2-2016

Ginkgo and *Welwitschia mitogenomes* reveal extreme contrasts in gymnosperm mitochondrial evolution

Wenhu Guo
University of Nebraska - Lincoln

Felix Grewe
University of Nebraska - Lincoln

Weishu Fan
University of Nebraska - Lincoln

Gregory J. Young
Indiana University

Volker Knoop
Universität Bonn

See next page for additional authors

Mol. Biol. Evol., Feb 2016; doi:10.1093/molbev/msw024

ARTICLE - DISCOVERIES

Ginkgo and *Welwitschia* mitogenomes reveal extreme contrasts in gymnosperm mitochondrial evolution

Wenhu Guo[1,2,3], Felix Grewe[1,4,5,6], Weishu Fan[1,4], Gregory J. Young[7,8], Volker Knoop[6], Jeffrey D. Palmer[7], Jeffrey P. Mower[1,4,*]

[1] Center for Plant Science Innovation, University of Nebraska, Lincoln, NE 68588, USA

[2] School of Biological Sciences, University of Nebraska, Lincoln, NE 68588, USA

[3] ACGT, Inc., Wheeling, IL 60090, USA

[4] Department of Agronomy and Horticulture, University of Nebraska, Lincoln, NE 68583, USA

[5] The Field Museum of Natural History, Chicago, IL 60605, USA

[6] IZMB—Institut für Zelluläre und Molekulare Botanik, Abteilung Molekulare Evolution, Universität Bonn, Kirschallee 1, 53115 Bonn, Germany

[7] Department of Biology, Indiana University, Bloomington, IN 47405, USA

[8] DuPont Pioneer,Wilmington, DE 19880, USA

*** Corresponding author:** Jeffrey P Mower. Address: 1901 Vine St, Beadle Center E128, University of Nebraska, Lincoln, NE 68588. Phone: +1 402-472-2130. Email: jpmower@unl.edu

Keywords: *Ginkgo biloba*, *Welwitschia mirabilis*, mitochondrial genomes, substitution rates, DNA turnover, RNA editing

Running Head: Contrasting evolution in gymnosperm mitogenomes

ABSTRACT

Mitochondrial genomes (mitogenomes) of flowering plants are well known for their extreme diversity in size, structure, gene content, and rates of sequence evolution and recombination. In contrast, little is known about mitogenomic diversity and evolution within gymnosperms. Only a single complete genome sequence is available, from the cycad *Cycas taitungensis*, while limited information is available for the one draft sequence, from Norway spruce (*Picea abies*). To examine mitogenomic evolution in gymnosperms, we generated complete genome sequences for the ginkgo tree (*Ginkgo biloba*) and a gnetophyte (*Welwitschia mirabilis*). There is great disparity in size, sequence conservation, levels of shared DNA, and functional content among gymnosperm mitogenomes. The *Cycas* and *Ginkgo* mitogenomes are relatively small, have low substitution rates, and possess numerous genes, introns, and edit sites; we infer that these properties were present in the ancestral seed plant. By contrast, the *Welwitschia* mitogenome has an expanded size coupled with accelerated substitution rates and extensive loss of these functional features. The *Picea* genome has expanded further, to more than 4 Mb. With regard to structural evolution, the *Cycas* and *Ginkgo* mitogenomes share a remarkable amount of intergenic DNA, which may be related to the limited recombinational activity detected at repeats in *Ginkgo*. Conversely, the *Welwitschia* mitogenome shares almost no intergenic DNA with any other seed plant. By conducting the first measurements of rates of DNA turnover in seed plant mitogenomes, we discovered that turnover rates vary by orders of magnitude among species.

INTRODUCTION

More than 70 angiosperm mitochondrial genomes (mitogenomes) have been sequenced, providing a rich picture of their prevailing features, ancestral characteristics, and evolutionary trends (Knoop 2012; Mower et al. 2012c). Genome size is extremely variable, ranging from 66 kb to 11.3 Mb, although most genomes are 200–800 kb in size (Sloan et al. 2012a; Skippington et al. 2015). The ancestral angiosperm mitogenome contained 41 protein genes; however, few angiosperms contain this full set due to episodically frequent functional transfer of genes to the nucleus (Adams et al. 2002). Angiosperm mitogenomes ancestrally contained 25 introns, five of them *trans*-spliced, with lineage-specific loss of introns occurring by retroprocessing (Ran et al. 2010; Grewe et al. 2011) or gene conversion with intron-less genes acquired horizontally (Hepburn et al. 2012; Wu and Hao 2014). C-to-U RNA editing occurs at some 200–800 mRNA sites, with the highest number ancestrally present and progressive loss of editing in most angiosperms (Mower 2008; Sloan et al. 2010; Rice et al. 2013; Richardson et al. 2013).

Angiosperm mitogenomes are also diverse in rates of sequence and structural evolution. The vast majority of angiosperm mitogenomes have very low synonymous substitution rates, on the order of 5 and 20 times lower than plastid and nuclear rates, respectively, while mutator-like mitochondrial lineages with 1–3 orders-of-magnitude higher rates have evolved repeatedly (Mower et al. 2007; Drouin et al. 2008; Richardson et al. 2013). Most angiosperm mitogenomes are rich in repeat sequences, with the larger repeats recombining at moderate to high frequency (Maréchal and Brisson 2010; Woloszynska 2010). This recombination generates multiple genomic arrangements of varying stoichiometry (Alverson et al. 2011; Davila et al. 2011; Mower et al. 2012a) and quickly erodes synteny even among closely related plants (Palmer and Herbon 1988; Allen et al. 2007; Sloan et al. 2012b). Angiosperm mitogenomes are also rich in foreign DNA, including plastid and nuclear DNA acquired via intracellular transfer and mitochondrial DNA (mtDNA) acquired via horizontal transfer (Mower et al. 2012b; Rice et al. 2013). In sum, one can infer with confidence that the ancestral angiosperm mitogenome possessed 41 protein genes, 20 *cis*-spliced and five *trans*-

spliced introns, and about 700–800 edit sites. As well, this genome was probably several hundred kb in size, rich in repeats and recombination, and slowly evolving in sequence.

In contrast to the heavily sequenced angiosperm clade, the number of mitogenomic sequences from other vascular plants is limited to a single gymnosperm, *Cycas taitungensis* (Chaw et al. 2008), and three lycophytes, *Isoetes engelmannii, Selaginella moellendorffii*, and *Huperzia squarrosa* (Grewe et al. 2009; Hecht et al. 2011; Liu et al. 2012). Similar to many angiosperms, the mitogenomes from *Cycas* and *Huperzia* are ~400 kb in length, whereas the draft mitogenome of Norway spruce (*Picea abies*) is much larger, at >4.3 Mb (Nystedt et al. 2013). The *Isoetes* and *Selaginella* mitogenomes are probably considerably smaller, although the highly complex network structures of these two genomes, the result of high-frequency recombination across many families of repeats, allows estimation of only their minimum sizes, of 58 and 183 kb, respectively. Gene and intron content in *Cycas* and *Picea* is very similar to angiosperms (Chaw et al. 2008; Nystedt et al. 2013), whereas the three lycophytes contain a more variable complement of genes and numerous novel introns, including several novel *trans*-splicing arrangements (Grewe et al. 2009; Hecht et al. 2011; Liu et al. 2012). RNA editing content is also variable, ranging from a few hundred sites in *Huperzia* (Liu et al. 2012) to over 2,100 sites in *Selaginella* (Hecht et al. 2011). Thus, while certain features of angiosperm mitogenomes are also present to some extent in other vascular plants, the limited number of non-angiosperm sequences and their variability preclude any strong conclusions about mitogenomic properties of the common ancestor of vascular plants.

Knowledge of mitogenomic diversity in gymnosperms, the sister lineage to angiosperms, could provide valuable insight on the origin and evolution of the complex mitogenomic features of angiosperms and whether these features are typical of other vascular plants. To this end, we sequenced mtDNA from two gymnosperms, the ginkgo tree (*Ginkgo biloba*) and a gnetophyte (*Welwitschia mirabilis*), and compared them with the published *Cycas* genome, thus sampling three of the five groups of extant gymnosperms (the other two being the two clades of conifers). Comparison of these

three mitogenomes revealed extensive diversity in size, structure, gene and intron content, and point mutation rates in gymnosperms. We also conducted the first measurements to compare amounts of shared mtDNA among seed plants as a function of divergence time. This analysis revealed enormous variation in rates of mtDNA turnover, with *Cycas* and *Ginkgo* exhibiting extremely low turnover rates compared with all other seed plants.

RESULTS AND DISCUSSION

Extensive variation in mitogenome size among gymnosperms

The *Ginkgo* mitogenome was assembled as a single circular molecule of length 347 kb (Figure S1A), which makes it similar in size to the 415-kb mitogenome of *Cycas* and the 414-kb mitogenome of the lycophyte *Huperzia* (Table 1). Approximately half of all sequenced angiosperm mitogenomes available in GenBank as of July 2015 also fall within a size range of 300 to 500 kb. The similarity in mitogenome size from multiple vascular plant lineages—including *Cycas*, *Ginkgo*, many angiosperms, and a lycophyte—suggests that the common ancestor of gymnosperms and angiosperms possessed a mitogenome of roughly 400 kb in size. In contrast, the *Welwitschia* mitogenome is 979 kb in size (Figure S1B), which is still smaller than the >4.3 Mb draft mitogenome from *Picea* (Nystedt et al. 2013). This indicates at least two major expansions in mitogenome size in the gnetophyte and conifer lineages. That a >10-fold range in mitogenome size has been discovered among only four examined gymnosperms raises the possibility that a substantially greater range will be revealed by more intensive sampling of the ~1,000 species of extant gymnosperms.

In certain angiosperm mitogenomes, most notably that of *Cucurbita pepo* (Alverson et al. 2010), genome expansion is largely the result of major increases in the abundance of repeats or mitochondrial DNA of plastid origin (MIPT, Mower et al. 2012b). This is decidedly not the case in *Welwitschia* compared with *Ginkgo* and *Cycas* (Table 1). In fact, the 415-kb *Cycas* mitogenome (Chaw et al. 2008) contains substantially more repeats (67 kb; covering 15.1% of the genome) and MIPTs (18 kb; 4.4%) than both the

979-kb *Welwitschia* genome (29 kb repetitive DNA, 2.9% coverage; 8.5 kb MIPT, 0.8%) and the 347-kb *Ginkgo* genome (33 kb, 9.5%; 0.5 kb, 0.1%).

Instead, we hypothesize that much of the massive amounts of unidentified DNA in these three gymnosperm genomes ultimately derives from the nuclear genome by intracellular transfer, as inferred for certain angiosperms, such as *Cucumis* and *Malus* (Alverson et al. 2011; Rodriguez-Moreno et al. 2011; Goremykin et al. 2012). In principle, testing this hypothesis requires nuclear genome sequences from at least these same gymnosperm species, none of which are currently available. In practice, however, it may be impossible to test this hypothesis for the *Cycas* and *Ginkgo* mitogenomes. This is because most of their unidentified DNA is shared between the two genomes (see below) and thus was present in the mitogenome of their common ancestor, some 343 million years ago (Ma). Unless their nuclear genomes are also extremely slowly evolving, this anciently shared mtDNA of putatively nuclear origin will no longer resemble the present-day nuclear genome of either species.

Numerous losses of protein and tRNA genes from the Welwitschia *mitogenome*

Comparison of mitochondrial gene content among gymnosperms reveals a pattern of evolutionary stasis for *Cycas* and *Ginkgo* contrasted with extensive loss in *Welwitschia* (Table 1; Figure 1A). The *Cycas* and *Ginkgo* mitogenomes contain 41 protein genes, which is identical to the set inferred for the common ancestor of angiosperms (Richardson et al. 2013). Also, the *Picea* mitogenome was reported to contain the same gene set as *Cycas*, although no further details were provided (Nystedt et al. 2013). Therefore, the common ancestor of seed plants probably contained this same set of 41 protein genes. In contrast, the *Welwitschia* mitogenome has lost 11 ribosomal protein genes and the *sdh3* gene. Presumably, most or all of these genes were functionally transferred to the nucleus prior to loss from the mitochondrial genome. These two patterns—marked, lineage-specific reduction in mitochondrial protein-gene content, with all losses restricted to ribosomal protein genes and *sdh* genes—mirror those found in angiosperms. A survey of 280 angiosperm mitogenomes by Southern blot hybridization revealed a highly punctuated tempo of gene loss (and nuclear transfer); most branches

on a tree of these angiosperms sustained no losses, while a small minority of branches sustained many losses (Adams et al. 2002). Extensive sampling of additional gymnosperms is needed to establish whether gene loss is widespread in other lineages or restricted to *Welwitschia*.

As in angiosperms, mitochondrial rRNA gene content is invariant in gymnosperms, whereas tRNA gene content varies widely (Table 1; Table S1). *Ginkgo* and *Cycas* have retained 23 and 27 tRNA genes for 16 and 18 amino acids, respectively, while the *Welwitschia* mitogenome contains only eight tRNA genes for eight amino acids. These numbers do not take into account the possibility that some of these tRNA genes—in particular, those derived from plastids—may not be functional. This issue is most relevant for *Welwitschia*, as two of its three plastid-derived tRNA genes are rarely present in angiosperm mtDNAs (See Figure S3 of Skippington et al. 2015). Thus, *Welwitschia* mtDNA may contain no more than six functional tRNA genes. *Ginkgo trnP-AGG* is unusual in that there are no orthologs in any other genome sequence currently in GenBank except *Cycas* (where it is unannotated), suggesting either a *de novo* tRNA originating within gymnosperms or a spurious prediction. It is possible that the adenosine is deaminated to inosine, enabling the recognition of both CCY proline codons. The presence of *Ginkgo trnG-UCC* is also unusual, as only a single ortholog can be found among vascular plants, in *Huperzia* (also unannotated). This tRNA is widespread among non-vascular plants and green algae, however, suggesting widespread loss among most vascular plant lineages.

Major loss of cis*-spliced introns and novel* trans*-spliced introns in* Welwitschia

Mitochondrial intron content shows much the same evolutionary pattern as gene content, with stasis in *Cycas* and *Ginkgo* but numerous losses in *Welwitschia* (Table 1; Figure 1B). The *Cycas* mitogenome contains 26 introns; this includes the entire set of introns inferred for the common ancestor of angiosperms plus an additional intron in *rps3* that was present in the common ancestor of vascular plants but lost in the lineage leading to angiosperms (Ran et al. 2010; Liu et al. 2012; Bonavita and Regina 2015). *Ginkgo* contains all but one of the introns found in *Cycas*, whereas the *Welwitschia*

mitogenome has retained only 10 introns, three *cis*-spliced and seven *trans*-spliced. The nad1i477 intron was previously shown to be retained in the gnetophytes *Gnetum* and *Ephedra*, indicating that this intron was lost specifically from the *Welwitschia* lineage (Gugerli et al. 2001). Complete mitogenomes from *Gnetum* and *Ephedra* will be necessary to assess the uniqueness of other *Welwitschia* intron losses. Although occasional cases of intron loss have been noted among the >70 sequenced angiosperm mitogenomes, only a single case of *Welwitschia*-like extensive loss of introns has been reported, in the common ancestor of the Geraniaceae (Park et al. 2015). Two of the three *cis*-spliced introns retained by *Welwitschia* correspond to two of the five retained by the Geraniaceae ancestor. This may be coincidence, or it may reflect the presence of some as-yet-unidentified functional element present widely in seed plants in these two introns that prevents their loss.

The *trans*-spliced arrangements for cox2i691 and nad1i728 are novel among gymnosperms, demonstrating a conversion from a *cis* to *trans* configuration in the *Welwitschia* lineage. As *matR* is a free-standing gene in *Welwitschia* but located within nad1i728 in *Ginkgo* (Figure S1), this intron presumably underwent two breakages in *Welwitschia* (Figure S2). Consistent with this hypothesis, the free-standing *Welwitschia matR* is still flanked by segments of the nad1i728 intron that are homologous to those flanking the gene in its *cis*-spliced arrangement in *Ginkgo*. This suggests that the *Welwitschia matR*-containing intron fragment still interacts with the 5′ and 3′ parts of the intron (i.e., the parts adjacent to exons 3 and 4, respectively) to fold into the proper secondary structure of this intron, although this remains to be tested experimentally. Double intron fractures, leading to tripartite *trans*-spliced introns, are extremely rare in evolution, with only two cases described previously, one in nad5i1477 in *Oenothera berteriana* (Knoop et al. 1997) and the other in a plastid intron in *Chlamydomonas reinhardtii* (Goldschmidt-Clermont et al. 1991). That being said, nad1i728 has been subject to far more single fractures than any other intron, having undergone 10 independent cases of fracture between its 5′ end and *matR* among over 400 examined land plants, and five fractures between *matR* and its 3′ end (Qiu and Palmer 2004). Thus, it is perhaps not surprising that this particular intron has fractured twice in a single lineage.

It is notable that no cases of intron gain have occurred in the three sequenced gymnosperm mitogenomes, especially considering that they represent three of the five lineages of extant gymnosperms. This parallels the situation in the far more extensively sampled angiosperms, where the only case of intron gain involves a homing group I intron in the *cox1* gene that has been horizontally transferred many times during angiosperm evolution (Cho et al. 1998; Sanchez-Puerta et al. 2008). The overall stasis in intron content in seed plants (with rare exception in *Welwitschia* and Geraniaceae) is in sharp contrast to the extensive intron flux, the result of many gains and losses, seen in lycophytes (Grewe et al. 2009; Hecht et al. 2011; Liu et al. 2012).

Ancestrally high editing in gymnosperms and massive loss of editing in Welwitschia

Variation in RNA editing content among gymnosperms (Table 1; Table S2) follows a similar pattern to gene and intron content. For each species, we predicted sites of C-to-U editing using the PREP-Mt online tool (Mower 2009) with a cutoff of 0.2. The number of predicted C-to-U edits across the entire coding regions of their shared 41 protein genes is similar for *Cycas* (1,214) and *Ginkgo* (1,306), whereas editing levels are much lower in *Welwitschia* (226 sites) due to lower editing frequency and gene loss (Table S2). Considering only the 29 protein genes shared by all three gymnosperms, there are 1,051 predicted edit sites in *Cycas* and 1,115 in *Ginkgo*, nearly five times as many as in *Welwitschia*. A total of 72% of *Cycas* sites, 68% of *Ginkgo* sites, and 61% of *Welwitschia* sites are shared with one or both of the other gymnosperms (Figure 2). The simplest interpretation of these results is that the majority of edit sites in each species were present in the gymnosperm common ancestor, while the species-specific sites are derived from specific gains in one lineage or convergent losses in the other two lineages. Additional sampling is needed to polarize the direction of change of this class of sites.

To assess the accuracy of these predictions, we compared them to empirical data generated by RT-PCR and cDNA sequencing (Table S3). For *Cycas*, 534 C-to-U edit sites were empirically identified in segments of 24 genes by Salmans et al. (2010), while no U-to-C sites were detected. In these same gene regions, 526 sites were predicted.

For *Ginkgo*, we empirically identified 659 C-to-U and no U-to-C edit sites in 17 genes, while 598 sites were predicted for the same regions. These comparisons demonstrate that edit-site prediction provides reliable estimates of RNA editing content in these two gymnosperm mitogenomes, consistent with results from previous studies (Mower 2005; Mower 2009; Salmans et al. 2010). The 9% under-counting of predicted edit sites in *Ginkgo* likely results from the inability to predict silent editing events, which comprise 8 to 15% of editing events in diverse angiosperms (Mower and Palmer 2006; Alverson et al. 2010; Picardi et al. 2010; Sloan et al. 2010). The much lower *Cycas* undercount (1.5%) reflects the unexpectedly low number of sites empirically identified for the *ccmC* gene by Salmans et al. (2010). The absence of U-to-C editing in all genes examined here provides further evidence that U-to-C editing may have been entirely lost, along with the underlying mechanism, in the common ancestor of seed plants (Guo et al. 2015).

We infer a high level of mitochondrial editing, between 700 and 1,300 sites, in the gymnosperm common ancestor. This is based on the >700 sites inferred for the angiosperm ancestor (Richardson et al. 2013) and the ~1,200–1,300 sites determined for *Ginkgo* and *Cycas*. The ancestral number in gymnosperms could be higher given the 1,700–2,100 edit sites determined for the lycophytes *Isoetes* and *Selaginella* (Grewe et al. 2011; Hecht et al. 2011) and the similarly heavy editing levels inferred for the hornworts *Megaceros* and *Phaeoceros* (Xue et al. 2010). If our inference is correct, then *Welwitschia* has experienced massive loss of editing. Loss of edit sites from an ancestrally high number has also occurred in many lineages of angiosperms (Mower 2008; Sloan et al. 2010; Richardson et al. 2013; Park et al. 2015). The massive loss of many edit sites and *cis*-spliced introns in *Welwitschia* is consistent with retroprocessing, the gene conversion of an unedited gene by an edited transcript (Ran et al. 2010; Grewe et al. 2011). This pattern is especially evident for five *Welwitschia* genes (*cox2, nad1, nad2, nad4,* and *nad5*) that have lost in aggregate nine *cis*-spliced introns (Figure 1) and which possess only 15 edit sites compared with the 247 sites shared between *Cycas* and *Ginkgo* (Table S2). Retroprocessing probably also affected the *Ginkgo cob* and *cox1* genes (which have only 10% the level of editing as the corresponding genes in *Cycas*), but data from other plants is required to confirm the evolutionary polarity of

these differences as loss in *Ginkgo* rather than gain in *Cycas*. Retroprocessing at a finer scale, by so-called “microconversions” (Sloan et al. 2010), may account for the many additional examples of edit site and/or intron loss.

Variation in nucleotide substitution rates and base composition

Using a concatenated set of 41 protein genes and a codon-based model of sequence evolution, we estimated relative and absolute rates of synonymous and nonsynonymous substitution for gymnosperm mitochondrial genes (Table 2). Following their divergence from the gymnosperm common ancestor, the *Welwitschia* lineage has experienced 5–6 times higher synonymous site divergence (d_S) and 7 times higher nonsynonymous site divergence (d_N) than the *Cycas* and *Ginkgo* lineages. Assuming a divergence time of 343 Ma for the crown-group age of gymnosperms (Table S4), absolute synonymous (R_S) and nonsynonymous (R_N) substitution rates for *Welwitschia* are estimated to be 0.87 and 0.37 substitutions per site per billion years, whereas the lower R_S and R_N values for *Cycas* and *Ginkgo* are comparable to the notoriously low rates of mitochondrial sequence evolution in most angiosperms (Wolfe et al. 1987; Palmer and Herbon 1988; Mower et al. 2007; Richardson et al. 2013). Thus, the higher rate of sequence evolution in *Welwitschia* is presumably the result of increased mutation pressure. Analyses of a few mitochondrial genes with denser taxonomic sampling (Mower et al. 2007) suggest that synonymous substitution rates in the mitochondria of some gymnosperms, such as *Ephedra* and *Podocarpus*, are likely to be even higher.

To assess intraspecific polymorphism in the *Welwitschia* mitogenome, we compared the Illumina assembly from the cultivated Nebraska individual to a partial (131 kb) fosmid-based assembly from a second individual cultivated in Germany (Figure S3). The fosmid and Illumina assemblies are identical in sequence except for four short tandem repeats with a varying number of repeat motifs that range in length from 3 to 22 bp. Inspection of the raw Illumina and fosmid reads verified that the number of repeat units was correctly inferred in the consensus of both individuals, indicating that these are *bona fide* copy-number variants segregating within the species. These results suggest that copy-number variants accumulate at an even higher rate than point mutations in the

Welwitschia genome. A similar result was obtained for the mitogenomes of two individuals of the lycophyte *Selaginella* (Hecht et al. 2011).

Because distinct rates of sequence evolution can drive changes in nucleotide composition, we examined patterns of guanine+cytosine content (GC%) in gymnosperms (Table 1). With rare exception, nearly all angiosperm mtDNAs have GC% of 43–46%. There are just three examined species that have higher values: *Butomus* (49.1%, Cuenca et al. 2013), *Viscum* (47.4%, Skippington et al. 2015), and *Liriodendron* (47.1%, Richardson et al. 2013). Among gymnosperms, *Picea's* GC% (44.7%) is similar to most angiosperms, whereas *Cycas* has a higher value (46.9%) than all but the above three angiosperms, and *Ginkgo* (50.4%) and *Welwitschia* (53.0%) have the most GC-rich mitogenomes of any land plants sequenced to date (Mower et al. 2012c; http://www.ncbi.nlm.nih.gov/genome/browse/?report=5). Given the general trend of AT-biased mutation that has been widely observed across eukaryotes (Haag-Liautard et al. 2008; Lynch et al. 2008; Denver et al. 2009; Keightley et al. 2009; Howe et al. 2010; Ossowski et al. 2010; Lee et al. 2012), the high GC% in the fast-evolving *Welwitschia* mitogenome is unexpected. This GC richness suggests that the mitochondrial mutational spectrum may instead be GC-biased in *Welwitschia*, although the mechanism driving this genome-wide increase in GC% is unclear.

Within protein-coding regions (Figure S4), 3rd position GC% is also higher in *Welwitschia* than in other gymnosperms, consistent with a mutational bias causing the higher genome-wide GC% in *Welwitschia*. However, GC% at 1st and 2nd codon positions is lower in *Welwitschia* compared with *Ginkgo* and *Cycas*. This is likely due to the considerable loss of editing in *Welwitschia*, which manifests as C to T changes within genes, predominantly at 1st and 2nd codon positions where most edit sites are located. In comparison with previous studies, which have demonstrated that high GC% correlates positively with editing frequency (Malek et al. 1996; Smith 2009; Hecht et al. 2011), the reduced editing in *Welwitschia* despite a higher genome-wide GC% is unexpected. On the other hand, the lower GC% at 1st and 2nd positions in *Welwitschia* than in *Cycas* and *Ginkgo* is consistent with the reduced editing in *Welwitschia*. This suggests that the action of retroprocessing is strong enough to maintain low editing

levels in *Welwitschia* even with a GC-biased mutational spectrum. More generally, it seems that GC% in coding regions, particularly at 1st and 2nd codon positions, may be a better indicator of editing abundance than genome-wide GC%.

Exceptionally low levels of mtDNA turnover in Ginkgo *and* Cycas

To investigate rates of DNA turnover in seed plants, we measured amounts of shared DNA among the three gymnosperm mtDNAs and many angiosperm mtDNAs (Figure 3A; Table S4). It is already well appreciated that angiosperm mtDNAs diverge very quickly at the structural level, resulting in a substantial loss of synteny and shared DNA among even closely related species (Knoop 2012; Mower et al. 2012c). For example, only 51% (205 kb) of the 401 kb mitogenome *of Vigna radiata* is homologous to the *Glycine max* mitogenome, even though the two species have a divergence time of just 19 Ma. Progressively lower amounts of DNA are shared when comparing more distantly related mitogenomes, as exemplified by the 154 kb of shared DNA between *Triticum* and *Zea* (with a divergence time of 54 Ma) and the 136–138 kb shared between *Liriodendron* and *Phoenix* (133 Ma). Consistent with this pattern, *Welwitschia* and *Cycas,* which diverged ~343 Ma, share only 36–38 kb of DNA. In stark contrast, the *Ginkgo* and *Cycas* mitogenomes have retained ~240 kb of shared DNA despite their similarly deep divergence time.

For broader perspective, we examined amounts of shared DNA in the quantitative context of divergence times (Figure 3B; Table S4) using phylogenetically independent pairs of seed plant taxa (Figure S5). We found a strong negative correlation (R^2 = 0.74) between divergence time and amount of shared DNA for 14 of the 16 pairs of seed plants examined. The other two pairs are such extreme outliers that they were not included in the regression calculation. One pair, as already suggested from the MAUVE plots, is *Ginkgo* and *Cycas*. Their genomes share more DNA than any other genome pair examined (Figure 3B, Table S4) despite their ancient split, which is 11–18 times deeper than for all other genome pairs with >200 kb of shared DNA (*Vigna*/*Glycine,* *Nicotiana*/*Capsicum*, *Citrullus*/*Cucurbita*). This low level of DNA turnover between *Ginkgo* and *Cycas* may be partly attributable to their woody tree habit and slow generation times; however, other woody perennials (*Liriodendron*/*Phoenix*,

Vitis/*Vaccinium*, *Malus*/*Ricinus*) do not exhibit similarly slow rates of DNA turnover. Furthermore, the also-woody gymnosperm *Welwitschia* fits the regression line quite well in genome-pair comparison with both *Cycas* (Figure 3B) and *Ginkgo* (not shown). The DNA shared by *Welwitschia* and the other two gymnosperms is restricted almost entirely to gene and intron sequences (<5 kb shared intergenic DNA), whereas *Ginkgo* and *Cycas* share ~170 kb of intergenic DNA.

The other outlier pair is extreme in the opposite direction. Rates of DNA turnover are exceptionally fast within *Silene* (Figure 3B, Table S4; Sloan et al. 2012a). Despite their divergence time of just 4.5 Ma, *S. latifolia* and *S. vulgaris* share only 80 kb of DNA, the majority of which (~50 kb) is shared gene or intron sequence, with only ~30 kb of shared intergenic DNA. This divergence time is 20–25 times more recent than in comparisons of other herbaceous plants with similar amounts of shared DNA (*Daucus*/*Helianthus*; *Butomus*/*Spirodela*) and approaching 100 times more recent than for *Ginkgo*/*Cycas*. Yet these two gymnosperms share three times as much DNA in total and six times as much intergenic DNA as the two *Silene* species. These analyses show that rates of sequence turnover are generally high in most seed plants, leading to progressively lower amounts of shared DNA over time, with dramatic exceptions involving exceptionally low rates of turnover in *Ginkgo* and *Cycas* and exceptionally high rates in *Silene*.

Limited repeat-mediated recombination in Ginkgo

The *Ginkgo* mitogenome contains two copies of three large, direct repeats of 5.3, 4.1, and 1.5 kb in size, and 31 medium repeats of 100–495 bp in size, of which all but two are also present in two copies (Figure S1A). To examine the recombinational activity of these repeats, we mapped reads from a 5-kb Illumina sequencing library to the genome and then calculated the proportion of read pairs that supported the alternative conformation (AC) of each repeat pair (Figure 4A). For the 5.3-kb and 4.1-kb repeats, roughly equal proportions of read pairs supported either the “master chromosome” genomic arrangement (Palmer and Shields 1984) or the AC resulting from homologous recombination. This suggests a multipartite structure for this genome involving a master

chromosome and three sets of subgenomic circles (Figure 4B), with most or all of these molecules potentially present at similar stoichiometries.

Recombinant forms of the smaller repeats were much less common (Figure 4A). For the 1.5-kb repeat, ~9% of read pairs supported the AC, while about half of the medium repeats had a very low level (<1%) of recombinants, suggesting that additional genomic arrangements exist at highly substoichiometric levels. Moreover, recombinant forms were not detected for the remaining medium repeats. Because the main genomic arrangements for all medium repeats were supported by >4,000 read pairs, the absence of detectable recombinants indicates that any recombination at these repeats must occur at <0.025% (1/4,000) frequency.

Welwitschia mtDNA contains 95 medium repeats up to 446 bp in length, all of them present in two copies, but no larger repeats (Figure S1B). Depth of sequencing coverage was much lower for this genome because total rather than mitochondrial-enriched DNA was used for sequencing. On average, only ~60 read pairs spanned the medium repeats. No recombinant forms were detected for any repeat, indicating that any recombination at these repeats must occur at <1.67% (1/60) frequency. No data are available on repeat recombination in *Cycas* (Chaw et al. 2008).

The frequency of recombination at medium repeats varies enormously among the relatively few angiosperm mtDNAs for which this has been measured. At one extreme, the tiny (66 kb) mtDNA of the parasitic plant *Viscum* is at more or less recombinational equilibrium (~1:1 stoichiometry) for its four medium repeats (of 387–593 bp), and even has relatively frequent recombination (AC frequency averages 5% and is as high as 24%) for its 33 short repeats of 50–100 bp (Skippington et al. 2015). At the other extreme, the enormous (6.7 and 11.3 Mb) mtDNAs of *Silene noctiflora* and *S. conica*, respectively, and the large (1.7 Mb) mtDNA of *Cucumis sativus* show no evidence of recombination at the vast majority of their thousands of medium repeats (Alverson et al. 2011; Sloan et al. 2012a). In between these extremes (in order from more to less recombinationally active) are the mtDNAs of *S. vulgaris, Vigna angularis,* and *Mimulus guttatus* (Mower et al. 2012a; Sloan et al. 2012a; Naito et al. 2013).

Relative to the angiosperm spectrum, the frequency of repeat-mediated recombination in *Ginkgo* is low, somewhere between that in *Mimulus* and the three recombinationally quiescent large-genome species. Another indication that recombination is relatively quiescent in *Ginkgo* is that its large repeat of 1.5 kb has an AC frequency of only 9%. Repeats larger than 1 kb are typically at or close to recombinational equilibrium in angiosperm mtDNAs (Maréchal and Brisson 2010; Woloszynska 2010). Indeed, the only genomes showing *Ginkgo*-like levels of recombination at repeats of ~1.5 kb in size are the notably inert genomes of *S. noctiflora*, *S. conica*, and *C. sativus* (Alverson et al. 2011; Sloan et al. 2012a). Both sets of comparisons indicate that the frequency of repeat-mediated recombination in *Ginkgo* mtDNA is rather low compared to all but the most recombination-deficient angiosperm mitogenomes. Our data raise the possibility of infrequent recombination in *Welwitschia* mtDNA too, but deeper sequencing is needed to properly test this hypothesis.

Independent proliferation of a putative transposable element in Ginkgo *and* Cycas

The *Cycas* mitogenome contains ~500 variants of a 36-bp, putatively mobile element, termed a Bpu element, which contains direct terminal repeats of length 4 bp (Chaw et al. 2008). Using the predominant Bpu sequence from *Cycas* as a BLAST query, we identified ~100 similar elements in *Ginkgo*, whereas *Welwitschia* contains only a single sequence of reduced sequence similarity (Figure 5A). The structure of the *Ginkgo* consensus Bpu (Figure 5B) is predicted to be very similar to the *Cycas* Bpu structure (Chaw et al. 2008), but they differ in the loop region. However, the additional 2-bp stem enclosing a 3-base loop that was proposed for the *Cycas* structure is unlikely to be sterically stable. If this 2-bp stem is opened, then the two structures would be identical.

Synteny analysis identified a single element, from *Cycas* and *Ginkgo*, whose location is shared among these three mitogenomes (Figure 5C). All other elements were located in unique positions in each genome; two representative cases are shown in Figure 5D. These findings suggest that the Bpu elements have expanded independently in *Cycas* and *Ginkgo*. Independent expansion is consistent with inferences by (Chaw et al. 2008) based on limited comparative data, as well as their hypothesis that Bpu sequences are

mobile, proliferative genetic elements. To our knowledge, major independent expansion of a repeat family in two or more lineages of land plant mitogenomes has not been reported. Whether this is for lack of such elements or lack of their investigation is unclear.

CONCLUSIONS

We have discovered extensive variation in many properties of gymnosperm mitogenomes. This level of diversity is particularly remarkable considering that only three gymnosperm mitogenomes have been completely sequenced. The *Ginkgo* and *Cycas* genomes exhibit a very conservative mode of evolution, as evidenced by moderate genome sizes, virtually no gene or intron losses, a large number of mostly shared RNA-edited sites, very low substitution rates, and extremely low rates of DNA turnover. Most of these properties could result, at least in part, from low levels of repeat-mediated rearrangements, as measured in *Ginkgo* but not yet investigated in *Cycas*. Reduced recombinational activity would be expected to limit the genome's capacity to shed DNA, whether intergenic regions, functional genes, or proliferating repeat elements. Reduced recombination would also limit the level of retroprocessing, thus hindering the ability to lose introns and edited sites.

The low rates of mitogenomic evolution in these two gymnosperms, together with data on angiosperms summarized in the Introduction, suggest that several of their features are primitive ones that characterized the ancestral seed plant mitogenome. This ancestor likely possessed a moderately sized (for seed plants) genome of ~400 kb, very low substitution rates, the same set of 41 protein-coding genes and 26 introns that are present in *Cycas*, and a large number (>700) of C-to-U edit sites.

In stark contrast to the conservative patterns of mitogenomic evolution in *Ginkgo* and *Cycas*, the *Welwitschia* mitogenome exhibits a number of distinctive and evolutionarily derived features. These include an expanded genome size, massive loss of genes and introns, novel *cis*-to-*trans* splicing intron arrangements, major reduction in RNA-editing abundance with concomitantly reduced GC content at 1st and 2nd codon positions, and accelerated substitution rates with an unusually biased spectrum that raised GC content

throughout the rest of the genome. It is noteworthy that some of these distinctive features (expanded size, elevated substitution rates, lost genes and introns, reduced RNA editing) have also been observed, to some degree or another, in other seed plant lineages, including Geraniaceae, *Plantago*, and certain species of *Silene* (Cho et al. 2004; Parkinson et al. 2005; Mower et al. 2007; Sloan et al. 2012a; Park et al. 2015). This syndrome of mitogenomic upheaval may be caused by perturbations of one or more common pathways involved in DNA replication, repair, and/or recombination, although the specific processes are unknown. However, some of these characteristics may evolve in an uncoupled fashion in response to unique underlying processes. In this regard, it is important to note that other seed plant genomes possess only a subset of these features. For example, the *Amborella* mitogenome is very large, but normal with respect to substitution rates and functional sequence content (Rice et al. 2013), while the high-rate and gene-poor *Viscum* mitogenome is unexceptional with respect to editing and is actually unusually small in size (Skippington et al. 2015).

With so much diversity found among so few examined gymnosperm mitogenomes, more gymnosperm mitogenome sequences are clearly needed to better understand the tempo and pattern of mtDNA evolution and recombination in this group, and to identify underlying mechanisms. There is real potential to uncover much larger genomes, particularly in Pinaceae, as already indicated by the multi-megabase draft mitogenome from *Picea* (Nystedt et al. 2013). It is likely that several lineages, such as *Ephedra* and *Podocarpus*, will prove to have even higher substitution rates than *Welwitschia*, as suggested by previous analysis of a few mitochondrial genes (Mower et al. 2007). Given the massive loss of functional complexity from the *Welwitschia* mitogenome, we may find even more extreme examples of gene, intron, or RNA editing loss in other gnetophytes. It is important to determine whether any gymnosperms have comparable or even lower rates of mtDNA turnover as *Ginkgo* and *Cycas taitungensis*. If so, then it may possible to reconstruct to a significant degree the gene order of the ancestral gymnosperm mitogenome.

It has long been recognized that rates of synonymous substitutions and gene loss/functional transfer to the nucleus vary enormously among lineages of angiosperms

(Adams et al. 2002; Cho et al. 2004; Parkinson et al. 2005; Mower et al. 2007). This study, which is the first to quantify rates of DNA turnover in seed plant mitogenomes, has uncovered equally provocative variation in turnover rates. Similar to the situation for substitution rates, rates of DNA turnover appear to be consistent in most seed plants with extreme departures for a small number of lineages. The low turnover rates in *Ginkgo* and *Cycas* contrasted with very high turnover rates in *Silene* should motivate more intensive examination of this phenomenon in many more seed plant mtDNAs, as well as the development of comparative methodology that allows turnover rates to be examined in a phylogenetic context. Finally, because recombination is a key factor that probably underlies many aspects of mtDNA evolution and diversity in seed plants, it is imperative that rates of genomic rearrangement and levels of repeat-mediated recombination be measured, across the fullest possible range of repeat sizes, in those genomes sequenced to a sufficiently high depth of coverage. This is particularly important in light of recent evidence that rearrangement and recombination rates may vary greatly in plant mitogenomes (Sloan et al. 2012a; Sloan et al. 2012b).

MATERIALS AND METHODS

Genome sequencing, assembly, and annotation

Leaf samples from *Ginkgo biloba* were collected from a single tree on the UNL campus. *Welwitschia mirabilis* leaves were collected from a single individual in the living collection at the Beadle Center greenhouses at UNL. Mitochondrial-enriched DNA of *Ginkgo* was isolated as described previously (Mower et al. 2010). Total cellular DNAs of *Ginkgo* and *Welwitschia* were extracted using a simplified CTAB procedure (Doyle 1987). The *Ginkgo* mtDNA sample was sequenced by BGI (Shenzhen, China) on an Illumina HiSeq 2000 machine from an 800-bp paired-end library, generating 23 million 2 x 100 bp reads. For *Ginkgo* total DNA, 80 million 2 x 100 bp reads were sequenced by BGI on an Illumina HiSeq 2000 machine from a 5-kb mate-pair library. Total DNA of *Welwitschia* was sequenced by ACGT, INC. (Wheeling, IL) on an Illumina NextSeq 500 machine, generating 38 million 2 x 150 bp reads from a 550 bp paired-end library.

Genome sequence data from both species were assembled with Velvet version 1.2.03 (Zerbino and Birney 2008). Multiple Velvet assemblies were constructed using different pairwise combinations of Kmer values and expected coverage values, as described previously (Grewe et al. 2014; Zhu et al. 2014). For all runs, the scaffolding option was turned off. For both *Ginkgo* and *Welwitschia*, mitochondrial contigs and plastid contigs were separately scaffolded into a single chromosome using either the paired-end or mate-pair library using SSPACE 3.0 (Boetzer et al. 2011). Remaining gaps in the *Ginkgo* mitogenome assembly were due to long mononucleotide repeats (10 to 20 bp in length) and were finished by Sanger sequencing.

Genes and introns were annotated using established procedures (Grewe et al. 2014; Zhu et al. 2014). Repeats and MIPTs were identified with a blast search using a word size of 7, an e-value of $1x10^{-6}$, and a minimal length of 100 bp. Blast results matching paralogs in the mitochondrial and plastid genomes (e.g., *atp1*/*atpA*, *rrn26*/*rrn23*, *rrn18*/*rrn16*) were excluded from the MIPT results. Tandem repeat sequences were identified using the web version (4.07b) of the Tandem Repeats Finder program with default parameters (Benson 1999). RNA edit sites were computationally predicted using the batch version of the PREP-Mt online server (Mower 2009), with a cutoff value of 0.2. Note that (Salmans et al. 2010) used different cutoffs for their edit-site predictions in *Cycas*, which probably accounts for the slight differences in their results vs. ours. Edit sites were experimentally determined for 15 genes from *Ginkgo* by RT-PCR and cDNA sequencing, as described previously (Hepburn et al. 2012; Rice et al. 2013; Richardson et al. 2013). All genomes generated in this study were deposited in GenBank under accession numbers KM672373 (*Ginkgo* mitogenome), KP099648 (*Ginkgo* plastome), KT313400 (*Welwitschia* mitogenome), and KT347148 (*Welwitschia* plastome).

Fosmid cloning

Green leaves of a second *Welwitschia mirabilis* individual were obtained from the University of Bonn's Botanical Garden. Total genomic DNA was isolated using the CTAB protocol. Fosmid cloning and the selection and sequencing of mitochondrial fosmid clones were performed as described previously (Grewe et al. 2009; Hecht et al. 2011). Five overlapping clones with insert sizes ranging from 28,334 to 41,104 bp were

sequenced. These were used to build a continuous sequence totaling 130,590 bp in length. The near identity (see Results and Discussion) of this sequence and the Illumina mitogenome assembly demonstrates the reliability of both approaches in generating high-quality genome sequences.

Estimation of nucleotide substitution rates

Forty-one mitochondrial protein genes were obtained from *Cycas*, *Ginkgo*, and *Welwitschia* and, as outgroups, three slowly evolving and gene-rich angiosperms (*Amborella trichopoda*, *Liriodendron tulipifera*, and *Vitis vinifera*) and two lycophytes (*Huperzia squarrosa* and *Isoetes engelmannii*). Genes were individually aligned with MUSCLE version 3.8.31 (Edgar 2004) using default parameters. To mitigate the confounding effects of C-to-U RNA editing on substitution-rate calculations, the predicted edited cytosines were converted to thymines in the sequence alignments. Poorly aligned regions were eliminated using Gblocks version 0.91 b (Castresana 2000) with relaxed settings (-t=c, -b1=h, -b2=h, -b4=5, -b5=h). Filtered alignments were concatenated with FASconCAT version 1.0 (Kück and Meusemann 2010), generating a final alignment of 34,080 bp. Trees representing synonymous (d_S) and nonsynonymous (d_N) branch lengths were calculated in HyPhy version 2.2 (Pond et al. 2005) with a local codon model by using the MG94CustomCF3x4 model crossed with a HKY85 nucleotide model. Topologies were constrained according to version 13 of the Angiosperm Phylogeny Website (http://www.mobot.org/mobot/research/apweb/). A divergence time of 343 Ma for the crown group age of gymnosperms was determined by averaging results from four studies (Table S4). Absolute rates of synonymous substitution (R_S) were calculated by summing the d_S branch lengths from the common ancestor of gymnosperms to the tips and dividing by 343 Ma. Absolute nonsynonymous rates (R_N) were calculated similarly by summing d_N branches and dividing by 343 Ma.

Genome alignments and shared DNA analysis

To identify mitochondrial synteny blocks between pairs of representative species, a total of 16 pairs of seed plant mitogenomes were aligned using Mauve version 2.3.1 with a LCB cutoff of 500 (Darling et al. 2010). To determine the amount of mtDNA shared between species, each pair of mitogenomes was aligned using blastn with a word size

of 7 and an e-value cutoff of $1x10^{-6}$. Using these parameters, the blastn searches should be able to detect homologous sequences as short as 30 bp (perfect match) or 33 bp (one mismatch).

Repeats and repeat-mediated recombination

The frequency of repeat-mediated recombination was evaluated for all repeats ≥100 bp in the *Ginkgo* and *Welwitschia* mitogenomes using previously described procedures (Mower et al. 2012a). Read pairs were classified as consistently mapping pairs if they mapped to the genome sequence in the expected orientation and distance (+/- 50%) based on the sequencing library type and insert size. To determine the number of read pairs supporting the arrangement of each repeat in the genome sequence, the number of consistently mapping pairs that spanned each repeat were counted. Inconsistently mapping read pairs were classified as those that did not map to the genome sequence in the expected orientation and/or distance. These inconsistently mapping read pairs were then checked to determine whether they map consistently to a recombinant form of the genome produced by recombination at a repeat. Read pairs supporting recombinant forms were counted for each repeat. Recombination frequency was calculated by dividing the number of recombinant read pairs by the total number of read pairs spanning the repeat in any arrangement.

Identification of Bpu elements

Bpu-like elements in the *Ginkgo* and *Welwitschia* mitogenomes were initially identified by blastn using the *Cycas* Bpu consensus sequence as a query and a word size of 7. A single match was identified in *Welwitschia*, while the top full-length matches from *Ginkgo* were used to establish a *Ginkgo* Bpu consensus sequence. The *Welwitschia* sequence and the *Ginkgo* consensus sequence were then used in a second blastn run to search for additional variants in both genomes. All Bpu-homologous sequences ≥27 bp in length were counted as Bpu-like elements. To compare the synteny of Bpu insertion sites among *Cycas*, *Ginkgo* and *Welwitschia*, the three mitogenomes were pairwise aligned using blastn with an e-value cut-off of 0.001 and a word size of 7. Bpu-like elements were then identified in each of the high-scoring segment pairs of length

≥100 bp. Secondary structures were predicted using the RNA-folding form of the mFold webserver (Zuker 2003).

ACKNOWLEDGMENTS

The authors gratefully acknowledge Patricia Viehoever and Bernd Weisshaar for support with fosmid sequencing and Gaven Nelson for assistance with annotation of the *Welwitschia* plastid genome. This work was supported by the National Science Foundation (awards IOS-1027529 to JPM and JDP and MCB-1125386 to JPM), the Negaunee Foundation (postdoctoral support to FG), and the University of Nebraska (start-up funds to JPM and a School of Biological Sciences Research Award to WG).

REFERENCES

Adams KL, Qiu YL, Stoutemyer M, Palmer JD. 2002. Punctuated evolution of mitochondrial gene content: High and variable rates of mitochondrial gene loss and transfer to the nucleus during angiosperm evolution. Proc Natl Acad Sci USA 99: 9905-9912.

Allen JO, Fauron CM, Minx P, Roark L, Oddiraju S, Lin GN, Meyer L, Sun H, Kim K, Wang C, et al. 2007. Comparisons among two fertile and three male-sterile mitochondrial genomes of maize. Genetics 177: 1173-1192.

Alverson AJ, Rice DW, Dickinson S, Barry K, Palmer JD. 2011. Origins and recombination of the bacterial-sized multichromosomal mitochondrial genome of cucumber. Plant Cell 23: 2499-2513.

Alverson AJ, Wei X, Rice DW, Stern DB, Barry K, Palmer JD. 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.

Benson G. 1999. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res 27: 573-580.

Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. 2011. Scaffolding pre-assembled contigs using SSPACE. Bioinformatics 27: 578-579.

Bonavita S, Regina T. 2015. The evolutionary conservation of *rps3* introns and *rps19-rps3-rpl16* gene cluster in *Adiantum capillus-veneris* mitochondria. Curr Genet: 1-12.

Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol 17: 540-552.

Chaw S-M, Chun-Chieh Shih A, Wang D, Wu Y-W, Liu S-M, Chou T-Y. 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.

Cho Y, Mower JP, Qiu YL, Palmer JD. 2004. Mitochondrial substitution rates are extraordinarily elevated and variable in a genus of flowering plants. Proc Natl Acad Sci USA 101: 17741-17746.

Cho Y, Qiu Y-L, Kuhlman P, Palmer JD. 1998. Explosive invasion of plant mitochondria by a group I intron. Proc Natl Acad Sci USA 95: 14244-14249.
Cuenca A, Petersen G, Seberg O. 2013. The complete sequence of the mitochondrial genome of *Butomus umbellatus* – a member of an early branching lineage of monocotyledons. PLoS ONE 8: e61552.
Darling AE, Mau B, Perna NT. 2010. progressiveMauve: Multiple genome alignment with gene gain, loss and rearrangement. PLoS ONE 5: e11147.
Davila JI, Arrieta-Montiel MP, Wamboldt Y, Cao J, Hagmann J, Shedge V, Xu YZ, Weigel D, Mackenzie SA. 2011. Double-strand break repair processes drive evolution of the mitochondrial genome in *Arabidopsis*. BMC Biol 9: 64.
Denver DR, Dolan PC, Wilhelm LJ, Sung W, Lucas-Lledó JI, Howe DK, Lewis SC, Okamoto K, Thomas WK, Lynch M, et al. 2009. A genome-wide view of *Caenorhabditis elegans* base-substitution mutation processes. Proc Natl Acad Sci USA 106: 16310-16314.
Doyle JJ. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull 19: 11-15.
Drouin G, Daoud H, Xia J. 2008. Relative rates of synonymous substitutions in the mitochondrial, chloroplast and nuclear genomes of seed plants. Mol Phylogenet Evol 49: 827-831.
Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32: 1792-1797.
Goldschmidt-Clermont M, Choquet Y, Girard-Bascou J, Michel F, Schirmer-Rahire M, Rochaix J-D. 1991. A small chloroplast RNA may be required for *trans*-splicing in *Chlamydomonas reinhardtii*. Cell 65: 135-143.
Goremykin VV, Lockhart PJ, Viola R, Velasco R. 2012. The mitochondrial genome of *Malus domestica* and the import-driven hypothesis of mitochondrial genome expansion in seed plants. Plant J 71: 615-626.
Grewe F, Edger PP, Keren I, Sultan L, Pires JC, Ostersetzer-Biran O, Mower JP. 2014. Comparative analysis of 11 Brassicales mitochondrial genomes and the mitochondrial transcriptome of *Brassica oleracea*. Mitochondrion 19, Part B: 135-143.
Grewe F, Herres S, Viehöver P, Polsakiewicz M, Weisshaar B, Knoop V. 2011. A unique transcriptome: 1782 positions of RNA editing alter 1406 codon identities in mitochondrial mRNAs of the lycophyte *Isoetes engelmannii*. Nucleic Acids Res 39: 2890-2902.
Grewe F, Viehoever P, Weisshaar B, Knoop V. 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.
Gugerli F, Sperisen C, Buchler U, Brunner I, Brodbeck S, Palmer JD, Qiu YL. 2001. The evolutionary split of Pinaceae from other conifers: evidence from an intron loss and a multigene phylogeny. Mol Phylogenet Evol 21: 167-175.
Guo W, Grewe F, Mower JP. 2015. Variable frequency of plastid RNA editing among ferns and repeated loss of uridine-to-cytidine editing from vascular plants. PLoS ONE 10: e0117075.
Haag-Liautard C, Coffey N, Houle D, Lynch M, Charlesworth B, Keightley PD. 2008. Direct estimation of the mitochondrial DNA mutation rate in *Drosophila melanogaster*. PLoS Biol 6: e204.

Hecht J, Grewe F, Knoop V. 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.
Hepburn NJ, Schmidt DW, Mower JP. 2012. Loss of two introns from the *Magnolia tripetala* mitochondrial *cox2* gene implicates horizontal gene transfer and gene conversion as a novel mechanism of intron loss. Mol Biol Evol 29: 3111-3120.
Howe DK, Baer CF, Denver DR. 2010. High rate of large deletions in *Caenorhabditis briggsae* mitochondrial genome mutation processes. Genome Biol Evol 2: 29-38.
Keightley PD, Trivedi U, Thomson M, Oliver F, Kumar S, Blaxter ML. 2009. Analysis of the genome sequences of three *Drosophila melanogaster* spontaneous mutation accumulation lines. Genome Res 19: 1195-1201.
Knoop V. 2012. Seed plant mitochondrial genomes: Complexity evolving. In: Bock R, Knoop V, editors. Genomics of Chloroplasts and Mitochondria: Springer Netherlands. p. 175-200.
Knoop V, Altwasser M, Brennicke A. 1997. A tripartite group II intron in mitochondria of an angiosperm plant. Mol Gen Genet 255: 269-276.
Kück P, Meusemann K. 2010. FASconCAT: Convenient handling of data matrices. Mol Phylogenet Evol 56: 1115-1118.
Lee H, Popodi E, Tang H, Foster PL. 2012. Rate and molecular spectrum of spontaneous mutations in the bacterium *Escherichia coli* as determined by whole-genome sequencing. Proc Natl Acad Sci USA 109: E2774-E2783.
Liu Y, Wang B, Cui P, Li L, Xue J-Y, Yu J, Qiu Y-L. 2012. The mitochondrial genome of the lycophyte *Huperzia squarrosa*: The most archaic form in vascular plants. PLoS ONE 7: e35168.
Lynch M, Sung W, Morris K, Coffey N, Landry CR, Dopman EB, Dickinson WJ, Okamoto K, Kulkarni S, Hartl DL, et al. 2008. A genome-wide view of the spectrum of spontaneous mutations in yeast. Proc Natl Acad Sci USA 105: 9272-9277.
Malek O, Lattig K, Hiesel R, Brennicke A, Knoop V. 1996. RNA editing in bryophytes and a molecular phylogeny of land plants. EMBO J 15: 1403-1411.
Maréchal A, Brisson N. 2010. Recombination and the maintenance of plant organelle genome stability. New Phytol 186: 299-317.
Micallef L, Rodgers P. 2014. euler*APE*: drawing area-proportional 3-Venn diagrams using ellipses. PLoS One 9: e101717.
Mower J. 2005. PREP-Mt: predictive RNA editor for plant mitochondrial genes. BMC Bioinformatics 6: 96.
Mower J, Palmer J. 2006. Patterns of partial RNA editing in mitochondrial genes of *Beta vulgaris*. Mol Gen Genet 276: 285-293.
Mower JP. 2008. Modeling sites of RNA editing as a fifth nucleotide state reveals progressive loss of edited sites from angiosperm mitochondria. Mol Biol Evol 25: 52-61.
Mower JP. 2009. The PREP suite: predictive RNA editors for plant mitochondrial genes, chloroplast genes and user-defined alignments. Nucleic Acids Res 37: W253-259.
Mower JP, Case AL, Floro ER, Willis JH. 2012a. 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.

Mower JP, Jain K, Hepburn NJ. 2012b. The role of horizontal transfer in shaping the plant mitochondrial genome. In: Laurence M-D, editor. Advances in Botanical Research: Academic Press. p. 41-69.
Mower JP, Sloan DB, Alverson AJ. 2012c. Plant mitochondrial genome diversity: the genomics revolution. In: Wendel JF, editor. Plant Genome Diversity Volume 1: Springer. p. 123-144.
Mower JP, Stefanovic S, Hao W, Gummow JS, Jain K, Ahmed D, Palmer JD. 2010. Horizontal acquisition of multiple mitochondrial genes from a parasitic plant followed by gene conversion with host mitochondrial genes. BMC Biol 8: 150.
Mower JP, Touzet P, Gummow JS, Delph LF, Palmer JD. 2007. Extensive variation in synonymous substitution rates in mitochondrial genes of seed plants. BMC Evol Biol 7: 135.
Naito K, Kaga A, Tomooka N, Kawase M. 2013. De novo assembly of the complete organelle genome sequences of azuki bean (*Vigna angularis*) using next-generation sequencers. Breeding Sci 63: 176-182.
Nystedt B, Street NR, Wetterbom A, Zuccolo A, Lin Y-C, Scofield DG, Vezzi F, Delhomme N, Giacomello S, Alexeyenko A, et al. 2013. The Norway spruce genome sequence and conifer genome evolution. Nature 497: 579-584.
Ossowski S, Schneeberger K, Lucas-Lledó JI, Warthmann N, Clark RM, Shaw RG, Weigel D, Lynch M. 2010. The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*. Science 327: 92-94.
Palmer JD, Herbon LA. 1988. Plant mitochondrial DNA evolved rapidly in structure, but slowly in sequence. J Mol Evol 28: 87-97.
Park S, Grewe F, Zhu A, Ruhlman TA, Sabir J, Mower JP, Jansen RK. 2015. Dynamic evolution of *Geranium* mitochondrial genomes through multiple horizontal and intracellular gene transfers. New Phytol: doi: 10.1111/nph.13467.
Parkinson CL, Mower JP, Qiu YL, Shirk AJ, Song KM, Young ND, dePamphilis CW, Palmer JD. 2005. Multiple major increases and decreases in mitochondrial substitution rates in the plant family Geraniaceae. BMC Evol Biol 5: 73.
Picardi E, Horner DS, Chiara M, Schiavon R, Valle G, Pesole G. 2010. Large-scale detection and analysis of RNA editing in grape mtDNA by RNA deep-sequencing. Nucleic Acids Res 38: 4755-4767.
Pond SLK, Frost SDW, Muse SV. 2005. HyPhy: hypothesis testing using phylogenies. Bioinformatics 21: 676-679.
Qiu Y-L, Palmer J. 2004. Many independent origins of *trans* splicing of a plant mitochondrial group II intron. J Mol Evol 59: 80-89.
Ran JH, Gao H, Wang XQ. 2010. Fast evolution of the retroprocessed mitochondrial *rps3* gene in Conifer II and further evidence for the phylogeny of gymnosperms. Mol Phylogenet Evol 54: 136-149.
Rice DW, Alverson AJ, Richardson AO, Young GJ, Sanchez-Puerta MV, Munzinger J, Barry K, Boore JL, Zhang Y, dePamphilis CW, et al. 2013. Horizontal transfer of entire genomes via mitochondrial fusion in the angiosperm *Amborella*. Science 342: 1468-1473.
Richardson AO, Rice DW, Young GJ, Alverson AJ, Palmer JD. 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.

Rodriguez-Moreno L, Gonzalez V, Benjak A, Marti MC, Puigdomenech P, Aranda M, Garcia-Mas J. 2011. Determination of the melon chloroplast and mitochondrial genome sequences reveals that the largest reported mitochondrial genome in plants contains a significant amount of DNA having a nuclear origin. BMC Genomics 12: 424.
Salmans ML, Chaw S-M, Lin C-P, Shih AC-C, Wu Y-W, Mulligan RM. 2010. Editing site analysis in a gymnosperm mitochondrial genome reveals similarities with angiosperm mitochondrial genomes. Curr Genet 56: 439-446.
Sanchez-Puerta MV, Cho Y, Mower JP, Alverson AJ, Palmer JD. 2008. Frequent, phylogenetically local horizontal transfer of the *cox1* group I intron in flowering plant mitochondria. Mol Biol Evol 25: 1762-1777.
Skippington E, Barkman TJ, Rice DW, Palmer JD. 2015. Miniaturized mitogenome of the parasitic plant *Viscum scurruloideum* is extremely divergent and dynamic and has lost all nad genes. Proc Natl Acad Sci USA 112: E3515-E3524.
Sloan DB, Alverson AJ, Chuckalovcak JP, Wu M, McCauley DE, Palmer JD, Taylor DR. 2012a. Rapid evolution of enormous, multichromosomal genomes in flowering plant mitochondria with exceptionally high mutation rates. PLoS Biol 10: e1001241.
Sloan DB, MacQueen AH, Alverson AJ, Palmer JD, Taylor DR. 2010. Extensive loss of RNA editing sites in rapidly evolving *Silene* mitochondrial genomes: selection vs. retroprocessing as the driving force. Genetics 185: 1369-1380.
Sloan DB, Müller K, McCauley DE, Taylor DR, Štorchová H. 2012b. Intraspecific variation in mitochondrial genome sequence, structure, and gene content in *Silene vulgaris*, an angiosperm with pervasive cytoplasmic male sterility. New Phytol 196: 1228-1239.
Smith DR. 2009. Unparalleled GC content in the plastid DNA of *Selaginella*. Plant Mol Biol 71: 627-639.
Wolfe KH, Li WH, Sharp PM. 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proc Natl Acad Sci USA 84: 9054-9058.
Woloszynska M. 2010. Heteroplasmy and stoichiometric complexity of plant mitochondrial genomes—though this be madness, yet there's method in't. J Exp Bot 61: 657-671.
Wu B, Hao W. 2014. Horizontal transfer and gene conversion as an important driving force in shaping the landscape of mitochondrial introns. G3 Genes Genom Genet 4: 605-612.
Xue JY, Liu Y, Li L, Wang B, Qiu YL. 2010. The complete mitochondrial genome sequence of the hornwort *Phaeoceros laevis*: retention of many ancient pseudogenes and conservative evolution of mitochondrial genomes in hornworts. Curr Genet 56: 53-61.
Zerbino DR, Birney E. 2008. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. Genome Res 18: 821-829.
Zhu A, Guo W, Jain K, Mower JP. 2014. Unprecedented heterogeneity in the synonymous substitution rate within a plant genome. Mol Biol Evol 31: 1228-1236.
Zuker M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res 31: 3406-3415.

TABLES

Table 1. General features of gymnosperm mitogenomes

	Cycas	*Ginkgo*	*Welwitschia*
Accession	AP009381	KM672373	KT313400
Size (bp)	414,903	346,544	978,846
GC%	46.9	50.4	53.0
Genes	71	67	40
tRNAs	27	23	8
rRNAs	3	3	3
Protein coding	41	41	29
Introns	26	25	10
Predicted edit sites	1214	1306	226
Repeats (kb)	67	33	29
Tandem repeats (kb)	22	3.6	24
Plastid-derived (kb)	18	0.5	8.5

Table 2. Synonymous and nonsynonymous substitution rates

Branch	Time (Ma)	d_N (sub/site)	R_N (sub/site/Ga)	d_S (sub/site)	R_S (sub/site/Ga)	d_N/d_S
Welwitschia	343	0.127	0.371	0.298	0.868	0.43
Cycas	343	0.017	0.050	0.048	0.140	0.36
Ginkgo	343	0.018	0.052	0.062	0.180	0.29

FIGURE LEGENDS

Figure 1. Mitochondrial gene and intron content in all three sequenced gymnosperms and in four angiosperms with ancestral-like gene and intron content. A) Gene content. "●" indicates presence of an intact gene, "Ψ" indicates a pseudogene, and "-" indicates gene loss. The first row marks 24 genes (*atp1*, *atp4*, *atp6*, *atp8*, *atp9*, *ccmB*, *ccmC*, *ccmFc*, *ccmFn*, *cob*, *cox1*, *cox2*, *cox3*, *matR, mttB*, *nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5*, *nad6*, *nad7*, and *nad9*) that are shared by all seven genomes and which are present universally, or nearly so, among >350 examined angiosperm mitogenomes (see Adams et al. (2002) and all sequenced genomes). B) Intron content. "●" and "ɵ" indicate presence of *cis*- and *trans*-spliced introns, respectively, with the one tripartite (double *trans*-spliced) intron highlighted by grey shading, "-" indicates intron loss *per se*, and "x" indicates intron loss due to gene loss (cf. Figure 1A).

Figure 2. Venn diagram depicting the numbers of shared and unique edit sites in gymnosperms. The circles and their overlaps were drawn proportional to the numbers in each category using eulerAPE v3 (Micallef and Rodgers 2014).

Figure 3. Shared mtDNA among seed plants. A) MAUVE alignments of five pairs of seed plants at varying evolutionary depths. Left curved arrows give the amount of mtDNA, in kb and on a percentage basis, in the plant listed that is shared with the second, while right arrows give the reciprocal values. B) Pairwise amounts of shared DNA as a function of divergence time. The regression line is based only on the blue points; the species in each of these comparisons, their genome sizes, and other relevant information are given in Table S4. The two red points: Gi-Cy, *Ginkgo vs*. *Cycas*; Sl-Sv, *Silene latifolia vs. S. vulgaris* SD2 (also see Table S4).

Figure 4. Recombinational activity in the *Ginkgo* mitogenome. A) Recombination frequency of all repeats >100 bp. B) Inferred multipartite structure due to high-frequency recombination of the two largest repeats in *Ginkgo*. Circular maps are exactly proportional to the size of the indicated master and subgenomic circles. Note that while these circles correspond to the genome assembly, they do not necessarily reflect the *in vivo* state of the *Ginkgo* mitogenome (Bendich 1996; Sloan 2013).

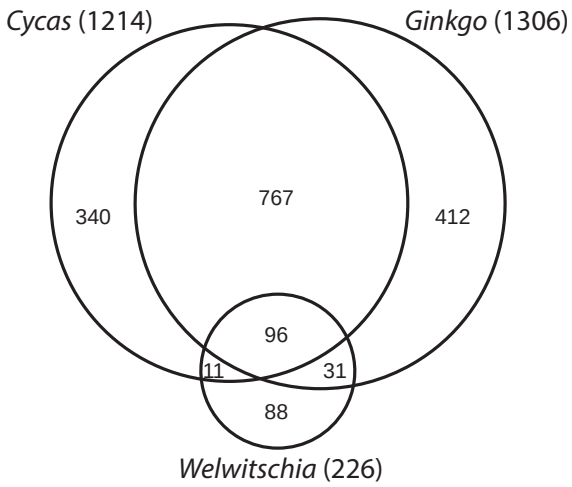
Figure 5. Bpu-like elements in gymnosperm mitogenomes. A) Comparison of consensus sequences of Bpu-like elements in *Cycas*, *Ginkgo,* and *Welwitschia*. B) Proposed secondary structure of the consensus Bpu sequences from *Ginkgo* (this study) and *Cycas* (Chaw et al. 2008, but see text). C) The only Bpu-like element that is present in the same location in *Cycas* and *Ginkgo.* D) Examples of genome-specific positions of Bpu-like elements in *Cycas* vs. *Ginkgo*. Numbers in parentheses are genome coordinates. Shading in C) and D) indicates Bpu-like elements.

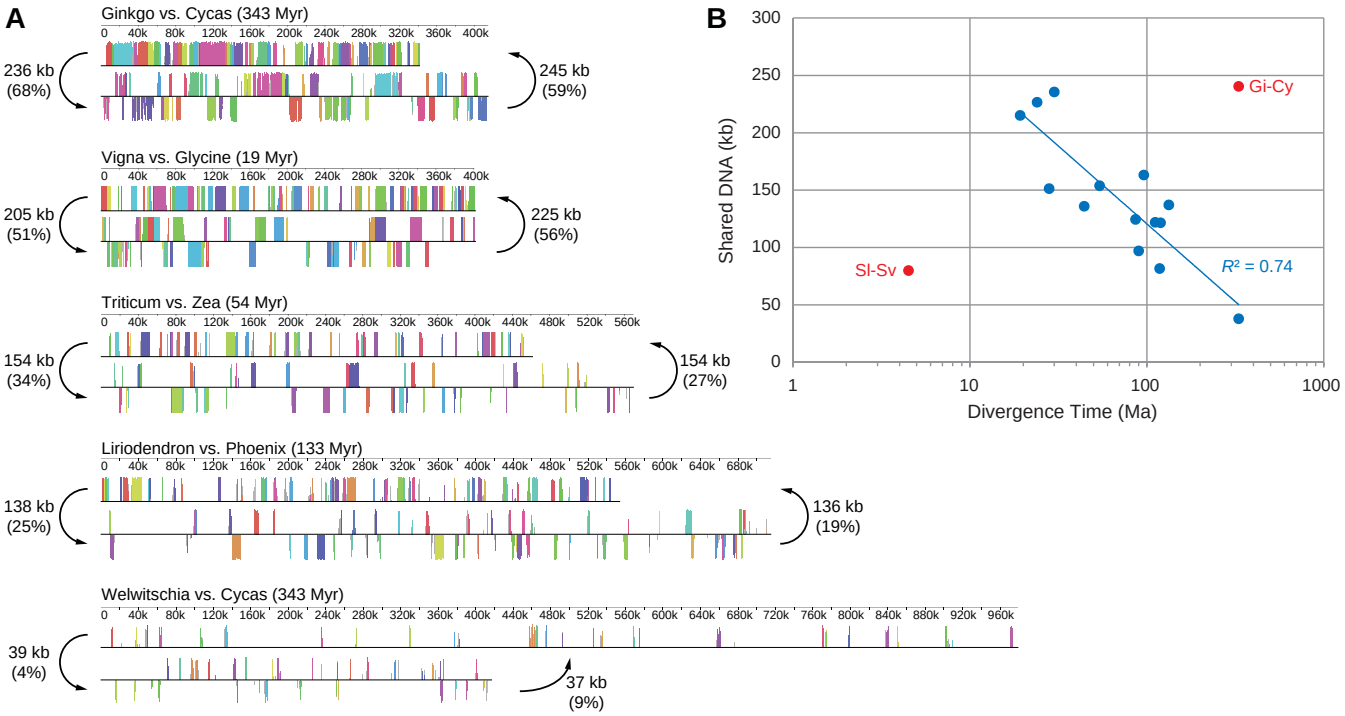
A

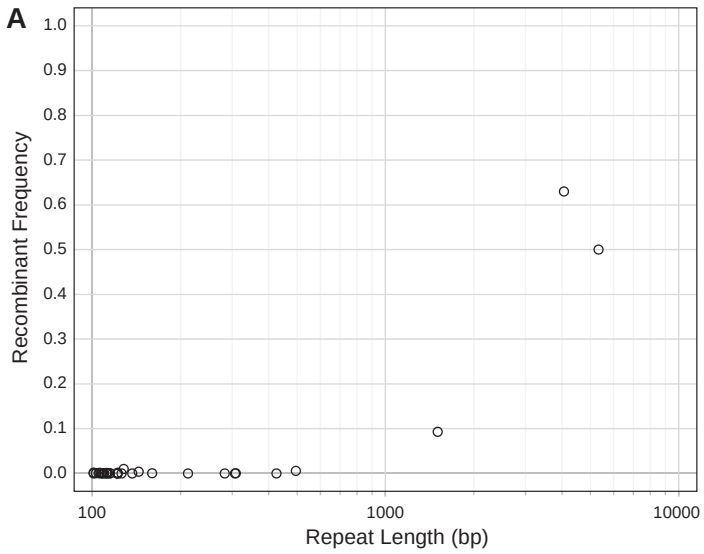
Gene	GYMNO			ANGIO			
	Cycas	Ginkgo	Welwitschia	Amborella	Liriodendron	Phoenix	Vitis
24 genes	●	●	●	●	●	●	●
rpl2	●	●	–	●	●	●	●
rpl5	●	●	–	●	●	●	●
rpl10	●	●	●	–	●	●	●
rpl16	●	●	ψ	●	●	●	●
rps1	●	●	–	●	●	●	●
rps2	●	●	–	●	●	●	–
rps3	●	●	●	●	●	●	●
rps4	●	●	●	●	●	●	●
rps7	●	●	–	●	●	●	●
rps10	●	●	–	●	●	●	●
rps11	●	●	–	●	●	●	–
rps12	●	●	●	●	●	●	●
rps13	●	●	–	●	●	●	●
rps14	●	●	–	●	●	●	●
rps19	●	●	–	●	●	●	●
sdh3	●	●	–	●	●	–	●
sdh4	●	●	●	●	●	ψ	●
total (●)	41	41	29	40	41	39	39

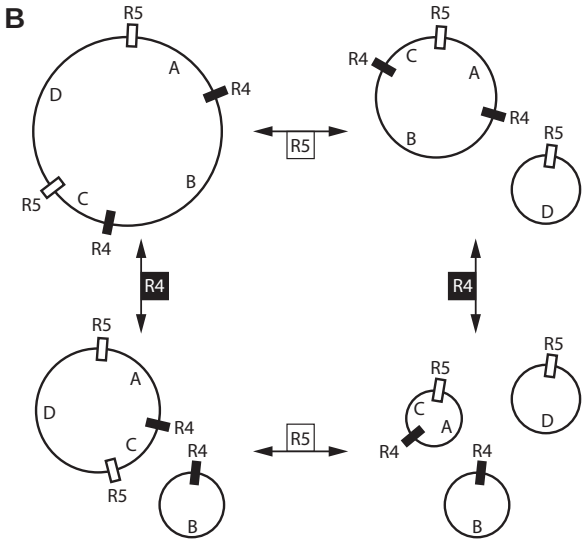
B

Intron	GYMNO			ANGIO			
	Cycas	Ginkgo	Welwitschia	Amborella	Liriodendron	Phoenix	Vitis
ccmFci829	●	●	–	●	●	●	●
cox2i373	●	●	–	●	●	●	●
cox2i691	●	●	ɵ	●	●	●	●
nad1i394	ɵ	ɵ	ɵ	ɵ	ɵ	ɵ	ɵ
nad1i477	●	●	–	●	●	●	●
nad1i669	ɵ	ɵ	ɵ	ɵ	ɵ	ɵ	ɵ
nad1i728	●	●	ɵ	ɵ	●	ɵ	●
nad2i156	●	●	–	●	●	●	●
nad2i542	ɵ	ɵ	ɵ	ɵ	ɵ	ɵ	ɵ
nad2i709	●	●	–	●	●	●	●
nad2i1282	●	●	–	●	●	●	●
nad4i461	●	●	●	●	●	●	●
nad4i976	●	●	–	●	●	●	●
nad4i1399	●	●	–	●	●	●	●
nad5i230	●	●	–	●	●	●	●
nad5i1455	ɵ	ɵ	ɵ	ɵ	ɵ	ɵ	ɵ
nad5i1477	ɵ	ɵ	ɵ	ɵ	ɵ	ɵ	ɵ
nad5i1872	●	●	–	●	●	●	●
nad7i140	●	●	–	●	●	●	●
nad7i209	●	●	●	●	●	●	●
nad7i676	●	●	–	●	●	●	●
nad7i917	●	●	●	●	●	●	●
rpl2i917	●	●	x	●	●	●	●
rps3i74	●	●	–	●	●	●	●
rps3i249	●	●	–	–	–	–	–
rps10i235	●	–	x	●	●	●	●
total *cis* (●)	21	20	3	19	20	19	20
total *trans* (ɵ)	5	5	7	6	5	6	5


Cycas (1214)
Ginkgo (1306)
340
767
412
96
11
31
88
Welwitschia (226)


A
Ginkgo vs. Cycas (343 Myr)
0 40k 80k 120k 160k 200k 240k 280k 320k 360k 400k
236 kb
(68%)
245 kb
(59%)
Vigna vs. Glycine (19 Myr)
0 40k 80k 120k 160k 200k 240k 280k 320k 360k 400k
205 kb
(51%)
225 kb
(56%)
Triticum vs. Zea (54 Myr)
0 40k 80k 120k 160k 200k 240k 280k 320k 360k 400k 440k 480k 520k 560k
154 kb
(34%)
154 kb
(27%)
Liriodendron vs. Phoenix (133 Myr)
0 40k 80k 120k 160k 200k 240k 280k 320k 360k 400k 440k 480k 520k 560k 600k 640k 680k
138 kb
(25%)
136 kb
(19%)
Welwitschia vs. Cycas (343 Myr)
0 40k 80k 120k 160k 200k 240k 280k 320k 360k 400k 440k 480k 520k 560k 600k 640k 680k 720k 760k 800k 840k 880k 920k 960k
39 kb
(4%)
37 kb
(9%)
B
Shared DNA (kb)
300
250
200
150
100
50
0
1
10
100
1000
Divergence Time (Ma)
Gi-Cy
Sl-Sv
R² = 0.74

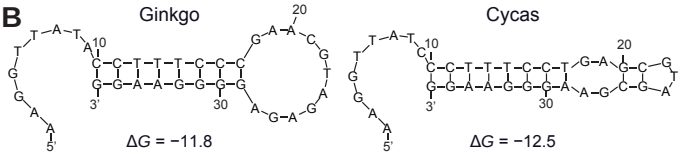

A
Recombinant Frequency
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
100
1000
10000
Repeat Length (bp)


B
R5
A
R4
D
B
C
R5
R4
R5
R4
C
A
R4
B
R5
D
R4
R4
R5
A
D
R4
C
R4
R5
B
R5
R5
C
A
R4
R4
B
R5
D

A

```
                  10        20        30
                  ....|....|....|....|....|....|....|.
Cycas             AAGGTTATCCCTTTCCTGAGCGTAGCGAAGGGAAGG
Ginkgo            ........A.......C..A.....A..G.......
Welwitschia       ..TA-...G.......CT.CACGC....G.......
```

B



C

```
Cycas     TCTTTTTCTTTCGAATCTCCAGGTTATCCCATAGTTTTACTCCGCGTAGCGAAGGGAAGGCATGGGCGGCGCTAAGTTCA   (208033–207924)
Ginkgo    .........C.T.................A..--------C.CGA.........G.....................   (45121–45195)
```

D

```
Cycas     GACAATCAACGGAACAATCTATTAGGTTATCCCTTTCCTGAGCGTAGCGAAGGGAAGGGCACGACGTCTCTTCTCTCCGG   (183842–183763)
Ginkgo    ..............C-------------------------------------C..A.....................   (122389–122349)

Cycas     TGGGATCAGA----CCATCCTTGA-------------------------------GGCAAACAATGCAATAGCTTCGGT   (185008–185051)
Ginkgo    ..A....GA.AACG.....A...GGTTATACCTTTCCCGAACGTAGAGAGGGGAA...GG........G.........   (123609–123688)
```