History of Plastid DNA Insertions Reveals Weak Deletion and AT Mutation Biases in Angiosperm Mitochondrial Genomes

Daniel B. Sloan* and Zhiqiang Wu

Department of Biology, Colorado State University, Fort Collins

*Corresponding author: E-mail: dbsloan@rams.colostate.edu.

Accepted: November 13, 2014

Data deposition: No new sequence data were generated, but GenBank accessions are provided for all analyzed genomes.

Abstract

Angiosperm mitochondrial genomes exhibit many unusual properties, including heterogeneous nucleotide composition and exceptionally large and variable genome sizes. Determining the role of nonadaptive mechanisms such as mutation bias in shaping the molecular evolution of these unique genomes has proven challenging because their dynamic structures generally prevent identification of homologous intergenic sequences for comparative analyses. Here, we report an analysis of angiosperm mitochondrial DNA sequences that are derived from inserted plastid DNA (*mtpts*). The availability of numerous completely sequenced plastid genomes allows us to infer the evolutionary history of these insertions, including the specific nucleotide substitutions and indels that have occurred because their incorporation into the mitochondrial genome. Our analysis confirmed that many *mtpts* have a complex history, including frequent gene conversion and multiple examples of horizontal transfer between divergent angiosperm lineages. Nevertheless, it is clear that the majority of extant *mtpt* sequence in angiosperms is the product of recent transfer (or gene conversion) and is subject to rapid loss/deterioration, suggesting that most *mtpts* are evolving relatively free from functional constraint. The evolution of mtpt sequences reveals a pattern of biased mutational input in angiosperm mitochondrial genomes, including an excess of small deletions over insertions and a skew toward nucleotide substitutions that increase AT content. However, these mutation biases are far weaker than have been observed in many other cellular genomes, providing insight into some of the notable features of angiosperm mitochondrial architecture, including the retention of large intergenic regions and the relatively neutral GC content found in these regions.

Key words: chloroplast, indel bias, intracellular gene transfer, mutational spectrum, plant mitochondria.

Introduction

A classic challenge in the field of molecular evolution is to identify the effects of mutation bias and separate them from other evolutionary forces that shape genome sequence and structure. For example, the nearly universal tendency for endosymbiotic and organelle genomes to shrink in size (McCutcheon and Moran 2012) has been interpreted as a consequence of the widespread mutation bias that favors deletions over insertions (Mira et al. 2001; Kuo and Ochman 2009) in combination with the relaxed selection pressures that accompany an obligately intracellular lifestyle. A related hypothesis is that variation in the magnitude of this deletion bias can be an important determinant of genome size (Petrov 2002). Mutation biases also act at the level of individual nucleotide substitutions, and it was long believed that mutation biases were a major determinant of genome-wide GC content—a view that has been brought into question by recent evidence of a widespread bias toward mutations that increase AT content even in species with relatively GC-rich genomes (Hershberg and Petrov 2010; Hildebrand et al. 2010; Van Leuven and McCutcheon 2012).

The enigmatic genomes of angiosperm mitochondria exhibit a number of unusual features and represent a particularly intriguing system for studying mechanisms of molecular evolution (Knoop et al. 2011; Mower et al. 2012). Their rates of nucleotide substitution in coding genes are among the slowest ever observed (Mower et al. 2007; Richardson et al. 2013), yet the rates of structural rearrangements and sequence gain/loss are so high that intergenic regions are often unrecognizable among or even within closely related species (Kubo and Newton 2008; Darracq et al. 2011; Sloan, Müller, et al. 2012). Angiosperm mitochondrial genomes also harbor

enormous quantities of intergenic DNA that contribute to their exceptionally large and variable genome sizes, which range from approximately 200 kb to over 10 Mb (Palmer and Herbon 1987; Sloan, Alverson, Chuckalovcak, et al. 2012). Remarkably, almost this entire range can be found within very closely related groups of species (Ward et al. 1981; Sloan, Alverson, Chuckalovcak, et al. 2012). Angiosperm mitochondrial DNA (mtDNA) also exhibits heterogeneous nucleotide composition. For reasons that are not understood, the GC content of synonymous sites in mitochondrial-coding genes (~33%) is far lower than in the copious intergenic regions (~44%) (Sloan and Taylor 2010).

Estimating mutation biases has been particularly difficult in angiosperm mitochondrial genomes, as neither of the two main methods to infer the rate and spectrum of mutations is particularly well suited to these genomes. The gold standard for measuring mutations is to observe changes appearing across generations in mutation accumulation (MA) lines, in which populations are repeatedly bottlenecked to remove/reduce the effects of selection (Denver et al. 2000). However, such studies are laborious and normally restricted to species with short generation times. An additional limitation of using MA lines to study mitochondrial genomes is that bottlenecking is only performed at the organismal level. Therefore, there is still opportunity for selection to act on the multiple copies of the mitochondrial genome that co-occur within a cell (Taylor et al. 2002; Clark et al. 2012). In the absence of MA studies, a second method for measuring mutation relies on the classic molecular evolution principle that the substitution rate in neutrally evolving sequences is simply equal to the rate of mutation (Kimura 1983). The challenge in this method lies in identifying suitable neutral sequences. Synonymous positions in protein-coding genes are the most commonly used class of sites. However, it is widely understood that these sites are not truly neutral (Chamary et al. 2006), and they are irrelevant for measuring any type of change other than point mutations. Intergenic regions have also been used as a source of relatively neutral sequence (Petrov et al. 2000; Kuo and Ochman 2009). However, the rapid structural evolution in angiosperm mitochondrial genomes makes it difficult to identify homologous intergenic regions across species and reconstruct the corresponding ancestral states.

One possible answer to these challenges is to take advantage of the frequent influx of "promiscuous" DNA into angiosperm mitochondrial genomes, which have been found to contain sequences from diverse foreign sources (Ellis 1982; Alverson et al. 2011; Rice et al. 2013). Previous studies have demonstrated the utility of analyzing insertions of mitochondrial and plastid DNA to identify nucleotide-substitution and indel biases in eukaryotic nuclear genomes (Bensasson, Petrov, et al. 2001; Huang et al. 2005; Noutsos et al. 2005; Rousseau-Gueutin et al. 2011; Hsu et al. 2014). Sequences of plastid origin are particularly abundant in angiosperm mtDNA, providing an opportunity to conduct similar analyses in mitochondrial genomes. In rare cases, mitochondrial sequences of plastid origin (known as *mtpts*) have taken on important mitochondrial functions or have been incorporated into existing mitochondrial genes (Dietrich et al. 1996; Nakazono et al. 1996; Hao and Palmer 2009; Sloan et al. 2010; Wang et al. 2012), but there is good reason to believe that most *mtpts* and other interorganellar DNA transfers are effectively neutral (Bensasson, Zhang, et al. 2001; Cummings et al. 2003). In particular, the large variation in the amount and identity of *mtpts* among and even within species suggests that they are frequently gained and lost (Allen et al. 2007; Alverson et al. 2010; Sloan, Müller, et al. 2012). Fortunately, even when these transferred sequences are not widely maintained in the mitochondrial genome across species, they can be compared against the plastid genomes themselves, which are highly conserved in flowering plants and have been subject to extensive sequencing efforts. Phylogenetic analysis of these data sets can be used to date individual transfers and infer the history of subsequent indels and nucleotide substitutions (Bensasson et al. 2003; Wang et al. 2007; Hazkani-Covo et al. 2010).

Here, we employ such an approach in analyzing angiosperm species with sequenced mitochondrial and plastid genomes. The evolution of *mtpt* sequences reveals evidence for mutational biases favoring deletions and substitutions that increase AT content, but the magnitude of these biases is relatively weak. We discuss the impact of these findings on our understanding of the unusual genome architecture of plant mitochondria.

Materials and Methods

Genome Sequences and Identification of *mtpts*

We identified 31 angiosperm species for which both mitochondrial and plastid genomes were available on GenBank as of November 2013 (table 1 and fig. 1). We also included the lone gymnosperm (*Cycas taitungensis*) for which both organelle genomes had been completely sequenced. In cases where multiple sequences were available from the same species, we arbitrarily chose the first published sequence. To identify *mtpts*, each mitochondrial genome was searched against the corresponding plastid genome (after removing the second copy of the large inverted repeat) with NCBI-BLASTN v2.2.24+, using the following parameters: –task blastn –dust no –word_size 7 –evalue 1e-10. Hits were filtered to exclude the mitochondrial genes *atp1*, *rrn18*, and *rrn26*, which retain detectable nucleotide sequence homology with their respective orthologs in plastid genomes (Hao and Palmer 2009). Adjacent BLAST hits were merged into a single fragment as long as they were in the same orientation and separated by a gap of no more than 100 bp in both the mitochondrial and plastid genomes.

Table 1
Summary of *mtpt* Content by Species

Species	GenBank Accessions		*mtpts* Fragments (minimum 200 bp)			
	Mitochondrial	Plastid	Count	Total Length (kb)	mtDNA Coverage (%)	Plastid DNA Coverage (%)[a]
Amborella trichopoda	KF754799–KF754803	NC_005086	70	130.5	3.4	87.2
Arabidopsis thaliana	NC_001284	NC_000932	5	2.9	0.8	2.3
Bambusa oldhamii	EU365401	NC_012927	27	40.2	7.9	32.5
Beta vulgaris	NC_002511	EF534108	4	6.8	1.8	5.5
Boea hygrometrica	NC_016741	NC_016468	43	52.6	10.3	40.9
Brassica napus	NC_008285	NC_016734	6	7.7	3.5	6.1
Brassica rapa	NC_016125	NC_015139	7	8.0	3.6	6.2
Carica papaya	NC_012116	NC_010323	13	21.3	4.5	16.1
Cucumis melo	JF412792, JF412800	NC_015983	21	30.4	1.3	22.7
Cucumis sativus	NC_016004-NC_016006	NC_007144	35	68.1	4.0	53.2
Cycas taitungensis	NC_010303	NC_009618	7	17.2	4.2	11.6
Daucus carota	NC_017855	NC_008325	8	6.6	2.4	4
Glycine max	NC_020455	NC_007942	7	2.6	0.6	1.1
Liriodendron tulipifera	NC_021152	NC_008326	17	26.3	4.7	18.4
Lotus japonicus	NC_016743	NC_002694	8	4.8	1.3	3
Millettia pinnata	NC_016742	NC_016708	3	2.3	0.5	1.9
Nicotiana tabacum	NC_006581	NC_001879	14	10.2	2.4	7.7
Oryza rufipogon	NC_013816	NC_017835	32	33.1	5.9	16.7
Oryza sativa	NC_007886	NC_008155	27	36.3	7.4	21.3
Phoenix dactylifera	NC_016740	NC_013991	35	68.4	9.6	51.2
Ricinus communis	NC_015141	NC_016736	6	4.9	1.0	3.7
Silene conica	JF750490-JF750629	NC_016729	42	24.4	0.2	18.2
Silene latifolia	NC_014487	NC_016730	3	1.1	0.4	0.8
Silene noctiflora	JF750431-JF750489	NC_016728	21	7.0	0.1	5.4
Silene vulgaris	JF750427-JF750430	NC_016727	6	9.0	2.1	7.2
Sorghum bicolor	NC_008360	NC_008602	17	26.7	5.7	22.2
Spirodela polyrhiza	NC_017840	NC_015891	15	8.1	3.5	6.2
Triticum aestivum	NC_007579	NC_002762	12	11.9	2.6	8.9
Vigna angularis	NC_021092	NC_021091	2	0.6	0.1	0.5
Vigna radiata	NC_015121	NC_013843	3	1.1	0.3	0.9
Vitis vinifera	NC_012119	NC_007957	23	66.8	8.6	47.3
Zea mays	NC_007982	NC_001666	12	22.9	4.0	19.4

[a]Plastid DNA coverage was calculated after excluding one copy of the large inverted repeat.

Alignment of *mtpts* and Homologous Sequences from Plastid Genomes

For each *mtpt* of at least 200 bp in length, homologous sequences in the set of 32 seed plant plastid genomes were identified and extracted based on NCBI-BLASTN searches. Sequences that covered less than 80% of the length of the *mtpt* were excluded. Each *mtpt* was aligned against the resulting set of extracted plastid sequences with MUSCLE v3.7 (Edgar 2004), using default parameters.

Phylogenetic Analysis

To infer the timing of plastid-to-mitochondrial transfers, each *mtpt*/plastid alignment was used to construct a maximum-likelihood tree with RAxML v8.0.0 under a GTRGAMMA model (Stamatakis 2014). To ensure sufficient signal for phylogenetic inference, only *mtpts* of at least 500 bp in length were included in the analysis. The resulting tree topologies were parsed to identify the location of the *mtpt* branch to infer when it diverged from the plastid genome. Horizontal gene transfer from other plants has also occurred in a number of angiosperm mitochondrial genomes (Bergthorsson et al. 2003; Rice et al. 2013). In cases in which such transfers involved plastid-derived sequence, the phylogenetic placement of the *mtpt* branch was also used to infer the donor lineages.

The above phylogenetic analyses examined each extant *mtpt* individually. To identify *mtpts* that were present in multiple species and potentially derived from a single ancestral event, we combined the phylogenetic data with an all-versus-all BLAST strategy. Using NCBI-BLASTN, each *mtpt*

from the phylogenetic analyses was searched against all other *mtpts* and each plastid genome, identifying clusters of *mtpts* that were more similar to each other than to any plastid genome. To avoid double-counting substitutions and indels, we only used a single sequence from families of shared *mtpts* in subsequent analysis of overall mutation biases in angiosperm mitochondrial genomes.

Indel and Substitution Analysis

We identified indels and nucleotide substitutions that have occurred in *mtpts* since their transfer to the mitochondrial genome by comparing each aligned *mtpt* against the corresponding set of plastid sequences. Only alignments with at least ten plastid sequences were included in this analysis. We excluded *mtpts* associated with the ancient transfer of the region containing the tRNA genes *trnW* and *trnP*, which are now expressed and functional in seed plant mitochondria (other transfers that are known to have taken on a functional role in mitochondrial genome were already excluded based on the 200 bp minimum threshold).

To ensure that we accurately identified substitutions that occurred in the *mtpts*, we excluded alignment positions that exhibited polymorphisms among the plastid genome sequences. The remaining alignment positions were screened for substitutions differentiating the *mtpt* from the conserved plastid genome sequence. The resulting data were used to produce *mtpt*-specific substitution matrices and to calculate predicted equilibrium GC content based on the following equation,

$$GC_{eq} = \frac{F_{AT\rightarrow GC}}{F_{AT\rightarrow GC} + F_{GC\rightarrow AT}}$$

in which $F_{AT\rightarrow GC}$ is the fraction of all ancestral A or T sites that were converted to a G or C in the *mtpt*, and $F_{GC\rightarrow AT}$ is the reverse.

The same set of alignments was used to identify derived small indels (<100 bp) in each *mtpt*. To avoid ambiguity in ancestral state reconstruction, we excluded *mtpt* indels that overlapped with an indel that was polymorphic among the set of plastid genome sequences. Excluding overlapping indels introduces a potential bias against detecting deletions. Because deletions have two breakpoints, they are more likely than insertions (which only have a single breakpoint) to overlap with polymorphic indels. To negate this bias, we also excluded insertions with a neighboring polymorphic indel found within half the length of the insertion in the flanking sequence on either side of the insertion site.

Pearson correlation analyses were performed in R v3.0.2 to assess the relationships between indel bias and mitochondrial genome size and between observed and equilibrium GC content (R Core Team 2014). Phylogenetically independent contrasts were generated with the APE package in R (Paradis et al. 2004).

Data Analysis Scripts

The main data analysis steps including downloading genome sequences from GenBank, parsing BLAST output, extracting sequences, and identification of variants in sequence alignments were performed with custom Perl scripts that incorporated BioPerl modules (Stajich et al. 2002). Graphics for select figures were generated in with custom R scripts. Code is available from the authors upon request.

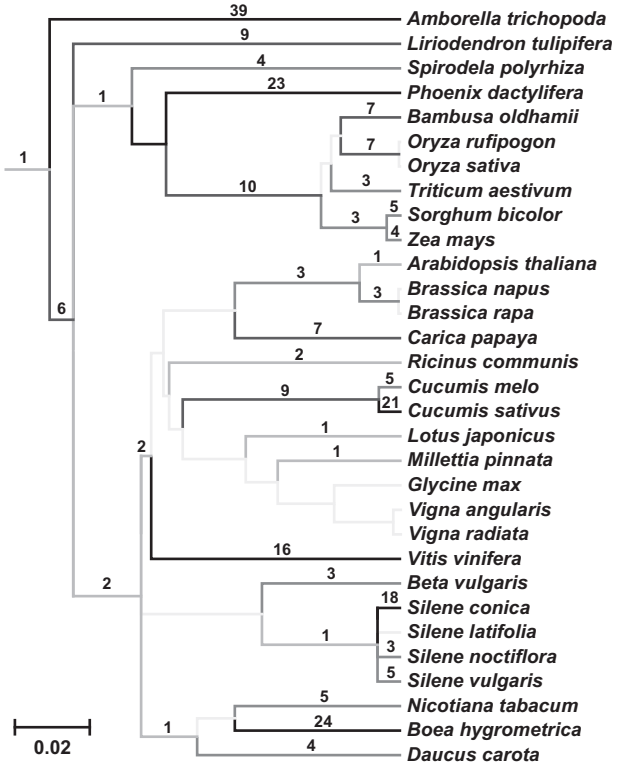

Fig. 1.—Phylogenetic origins of *mtpt* fragments (minimum 500 bp). Darker shading indicates a larger total number of *mtpt* fragments, with the specific count noted above each branch. Branch lengths were estimated based on a concatenation of four plastid genes (*matK*, *psaA*, *psaB*, and *rbcL*) using a GTR (REV) substitution model and a molecular clock constraint in baseml (Yang 2007). The reference tree for this figure was based on a maximum-likelihood topology with splits among *Silene* species and among asterids, caryophyllids, and rosids both collapsed into polytomies.

Results

mtpt Content in Angiosperms

Our analysis of mitochondrial and plastid genome sequences confirmed that there is tremendous variation in the amount of *mtpts* found in different angiosperm species (table 1 and fig. 2). The total length of *mtpt* sequence ranged from less than 1 kb in *Vigna angularis* to more than 130 kb in *Amborella*

trichopoda. When mapped back against their corresponding plastid DNA sequences, these fragments cover anywhere from 0.5% to 87.2% of the plastid genome (table 1 and fig. 2). Plastid-derived sequences accounted for less than 1% of many angiosperm mitochondrial genomes. At the other extreme, they represent 10.3% of the *Boea hygrometrica* mitochondrial genome. An even higher percentage has been reported for the mitochondrial genome of *Cucurbita pepo* (Alverson et al. 2010), but this species was not included in our study because it lacks a sequenced plastid genome. These values are based on identified fragments of at least 200 bp in length, but including smaller fragments does not increase the totals substantially.

The largest *mtpt* fragment was 12.6 kb in length (found in *Zea mays*), but there was clear evidence that some of the existing sequences were part of larger transfers that were subsequently broken up by large deletions and rearrangements. For example, the *Silene conica* mitochondrial genome contains two *mtpt* fragments from a 35-kb region of plastid DNA (fig. 3). Although these fragments are now located on different chromosomes in the *S. conica* mitochondrial genome, their corresponding boundaries precisely abut in the plastid genome, suggesting that they were derived from a single transfer that was subsequently split by a rearrangement. This transfer appears to have occurred relatively recently because it shares the derived inversion found in the *S. conica* plastid genome (Sloan, Alverson, Wu, et al. 2012). The transferred 35 kb sequence has been reduced to only 18 kb by a series of 23 large deletions ranging from 96 to 4,615 bp in size. Many of these were likely associated with a microhomology-mediated repair process (Deriano and Roth 2013), as 14 of the 23 deletions show small regions (7–18 bp) of sequence similarity between the pair of deletion breakpoints (fig. 3).

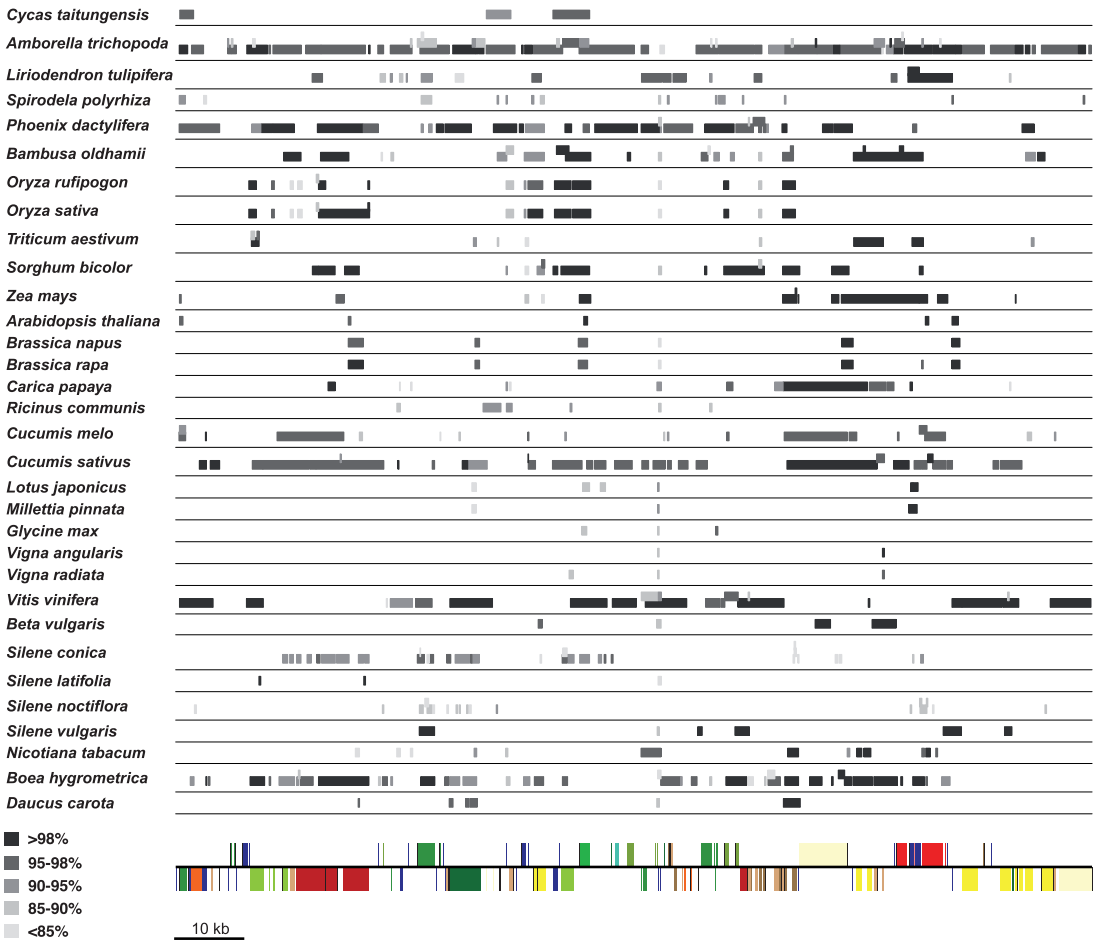

Fig. 2.—Origins of *mtpts* from the plastid genome. The location of each *mtpt* fragment (minimum 200 bp) within the plastid genome. Shading indicates nucleotide sequence identity (excluding gaps) relative to the corresponding plastid sequence. The *Nicotiana tabacum* plastid genome was used as a reference for defining position. The map of the *N. tabacum* plastid genome at the bottom of the figure was generated with OGDRAW v1.2 (Lohse et al. 2007) after removing the second copy of the inverted repeat.

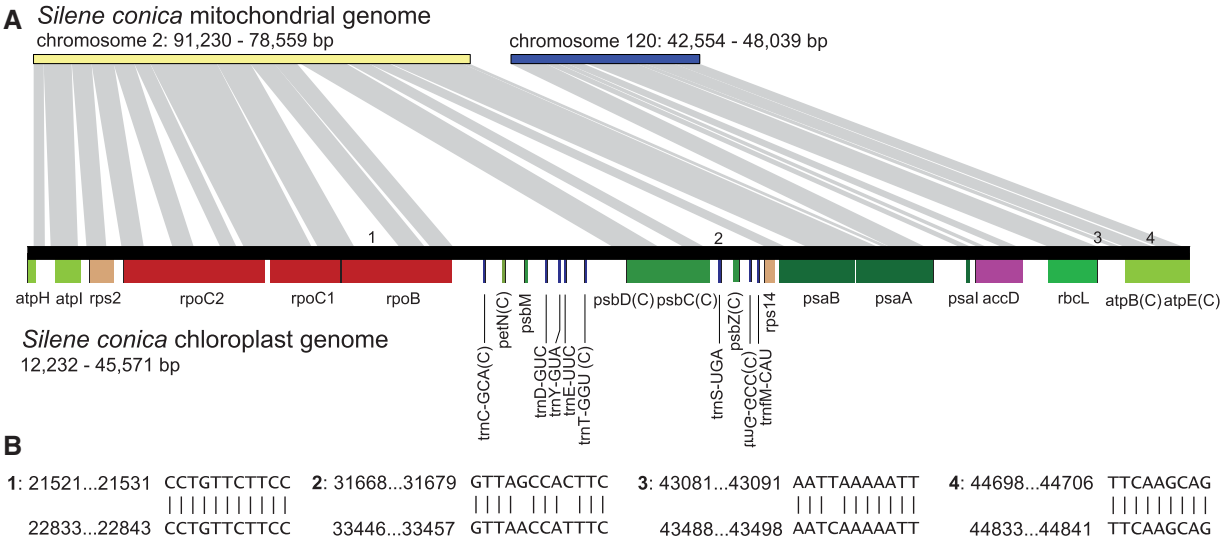

Fig. 3.—Structural rearrangements and large deletions in *S. conica mtpt*. (*A*) Gray connections indicate stretches of homology between two different mitochondrial chromosomes and a contiguous region in the plastid genome. Plastid gene names followed by "(C)" are found on the complementary strand. Only *mtpt* fragments of at least 200 bp in length are shown, but adjacent small fragments extend the boundaries to positions 10,764 and 91,440 in the plastid genome and mitochondrial chromosome 2, respectively. The plastid gene map was generated with OGDRAW v1.2 (Lohse et al. 2007). (*B*) Alignments show representative examples of microhomology in the plastid sequences that correspond to the *mtpt* deletion breakpoints. Numbering (1–4) corresponds to the deletions labeled in part (*A*).

These results are consistent with the view that large stretches of DNA can undergo intracellular transfer followed by a process of fragmentation and decay (Clifton et al. 2004; Richly and Leister 2004; Wang et al. 2007). The relationship between *mtpt* length and sequence identity with the plastid genome (fig. 4) provides additional support for this interpretation. The largest *mtpt* fragments all remain nearly identical to the corresponding plastid genome sequence, suggesting they were the products of very recent transfers. However, the relationship between *mtpt* length and sequence identity is not a simple positive one (fig. 4). Instead, it appears to follow a bounded distribution, in which fragments of any size can exhibit high levels of sequence conservation (but only short fragments exhibit high divergence). This pattern suggests that the initial transfers from plastid to mitochondrial genomes can span a wide size range.

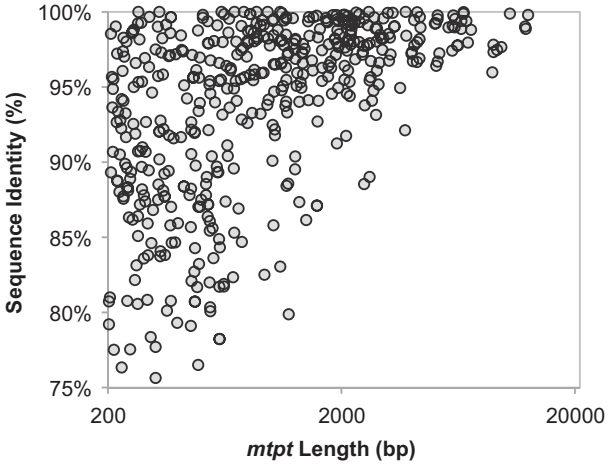

Fig. 4.—The relationship between the length of *mtpt* fragments and their nucleotide sequence identity (excluding gaps) with the corresponding plastid sequence.

History of *mtpt* Transfers

A phylogenetic analysis of all *mtpts* of at least 500 bp in length mapped the majority of these fragments to terminal branches within the plastid tree (fig. 1), suggesting that most of the extant *mtpts* are of relatively recent origin. However, we also found evidence of a number of more ancient transfers (fig. 1), which is consistent with previous studies (Wang et al. 2007).

Although phylogenetic placement can be used to date the timing of DNA transfers (Bensasson et al. 2003; Cummings et al. 2003; Hazkani-Covo et al. 2010), these analyses can be complicated by subsequent gene conversion between mitochondrial and plastid genomes. Clear evidence of gene conversion has already been documented in angiosperm *mtpts*. For example, Clifton et al. (2004) concluded that ongoing copy correction could explain the discordant lines of evidence from sequence versus structural data about the origins of a *mtpt* that is shared by multiple grass species. Our detailed

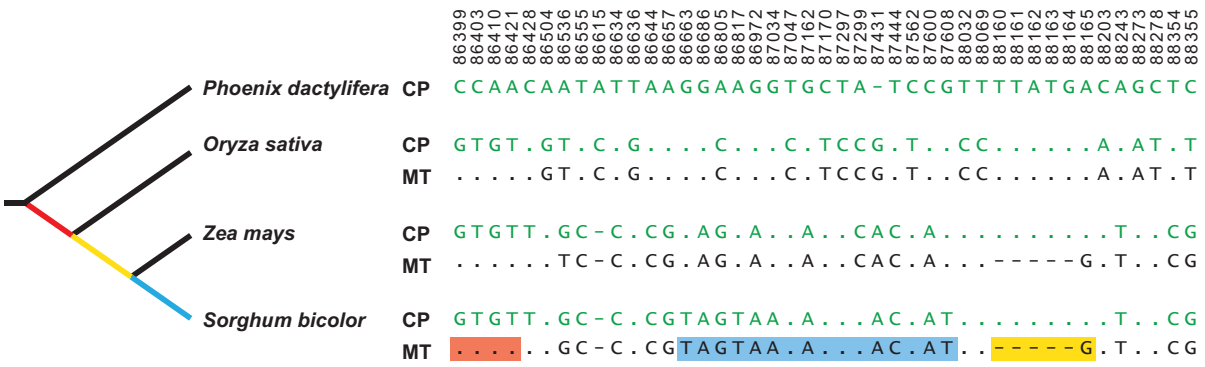

Fig. 5.—Complex history of gene conversion in a *Sorghum bicolor mtpt*. The aligned characters represent all parsimony-informative sites from a 2-kb *mtpt* fragment found in grasses along with the corresponding plastid sequences. Shading in the *S. bicolor mtpt* sequence indicates regions with different phylogenetic histories resulting from different timing of gene conversion events (as shown by corresponding color coding in the species tree). Numbering corresponds to nucleotide position in the *Phoenix dactylifera* plastid genome.

analysis of this family of *mtpts* and their plastid homologs revealed an even more complex history of gene conversion, in which a series of events have created a mosaic of different genealogical histories within a single *mtpt* (fig. 5). The copy of *rps19* at one end of the shared *mtpt* retains evidence of the previously hypothesized ancient transfer that preceded the divergence of the grasses, with multiple derived substitutions shared by the mitochondrial copies but not found in plastid genomes (fig. 5). Other parts of the *mtpt* exhibit more recent divergence from their plastid counterparts. For example, an internal region containing the 3′-end of *rpl2* in *Sorghum bicolor* was apparently subject to a recent conversion event that occurred since the divergence of *Sorghum* and *Zea*, whereas yet another region of this same *mtpt* (containing the 5′-end of the *rp2*) appears to have experienced a separate gene conversion event after the divergence between *Sorghum* and *Oryza* but before the *Sorghum-Zea* split (fig. 5). Because of the recurrent history of gene conversion, our phylogenetic analysis (fig. 1) does not necessarily reflect the timing of the initial transfer of a plastid sequence into the mitochondrial genome. Instead, it may indicate the point at which a *mtpt* began to diverge from its plastid counterpart based on its most recent bout of gene conversion—something that can become quite complicated when gene conversion creates a mosaic of different genealogical histories within the same fragment (Hao et al. 2010).

Inferring the history of *mtpts* is further complicated by horizontal transfer of mtDNA between species, which appears to be a rampant process in many angiosperm mitochondrial genomes (Bergthorsson et al. 2003; Rice et al. 2013). When *mtpts* are among the set of transferred sequences, the net effect is the movement of plastid DNA from one species to the mitochondrial genome of another (Woloszynska et al. 2004; Rice et al. 2013). Our phylogenetic analysis identified a number of examples of this pattern, including previously documented cases in *A. trichopoda* (Rice et al. 2013). In addition, the mitochondrial genomes of *Glycine max* and *Lotus japonicus* have each acquired a partial copy of the plastid gene *rbcL* from (independent) donors within the asterids, and the *Phoenix dactylifera* mitochondrial genome contains a fragment with partial copies of the plastid tRNA genes *trnI-GAU* and *trnA-UGC* that was acquired from a eudicot donor (supplementary fig. S1, Supplementary Material online). The *rbcL mtpt* in *L. japonicus* is highly similar to plastid sequences in the parasitic genus *Cuscuta*, suggesting some role of host-parasite interactions, as observed in other types of horizontal transfers (Mower et al. 2004; Xi et al. 2013).

The direction of DNA transfer between mitochondrial and plastid genomes is overwhelmingly biased, with almost all movement occurring from the plastids to the mitochondria. However, recent evidence has indicated that transfers in the opposite direction do occur, albeit very rarely (Iorizzo et al. 2012; Straub et al. 2013). Because our analysis required that *mtpts* had homologs in at least ten plastid genomes, the previously characterized example of mitochondria-to-plastid transfer in *Daucus carota* was filtered during data processing (and a second example in *Asclepias syriaca* was excluded entirely from this study because the corresponding mitochondrial genome was not yet available at the time of data collection).

Indel Spectrum in Angiosperm *mtpts*

Our analysis of small indels in angiosperm *mtpts* identified more deletions (819) than insertions (634). The average deletion size of 6.4 bp also exceeded the average insertion size of 4.0 bp (fig. 6*A*). These results provide clear evidence of a deletion bias with roughly twice the amount of sequence being removed by small deletions than being added by small insertions. However, the magnitude of this bias is much smaller than previous estimates from prokaryotic genomes (fig. 6*B*).

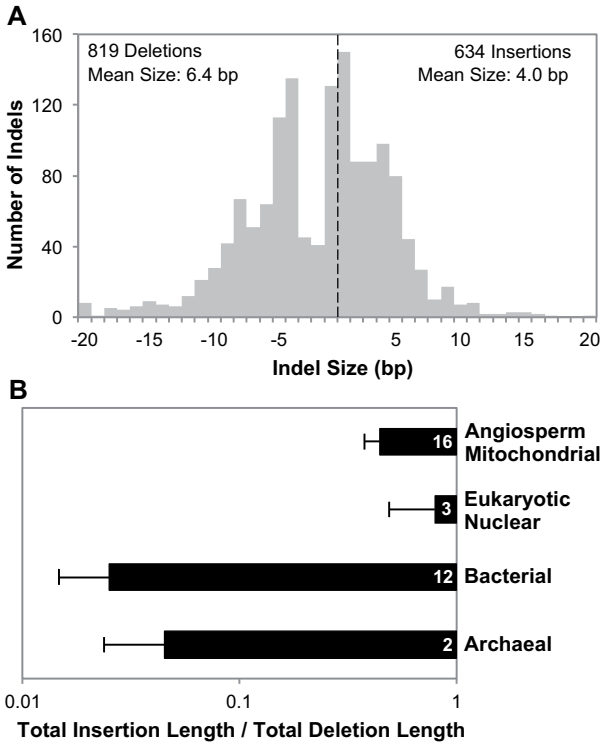

Fig. 6.—Weak deletion bias in *mtpts*. The distribution of indel sizes pooled across all angiosperm genomes in this study is skewed toward deletions (*A*). However, the observed skew is much weaker than those reported in bacterial and archaeal genomes and, instead, is more in line with estimates from eukaryotic nuclear genome (*B*). Indel bias measurements are from Kuo and Ochman (2009). Sample sizes (number of species) are indicated at the base of each bar. For the angiosperm mitochondrial genomes, only species with at least 20 *mtpt* indels were included. Error bars represent the standard error of the mean calculated based on log-transformed values.

The skew toward deletions is more in line with the weaker deletion biases observed in some eukaryotic nuclear genomes (fig. 6*B*). Analyzing the subset of species with at least 20 documented indels (supplementary table S1, Supplementary Material online), we found only a weak and nonsignificant correlation between deletion bias and genome size (fig. 7*A*). Performing the same correlation analysis on phylogenetically independent contrasts also yielded a weakly positive but nonsignificant relationship (data not shown).

Substitution Patterns in Angiosperm *mtpts*

We analyzed a total 283,741 sites that met our filtering criteria (see Materials and Methods) and identified 5,619 substitutions in *mtpt* sequences (table 2). Unlike many genomes in which transitions greatly outnumber transversions, the observed transition:transversion was only 0.54. In fact, in most angiosperm species, the frequency of transitions was even lower, as this average was inflated by an unusually high rate of transitions in two rapidly evolving *Silene* species, *S. conica* and *S. noctiflora* (supplementary table S2, Supplementary Material online). We found that 1.38% of sites at which the ancestral plastid sequence contained an A or T experienced a substitution to a G or C. The opposite pattern was more frequent, as 1.78% of G and C sites were changed to an A or T. In the absence of any selection, this mutational asymmetry would produce an equilibrium GC content of 43.7%.

Although, on a per site basis, we found a bias in favor of substitutions toward A or T, we actually observed a larger total number of substitutions in the reverse direction. There were a total of 2,395 A/T-to-G/C substitutions and only 1,971 G/C-to-A/T substitutions (table 2). This result reflects the fact that plastid genomes are AT-rich, so there are fewer opportunities for mutations toward A or T to occur in sequences of plastid origin. The ancestral GC content of the plastid sites analyzed in this study was 39.0% (below the equilibrium GC content of 43.7% predicted from observed *mtpt* substitutions).

To assess lineage-specific substitution patterns, we analyzed the subset of species in our data set with at least 100 *mtpt* substitutions (supplementary table S2, Supplementary Material online). These species-specific data produced a wide range of predicted equilibrium GC values, from 23.5% in *B. hygrometrica* to 63.6% in *Liriodendron tulipifera*. We found a positive but nonsignificant correlation between equilibrium GC content predicted from *mtpt* substitutions and observed genome-wide GC content (fig. 7*B*). Analysis of phylogenetically independent contrasts also yielded a weakly positive but nonsignificant relationship (data not shown).

Discussion

The bulk movement of DNA sequence between genomic compartments creates an opportunity to estimate mutational parameters in eukaryotic genomes (Bensasson, Petrov, et al. 2001; Huang et al. 2005; Noutsos et al. 2005; Rousseau-Gueutin et al. 2011; Hsu et al. 2014). In angiosperm mitochondrial genomes, *mtpts* arguably represent the best of class of sequences to measure indel and substitution biases because of 1) their sheer abundance, 2) their general lack of conservation or apparent functional constraint, and 3) the availability of numerous highly conserved plastid genomes for comparative analyses. By taking advantage of these properties of *mtpts*, we have provided some of the first detailed estimates of indel size distributions and nucleotide substitution patterns in angiosperm mtDNA.

Deletion Bias and Genome Size

Plant mitochondria have reversed the nearly universal pattern of reductive evolution in organelle and endosymbiont genomes and experienced a proliferation of noncoding and intergenic sequence content, raising the question as to

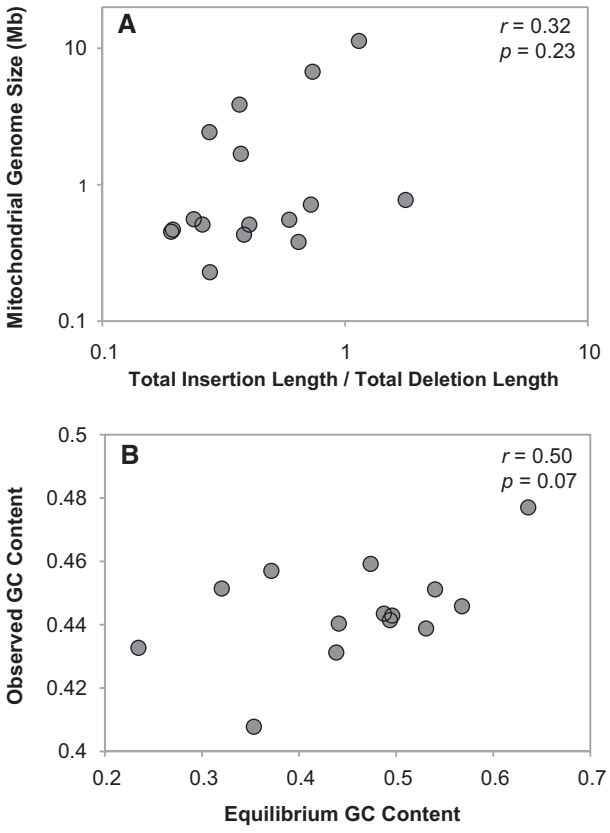

Fig. 7.—Mutational biases an interspecific variation in mitochondrial genome size (*A*) and GC content (*B*) across species. Each point represents a species with a minimum of 20 *mtpt* indels or 100 *mtpt* substitutions.

Table 2
mtpt Nucleotide Substitution Matrix

Ancestral Nucleotide	*mtpt* Nucleotide			
	A	C	G	T
A	85,085 (0.9815)	751 (0.0087)	483 (0.0056)	369 (0.0043)
C	508 (0.0092)	53,932 (0.9775)	217 (0.0039)	515 (0.0093)
G	502 (0.0091)	284 (0.0051)	54,235 (0.9778)	446 (0.0080)
T	383 (0.0044)	477 (0.0055)	684 (0.0079)	84,870 (0.9821)

Note.—Relative frequencies for each ancestral nucleotide are indicated in parentheses such that the rows sum to one.

whether the mutational bias that generally favors small deletions over small insertions may also have been reversed in these genomes. We did not find evidence of such a reversal in angiosperm mtDNA, but the deletion bias that does exist appears to be very weak (fig. 6). Could the relaxed deletion bias in angiosperm mtDNA be responsible for the mitochondrial genome expansion in this group?

On the one hand, the observation of a weak deletion bias in angiosperm mtDNA is grossly consistent with the mutational equilibrium model of genome size evolution (Petrov 2002). Under this model, selection is more likely to tolerate large insertions than large deletions because the probability of disrupting a functional element increases with deletion size but not with insertion size. Mutation biases that favor small deletions are expected to counteract the expansion resulting from the excess of large insertions and thereby determine an equilibrium genome size, so weaker deletion biases would be associated with larger genome size.

On the other hand, small indel biases appear to have very little power to explain the variation in mitochondrial genome size among angiosperms (fig. 7). The mutational equilibrium model of genome size evolution has been criticized based on the argument that the cumulative effect of small deletions may be unrealistically slow to counter much more rapid mechanisms of genome expansion (Gregory 2004). Indeed, even in the initial formalization of this model, it was viewed as contributing to long-term equilibrium genome sizes but not necessarily as an explanation for rapid fluctuations in genome size driven by bursts of transposable element activity, polyploidy, etc. (Petrov 2002). This is particularly important given the enormous size variation in angiosperm mitochondrial genomes, in which species within a family or even a genus can differ by more than an order of magnitude in size (Ward et al. 1981; Sloan, Alverson, Chuckalovcak, et al. 2012). In the short-run, the effects of small indels are likely overwhelmed by the much larger structural changes that make angiosperm mtDNA so dynamic. For example, figure 3 illustrates how a series of large deletions can rapidly reduce the size of a *mtpt* fragment. The promiscuous sequences in angiosperm mitochondria also indicate that these genomes experience insertions of large DNA fragments or even entire foreign mitochondrial genomes (Rice et al. 2013), raising the possibility that an excess of very large insertions could be a major determinant of genome expansion in this group.

Therefore, while a weak deletion bias might have contributed to the long-term increase in size and the retention of large intergenic regions in plant mitochondrial genomes, it is highly unlikely that more recent changes in small indel bias can explain the remarkable diversity in genome size within this group.

Nucleotide Substitution Bias and GC Content

One of the many unanswered questions about the molecular evolution of angiosperm mtDNA relates to the variation in nucleotide composition across the genome—particularly the higher GC content in intergenic regions than in synonymous positions in protein-coding genes. Although no clear explanation for this phenomenon has been provided, it was hypothesized that the low GC content at synonymous sites might reflect the long-term equilibrium associated with mutation biases in the mitochondrial genome, whereas the intergenic content (much of which is derived from relatively recent transfers of promiscuous DNA) might not yet have reached equilibrium (Sloan and Taylor 2010). Our results are inconsistent

with this hypothesis. The substitution matrix estimated from *mtpt* sequences (table 2) predicts an equilibrium GC content that is substantially higher than observed values at synonymous sites of mitochondrial genes but right in line with values from intergenic regions. This finding is supported by the observation that older *mtpts* have higher GC content (Fang et al. 2012), indicating that substitution biases bring the nucleotide composition of horizontally transferred sequences into balance with the rest of their new host genome over time (Lawrence and Ochman 1997).

Therefore, our results suggest that selection on synonymous substitutions is strong enough to substantially alter nucleotide composition in mitochondrial protein genes. There is evidence of selection on biased codon usage and recognition motifs for RNA editing sites in plant mitochondrial genomes, but an earlier analysis pointed to relatively weak effects (Sloan and Taylor 2010). The conclusion that selection might be reducing GC content in angiosperm mitochondrial genes is intriguing because it runs opposite the emerging pattern that selection generally acts to increase GC content and counteract the widespread phenomenon of AT-biased mutation (Hershberg and Petrov 2010; Hildebrand et al. 2010; Van Leuven and McCutcheon 2012).

Methodological Limitations in the Analysis of mtpts

Although we have made the argument that *mtpts* provide a valuable opportunity to estimate mutational parameters in angiosperm mitochondrial genomes, it is important to recognize some of the assumptions and limitations of these analyses. First, like other indirect approaches to measuring mutation, our analysis relies on the crucial assumption that substitutions and indels in *mtpts* are effectively neutral, which might not always be the case even in pseudogenes (Denver et al. 2004).

Second, we are also assuming that *mtpts* are broadly representative of mtDNA such that estimated mutation parameters can be applied to the rest of the genome. Christensen (2013) has recently hypothesized that plant mitochondrial genes experience lower mutation rates than surrounding intergenic regions possibly because of transcription-coupled repair. Although subsequent analysis did not find evidence for this repair mechanism (Christensen 2014), there is evidence of highly localized substitution rate variation in some angiosperm organelle genomes (Sloan et al. 2009; Magee et al. 2010; Zhu et al. 2014). The possibility that mutation patterns systematically differ between genic and intergenic regions (Christensen 2013) provides an alternative explanation for the discrepancy between the equilibrium GC content predicted from *mtpt* substitutions and the observed values at synonymous sites.

Third, our estimates of mutational parameters depend on accurate reconstruction of ancestral states from phylogenetic data. To minimize bias in our estimates, we restricted our analysis to sites for which we could be extremely confident in identifying the ancestral state—namely those that were completely conserved among the aligned plastid genomes. This approach, however, comes at the cost of excluding large quantities of data. For example, approximately half of the alignment positions were excluded from the substitution analysis because of this requirement. This loss of data reduces the statistical power of our analysis and could potentially introduce a bias itself if the sites that are conserved among plastid genomes experience nonrepresentative mutation patterns after being transferred to the mitochondrial genome. Additional analyses that employ ancestral state reconstructions for variable site have the potential to extract additional information about *mtpt* sequence evolution, but these should be undertaken with caution because estimates of mutation parameters will be highly sensitive to errors in ancestral state reconstruction.

Fourth, the history of gene conversion between the plastid genome and *mtpts* following their initial transfer has the potential to bias our estimates of mutation parameters. Copy correction by itself does not necessarily present a problem, as it simply "erases" mutations that have occurred in *mtpts*. However, if certain types of mutations differentially reduce the probability that an *mtpt* undergoes gene conversion, that would affect our ability to detect those changes and thereby alter the inferred spectrum of mutations. There is an extensive literature on how gene conversion can be biased with respect to nucleotide substitutions (e.g., Marais 2003; Khakhlova and Bock 2006; Duret and Galtier 2009) and more recent evidence for bias associated with indels (Assis and Kondrashov 2012; Leushkin and Bazykin 2013), but these issues have not been explored in plant mitochondrial genomes. It is, therefore, possible that spectra of indels and nucleotide substitutions found in *mtpts* represent a composite of biased mutation and biased gene conversion.

Finally, except in the most *mtpt*-rich genomes, we have limited statistical precision in generating species-specific estimates of indel and substitution parameters, which may contribute to the high variance in these estimates (supplementary tables S1 and S2, Supplementary Material online) and hinder efforts to explain mitochondrial genome variation within angiosperms (fig. 7). In many ways, these limitations are an illustration of why isolating the effects of mutational biases has posed such a longstanding challenge to the field of molecular evolution. Despite these difficulties, however, we find that *mtpts* are a particularly valuable tool for dissecting the mechanisms shaping the evolution of the enigmatic mitochondrial genomes found in angiosperms.

Supplementary Material

Supplementary figure S1 and tables S1 and S2 are available at *Genome Biology and Evolution* online (http://www.gbe.oxfordjournals.org/).

Acknowledgments

The authors thank Jocelyn Cuthbert for assistance with data collection and Danny Rice and Jeff Palmer for sharing the *A. trichopoda* mitochondrial genome sequence prior to its publication. This work was supported by the National Science Foundation [grant number MCB-1412260] and Colorado State University.

Literature Cited

Allen JO, et al. 2007. Comparisons among two fertile and three male-sterile mitochondrial genomes of maize. Genetics 177:1173–1192.

Alverson AJ, et al. 2010. Insights into the evolution of plant mitochondrial genome size from complete sequences of *Citrullus lanatus* and *Cucurbita pepo* (Cucurbitaceae). Mol Biol Evol. 27:1436–1448.

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.

Assis R, Kondrashov AS. 2012. A strong deletion bias in nonallelic gene conversion. PLoS Genet. 8:e1002508.

Bensasson D, Feldman MW, Petrov DA. 2003. Rates of DNA duplication and mitochondrial DNA insertion in the human genome. J Mol Evol. 57:343–354.

Bensasson D, Petrov DA, Zhang DX, Hartl DL, Hewitt GM. 2001. Genomic gigantism: DNA loss is slow in mountain grasshoppers. Mol Biol Evol. 18:246–253.

Bensasson D, Zhang D, Hartl DL, Hewitt GM. 2001. Mitochondrial pseudogenes: evolution's misplaced witnesses. Trends Ecol Evol. 16: 314–321.

Bergthorsson U, Adams KL, Thomason B. 2003. Widespread horizontal transfer of mitochondrial genes in flowering plants. Nature 424: 197–201.

Chamary JV, Parmley JL, Hurst LD. 2006. Hearing silence: non-neutral evolution at synonymous sites in mammals. Nat Rev Genet. 7:98–108.

Christensen AC. 2013. Plant mitochondrial genome evolution can be explained by DNA repair mechanisms. Genome Biol Evol. 5:1079–1086.

Christensen AC. 2014. Genes and junk in plant mitochondria-repair mechanisms and selection. Genome Biol Evol. 6:1448–1453.

Clark KA, et al. 2012. Selfish little circles: transmission bias and evolution of large deletion-bearing mitochondrial DNA in *Caenorhabditis briggsae* nematodes. PLoS One 7:e41433.

Clifton SW, et al. 2004. Sequence and comparative analysis of the maize NB mitochondrial genome. Plant Physiol. 136:3486–3503.

Cummings MP, Nugent JM, Olmstead RG, Palmer JD. 2003. Phylogenetic analysis reveals five independent transfers of the chloroplast gene *rbcL* to the mitochondrial genome in angiosperms. Curr Genet. 43: 131–138.

Darracq A, et al. 2011. Structural and content diversity of mitochondrial genome in beet: a comparative genomic analysis. Genome Biol Evol. 3: 723–736.

Denver DR, Morris K, Lynch M, Thomas WK. 2004. High mutation rate and predominance of insertions in the *Caenorhabditis elegans* nuclear genome. Nature 430:679–682.

Denver DR, Morris K, Lynch M, Vassilieva LL, Thomas WK. 2000. High direct estimate of the mutation rate in the mitochondrial genome of *Caenorhabditis elegans*. Science 289:2342–2344.

Deriano L, Roth DB. 2013. Modernizing the nonhomologous end-joining repertoire: alternative and classical NHEJ share the stage. Annu Rev Genet. 47:433–455.

Dietrich A, Small I, Cosset A, Weil JH, Marechal-Drouard L. 1996. Editing and import: strategies for providing plant mitochondria with a complete set of functional transfer RNAs. Biochimie. 78:518–529.

Duret L, Galtier N. 2009. Biased gene conversion and the evolution of mammalian genomic landscapes. Annu Rev Genomics Hum Genet. 10:285–311.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32:1792–1797.

Ellis J. 1982. Promiscuous DNA—chloroplast genes inside plant mitochondria. Nature 299:678–679.

Fang Y, et al. 2012. A complete sequence and transcriptomic analyses of date palm (*Phoenix dactylifera* L.) mitochondrial genome. PLoS One 7: e37164.

Gregory TR. 2004. Insertion-deletion biases and the evolution of genome size. Gene 324:15–34.

Hao W, Palmer JD. 2009. Fine-scale mergers of chloroplast and mitochondrial genes create functional, transcompartmentally chimeric mitochondrial genes. Proc Natl Acad Sci U S A. 106:16728–16733.

Hao W, Richardson AO, Zheng Y, Palmer JD. 2010. Gorgeous mosaic of mitochondrial genes created by horizontal transfer and gene conversion. Proc Natl Acad Sci U S A. 107:21576–21581.

Hazkani-Covo E, Zeller RM, Martin W. 2010. Molecular poltergeists: mitochondrial DNA copies (*numts*) in sequenced nuclear genomes. PLoS Genet. 6:e1000834.

Hershberg R, Petrov DA. 2010. Evidence that mutation is universally biased towards AT in bacteria. PLoS Genet. 6:e1001115.

Hildebrand F, Meyer A, Eyre-Walker A. 2010. Evidence of selection upon genomic GC-content in bacteria. PLoS Genet. 6:e1001107.

Hsu C-Y, Wu C-S, Chaw S-M. 2014. Ancient nuclear plastid DNA in the yew family (Taxaceae). Genome Biol Evol. 6:2111–2121.

Huang CY, Grunheit N, Ahmadinejad N, Timmis JN, Martin W. 2005. Mutational decay and age of chloroplast and mitochondrial genomes transferred recently to angiosperm nuclear chromosomes. Plant Physiol. 138:1723–1733.

Iorizzo M, et al. 2012. Against the traffic: the first evidence for mitochondrial DNA transfer into the plastid genome. Mob Genet Elements. 2: 261–266.

Khakhlova O, Bock R. 2006. Elimination of deleterious mutations in plastid genomes by gene conversion. Plant J. 46:85–94.

Kimura M. 1983. The neutral theory of molecular evolution. Cambridge: Cambridge University Press.

Knoop V, Volkmar U, Hecht J, Grewe F. 2011. Mitochondrial genome evolution in the plant lineage. In: Kempken F, editor. Plant Mitochondria. New York: Springer. p. 3–29.

Kubo T, Newton KJ. 2008. Angiosperm mitochondrial genomes and mutations. Mitochondrion 8:5–14.

Kuo CH, Ochman H. 2009. Deletional bias across the three domains of life. Genome Biol Evol. 1:145–152.

Lawrence JG, Ochman H. 1997. Amelioration of bacterial genomes: rates of change and exchange. J Mol Evol. 44:383–397.

Leushkin EV, Bazykin GA. 2013. Short indels are subject to insertion-biased gene conversion. Evolution 67:2604–2613.

Lohse M, Drechsel O, Bock R. 2007. OrganellarGenomeDRAW (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. Curr Genet. 52:267–274.

Magee AM, et al. 2010. Localized hypermutation and associated gene losses in legume chloroplast genomes. Genome Res. 20:1700–1710.

Marais G. 2003. Biased gene conversion: implications for genome and sex evolution. Trends Genet. 19:330–338.

McCutcheon JP, Moran NA. 2012. Extreme genome reduction in symbiotic bacteria. Nat Rev Microbiol. 10:13–26.

Mira A, Ochman H, Moran NA. 2001. Deletional bias and the evolution of bacterial genomes. Trends Genet. 17:589–596.

Mower JP, Sloan DB, Alverson AJ. 2012. Plant mitochondrial diversity—the genomics revolution. In: Wendel JF, editor. Plant genome diversity. Vienna: Springer. p. 123–144.

Mower JP, Stefanovic S, Young GJ, Palmer JD. 2004. Gene transfer from parasitic to host plants. Nature 432:165–166.
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.
Nakazono M, Nishiwaki S, Tsutsumi N, Hirai A. 1996. A chloroplast-derived sequence is utilized as a source of promoter sequences for the gene for subunit 9 of NADH dehydrogenase (*nad9*) in rice mitochondria. Mol Gen Genet. 252:371–378.
Noutsos C, Richly E, Leister D. 2005. Generation and evolutionary fate of insertions of organelle DNA in the nuclear genomes of flowering plants. Genome Res. 15:616–628.
Palmer JD, Herbon LA. 1987. Unicircular structure of the *Brassica hirta* mitochondrial genome. Curr Genet. 11:565–570.
Paradis E, Claude J, Strimmer K. 2004. APE: analyses of phylogenetics and evolution in R language. Bioinformatics 20:289–290.
Petrov DA. 2002. Mutational equilibrium model of genome size evolution. Theor Popul Biol. 61:531–544.
Petrov DA, Sangster TA, Johnston JS, Hartl DL, Shaw KL. 2000. Evidence for DNA loss as a determinant of genome size. Science 287:1060.
R Core Team. 2014. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.
Rice DW, 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.
Richly E, Leister D. 2004. NUPTs in sequenced eukaryotes and their genomic organization in relation to NUMTs. Mol Biol Evol. 21:1972–1980.
Rousseau-Gueutin M, Ayliffe MA, Timmis JN. 2011. Conservation of plastid sequences in the plant nuclear genome for millions of years facilitates endosymbiotic evolution. Plant Physiol. 157:2181–2193.
Sloan DB, Alverson AJ, Chuckalovcak JP, et al. 2012. Rapid evolution of enormous, multichromosomal genomes in flowering plant mitochondria with exceptionally high mutation rates. PLoS Biol. 10:e1001241.
Sloan DB, Alverson AJ, Storchova H, Palmer JD, Taylor DR. 2010. Extensive loss of translational genes in the structurally dynamic mitochondrial genome of the angiosperm *Silene latifolia*. BMC Evol Biol. 10:274.
Sloan DB, Alverson AJ, Wu M, Palmer JD, Taylor DR. 2012. Recent acceleration of plastid sequence and structural evolution coincides with extreme mitochondrial divergence in the angiosperm genus *Silene*. Genome Biol Evol. 4:294–306.
Sloan DB, Müller K, McCauley DE, Taylor DR, Štorchová H. 2012. 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.
Sloan DB, Oxelman B, Rautenberg A, Taylor DR. 2009. Phylogenetic analysis of mitochondrial substitution rate variation in the angiosperm tribe *Sileneae* (Caryophyllaceae). BMC Evol Biol. 9:260.
Sloan DB, Taylor DR. 2010. Testing for selection on synonymous sites in plant mitochondrial DNA: the role of codon bias and RNA editing. J Mol Evol. 70:479–491.
Stajich JE, et al. 2002. The bioperl toolkit: Perl modules for the life sciences. Genome Res. 12:1611–1618.
Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30: 1312–1313.
Straub SC, Cronn RC, Edwards C, Fishbein M, Liston A. 2013. Horizontal transfer of DNA from the mitochondrial to the plastid genome and its subsequent evolution in milkweeds (Apocynaceae). Genome Biol Evol. 5:1872–1885.
Taylor DR, Zeyl C, Cooke E. 2002. Conflicting levels of selection in the accumulation of mitochondrial defects in *Saccharomyces cerevisiae*. Proc Natl Acad Sci U S A. 99:3690–3694.
Van Leuven JT, McCutcheon JP. 2012. An AT mutational bias in the tiny GC-rich endosymbiont genome of *Hodgkinia*. Genome Biol Evol. 4: 24–27.
Wang D, et al. 2007. Transfer of chloroplast genomic DNA to mitochondrial genome occurred at least 300 million years ago. Mol Biol Evol. 24: 2040–2048.
Wang D, Rousseau-Gueutin M, Timmis JN. 2012. Plastid sequences contribute to some plant mitochondrial genes. Mol Biol Evol. 29: 1707–1711.
Ward BL, Anderson RS, Bendich AJ. 1981. The mitochondrial genome is large and variable in a family of plants (Cucurbitaceae). Cell 25: 793–803.
Woloszynska M, Bocer T, Mackiewicz P, Janska H. 2004. A fragment of chloroplast DNA was transferred horizontally, probably from non-eudicots, to mitochondrial genome of *Phaseolus*. Plant Mol Biol. 56: 811–820.
Xi Z, et al. 2013. Massive mitochondrial gene transfer in a parasitic flowering plant clade. PLoS Genet. 9:e1003265.
Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 24:1586–1591.
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.

Associate editor: Sarah Schaack