Mitochondrial Retroprocessing Promoted Functional Transfers of *rpl5* to the Nucleus in Grasses

Zhiqiang Wu,[1] Daniel B. Sloan,[1] Colin W. Brown,[2] Mónica Rosenblueth,[3] Jeffrey D. Palmer,[4] and Han Chuan Ong*[,5]

[1]Department of Biology, Colorado State University, Fort Collins, CO

[2]Institute for Cellular and Molecular Biology, University of Texas, Austin, TX

[3]Center for Genomics Sciences, UNAM, Cuernavaca, Mexico

[4]Department of Biology, Indiana University, Bloomington, IN

[5]Department of Biology, King University, Bristol, TN

***Corresponding author:** E-mail: hanchuanong@king.edu.

Associate editor: Michael Purugganan

Abstract

Functional gene transfers from the mitochondrion to the nucleus are ongoing in angiosperms and have occurred repeatedly for all 15 ribosomal protein genes, but it is not clear why some of these genes are transferred more often than others nor what the balance is between DNA- and RNA-mediated transfers. Although direct insertion of mitochondrial DNA into the nucleus occurs frequently in angiosperms, case studies of functional mitochondrial gene transfer have implicated an RNA-mediated mechanism that eliminates introns and RNA editing sites, which would otherwise impede proper expression of mitochondrial genes in the nucleus. To elucidate the mechanisms that facilitate functional gene transfers and the evolutionary dynamics of the coexisting nuclear and mitochondrial gene copies that are established during these transfers, we have analyzed *rpl5* genes from 90 grasses (Poaceae) and related monocots. Multiple lines of evidence indicate that *rpl5* has been functionally transferred to the nucleus at least three separate times in the grass family and that at least seven species have intact and transcribed (but not necessarily functional) copies in both the mitochondrion and nucleus. In two grasses, likely functional nuclear copies of *rpl5* have been subject to recent gene conversion events via secondarily transferred mitochondrial copies in what we believe are the first described cases of mitochondrial-to-nuclear gene conversion. We show that *rpl5* underwent a retroprocessing event within the mitochondrial genome early in the evolution of the grass family, which we argue predisposed the gene towards successful, DNA-mediated functional transfer by generating a "pre-edited" sequence.

Key words: **endosymbiotic gene transfer, intracellular gene transfer, mtDNA, pseudogene, reverse transcription.**

Article

Introduction

The evolution of mitochondrial genomes (mitogenomes) has been characterized by massive gene loss ever since the endosymbiotic origin of mitochondria, but the extent of gene loss across eukaryotic lineages is heterogeneous (Gray et al. 1999). Angiosperm mitogenomes contain many more protein genes than animal and fungal mitogenomes. Mitochondrial protein gene content is also much more variable in angiosperms, reflecting vastly more frequent functional transfer of mitochondrial genes to the nucleus in angiosperms compared with animals and fungi (Adams et al. 2002b). Therefore, angiosperms are an outstanding system to study the process of intracellular gene transfer—one of the major modes of eukaryotic genome evolution (Adams and Palmer 2003; Timmis et al. 2004). The overwhelming majority of functional transfers in angiosperms are restricted to 17 of the 41 protein genes present in the angiosperm common ancestor; these include all 15 ribosomal protein genes and both succinate dehydrogenase genes. A survey of 16 of these frequently transferred genes among 280 diverse angiosperms revealed an average of 23 independent mitochondrial losses—and likely transfers—per gene (Adams et al. 2002b).

The predominant mechanism by which mitochondrial coding sequence information is physically transferred to the nucleus in plants has been a source of uncertainty and controversy. On one hand, it has been argued that an RNA-mediated mechanism is involved because many characterized examples of transferred mitochondrial genes resemble reverse transcribed mRNAs, in that introns have been removed and/or sites of mitochondrial C-to-U editing possess the corresponding (T) edited nucleotide (Nugent and Palmer 1991; Covello and Gray 1992; Brennicke et al. 1993; Wischmann and Schuster 1995; Knoop et al. 1995; Adams et al. 2000). On the other hand, genomic evidence has clearly demonstrated that direct insertion of DNA from organelle genomes into the nucleus is very common (Stupar et al. 2001; Lough et al. 2008; Michalovova et al. 2013), leading some authors to conclude that most or all functional gene transfers are DNA-mediated (Henze and Martin 2001; Timmis et al. 2004). Experimental reconstructions of plastid-to-nucleus transfers

have also shown that the frequency of DNA-mediated insertions greatly exceeds that of RNA-mediated insertions, which appear to be too rare to detect over short timescales in the lab (Fuentes et al. 2012).

Although DNA-mediated transfer of organellar sequences to the nucleus is undoubtedly frequent, most of these transferred sequences are presumably "dead on arrival" and decay as pseudogenes, owing to the presence of compartment-specific RNA processing barriers (i.e., group II introns and C-to-U RNA edits in plant mitochondria; Hammani and Giegé 2014). For example, the inserted copy of a nearly complete mitogenome in nuclear chromosome 2 of *Arabidopsis* is not expressed and is unlikely to be functional (Stupar et al. 2001). Experimental studies have demonstrated that the existence of a single intron or RNA editing site within a gene does not necessarily prohibit functional transfer to the nucleus because genetic mechanisms such as promiscuous splicing or the use of alternative translation initiation sites may obviate the need to remove introns and editing sites from the gene itself (Sheppard et al. 2011; Fuentes et al. 2012). However, contrary to observations from phylogenetic studies, these mechanisms do not result in transferred genes that resemble reverse transcribed mRNAs, and it remains difficult to explain how a mitochondrial gene containing numerous introns and/or editing sites could remain functional after a direct DNA-mediated transfer to the nucleus.

A two-step mechanism could potentially reconcile these seemingly conflicting observations about RNA-mediated versus DNA-mediated transfers (Henze and Martin 2001). The first step would involve retroprocessing *within* the mitogenome. In other words, a mitochondrial mRNA would be reverse transcribed into cDNA followed by homologous recombination with mitochondrial genomic DNA, effectively eliminating introns and RNA editing sites from a gene that would still reside in the mitogenome. After retroprocessing had occurred within the mitogenome, the resulting gene would then be "ready-to-go" and amenable to functional transfer to the nucleus via DNA-mediated mechanisms because it would no longer have mitochondrion-specific RNA processing requirements that impede function in the nucleus. Therefore, the second step in this mechanism would be the DNA-mediated duplication/transfer of the retroprocessed mitochondrial gene to the nuclear genome. Despite the logical appeal of this two-step transfer mechanism, evidence for it is scarce.

Incipient gene transfers face an additional hurdle in that their gene products need to be targeted back to the mitochondria. Transferred genes often acquire regulatory and mitochondrial-targeting peptides by recruiting preexisting sequences (Figueroa et al. 1999; Kubo et al. 1999; Adams et al. 2000; Sandoval et al. 2004; Liu et al. 2009) or by 5′-end exon shuffling (Long et al. 1996; Bonen and Calixte 2006). Despite the rarity of functional activation following mitochondrion-to-nucleus gene insertions (Rand et al. 2004), most mitochondrial ribosomal protein genes have been transferred repeatedly to the nucleus in flowering plants (Adams et al. 2000, 2002b; Liu et al. 2009). For example, intracellular transfer of the mitochondrial ribosomal protein gene *rpl5* was inferred to have occurred 19 separate times based on Southern hybridization analyses of 280 angiosperm species (Adams et al. 2002b), including a single transfer within the grasses (only three grasses were sampled: *Hordeum*, *Triticum*, and *Zea*). However, subsequent characterization of the putative targeting sequences from nuclear *rpl5* genes of *Triticum* and *Zea* suggested that they arose from two separate transfers from the mitochondrion (Sandoval et al. 2004). Interestingly, despite having a functional copy of *rpl5* in the nucleus, *Triticum* retains an intact and expressed copy in its mitogenome (Sandoval et al. 2004; Ogihara et al. 2005), whereas the gene is absent from the mitogenomes of *Zea* and its close relatives *Tripsacum* and *Sorghum* (Clifton et al. 2004; Allen et al. 2007).

To more fully reconstruct the history of *rpl5* transfers to the nucleus in grasses and gain mechanistic insights into intracellular gene transfer processes in general, we analyzed published grass genomes and transcriptomes and generated *rpl5* gene sequences for dozens of additional grasses and nongrass monocots. Based on a data set of 90 taxa, we found evidence that *rpl5* has been functionally transferred to the nucleus at least three separate times within the grass family. A broad range of grasses possess intact and transcribed copies of *rpl5* in both the mitochondrial and nuclear genomes, indicating persistence of a state of transcompartmental duplication for some 40–60 million years, although whether the mitochondrial copies are actually functional is not clear. In at least two cases, the mitochondrial copy has replaced part of the nuclear copy via the first documented cases of mitochondrial-to-nuclear gene conversion. Finally, we show that *rpl5* likely underwent retroprocessing within the mitogenome early in grass evolution, resulting in the loss of all but one editing site and providing evidence in support of the hypothesis that retroprocessing within the mitochondria can open the door to subsequent transfers to the nucleus via DNA intermediates.

Results

Phylogenetic Distribution of Nuclear and Mitochondrial Copies of *rpl5* in Grasses

We determined the sequence and locations of *rpl5* genes in 80 grass taxa, six related nongrass species within the Poales, and four additional monocots (supplementary table S1, Supplementary Material online). When sequenced mitochondrial and/or nuclear genomes were available, we could confidently infer gene presence/absence. In species for which only PCR-based data were available, we provisionally concluded that a gene was absent from a genome only when all possible primer combinations (supplementary table S2, Supplementary Material online) failed to produce an amplified product. Mitochondrial *rpl5* copies were generally easier to amplify than nuclear homologs because of the high copy number of mitochondrial genomes and the extremely low substitution rates that are typical of mitochondrial DNA in angiosperms, including grasses (Wolfe et al. 1987; Palmer and Herbon 1988; Gaut 1998; Drouin et al. 2008). Thus, the failure to amplify a mitochondrial-like *rpl5* sequence from a given

species is a good indication of its absence from the mitogenome. In species for which nuclear or mitochondrial genome sequences were not available, two lines of evidence were used to identify amplified *rpl5* sequences as being nuclear in origin: 1) the presence of 5′ extensions that code for N-terminal sequences predicted (see Methods) to confer mitochondrial targeting and 2) an increased nucleotide substitution rate, which is characteristic of nuclear DNA in plants (Wolfe et al. 1987; Drouin et al. 2008).

We found that 18 grasses, six nongrass species within the Poales, and the four other monocots in our sample have only a single detectable *rpl5*, an intact mitochondrial gene; ten grasses have both an intact mitochondrial *rpl5* and an intact nuclear *rpl5*; three grasses have both a mitochondrial *rpl5* pseudogene (identified by the presence of frameshift mutations) and an intact nuclear *rpl5*; 45 grasses have an intact nuclear *rpl5* but no (amplifiable) mitochondrial copy; and four grasses have only a single detectable *rpl5*, a mitochondrial pseudogene (fig. 1; supplementary table S1, Supplementary Material online). In these four cases, it is likely that *rpl5* has been functionally transferred to the nucleus, but we were unable to verify this by amplifying the functional nuclear copy. In these cases, it is also conceivable that the mitochondrial genes have been functionally replaced by duplicated or retargeted copies of a nuclear gene that encodes the homologous subunit in cytosolic or plastid ribosomes (Adams, et al. 2002a). Twenty-three of the 28 species for which only an intact mitochondrial *rpl5* was detected lack a sequenced nuclear genome. Some of these might also possess an intact nuclear *rpl5* copy, one that simply failed to amplify.

The availability of sequenced nuclear genomes from 17 grass species (supplementary table S1, Supplementary Material online) revealed that even species that lack a functional nuclear copy of *rpl5* are not devoid of *rpl5*-like sequences in the nucleus. For example, the nuclear chromosomes of *Oryza sativa* (International Rice Genome Sequencing Project 2005) are replete with fragments of *rpl5* homologs as well as one full-length *rpl5* sequence on chromosome 9 that has 99.6% similarity (two changes out of 567 nucleotides) to the mitochondrial copy (see top of fig. 2). However, these recent *rpl5* insertions in the nucleus as well as other *rpl5* fragments are likely nonfunctional. Nuclear insertions of mitochondrial origins ("*numts*") are common in plant genomes (Lough et al. 2008; Hazkani-Covo et al. 2010), and the vast majority are nonfunctional (Richly and Leister 2004). Figure 2 includes three other putatively nonfunctional *rpl5 numts*, two from *Aegilops tauschii* and one from *Hordeum vulgare*, but most recent *numts* were excluded from these analyses.

Independent Functional Transfers of Mitochondrial *rpl5* to the Nucleus in Grasses

The majority of grass species are found in two large clades, which are referred to as BEP (Bambusoideae, Ehrhartoideae, and Pooideae) and PACMAD (Panicoideae, Arundinoideae, Chloridoideae, Micraioideae, Aristidoideae, and Danthonioideae). The highly intermixed phylogenetic distribution of intact mitochondrial *rpl5* genes, mitochondrial *rpl5* pseudogenes, and mitochondrial gene losses across the BEP and PACMAD clades (fig. 1) could indicate that mitochondrial *rpl5* was repeatedly inactivated following a single functional transfer to the nucleus in the common ancestor of these two clades. This would be similar to the pattern observed for mitochondrial *cox2* in legumes (Adams et al. 1999) and *rps19* in grasses (Atluri et al. 2015). Alternatively, *rpl5* could have been independently transferred multiple times. If mitochondrial *rpl5* experienced only a single ancient loss and transfer to the nucleus, then all grass nuclear *rpl5* genes should form a single well-supported clade that mirrors the species relationships within the grass family. In contrast, if mitochondrial *rpl5* was transferred repeatedly during grass evolution, the identified *rpl5* copies should exhibit topological conflicts with the grass species tree.

We performed phylogenetic analyses to differentiate between these alternative hypotheses and found substantial heterogeneity in rates of sequence evolution between nuclear and mitochondrial copies of *rpl5*, reflecting well-established mutation-rate differences between these genomic compartments in angiosperms, including grasses (Wolfe et al. 1987; Palmer and Herbon 1988; Gaut 1998; Drouin et al. 2008). Analyses based on Bayesian (fig. 2) and maximum-likelihood (supplementary fig. S1, Supplementary Material online) methods both tended to group nuclear *rpl5* sequences together, but neither resolved all of the identified nuclear copies as a single monophyletic clade, even despite the expectation that rapidly evolving sequences will group together spuriously because of long-branch attraction. This bias is a grave problem in phylogenetics generally (Bergsten 2005; Yang and Rannala 2012) and has been shown to affect plant nuclear sequences that are highly divergent relative to mitochondrial homologs (Adams et al. 2000). With one notable exception, the phylogenetic placement of the nuclear *rpl5* sequences provided strong support for monophyly of each of the six grass subfamilies for which at least two species were sampled (fig. 2 and supplementary fig. S1, Supplementary Material online), suggesting that the species within each of these subfamilies shared a single functional transfer of *rpl5*. The one exception was the failure of the nuclear sequence in *Gynerium sagittatum* to group with other members of the Panicoideae (fig. 2). We describe the putative cause of the unexpected placement of *Gynerium* in the next section. Although none of the relationships among the six subfamilies were even moderately well supported, the failure of the subfamilies to form a single monophyletic group and their incongruence with estimates of grass phylogeny from plastid and nuclear data (fig. 1; Grass Phylogeny Working Group II 2012; Wu and Ge 2012; Cotton et al. 2015) nonetheless raise the possibility of independent *rpl5* transfers in some or even all of these subfamilies.

To further assess the possibility of single versus multiple *rpl5* transfers, we examined the N-terminal targeting sequences of identified nuclear copies. As previously reported (Sandoval et al. 2004), the targeting sequences of *Zea* and *Triticum* have no detectable similarity and thus appear to be the product of independent transfers (fig. 3 and

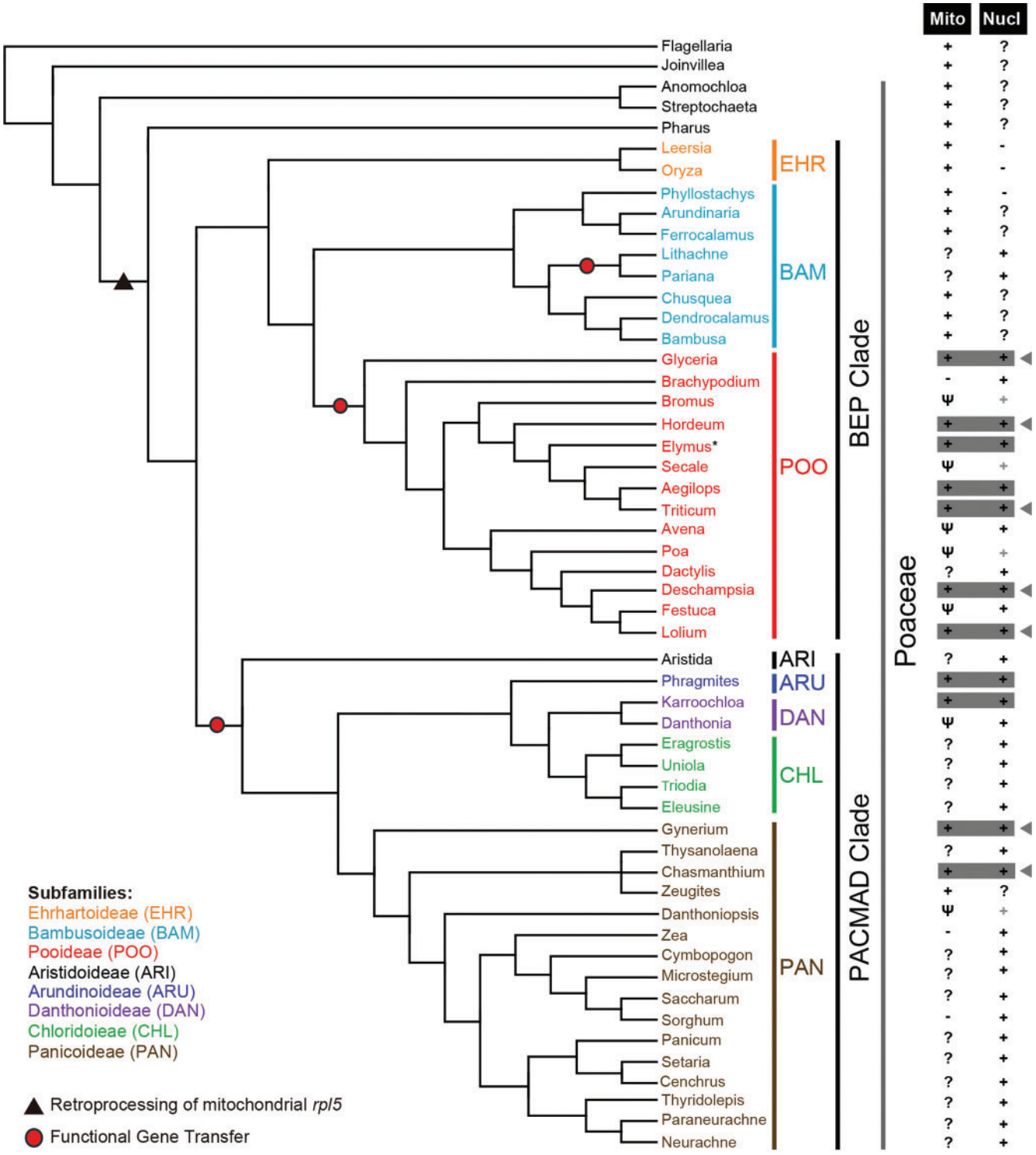

Fig. 1. *rpl5* gene status in 53 genera of Poales. The status of *rpl5* in the mitochondrial and nuclear genomes is indicated as follows: A dark "+" indicates that *rpl5* is an intact open reading frame over the region sequenced. A light "+" in the nuclear column indicates that a nuclear *rpl5* sequence was not obtained by PCR amplification but is inferred to be present because the only mitochondrial copy of *rpl5* found is a pseudogene. A "Ψ" indicates the presence of an *rpl5* pseudogene. A "−" indicates a confirmed absence of an intact copy of *rpl5* in a sequenced mitochondrial or nuclear genome. A "?" indicates that a copy of *rpl5* was not found but its absence cannot be confirmed due to the lack of a genome sequence. Genera with intact copies of *rpl5* in both genomes are highlighted with gray boxes, and those cases in which transcription of both copies has been shown are marked with arrowheads. In some cases, multiple species were sampled from the same genus (supplementary table S1, Supplementary Material online). The asterisk indicates that the inferred presence of intact genes in both genomes in *Elymus* is based on mitochondrial data from one species and nuclear data from another (supplementary table S1, Supplementary Material online). Additional outgroup sequences were surveyed, all of which showed evidence for an intact mitochondrial copy only (supplementary table S1, Supplementary Material online). Poaceae subfamily designations and phylogenetic relationships are based on published sources (Bouchenak-Khelladi et al. 2008; Cotton et al. 2015; Grass Phylogeny Working Group II 2012; Wysocki et al. 2015).

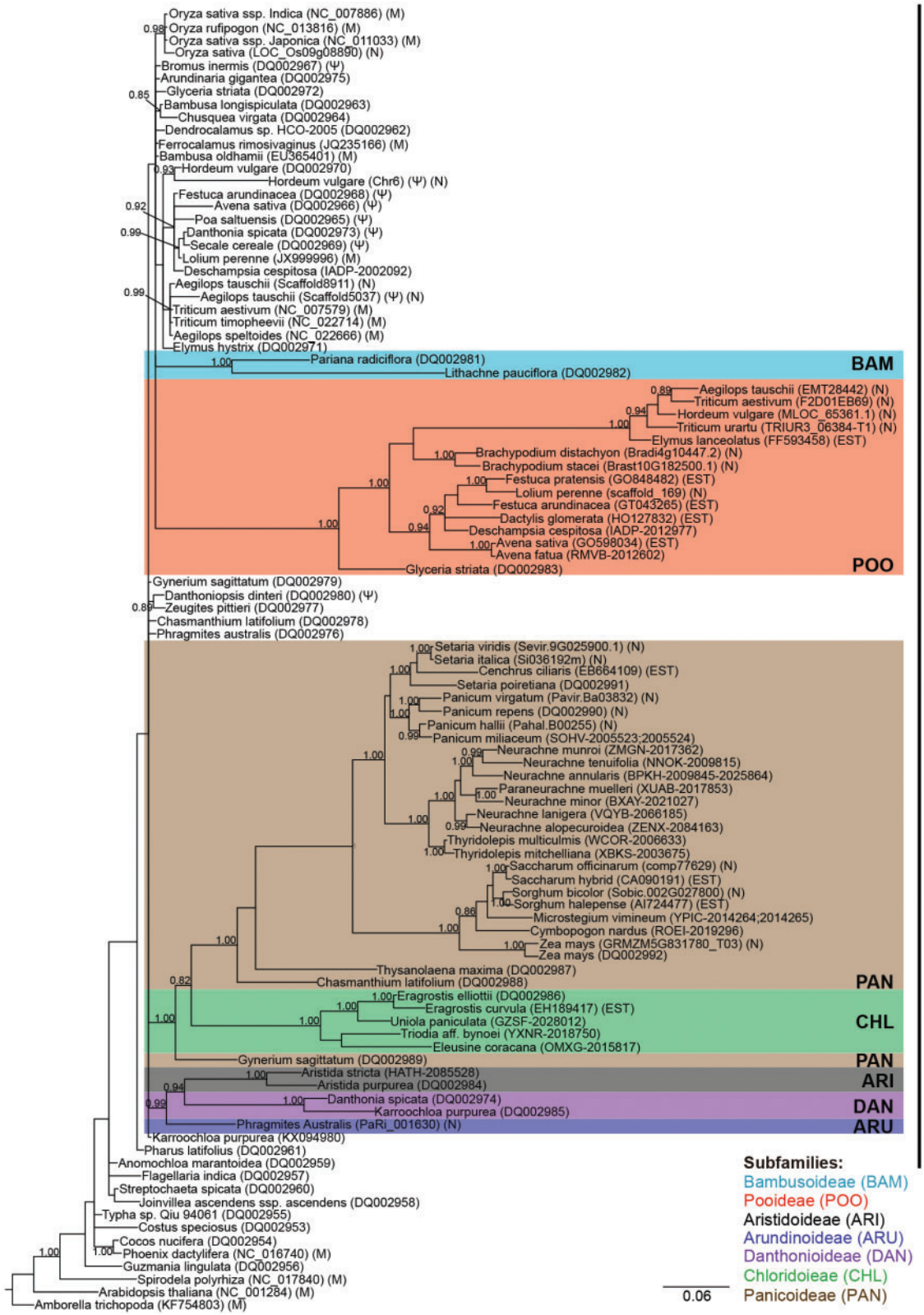

Fig. 2. Bayesian phylogenetic analysis of *rpl5* sequences. Included are 101 mitochondrial and nuclear *rpl5* sequences from 84 taxa. Shading indicates Poaceae subfamilies (see bottom right of figure) and is shown only for intact nuclear genes, i.e., all unshaded sequences are mitochondrial (pseudo)genes or *numt* pseudogenes. Tips are labeled to indicate sequences from complete mitogenomes (M), sequenced nuclear genomes (N), and the NCBI EST database (EST). Accession numbers are specified for GenBank, 1KP, and EST sequences. For species with sequenced nuclear genomes, the gene/locus name is also noted in parentheses. Pseudogenes are labeled as "Ψ", and for pseudogenes from sequenced nuclear genomes, the chromosome/contig/scaffold name is noted in parentheses. Bayesian posterior probability values >0.75 are shown. The scale bar indicates number of substitutions per site.

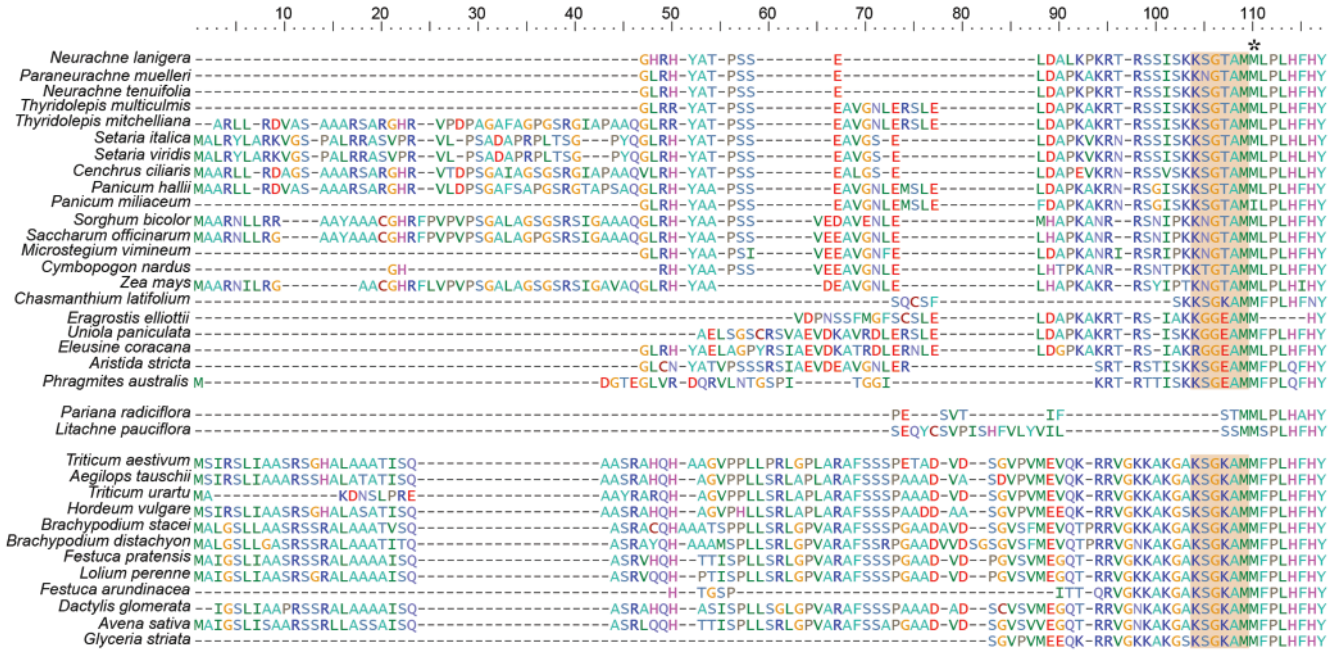

Fig. 3. N-terminal sequences of nuclear-encoded *rpl5* proteins. The sequences are ordered according to the three groups of grasses with distinct N-terminal sequences, indicative of independent functional transfers (also see fig. 1). The asterisk indicates the typical position of the initiator methionine when the *rpl5* gene is in the mitogenome. The shaded amino acids correspond to sequence that is upstream of the mitochondrial gene. In many cases, only partial sequences were available, and no start methionine was identified. Other taxa were omitted because no N-terminal sequence was available. All sequences were aligned with MAFFT.

supplementary figs. S2–S5, Supplementary Material online). We found that the *Triticum*-like targeting sequence is shared with the rest of the Pooideae, including the deepest branching member *Glyceria* (fig. 3 and supplementary fig. S2, Supplementary Material online), indicating that the transfer occurred in a common ancestor of this subfamily. Notably, however, this targeting sequence appears to be unrelated to the N-terminal sequences of *Lithacne* and *Pariana* (fig. 3), members of the subfamily (Bambusoideae) that is sister to Pooideae, indicating that there were independent transfers of *rpl5* in the two groups (fig. 1). Although the N-terminal sequences available for *Lithacne* and *Pariana* are short and of uncertain homology, the strong grouping (98% bootstrap support; supplementary fig. S1, Supplementary Material online) of the mature portions of these two nuclear genes indicates a single ancestral transfer of *rpl5* before the split between *Lithacne* and *Pariana* within the Bambusoideae (fig. 1). The apparent absence of a similar nuclear gene in other members of the Bambusoideae suggests the transfer most likely occurred in an immediate ancestor of *Lithacne* and *Pariana* (however, an older transfer is also possible if there were subsequent losses or if our methods missed related copies in other Bambusoideae lineages).

The *Zea*-like targeting sequence is shared across multiple subfamilies, including the Aristidoideae, Arundinoideae, Chloridoideae, and Panicoideae (fig. 3 and supplementary figs. S4 and S5, Supplementary Material online), which suggests that there was a functional *rpl5* transfer at the base of the PACMAD clade (fig. 1; Grass Phylogeny Working Group II 2012; Cotton et al. 2015). Although targeting sequences were not available from members of the Danthonioideae subfamily, the nuclear *rpl5* sequences from this subfamily share a derived 6-bp deletion with other sampled PACMAD subfamilies (supplementary fig. S6, Supplementary Material online). Therefore, the Danthonioideae nuclear *rpl5* sequences are likely the product of the same ancestral PACMAD transfer event (fig. 1).

Multiple lines of evidence support the interpretation that the *rpl5* nuclear genes identified as being intact in figure 1 are functional. First, we have confirmed that many of them are transcribed. Using RT-PCR and/or 5′ RACE, we succeeded in producing *rpl5* cDNAs for both members of the Bambusoideae transfer clade and for numerous members of the Pooideae and PACMAD transfer clades (supplementary table S1, Supplementary Material online). For many other members of these two clades, *rpl5* sequences were identified in published transcriptome and EST data sets (supplementary table S1, Supplementary Material online), and have acquired intron sequences that are actively spliced out (supplementary figs. S4 and S5, Supplementary Material online; cf. Bonen and Calixte 2006; Liu et al. 2009). Second, in some species, the nuclear copy appears to be the only functional *rpl5* gene based on the absence of an intact *rpl5* gene in sequenced mitogenomes or the failure to amplify an intact mitochondrial *rpl5* sequence with PCR (fig. 1; Clifton et al. 2004; Allen et al. 2007). Third, the nuclear copies of *rpl5* have generally experienced a large excess of synonymous substitutions ($d_N/d_S < 0.25$ in 21 of 27 cases analyzed, with $d_N/d_S < 1.0$ in the other six cases; supplementary table S3, Supplementary Material online) and have maintained reading frames free from frameshifts or internal stop codons. These findings indicate that these genes have been subject

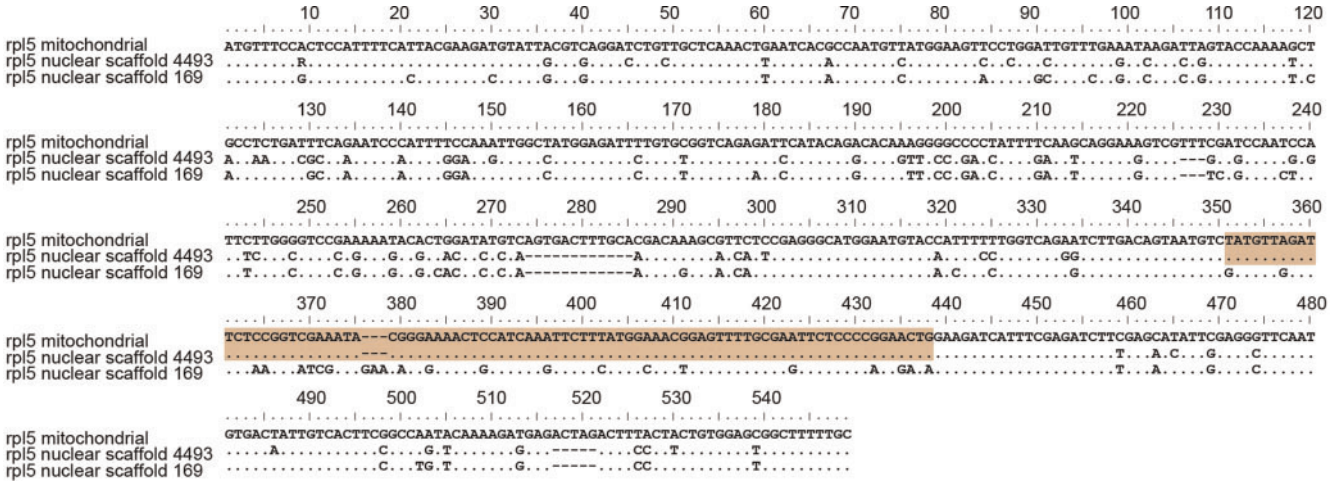

FIG. 4. Nucleotide alignment showing a chimeric *rpl5* sequence in *Lolium perenne*. The region shared by the mitochondrial sequence and the chimeric nuclear sequence is highlighted. All three sequences shown are from *L. perenne*.

to purifying selection over the millions of years since their respective transfers, presumably resulting from functional constraint. Finally, a number of the N-terminal extensions of nuclear copies of *rpl5* were predicted by TargetP and MitoFates to function as mitochondrial targeting sequences (supplementary table S4, Supplementary Material online).

There are seven alignment positions at which the same amino acid substitution or deletion appears to have occurred in parallel in three lineages of nuclear transfers (supplementary fig. S7 and table S5, Supplementary Material online). None of these changes were universal within all three clades, but they all occurred on at least one branch in each of the three lineages. Although such changes may have simply occurred by chance at sites that are not highly constrained, they are also candidate targets for adaptation to nuclear expression and transport into mitochondria.

Coexistence of Intact and Transcribed Nuclear and Mitochondrial *rpl5* Copies and Evidence for Intergenomic Gene Conversion

In ten phylogenetically disparate members of the Pooideae and PACMAD transfer clades, intact *rpl5* copies are present in both the mitochondrial and nuclear genomes (fig. 1 and supplementary table S1, Supplementary Material online; we also found intact mitochondrial and nuclear copies within the genus *Elymus* but have not shown that these cooccur in the same species in this genus). In all ten species, there is evidence for transcription of the nuclear copy, and in all seven of those examined there is also evidence for transcription of the mitochondrial copy (fig. 1; supplementary table S1, Supplementary Material online). Therefore, it is clear that mitochondrial and nuclear gene copies, as well as their expression at the RNA level, have persisted for extensive periods of time after initial duplication and incorporation into the nucleus.

One potential consequence of the coexistence of mitochondrial and nuclear copies, is the opportunity for gene conversion to occur on a transcompartmental scale. We identified two cases of mitochondrial-to-nuclear gene conversion in grass *rpl5* genes, one in *Lolium perenne* and the other in *Gynerium sagittatum*. The assembled nuclear genome of *L. perenne* (Byrne et al. 2015) contains two copies of an *rpl5* gene of mitochondrial origin. One copy (on scaffold 169) is very similar to its sister taxon in our sampling, *Festuca pratensis* (fig. 2 and supplementary fig. S1, Supplementary Material online). The second copy (scaffold 4493) is so similar to the scaffold-169 copy over much of their lengths that the two copies are very likely the result of a single transfer followed by gene duplication within the nuclear genome (fig. 4). However, an 88-bp region within the *rpl5* alignment (fig. 4) shows a strikingly different pattern of sequence relatedness; here the two nuclear copies in *L. perenne* differ substantially from one another (24% different at alignable positions, plus a 3-bp indel) and the scaffold-4493 copy is identical to mitochondrial *rpl5* from *Lolium* (fig. 4). The clearly chimeric nature of the scaffold-4493 nuclear *rpl5* sequence is strong evidence of its relatively recent partial conversion by the mitochondrial sequence. A second such conversion event is readily apparent in the putative nuclear *rpl5* gene of *G. sagittatum*, whose 3′ half closely resembles the transferred sequences found in other members of the Panicoideae, but whose 5′ half is identical to the *Gynerium* mitochondrial sequence (supplementary fig. S8, Supplementary Material online). This observation accounts for the unexpected phylogenetic placement and relatively short branch of the *Gynerium* nuclear sequence (fig. 2 and supplementary fig. S1, Supplementary Material online; also see Hao and Palmer 2011) and also explains why it lacks the 6-bp deletion found in other nuclear sequences from the PACMAD clade (supplementary fig. S6, Supplementary Material online). The nuclear *rpl5* gene from *Phragmites australis*, whose N-terminal region (fig. 3) indicates that it is probably derived from the same transfer as the rest of the PACMAD clade (fig. 1), also lacks this 6-bp deletion (supplementary fig. S6, Supplementary Material online) and exhibits a

relatively short branch length (fig. 2). Unlike *Gynerium*, however, there is no evidence elsewhere in the *Phragmites* sequence for a chimeric origin, and the 6-bp indel region in *Phragmites* differs at two positions from the corresponding region in the *Gynerium* nuclear sequence and all grass mitochondrial sequences (supplementary fig. S6, Supplementary Material online). The evolutionary history of this indel region in *Phragmites* is therefore unclear.

Repeated Acquisitions of *HSP70*-Derived Targeting Sequences by Transferred Mitochondrial Genes

It was previously reported that the targeting sequence associated with the nuclear *rpl5* gene in the Pooideae was likely acquired from another mitochondrially targeted ribosomal protein gene (*rpl4*) based on the similarity between these two sequences in *Triticum aestivum* (Sandoval et al. 2004). However, our broader taxon sampling revealed that the direction of transfer was actually the opposite; *rpl4* genes in *T. aestivum* acquired their targeting sequence from *rpl5*. All Pooideae species have a nuclear *rpl5* gene with a similar targeting sequence, whereas only *T. aestivum* has a copy of *rpl4* with this targeting sequence (supplementary fig. S9, Supplementary Material online). The *rpl4* gene is absent from all sequenced mitogenomes and was presumably transferred to the nucleus in the common ancestor of eukaryotes (Gray et al. 1998). *Triticum aestivum* has additional copies of *rpl4* that carry the targeting sequence that is typically associated with this gene in other grasses (supplementary fig. S10, Supplementary Material online). Therefore, there appears to have been a duplication of the nuclear *rpl4* gene within the *T. aestivum* lineage, with one of the duplicated copies obtaining a targeting sequence from the recently transferred *rpl5* gene in an intriguing process of serial acquisition of targeting sequences.

A genome-wide search for genes with similar targeting sequences identified the mitochondrially targeted chaperonin gene *HSP70* as the ultimate source of the sequence that was acquired by *rpl5* prior to the divergence of the Pooideae subfamily and then transferred to a copy of *rpl4* in the *T. aestivum* lineage (supplementary figs. S9 and S11, Supplementary Material online). *HSP70* has been identified as a relatively frequent donor of targeting sequences for other transferred mitochondrial genes in angiosperms (Liu et al. 2009), including the ribosomal protein gene *rps19* in grasses, which was transferred to the nucleus prior to the diversification of the major grass lineages (Atluri et al. 2015). The *rpl5* targeting sequence in Pooideae species appears to have been acquired directly from *HSP70* rather than via an *rps19* intermediate because the *rpl5* sequence includes all three exons found in *HSP70*. In contrast, the third exon is missing from *rps19* genes throughout the grasses (supplementary fig. S9, Supplementary Material online). Unlike the Pooideae set of nuclear *rpl5* genes, BLAST searches failed to identify source sequences for the putative targeting sequences of the nuclear *rpl5* genes associated with the two other *rpl5* transfers in grasses.

Retroprocessing of Mitochondrial *rpl5* Early in Grass Evolution

The distribution of RNA editing sites in mitochondrial copies of *rpl5* across angiosperms revealed the simultaneous loss of all but one editing site early in the evolution of grasses. A sample of diverse monocots as well as the eudicot *Arabidopsis* and the "basal" angiosperm *Amborella* indicate that angiosperms typically have ~10 C-to-U editing sites in this gene (fig. 5; Giegé and Brennicke 1999; Rice et al. 2013). However, within the Poaceae, only representatives of the Anomochlooideae subfamily (*Anomochloa* and *Streptochaeta*), which is sister to the rest of the grasses (Bouchenak-Khelladi et al. 2008), appear to maintain this high level of RNA editing. The rest of the grasses have lost, via C-to-T substitutions at the DNA level, all but one of the ten RNA editing sites inferred for the ancestral grass (fig. 5). Active editing of the single retained editing site has been shown by empirical studies in *Oryza* and *Triticum* (Kubo et al. 1999; Sandoval et al. 2004), and *Triticum* has gained an additional editing site that is not found in other angiosperms (fig. 5). Retroprocessing has been identified as a mechanism responsible for the simultaneous loss of multiple introns and/or RNA editing sites in mitochondrial genes of seed plants (Lamattina and Grienenberger 1991; Geiss et al. 1994; Krishnasamy et al. 1994; Lu et al. 1998; Itchoda et al. 2002; Parkinson et al. 2005; Petersen et al. 2006; Lopez et al. 2007; Ran et al. 2010; Sloan et al. 2010b; Cuenca et al. 2010, 2012, 2016) and other eukaryotes (Lavrov et al. 2016). Therefore, it appears that the mitochondrial copy of *rpl5* underwent a retroprocessing event (or events) in the common ancestor of all grasses exclusive of the "basal" Anomochlooideae (fig. 1 and 5).

Discussion

The Dynamics and Functionality of Transcompartmental Gene Duplicates

Our analysis highlights the dynamic nature of mitochondrial-to-nuclear transfer of *rpl5* in grasses, which appears to include at least three distinct functional transfers and at least seven species that retain intact and transcribed copies of the gene in both genomes (fig. 1). Coexistence of intact and transcribed copies in the nucleus and mitochondrion has also been reported for *cox2* in legumes (Adams et al. 1999), *rps13* in *Silene vulgaris* (Sloan et al. 2012), *rps19* in grasses (Atluri et al. 2015), *sdh4* in Salicaceae and *Lupinus* (Choi et al. 2006; Havird and Sloan 2016), and *orf164* in Brassicaceae (Qiu et al. 2014), with the latter being an intriguing example of reverse, nuclear-to-mitochondrial transfer. A strong case can be made that functional transfer of genes from mitogenomes to the nucleus generally requires a period of dual functionality in both genomes. In particular, there is evidence that this transitional period persisted for many millions of years in cases where mitochondrial and nuclear genes have been differentially lost following an initial duplicative transfer event (Adams et al. 1999; Atluri et al. 2015).

There is, however, no direct evidence that both the mitochondrial and nuclear copies are currently functional in any one of the many extant plants shown to possess transcompartmentally duplicated and transcribed genes. It is not unlikely that many intact mitochondrial *rpl5* genes in grasses are

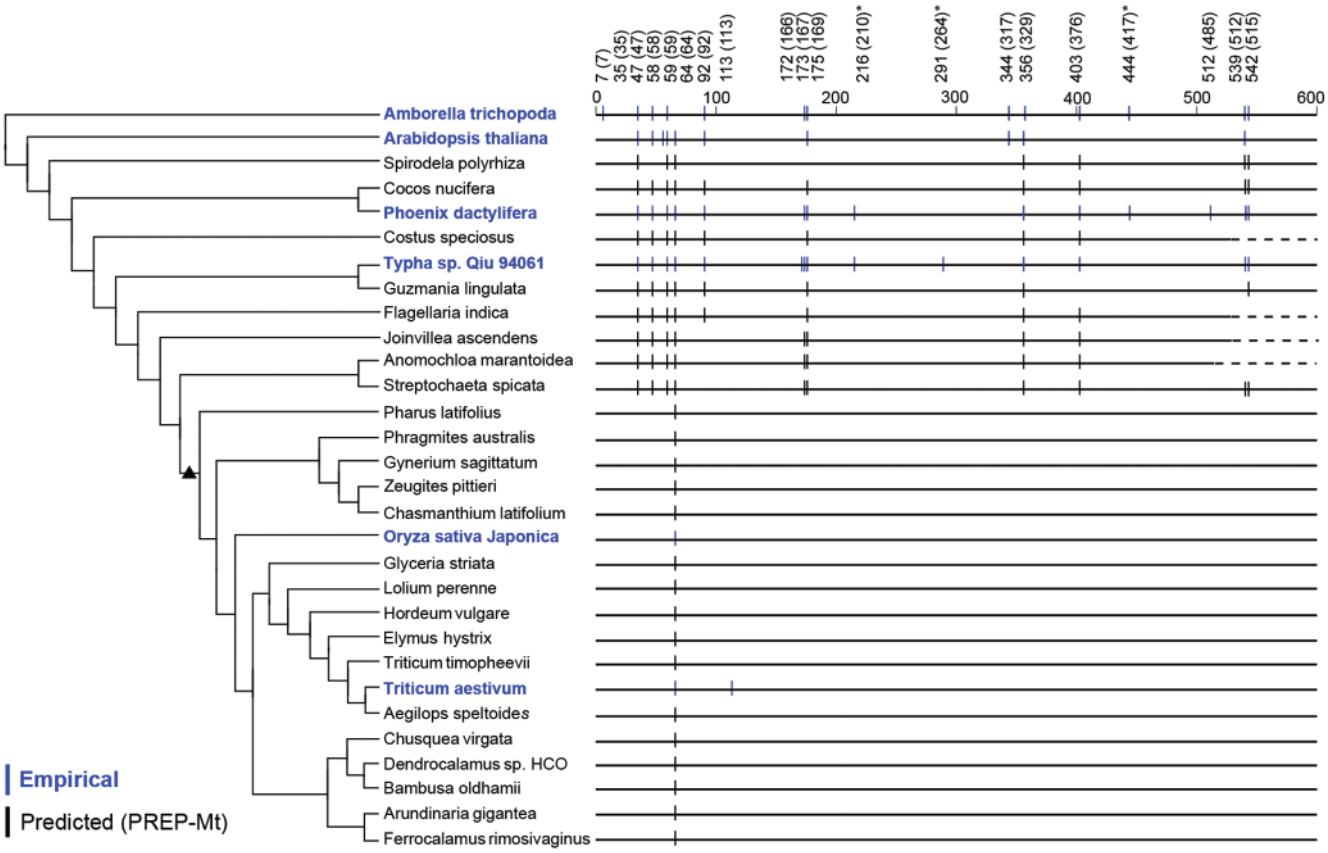

Fig. 5. RNA editing of mitochondrial *rpl5* in grasses and other angiosperms. RNA editing sites are indicated by vertical lines. Numbering indicates the position of each editing site in a multiple alignment of these sequences, whereas the position with respect to the *Oryza sativa* sequence is given in parentheses. Asterisks indicate synonymous editing sites. The species for which editing was empirically determined (see Materials and Methods) are shown in bolded text. Dashed lines indicate missing data. Two predicted editing sites (alignment position 208 in all species and alignment position 257 in *Costus speciosus*) that were not validated in any of the empirical data sets were excluded from the figure. The triangle indicates the timing of the inferred retroprocessing event(s).

actually nonfunctional, as the extremely slow mutation rates in most plant mtDNAs, coupled with the short length of *rpl5*, make the accrual of even a single DNA-level hallmark of pseudogene status (such as a frameshift indel, an unexpectedly large indel or truncation, or a highly deleterious missense mutation) a relatively uncommon event. Furthermore, in plant mitogenomes, intergenic regions are often transcribed and edited without any apparent functional significance (Grewe et al. 2014; Grimes et al. 2014; Wu et al. 2015), and numerous cases are known of plant mitochondrial pseudogenes that are nonetheless transcribed and RNA edited (Schuster and Brennicke 1993; Aubert et al. 1992; Brandt et al. 1993; Sutton et al. 1993; Giegé and Brennicke 1999; Notsu et al. 2002; Mower and Palmer 2006; Ong and Palmer 2006; Sloan et al. 2010b; Grewe et al. 2014; Wu et al. 2015). In some cases, these pseudogenes have persisted for remarkably long periods of time (up to roughly 80 My; Ong and Palmer 2006). Therefore, despite the persistence of intact mitochondrial *rpl5* copies since duplicative transfers to the nucleus at the bases of the PACMAD and Pooideae clades some 40–60 Myr ago (Prasad et al. 2011; Christin et al. 2014), it is possible that many or even all of these mitochondrial copies are actually nonfunctional.

Chimeric Genes Created by Recurrent Gene Transfer Coupled with Mitochondrial-to-Nuclear Gene Conversion

Regardless of whether the surviving mitochondrial copies in the PACMAD and Pooideae clades are still functional, one consequence of the coexistence of mitochondrial and nuclear *rpl5* genes in many grasses is evident in the two examples of chimeric nuclear genes that we identified (fig. 4 and supplementary fig. S8, Supplementary Material online). In both cases, a portion of a transferred nuclear *rpl5* gene has apparently been replaced by sequence from its mitochondrial homolog. To our knowledge, these are the first described cases of mitochondrial-to-nuclear gene conversion. Transcompartmental gene conversion could occur by two related pathways with respect to the converting sequence. This sequence could integrate into the nucleus apart from the ancestrally transferred gene, followed by gene conversion between the two nuclear copies. Alternatively, the converting copy could recombine directly and only with the ancestrally transferred copy, leaving no other trace of the secondary transfer event.

Owing to the much higher substitution rate in grass nuclear genomes compared with their mitochondrial genomes, the

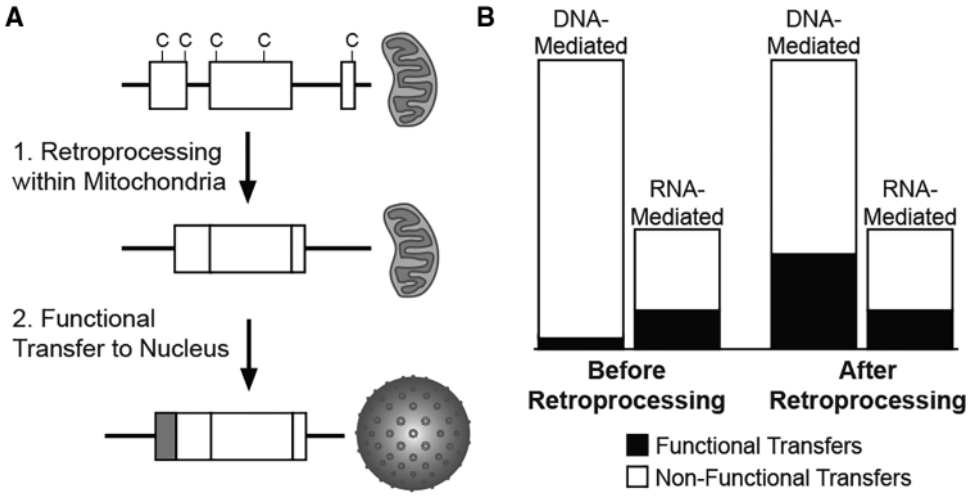

Fig. 6. Conceptual model of the effect of retroprocessing on the relative frequency of DNA- and RNA-mediated functional gene transfers. (*A*) Functional transfer of an organelle gene with many nonsynonymous edit sites and/or introns may be facilitated by a two-step mechanism, in which the gene is first reverse-transcribed and reintegrated into the mitochondrial genome (i.e., retroprocessing) before being physically moved to the nucleus. Introns are drawn as lines between exons (boxes), and sites that require RNA editing are marked with "C." These features are removed by the initial retroprocessing step. The processed gene is then more amenable to DNA-mediated functional transfer to the nucleus, but acquisition of nuclear-specific elements such as an N-terminal targeting sequence (gray box) is still required in most cases. (*B*) Prior to retroprocessing, the frequency of functional transfer to the nucleus is expected to be low and dominated by directly RNA-mediated events, i.e., events in which an edited and/or spliced mitochondrial cDNA is transferred to the nucleus. After wholesale retroprocessing of such a gene, the frequency of functional transfer is expected to increase significantly and be dominated by DNA-mediated events. Note that this figure is meant only to illustrate a conceptual model. The actual proportions of functional versus nonfunctional and DNA- versus directly RNA-mediated transfers have not been quantified.

two chimeric genes are a decided mixture of relatively derived (nuclear) and relatively ancestral (mitochondrial) *rpl5* sequences. Therefore, the process of mitochondrial-to-nuclear gene conversion has the potential to obscure the history of posttransfer sequence divergence. In fact, it may even erase signatures of the transfer process itself. As we discuss in the next section, transfer events that are RNA-mediated (either directly or indirectly) result in C-to-T changes at the DNA level corresponding to sites of mRNA editing. But subsequent mitochondrial-to-nuclear gene conversion events could result in T-to-C reversions.

Retroprocessing and the High Rate of DNA-Mediated *rpl5* Gene Transfer in Grasses

The occurrence of multiple functional transfers of a mitochondrial gene within a single plant family raises important questions about the factors than govern the frequency of gene transfer to the nucleus. We found that the three functional *rpl5* transfers in grasses were preceded by retroprocessing that eliminated all but one RNA editing site in this gene (figs. 1 and 4). This finding has important implications for the relative contributions of DNA-mediated and RNA-mediated intracellular gene duplications to functional organellar gene transfer. Many cases of functional transfer of plant mitochondrial genes are thought to involve processed RNA intermediates because the transferred genes lack the RNA editing sites and/or introns present in the mostly closely related mitochondrial genes (Nugent and Palmer 1991; Covello and Gray 1992; Brennicke et al. 1993; Wischmann and Schuster 1995; Knoop et al. 1995; Adams et al. 2000; Subramanian and Bonen 2006; Hazle and Bonen 2007; Liu et al. 2009). However, the physical form of the nucleic acids (i.e., RNA, cDNA, or genomic DNA) that move between genomic compartments and result in functional transfers remains contentious (Henze and Martin 2001; Timmis et al. 2004; Timmis 2012). In particular, the evidence for frequent transfer of genomic DNA from organelles to the nucleus (Blanchard and Schmidt 1995; Stupar et al. 2001; Huang et al. 2003; Stegemann et al. 2003; Richly and Leister 2004; Noutsos et al. 2005; Lough et al. 2008; Hazkani-Covo et al. 2010; Michalovova et al. 2013) has generated uncertainty regarding the role of RNA intermediates.

Despite experimental evidence that some plastid-to-nuclear functional transfers can occur without eliminating introns or editing sites (Sheppard et al. 2011; Fuentes et al. 2012), the mitochondrial genes contained in large-scale transfers of DNA fragments are generally expected to be "dead on arrival" because of their inability to be expressed in the nucleus. This is especially true for genes that contain large numbers of introns and/or RNA editing sites, the latter being abundant in most mitochondrial protein genes of angiosperms (Mower et al. 2012). Therefore, even though RNA-mediated transfer is far less frequent than DNA-mediated transfer (Fuentes et al. 2012), it is likely that a much larger proportion of RNA-mediated transfers yields functional genes in the nucleus (fig. 6). The retroprocessing of mitochondrial genes, as in the case of grass *rpl5*, should significantly increase the proportion of DNA-mediated (but not RNA-mediated) transfers that become functional (fig. 6). Thus, the relatively high frequency of functional intracellular transfers of grass

mitochondrial *rpl5* was probably facilitated by retroprocessing that occurred near the base of the grass family. In effect, this model separates an RNA-mediated functional transfer into a two-step process: 1) RNA-mediated retroprocessing *within* the mitogenome and 2) subsequent DNA-mediated physical transfer to the nuclear genome (Henze and Martin 2001). It has been shown that retroprocessed sequences can coexist (at least transiently) in the mitogenome with ancestral copies of the same gene (Cuenca et al. 2012). Therefore, it is likely that this two-step transfer process can also act in a cryptic fashion in lineages in which there has not been wholesale replacement of the native gene within the mitogenome. Another potential example of this two-step process is the recent report of *matR* functional transfer to the nucleus in *Pelargonium* following a more ancient retroprocessing event within the mitogenome during Geraniaceae evolution (Grewe et al. 2016).

Importantly, the retroprocessing of mitochondrial *rpl5* early in grass evolution did not result in the permanent elimination of one of the ancestral editing sites, the nonsynonymous site at position 64 (fig. 5). There are at least three plausible explanations for this observation. First, this site may have been restored—in the mitochondrial genome—by a T-to-C back-mutation relatively soon after retroprocessing occurred. Second, retroprocessing may have missed this site, as a result of either multiple partial gene-conversion events, none of which overlapped the site, or a single discontinuous conversion that skipped it. There is ample evidence that gene conversion can affect localized regions within plant mitochondrial genes (Bergthorsson et al. 2003; Barkman et al. 2007; Hao and Palmer 2009; Hao et al. 2010; Mower et al. 2010; Sloan et al. 2010a, 2010b; Hepburn et al. 2012). Third, retroprocessing may have involved an RNA transcript that happened not to have been edited at position 64, as partial editing is common in plant mitochondrial genes particularly at sites that are not functionally important (Mower and Palmer 2006; Wu et al. 2015; Guo et al. 2017). Regardless of the explanation, the presence of this one editing site is fortuitous because it provides important insight into the of role of DNA-mediated versus RNA-mediated mechanisms in *rpl5* gene transfer in grasses. Notably, all identified nuclear copies of *rpl5* across the Poaceae have a C at this position in the gene, supporting the interpretation that they are the result of DNA-mediated transfers. The fact that copies of *rpl5* have repeatedly become functional in the nucleus without the His-to-Tyr amino acid change created by RNA editing at this position suggests that editing at this particular site is not essential and helps explain why partial editing could be tolerated.

The extent to which retroprocessing increases the frequency of functional intracellular transfer of other genes and in other taxa remains to be seen. One intriguing possibility is that genome-wide increases in rates of retroprocessing are responsible for the highly reduced mitochondrial gene content seen in the genomes of plants such as *Lachnocaulon*, *Allium*, *Erodium*, and *Phlox* (Adams et al. 2002b; Park et al. 2015). As the number of sequenced angiosperm mitogenomes continues to increase, it should become possible to identify historical retroprocessing events and further test this hypothesized relationship between retroprocessing and functional gene transfer.

Materials and Methods

Taxon Sampling

We generated and collected data from multiple sources, resulting in a total of 80 grasses (Poaceae), six related nongrass species within the Poales, and four additional monocots (supplementary table S1, Supplementary Material online). These included 32 taxa with sequenced nuclear and/or mitochondrial genomes, 20 species from the 1000 Plants (1KP) initiative (http://www.onekp.com/) and ten species with expressed-sequence-tag (EST) data in the NCBI EST database (http://www.ncbi.nlm.nih.gov/nucest/). We also performed PCR and Sanger sequencing of *rpl5* from 36 species. Six of the taxa were duplicated in two different data sets, and one species (*Zea mays*) was present in three different data sets, resulting in a total of 90 unique taxa (including subspecies). *Arabidopsis thaliana* and *Amborella trichopoda* were also included as outgroups.

Nucleic Acid Extraction and cDNA Synthesis

Total genomic DNA was isolated from H_2O-rinsed leaf-tissue samples using the CTAB isolation procedure of Doyle and Doyle (1987). Total plant RNAs were isolated with the RNeasy Mini Kit (Qiagen, CA) and treated twice with DNase I (TaKaRa Bio Inc., Japan). Reverse transcription reactions were performed using Moloney Murine Leukemia Virus reverse transcriptase (New England Biolabs, MA) and random 9-mer oligonucleotides (Invitrogen, CA) followed by a final RNase treatment. Control templates were generated in parallel without reverse transcriptase to test for genomic DNA amplification.

PCR Amplification and Sanger Sequencing

Genomic DNA and cDNA samples were used to amplify mitochondrial and nuclear copies of *rpl5*. The six PCR primers used for amplification are given in supplementary table S2, Supplementary Material online. Thermal cycling was performed with the following parameters: 94 °C for 2 min, followed by 35 cycles of 94 °C for 1 min, 56 °C for 30 s, and 72 °C for 1 min. After these cycles, a polishing extension at 72 °C for 10 min was performed. Amplified products were directly sequenced or cloned with TOPO TA Cloning Kit (Invitrogen, San Diego, CA) prior to sequencing. Both strands were sequenced using an Applied Biosystems 3700 automated fluorescence system operated by the Indiana University Molecular Biology Institute. Characterization and identification of upstream sequences were performed with a combination of the Vectorette II kit (Sigma-Genosys, MO) and RLM-RACE kit (Ambion, TX) according to the manufacturers' protocols. The resulting sequences were deposited in GenBank (accession numbers are given in supplementary table S1, Supplementary Material online).

Plant Genomes and Transcriptomes

We extracted *rpl5* sequences from publicly available plant genomes and transcriptomes (supplementary table S1, Supplementary Material online) with NCBI TBLASTN, using the amino acid sequence from the *Oryza sativa* mitogenome (GenBank NC_011033) as a query. BLAST searches were performed with default parameters. All hits covering >50% of the query sequence were retained. For searches of the 1KP data set, partial and overlapping hits from the same species were manually merged.

Phylogenetic Analyses

The set of *rpl5* nucleotide sequences was trimmed to exclude all extensions of the nuclear genes relative to the mitochondrial ones and aligned using MAFFT v7.221 (Katoh and Standley 2013). The resulting alignment was inspected with BioEdit v7.1.9 and adjusted manually. Bayesian and maximum-likelihood phylogenetic analyses were performed with MrBayes v3.1.2 (Ronquist and Huelsenbeck 2003) and PhyML v2.4.4 (Guindon and Gascuel 2003), respectively. For Bayesian analyses, the MCMC algorithm was run for 10,000,000 generations with four incrementally heated chains, starting from random trees and sampling one out of every 1000 generations. When the log-likelihood scores stabilized, a consensus tree was calculated after discarding the first 25% of the trees as burn-in. The maximum-likelihood analyses were performed with 1000 bootstrap replicates, the GTR substitution model, four rate categories, and an initial BIONJ tree. The transition/transversion ratio, proportion of invariable sites, and gamma distribution parameter were estimated from the data. The jModelTest v2.1.7 software package (Darriba et al. 2012) was used to inform model selection. RNA editing sites were not excluded from phylogenetic analyses.

Identification of N-Terminal Targeting Sequences and RNA-Editing Sites

N-terminal targeting sequences in nuclear-encoded copies of *rpl5* were predicted from primary amino acid sequences using TargetP v1.01 (Emanuelsson et al. 2000) and MitoFates (Fukasawa et al. 2015). For mitochondrial-encoded copies, PREP-Mt (Mower 2009) was used to predict editing sites based on genomic DNA sequences, using a cutoff value of 0.5. In addition, empirically determined RNA-editing sites were taken from published sources (Giegé and Brennicke 1999; Kubo et al. 1999; Sandoval et al. 2004; Rice et al. 2013). We identified RNA-editing sites in *Phoenix dactylifera* by comparing the *rpl5* genomic sequence with a published cDNA library (Fang et al. 2012). We also tentatively identified RNA-editing sites in *Typha* by comparing our genomic sequence from *Typha* sp. Qiu 94061 with cDNA sequences of putative mitochondrial origin from two *Typha* species in the 1KP data set.

Estimation of d_N/d_S Ratios

The codeml program within PAML v4.7 (Yang 2007) was used to calculate d_N/d_S ratios for nuclear copies of *rpl5*. A topology inferred from Bayesian phylogenetic analysis as above (supplementary fig. S12, Supplementary Material online) was used as a constraint tree, and a separate d_N/d_S value was calculated for each branch based on an F1 × 4 codon-frequency model. Partial sequences that covered <90% of the alignment were excluded from the analysis, and only a single representative sequence from each genus was used, resulting in a total of 27 sequences, including the mitochondrial copy of *rpl5* from *Phoenix dactylifera*, which was used as an outgroup.

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

We thank E. Kellogg, L. Clark, and K. Clay for generously contributing DNA samples and/or plant materials; Gane Ka-Shu Wong and the 1KP consortium for providing access to the 1KP data set; M.P. Simmons for helpful discussion about phylogenetic analysis; and two anonymous reviewers and the lab groups of DBS and Rachel Mueller for comments on an earlier version of this manuscript. This research was supported by NIH GM-35087 and GM-70612 to JDP and by NSF MCB-1412260 and start-up funds from Colorado State University to DBS.

References

Adams KL, Daley DO, Qiu YL, Whelan J, Palmer JD. 2000. Repeated, recent and diverse transfers of a mitochondrial gene to the nucleus in flowering plants. *Nature* 408:354–357.

Adams KL, Palmer JD. 2003. Evolution of mitochondrial gene content: gene loss and transfer to the nucleus. *Mol Phylogenet Evol*. 29:380–395.

Adams KL, Daley DO, Whelan J, Palmer JD. 2002a. Genes for two mitochondrial ribosomal proteins in flowering plants are derived from their chloroplast or cytosolic counterparts. *Plant Cell* 14:931–943.

Adams KL, Qiu YL, Stoutemyer M, Palmer JD. 2002b. 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 U S A*. 99:9905–9912.

Adams KL, Song K, Roessler PG, Nugent JM, Doyle JL, Doyle JJ, Palmer JD. 1999. Intracellular gene transfer in action: dual transcription and multiple silencing of nuclear and mitochondrial *cox2* genes in legumes. *Proc Natl Acad Sci U S A*. 96:13863–13868.

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

Atluri S, Rampersad SN, Bonen L. 2015. Retention of functional genes for S19 ribosomal protein in both the mitochondrion and nucleus for over 60 million years. *Mol Genet Genomics* 290:2325–2333.

Aubert D, Bisanz-Seyer C, Herzog M. 1992. Mitochondrial *rps14* is a transcribed and edited pseudogene in *Arabidopsis thaliana*. *Plant Mol Biol*. 20:1169–1174.

Barkman TJ, McNeal JR, Lim SH, Coat G, Croom HB, Young ND, Depamphilis CW. 2007. Mitochondrial DNA suggests at least 11 origins of parasitism in angiosperms and reveals genomic chimerism in parasitic plants. *BMC Evol Biol*. 7:248.

Bergsten J. 2005. A review of long-branch attraction. *Cladistics* 21:163–193.

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

Blanchard JL, Schmidt GW. 1995. Pervasive migration of organellar DNA to the nucleus in plants. *J Mol Evol*. 41:397–406.

Bonen L, Calixte S. 2006. Comparative analysis of bacterial-origin genes for plant mitochondrial ribosomal proteins. *Mol Biol Evol*. 23:701–712.
Bouchenak-Khelladi Y, Salamin N, Savolainen V, Forest F, Bank M, Chase MW, Hodkinson TR. 2008. Large multi-gene phylogenetic trees of the grasses (Poaceae): progress towards complete tribal and generic level sampling. *Mol Phylogenet Evol*. 47:488–505.
Brandt P, Unseld M, Eckert-Ossenkopp U, Brennicke A. 1993. An *rps14* pseudogene is transcribed and edited in *Arabidopsis* mitochondria. *Curr Genet*. 24:330–336.
Brennicke A, Grohmann L, Hiesel R, Knoop V, Schuster W. 1993. The mitochondrial genome on its way to the nucleus: different stages of gene transfer in higher plants. *FEBS Lett*. 325:140–145.
Byrne SL, Nagy I, Pfeifer M, Armstead I, Swain S, et al. 2015. A synteny-based draft genome sequence of the forage *grass Lolium perenne*. *Plant J*. 84:816–826.
Choi C, Liu Z, Adams KL. 2006. Evolutionary transfers of mitochondrial genes to the nucleus in the Populus lineage and coexpression of nuclear and mitochondrial *Sdh4* genes. *New Phytol*. 172:429–439.
Christin PA, Spriggs E, Osborne CP, Strömberg CA, Salamin N, Edwards EJ. 2014. Molecular dating, evolutionary rates, and the age of the grasses. *Syst Biol*. 63:153–165.
Clifton SW, Minx P, Fauron CMR, Gibson M, Allen JO, Sun H, Thompson M, Barbazuk WB, Kanuganti S, Tayloe C, et al. 2004. Sequence and comparative analysis of the maize NB mitochondrial genome. *Plant Physiol*. 136:3486–3503.
Cotton JL, Wysocki WP, Clark LG, Kelchner SA, Pires JC, Edger PP, Mayfield-Jones D, Duvall MR. 2015. Resolving deep relationships of PACMAD grasses: a phylogenomic approach. *BMC Plant Biol*. 15:178.
Covello PS, Gray MW. 1992. Silent mitochondrial and active nuclear genes for subunit 2 of cytochrome c oxidase (*cox2*) in soybean: evidence for RNA-mediated gene transfer. *EMBO J*. 11:3815–3820.
Cuenca A, Petersen G, Seberg O, Davis JI, Stevenson DW. 2010. Are substitution rates and RNA editing correlated?. *BMC Evol Biol*. 10:349.
Cuenca A, Petersen G, Seberg O, Jahren AH. 2012. Genes and processed paralogs coexist in plant mitochondria. *J Mol Evol*. 74:158–169.
Cuenca A, Ross TG, Graham SW, Barrett CF, Davis JI, Seberg O, Petersen G. 2016. Localized retroprocessing as a model of intron loss in the plant mitochondrial genome. *Genome Biol Evol*. 8:2176–2189.
Darriba D, Taboada GL, Doallo R, Posada D. 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods* 9:772.
Doyle JJ, Doyle JL. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissues. *Phytochem Bull*. 19:11–15.
Drouin G, Daoud H, Xia JN. 2008. Relative rates of synonymous substitutions in the mitochondrial, chloroplast and nuclear genomes of seed plants. *Mol Phylogenet Evol*. 49:827–831.
Emanuelsson O, Nielsen H, Brunak S, von Heijne G. 2000. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J Mol Biol*. 300:1005–1016.
Fang Y, Wu H, Zhang T, Yang M, Yin Y, Pan L, Yu X, Zhang X, Hu S, Al-Mssallem IS, Yu J. 2012. A complete sequence and transcriptomic analyses of date palm (*Phoenix dactylifera* L.) mitochondrial genome. *PLoS One* 7:E37164.
Figueroa P, Gomez I, Holuigue L, Araya A, Jordana X. 1999. Transfer of *rps14* from the mitochondrion to the nucleus in maize implied integration within a gene encoding the iron-sulphur subunit of succinate dehydrogenase and expression by alternative splicing. *Plant J*. 18:601–609.
Fuentes I, Karcher D, Bock R. 2012. Experimental reconstruction of the functional transfer of intron-containing plastid genes to the nucleus. *Curr Biol*. 22:763–771.
Fukasawa Y, Tsuji J, Fu S-C, Tomii K, Horton P, Imai K. 2015. MitoFates: Improved prediction of mitochondrial targeting sequences and their cleavage sites. *Mol Cell Proteomics* 14:1113–1126.
Gaut BS. 1998. Molecular clocks and nucleotide substitution rates in higher plants. *Evol Biol*. 30:93–120.
Geiss KT, Abbas GM, Makaroff CA. 1994. Intron loss from the NADH dehydrogenase subunit 4 gene of lettuce mitochondrial DNA: evidence for homologous recombination of a cDNA intermediate. *Mol Gen Genet*. 243:97–105.
Giegé P, Brennicke A. 1999. RNA editing in *Arabidopsis* mitochondria effects 441 C to U changes in ORFs. *Proc Natl Acad Sci U S A*. 96:15324–15329.
GPWG II (Grass Phylogeny Working Group II). 2012. New grass phylogeny resolves deep evolutionary relationships and discovers C4 origins. *New Phytol*. 193:304–312.
Gray MW, Burger G, Lang BF. 1999. Mitochondrial evolution. *Science* 283:1476–1481.
Gray MW, Lang BF, Cedergren R, Golding GB, Lemieux C, Sankoff D, Turmel M, Brossard N, Delage E, Littlejohn TG, et al. 1998. Genome structure and gene content in protist mitochondrial DNAs. *Nucleic Acids Res*. 26:865–878.
Grewe F, Edger PP, Keren I, Sultan L, Pires JC, Ostersetzer-Biran O, et al. 2014. Comparative analysis of 11 Brassicales mitochondrial genomes and the mitochondrial transcriptome of *Brassica oleracea*. *Mitochondrion* 19:135–143.
Grewe F, Zhu A, Mower JP. 2016. Loss of a trans-splicing *nad1* intron from Geraniaceae and transfer of the maturase gene matR to the nucleus in *Pelargonium*. *Genome Biol Evol*. 8:3193–3201.
Grimes BT, Sisay AK, Carroll HD, Cahoon AB. 2014. Deep sequencing of the tobacco mitochondrial transcriptome reveals expressed ORFs and numerous editing sites outside coding regions. *BMC Genomics* 15:31.
Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol*. 52:696–704.
Guo W, Zhu A, Fan W, Mower JP. 2017. Complete mitochondrial genomes from the ferns *Ophioglossum californicum* and *Psilotum nudum* are highly repetitive with the largest organellar introns. *New Phytol*. 213:391–403.
Hammani K, Giegé P. 2014. RNA metabolism in plant mitochondria. *Trends Plant Sci*. 19:380–389.
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, Palmer JD. 2011. HGT turbulence: confounding phylogenetic influence of duplicative horizontal transfer and differential gene conversion. *Mob Genet Elements* 1:256–304.
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.
Havird JC, Sloan DB. 2016. The roles of mutation, selection, and expression in determining relative rates of evolution in mitochondrial vs. nuclear genomes. *Mol Biol Evol*. 33:3042–3053.
Hazkani-Covo E, Zeller RM, Martin W. 2010. Molecular poltergeists: mitochondrial DNA copies (*numts*) in sequenced nuclear genomes. *PLoS Genet*. 6:e1000834.
Hazle T, Bonen L. 2007. Status of genes encoding the mitochondrial S1 ribosomal protein in closely-related legumes. *Gene* 405:108–116.
Henze K, Martin W. 2001. How do mitochondrial genes get into the nucleus?. *Trends Genet*. 17:383–387.
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.
Huang CY, Ayliffe MA, Timmis JN. 2003. Direct measurement of the transfer rate of chloroplast DNA into the nucleus. *Nature* 422:72–76.
International Rice Genome Sequencing Project. 2005. The map-based sequence of the rice genome. *Nature* 436:793–800.
Itchoda N, Nishizawa S, Nagano H, Kubo T, Mikami T. 2002. The sugar beet mitochondrial nad4 gene: an intron loss and its phylogenetic implication in the Caryophyllales. *Theor Appl Genet*. 104:209–213.
Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 30:772–780.

Knoop V, Ehrhardt T, Lättig K, Brennicke A. 1995. The gene for ribosomal protein S10 is present in mitochondria of pea and potato but absent from those of *Arabidopsis* and *Oenothera*. *Curr Genet*. 27:559–564.

Krishnasamy S, Grant RA, Makaroff CA. 1994. Subunit 6 of the Fo-ATP synthase complex from cytoplasmic male-sterile radish: RNA editing and NH2-terminal protein sequencing. *Plant Mol Biol*. 24:129–141.

Kubo N, Harada K, Hirai A, Kadowaki K. 1999. A single nuclear transcript encoding mitochondrial RPS14 and SDHB of rice is processed by alternative splicing: Common use of the same mitochondrial targeting signal for different proteins. *Proc Natl Acad Sci U S A*. 96:9207–9211.

Lamattina L, Grienenberger JM. 1991. RNA editing of the transcript coding for subunit 4 of NADH dehydrogenase in wheat mitochondria: uneven distribution of the editing sites among the four exons. *Nucleic Acids Res*. 19:3275–3282.

Lavrov DV, Adamski M, Chevaldonné P, Adamska M. 2016. Extensive mitochondrial mRNA editing and unusual mitochondrial genome organization in calcaronean sponges (Calcarea, Porifera). *Curr Biol*. 26:86–92.

Liu SL, Zhuang Y, Zhang P, Adams KL. 2009. Comparative analysis of structural diversity and sequence evolution in plant mitochondrial genes transferred to the nucleus. *Mol Biol Evol*. 26:875–891.

Long M, de Souza SJ, Rosenberg C, Gilbert W. 1996. Exon shuffling and the origin of the mitochondrial targeting function in plant cytochrome c1 precursor. *Proc Natl Acad Sci U S A*. 93:7727–7731.

Lopez L, Picardi E, Quagliariello C. 2007. RNA editing has been lost in the mitochondrial cox3 and rps13 mRNAs in *Asparagales*. *Biochimie* 89:159–167.

Lough AN, Roark LM, Kato A, Ream TS, Lamb JC, Birchler JA, Newton KJ. 2008. Mitochondrial DNA transfer to the nucleus generates extensive insertion site variation in maize. *Genetics* 178:47–55.

Lu MZ, Szmidt AE, Wang XR. 1998. RNA editing in gymnosperms and its impact on the evolution of the mitochondrial *coxI* gene. *Plant Mol Biol*. 37:225–234.

Michalovova M, Vyskot B, Kejnovsky E. 2013. Analysis of plastid and mitochondrial DNA insertion in the nucleus (*NUPTs* and *NUMTs*) of six plant species: size, relative age and chromosomal localization. *Heredity* 111:314–320.

Mower JP. 2009. The PREP Suite: Predictive RNA editors for plant mitochondrial genes, chloroplast genes and user-defined alignments. *Nucleic Acids Res*. 37:W253–W259.

Mower JP, Palmer JD. 2006. Patterns of partial RNA editing in mitochondrial genes of *Beta vulgaris*. *Mol Genet Genomics* 276:285–293.

Mower JP, Stefanović 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, Sloan DB, Alverson AJ. 2012. Plant mitochondrial genome diversity: the genomics revolution. In: Wendel JF, Greilhuber J, Dolezel J, Leitch IJ, editors. Plant Genome Diversity, Vol 1. Vienna: Springer; 123–144.

Notsu Y, Masood S, Nishikawa T, Kubo N, Akiduki G, Nakazono M, et al. 2002. The complete sequence of the rice (*Oryza sativa* L.) mitochondrial genome: frequent DNA sequence acquisition and loss during the evolution of flowering plants. *Mol Genet Genomics* 268:434–445.

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.

Nugent JM, Palmer JD. 1991. RNA-mediated transfer of the gene *coxII* from the mitochondrion to the nucleus during flowering plant evolution. *Cell* 66:473–481.

Ogihara Y, Yamazaki Y, Murai K, Kanno A, Terachi T, Shiina T, Miyashita N, Nasuda S, Nakamura C, Mori N, et al. 2005. Structural dynamics of cereal mitochondrial genomes as revealed by complete nucleotide sequencing of the wheat mitochondrial genome. *Nucleic Acids Res*. 33:6235–6250.

Ong HC, Palmer JD. 2006. Pervasive survival of expressed mitochondrial rps14 pseudogenes in grasses and their relatives for 80 million years following three functional transfers to the nucleus. *BMC Evol Biol*. 6:55.

Palmer JD, Herbon LA. 1988. Plant mitochondrial DNA evolves 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*. 208:570–583.

Parkinson CL, Mower JP, Qiu YL, Shirk AJ, Song K, 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.

Petersen G, Seberg O, Davis JI, Stevenson DW. 2006. RNA editing and phylogenetic reconstruction in two monocot mitochondrial genes. *Taxon* 55:871–886.

Prasad V, Strömberg CA, Leaché AD, Samant B, Patnaik R, Tang L, Mohabey DM, Ge S, Sahni A. 2011. Late Cretaceous origin of the rice tribe provides evidence for early diversification in Poaceae. *Nat Commun*. 2:480.

Qiu Y, Filipenko SJ, Darracq A, Adams KL. 2014. Expression of a transferred nuclear gene in a mitochondrial genome. *Curr Plant Biol*. 1:68–72.

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.

Rand DM, Haney RA, Fry AJ. 2004. Cytonuclear coevolution: the genomics of cooperation. *Trends Ecol Evol*. 19:645–653.

Rice DW, Alverson AJ, Richardson AO, Young GJ, Sanchez-Puerta MV, Munzinger J, Barry K, Boore JL, Zhang Y, Knox EB, et al. 2013. Horizontal transfer of entire genomes via mitochondrial fusion in the angiosperm *Amborella*. *Science* 342:1468–1473.

Richly E, Leister D. 2004. NUMTs in sequenced eukaryotic genomes. *Mol Biol Evol*. 21:1081–1084.

Ronquist F, Huelsenbeck JP. 2003. MrBAYES 3, Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572–1574.

Sandoval P, Leon G, Gomez I, Carmona R, Figueroa P, Holuigue L, Araya A, Jordana X. 2004. Transfer of RPS14 and RPL5 from the mitochondrion to the nucleus in grasses. *Gene* 324:139–147.

Schuster W, Brennicke A. 1993. RNA editing makes mistakes in plant mitochondria: editing loses sense in transcripts of a *rps19* pseudogene and in creating stop codons in *coxI* and *rps3* mRNAs of *Oenothera*. *Nucleic Acids Res*. 19:6923–6928.

Sheppard AE, Madesis P, Lloyd AH, Day A, Ayliffe MA, Timmis JN. 2011. Introducing an RNA editing requirement into a plastid-localised transgene reduces but does not eliminate functional gene transfer to the nucleus. *Plant Mol Biol*. 76:299–309.

Sloan DB, Alverson AJ, Štorchová H, Palmer JD, Taylor DR. 2010a. Extensive loss of translational genes in the structurally dynamic mitochondrial genome of *Silene latifolia*. *BMC Evol Biol*. 10:274.

Sloan DB, MacQueen AH, Alverson AJ, Palmer JD, Taylor DR. 2010b. 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. 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.

Stegemann S, Hartman S, Ruf S, Bock R. 2003. High-frequency gene transfer from the chloroplast genome to the nucleus. *Proc Natl Acad Sci U S A*. 100:8828–8833.

Stupar RM, Lilly JW, Town CD, Cheng Z, Kaul S, Buell CR, Jiang J. 2001. Complex mtDNA constitutes an approximate 620-kb insertion on *Arabidopsis thaliana* chromosome 2: implication of potential sequencing errors caused by large-unit repeats. *Proc Natl Acad Sci U S A*. 98:5099–5103.

Subramanian S, Bonen L. 2006. Rapid evolution in sequence and length of the nuclear-located gene for mitochondrial L2 ribosomal protein in cereals. *Genome* 49:275–281.

Sutton CA, Conklin PL, Pruitt KD, Calfee AJ, Cobb AG, Hanson MR. 1993. Editing of *rps3*/*rpl16* transcripts creates a premature truncation of the *rpl16* open reading frame. *Curr Genet*. 23:472–476.

Timmis JN, Ayliffe MA, Huang CY, Martin W. 2004. Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nat Rev Genet*. 5:123–135.

Timmis JN. 2012. Endosymbiotic evolution: RNA intermediates in endosymbiotic gene transfer. *Curr Biol*. 22:R296–R298.

Wischmann C, Schuster W. 1995. Transfer of *rps10* from the mitochondrion to the nucleus in *Arabidopsis thaliana*: evidence for RNA-mediated transfer and exon shuffling at the integration site. *FEBS Lett*. 374:152–156.

Wolfe KH, Li WH, Sharp PM. 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proc Natl Acad Sci U S A*. 84:9054–9058.

Wu ZQ, Ge S. 2012. The phylogeny of the BEP clade in grasses revisited: Evidence from the whole-genome sequences of chloroplasts. *Mol Phylogenet Evol*. 62:573–578.

Wu Z, Stone JD, Štorchová H, Sloan DB. 2015. High transcript abundance, RNA editing, and small RNAs in intergenic regions within the massive mitochondrial genome of the angiosperm *Silene noctiflora*. *BMC Genomics* 16:938.

Wysocki WP, Clark LG, Attigala L, Ruiz-Sanchez E, Duval MR. 2015. Evolution of the bamboos (Bambusoideae; Poaceae): a full plastome phylogenomic analysis. *BMC Evol Biol*. 15:50.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol*. 24:1586–1591.

Yang Z, Rannala B. 2012. Molecular phylogenetics: principles and practice. *Nat Rev Genet*. 13:303–314.