1093/molbev/msu199>
- Lynch, M., B. Koskella, and S. Schaack, 2006 Mutation Pressure and the Evolution of Organelle Genomic Architecture. *Science* 311: 1727–1730. <https://doi.org/10.1126/science.1118884>
- 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. <https://doi.org/10.1128/jb.178.3.633-637.1996>
- 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. <https://doi.org/10.1104/pp.112.194720>
- 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. <https://doi.org/10.1093/gbe/evs042>
- 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. <https://doi.org/10.1093/nar/14.24.9755>
- 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. <https://doi.org/10.1007/BF02143500>
- Palmer, J. D., and C. R. Shields, 1984 Tripartite structure of the Brassica campestris mitochondrial genome. *Nature* 307: 437–440. <https://doi.org/10.1038/307437a0>
- 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. <https://doi.org/10.1111/nph.13467>
- 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. <https://doi.org/10.1186/1471-2148-5-73>
- Paszkiwicz, 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. <https://doi.org/10.1105/tpc.16.00700>
- 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. <https://doi.org/10.1534/genetics.116.196154>
- 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. <https://doi.org/10.1186/1741-7007-11-29>
- Sakamoto, W., H. Kondo, M. Murata, and F. Motoyoshi, 1996 Altered mitochondrial gene expression in a maternal distorted leaf mutant of *Arabidopsis* induced by *chloroplast mutator*. *Plant Cell* 8: 1377–1390. <https://doi.org/10.1105/tpc.8.8.1377>
- Schuster, W., and A. Brennicke, 1994 The Plant Mitochondrial Genome: Physical Structure, Information Content, RNA Editing, and Gene Migration to the Nucleus. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 45: 61–78. <https://doi.org/10.1146/annurev.pp.45.060194.000425>
- Sharma, A., E. K. Gaidamakova, O. Grichenko, V. Y. Matrosova, V. Hoeke *et al.*, 2017 Across the tree of life, radiation resistance is governed by antioxidant Mn(2+), gauged by paramagnetic resonance. *Proc. Natl. Acad. Sci. USA* 114: E9253–E9260. <https://doi.org/10.1073/pnas.1713608114>
- 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. <https://doi.org/10.1038/srep31533>
- 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. <https://doi.org/10.1105/tpc.106.048355>
- Siculella, L., F. Damiano, M. R. Cortese, E. Dassisti, G. Rainaldi *et al.*, 2001 Gene content and organization of the oat mitochondrial genome. *Theor. Appl. Genet.* 103: 359–365. <https://doi.org/10.1007/s001220100568>
- 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. <https://doi.org/10.1111/nph.12395>
- Sloan, D. B., A. J. Alverson, J. P. Chackalovcak, 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. <https://doi.org/10.1371/journal.pbio.1001241>
- 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. <https://doi.org/10.1186/1471-2148-10-274>
- Smith, D. R., 2016 The mutational hazard hypothesis of organelle genome evolution: 10 years on. *Mol. Ecol.* 25: 3769–3775. <https://doi.org/10.1111/mec.13742>
- 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. <https://doi.org/10.1186/1471-2229-10-83>
- 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. <https://doi.org/10.1093/nar/14.14.5651>
- Sugiyama, Y., Y. Watase, M. Nagase, N. Makita, S. Yagura *et al.*, 2005 The complete nucleotide sequence and multipartite organization of the tobacco mitochondrial genome: comparative analysis of mitochondrial genomes in higher plants. *Mol. Genet. Genomics* 272: 603–615. <https://doi.org/10.1007/s00438-004-1075-8>
- 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. <https://doi.org/10.1105/tpc.15.00680>
- 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. [https://doi.org/10.1016/0092-8674\(81\)90187-2](https://doi.org/10.1016/0092-8674(81)90187-2)
- 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. USA* 84: 9054–9058. <https://doi.org/10.1073/pnas.84.24.9054>

Communicating editor: B. Andrews