

RESEARCH

Open Access



Comparative analyses of three complete *Primula* mitogenomes with insights into mitogenome size variation in Ericales

Lei Wei^{1,2†}, Tong-Jian Liu^{1*†}, Gang Hao³, Xue-Jun Ge¹  and Hai-Fei Yan^{1*} 

Abstract

Background: Although knowledge of the sizes, contents, and forms of plant mitochondrial genomes (mitogenomes) is increasing, little is known about the mechanisms underlying their structural diversity. Evolutionary information on the mitogenomes of *Primula*, an important ornamental taxon, is more limited than the information on their nuclear and plastid counterparts, which has hindered the comprehensive understanding of *Primula* mitogenomic diversity and evolution. The present study reported and compared three *Primula* mitogenomes and discussed the size expansion of mitogenomes in Ericales.

Results: Mitogenome master circles were sequenced and successfully assembled for three *Primula* taxa and were compared with publicly available Ericales mitogenomes. The three mitogenomes contained similar gene contents and varied primarily in their structures. The *Primula* mitogenomes possessed relatively high nucleotide diversity among all examined plant lineages. In addition, high nucleotide diversity was found among *Primula* species between the Mediterranean and Himalaya-Hengduan Mountains. Most predicted RNA editing sites appeared in the second amino acid codon, increasing the hydrophobic character of the protein. An early stop in *atp6* caused by RNA editing was conserved across all examined Ericales species. The interfamilial relationships within Ericales and interspecific relationships within *Primula* could be well resolved based on mitochondrial data. Transfer of the two longest mitochondrial plastid sequences (MTPTs) occurred before the divergence of *Primula* and its close relatives, and multiple independent transfers could also occur in a single MTPT sequence. Foreign sequence [MTPTs and mitochondrial nuclear DNA sequences (NUMTs)] uptake and repeats were to some extent associated with changes in Ericales mitogenome size, although none of these relationships were significant overall.

Conclusions: The present study revealed relatively conserved gene contents, gene clusters, RNA editing, and MTPTs but considerable structural variation in *Primula* mitogenomes. Relatively high nucleotide diversity was found in the *Primula* mitogenomes. In addition, mitogenomic genes, collinear gene clusters, and locally collinear blocks (LCBs) all showed phylogenetic signals. The evolutionary history of MTPTs in *Primula* was complicated, even in a single MTPT sequence. Various reasons for the size variation observed in Ericales mitogenomes were found.

[†]Lei Wei and Tong-Jian Liu contributed equally to this work.

*Correspondence: liutongjian@scbg.ac.cn; yanhaifei@scbg.ac.cn

¹ Key Laboratory of Plant Resources Conservation and Sustainable Utilization, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou, China

Full list of author information is available at the end of the article



Keywords: Genome skimming, Intracellular gene transfer event, Mitochondrial genome, Phylogenetic relationship, Primulaceae

Background

The mitochondrion is one of three genetic compartments in plant cells that play a fundamental role in encoding several essential mitochondrial electron transfer chains [1, 2]. Only 321 plant mitochondrial genomes (mitogenomes) have been released in the NCBI organelle database (organelle genome resources), while more than 7000 plant plastid genomes have been assembled and published to date (accessed on Jan. 18, 2022). Advancements in sequencing technologies and subsequently applied assembly algorithms have resulted in abundant plant mitogenomes (especially for closely related plants), which could contribute to revealing the evolutionary mechanisms of mitogenomes [3] and the domestication history of plants [4].

Plant mitogenomes vary in size but show conserved gene contents. For example, a core set of 24 protein-coding genes are generally shared among angiosperm mitogenomes [5], whereas the mitogenome length varies from 66kb in *Viscum scurruloideum* Barlow [6] to 11.3Mb in *Silene conica* L. [7]. RNA editing can result in amino acid sequences that differ from the corresponding conserved gene templates, which is often pervasive in the mitochondria of diverse eukaryotes (especially in land plants, [8–10]) and frequently causes the conversion of cytosine to uridine [8, 9, 11]. Another essential characteristic of plant mitogenomes is the great variation in their structures and noncoding contents [1, 5, 12]. Plant mitogenomes contain orders of magnitude more noncoding nucleotides (including introns, repetitive elements, and foreign DNA) than their metazoan counterparts [13], which contribute to the large size or the size variation of plant mtDNA genomes. First, mitochondrial introns in vascular plants play essential roles in the regulation of mitochondrial genes, which can reach 11.4kb in *Nymphaea colorata* [14]. Group II introns are particularly prevalent within plant mitogenomes; these introns form a lariat-like structure via splicing, in contrast to group I introns. Second, the presence of repeats of various sizes in plant mitogenomes may result in high rates of genome rearrangements and even promote alternative genomic forms via recombination [12]. Third, the uptake of foreign sequences from the nuclear and plastid genomes is a vital source of noncoding sequences and might contribute to angiosperm mitogenome size expansion. For example, mitochondrial plastid sequences (MTPTs), which were first identified in maize, contribute 1 to 10% of the mitochondrial genomes of higher plants [15]. The majority of

MTPTs are nonfunctional, with the exception of several tRNA genes and partial genes such as *ccmC* in *Vitis vinifera* [16]. In addition, a large plant mitogenome often includes larger mitochondrial nuclear DNA sequences (NUMTs). For example, approximately 20 and 33% of mtDNA was shown to be imported from the nucleus in apple [17] and watermelon [18], respectively. In short, the exploration of these noncoding sequences is essential to reveal the evolution of plant mitogenomes.

To date, most plant mitogenomes have been reconstructed as master circles containing the complete mitochondrial gene set [1, 7, 19–21]. Several hypotheses have been proposed to explain the size and/or structural variations of plant mitogenomes, such as different DNA repair mechanisms [22] and the integration of foreign sequences [3, 23]. However, the mechanisms underlying genome size variation and rearrangements remain unclear due to the small number of publicly available mitogenomes in plants [24–26].

The genus *Primula* L. (Primulaceae, Ericales) is an important ornamental and alpine plant group comprising more than 500 species worldwide [27–29]. The genus *Primula* has experienced rapid radiation (~30 Mya, [30]) and is mainly distributed in two geographically distant hotspots, the Himalaya–Hengduan Mountains in Asia and the Caucasus–Alps–Pyrenees regions in Europe [29]. The mechanisms underlying mitogenome evolution remain largely unknown because of the scarcity of well-studied mitochondrial data in *Primula* compared with the continuing progress in understanding nuclear and plastid genomes [31–33]. To our knowledge, only the *Aegiceras corniculatum* (L.) Blanco mitogenome from the family Primulaceae has been assembled thus far [34]. Thus, mitogenomes of species within *Primula* should be further studied, which may facilitate the discovery of their evolutionary dynamics on different scales, such as at the genus, family, or order level.

In this study, we aimed to explore the evolutionary history of mitogenomes among three *Primula* taxa: *Primula valentiniana* Hand.-Mazz., *Primula smithiana* Craib, and *Primula palinuri* Petagna. The first two species are restricted to alpine regions of the Himalaya–Hengduan Mountains. In contrast, *P. palinuri* is endemic to southern Italy and is adapted to the Mediterranean climate. We first sequenced and assembled the complete mitogenomes and plastomes of the three *Primula* species. Sequence repeats, foreign DNA fragments, RNA editing sites, and structural variation were analyzed in the

three assembled mitogenomes. We also compared the nucleotide diversity of *Primula* taxa with that of other plant lineages. In addition, we discuss the phylogenetic application of mitogenomes and the potential causes of size variation in Ericales mitogenomes. In conclusion, our study provides insights into how the mitogenome has evolved in Ericales.

Results

Complete mitogenome size and gene content

The master circles of the three *Primula* mitogenomes were successfully reconstructed using two strategies (details in the Methods). Several putative mitochondrial contigs were identified for the de novo assembly of the *P. palinuri*, *P. valentiniana*, and *P. smithiana* sequences (Table S1). Further extension and merging of these contig sequences generated complete mitogenomes with master circles of 407,597 bp, 349,360 bp, and 426,527 bp, respectively (Fig. 1; Table 1). These draft genomes were verified by complete read remapping and/or polymerase chain reaction (PCR) amplification, since all the target sequences in the merging region of raw contigs were confirmed (Fig. S1). These mitogenomes exhibited similar GC contents, ranging from 45.2 to 45.5% (Fig. 1; Table 1).

Thirty-eight protein-coding genes, three rRNA genes, and 19 tRNA genes were shared among the three species (Table S2). These results suggested that the three mitogenomes presented relatively conserved gene contents. Most protein-coding genes occurred in single copies within the three mitogenomes. However, *atp8*, *cox1*, *cox3*, *nad4*, *rps10*, *sdh3*, and *sdh4* presented two gene copies in *P. smithiana* (Table S2). These mitogenomes also contained the same RNA genes (i.e., *rrn5*, *rrn18*, and *rrn26*). It was notable that although most of the 19 shared tRNA genes occurred in single copies, the *P. smithiana* mitogenome possessed two copies at the *trnC-GCA*, *trnD-GUC*, *trnM-CAU*, *trnN-GUU*, *trnS-GCU*, *trnS-UGA*, and *trnY-GUA* genes and four copies at the *trnH-GUG* gene. All *Primula* mitogenomes lacked the group I intron, whereas 30, 22, and 23 group II introns were identified in *P. smithiana*, *P. palinuri*, and *P. valentiniana*, respectively (Table 1).

RNA editing in *Primula* mitogenomes

The three mitogenomes exhibited similar editing site numbers predicted with the PREP-Mt server [35]. For example, *P. palinuri*, *P. valentiniana*, and *P. smithiana* contained 467, 460, and 456 C-U editing sites, respectively (Fig. 2a). Approximately 300 RNA editing sites occurred at the 2nd base of the codon in each of the three *Primula* species. In contrast, approximately 140 sites occurred at the 1st base of the codon (Fig. 2a). Although the RNA editing sites were not evenly distributed among

coding genes, the hydrophobic properties of all proteins were increased after editing (Fig. 2b). These amino acid changes mainly involved Ser-Leu, Pro-Leu, Arg-Cys, Pro-Ser, and His-Tyr transitions (Fig. 2c). For example, the average hydrophobic character per site in *ccmB* increased from 0.28 to 0.93 (Fig. 2d).

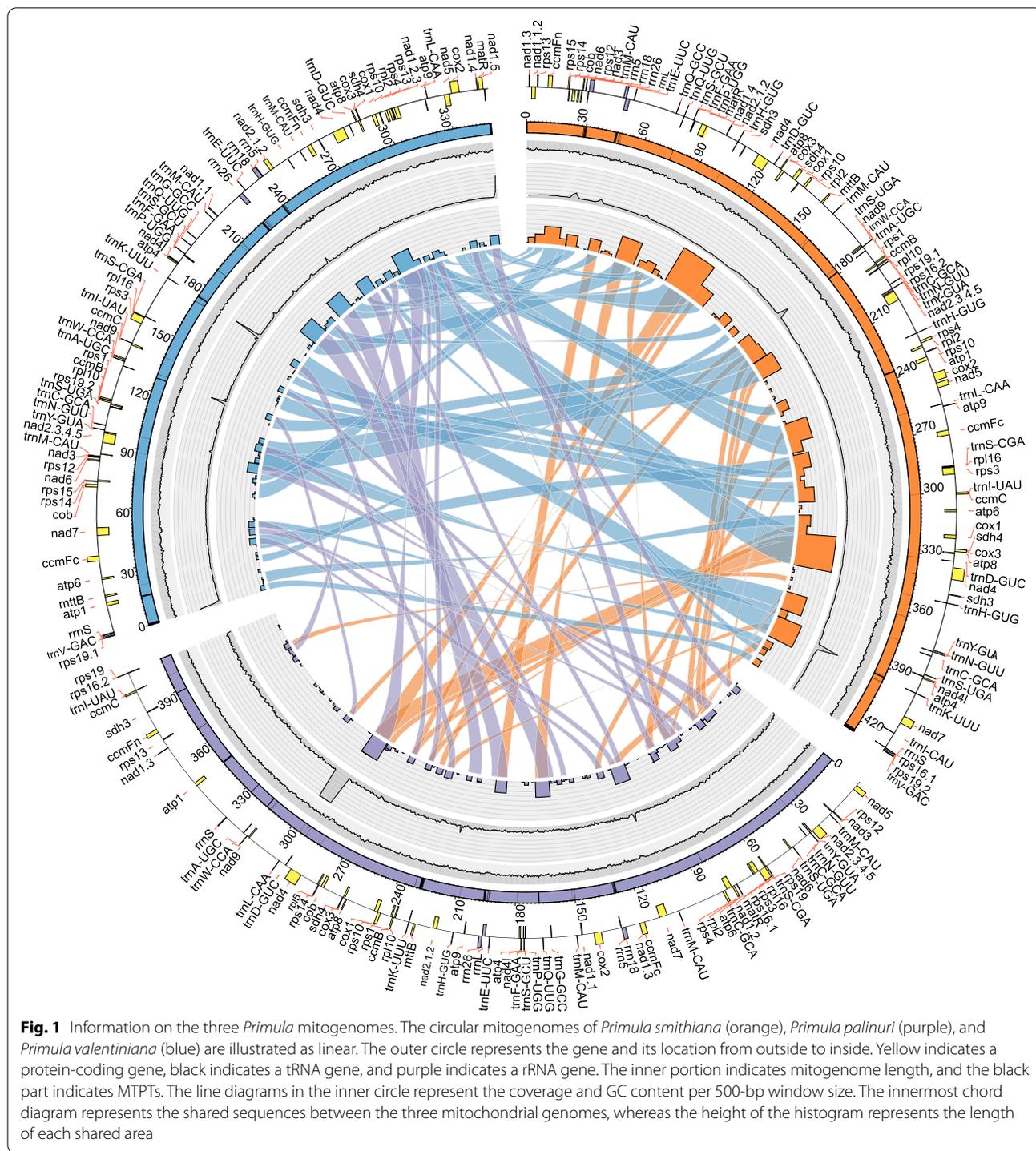
Two RNA editing sites in *atp6* and *ccmFc* produced stop codons in all *Primula* taxa. An editing-derived stop codon resulted in a 107-bp deletion in *atp6* (Fig. S2), and an RNA editing site occurred at the penultimate amino acid codon in *ccmFc* (Fig. S3). Notably, both editing sites were also found across the other eight Ericales mitogenomes (Figs. S2–3). The editing efficiency of the edited-introducing stop codon in *atp6* ranged from 50% (*C. sinensis*) to 100% (*R. simsii*), suggesting effective termination. However, the editing efficiency for *ccmFc* was 0, indicating that the predicted RNA editing site was a false positive.

Repeats and foreign-derived sequence identification

The total length of repeat regions in the three *Primula* mitogenomes varied dramatically. The *P. smithiana* mitogenome contained repeats with a total length of 181,003 bp (Fig. 3a), accounting for over 30% of the whole mitogenome sequence. *P. palinuri* and *P. valentiniana* exhibited relatively compact mitogenomes with total repeat lengths of 8105 bp and 4036 bp, respectively (Fig. 3b–c). The *Primula* mitogenomes lacked large tandem repeats (Tables S3–5).

Eleven MTPTs were identified within the three *Primula* taxa, and seven were commonly identified in the three species (Table S6). In addition, MTPT2 and MTPT6 were specific to *P. palinuri*, whereas MTPT8 and MTPT9 were specific to *P. smithiana* and *P. valentiniana*. Only the first two largest MTPTs (MTPT7 and MTPT10, both longer than 500 bp) were extracted for further analysis, as the *Primula* mitogenomes lacked large MTPTs. The depths of these MTPTs ranged from 44× to 80×, whereas those of their plastid counterparts ranged from 147.5× to 828.5× (Fig. S4). The presence of the two MTPTs was further verified through PCR amplification (Fig. S1). The GC contents of MTPT7 were 38.5% (*P. smithiana*), 38.7% (*P. valentiniana*), and 40.9% (*P. palinuri*), whereas those of MTPT10 were approximately 50–51% among the three species.

The phylogenetic relationships indicated that MTPT7 might have experienced three independent transfer events (Fig. 4a). One may have occurred after the speciation of *P. palinuri*, as its MTPT7 was grouped with its plastid counterpart. In contrast, the other two were distributed at distant positions in the phylogenetic tree (Fig. 4a). In particular, MTPT7 occurred very recently, with a unique insertion (approximately 1000 bp, Fig. 4a)



in *P. palinuri*, exhibiting a higher GC content (40.9%). The oldest MTPT7 event occurred around the time of divergence between *Androsace* L. and *Bryocarpum* Hook. f. & Thomson (Fig. 4a), whereas MTPT10 probably appeared before the diversification of *Primula* (Fig. 4b).

NUMTs were searched against the *P. veris* L. nuclear genome (GCA_000788445.1). The total length of nuclear-shared sequences ranged from 210kb to 266kb. The longest nuclear-shared sequence was 9928bp, which occurred in *P. smithiana* (Table S7).

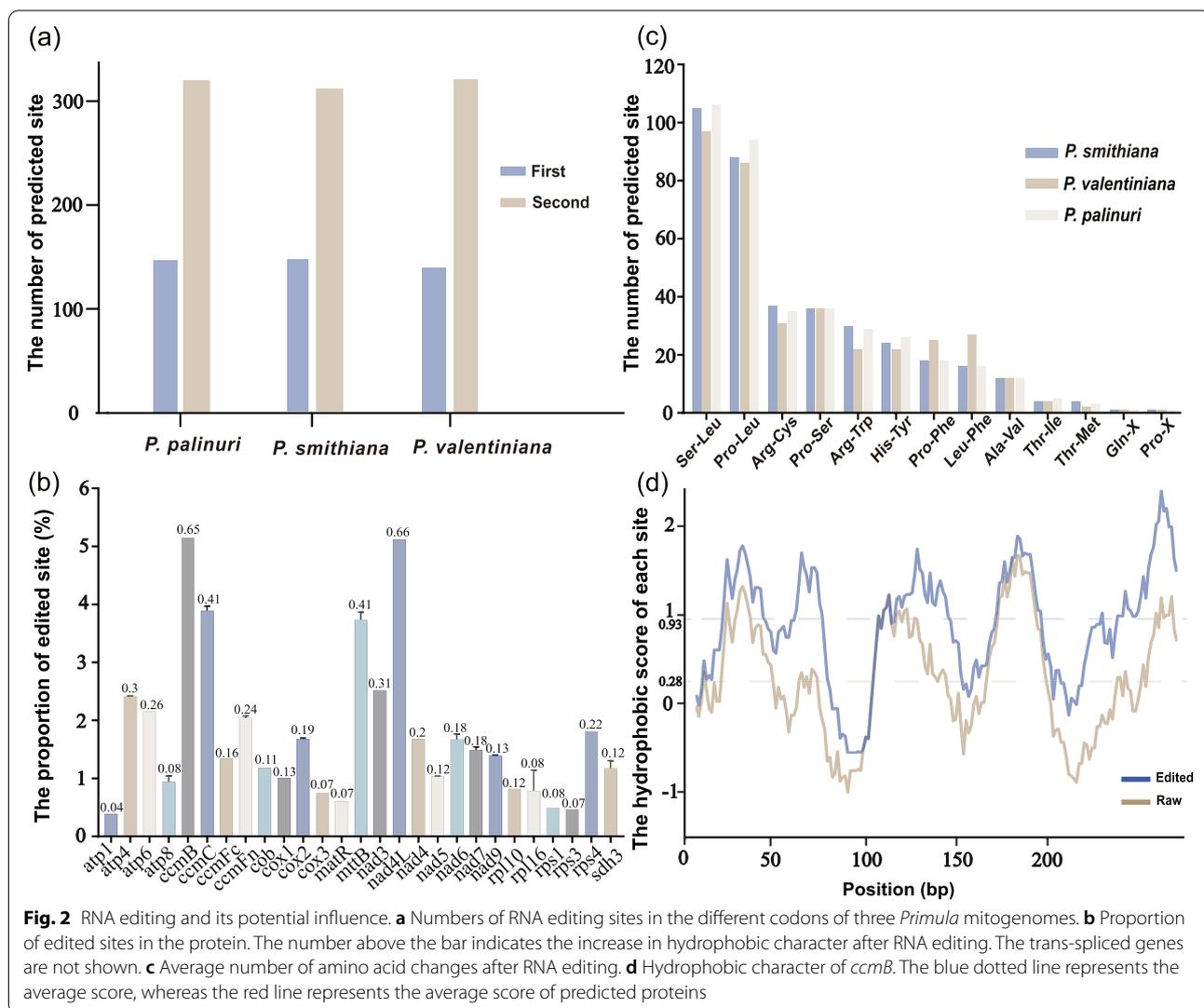
Table 1 General features of the three *Primula* mitogenomes

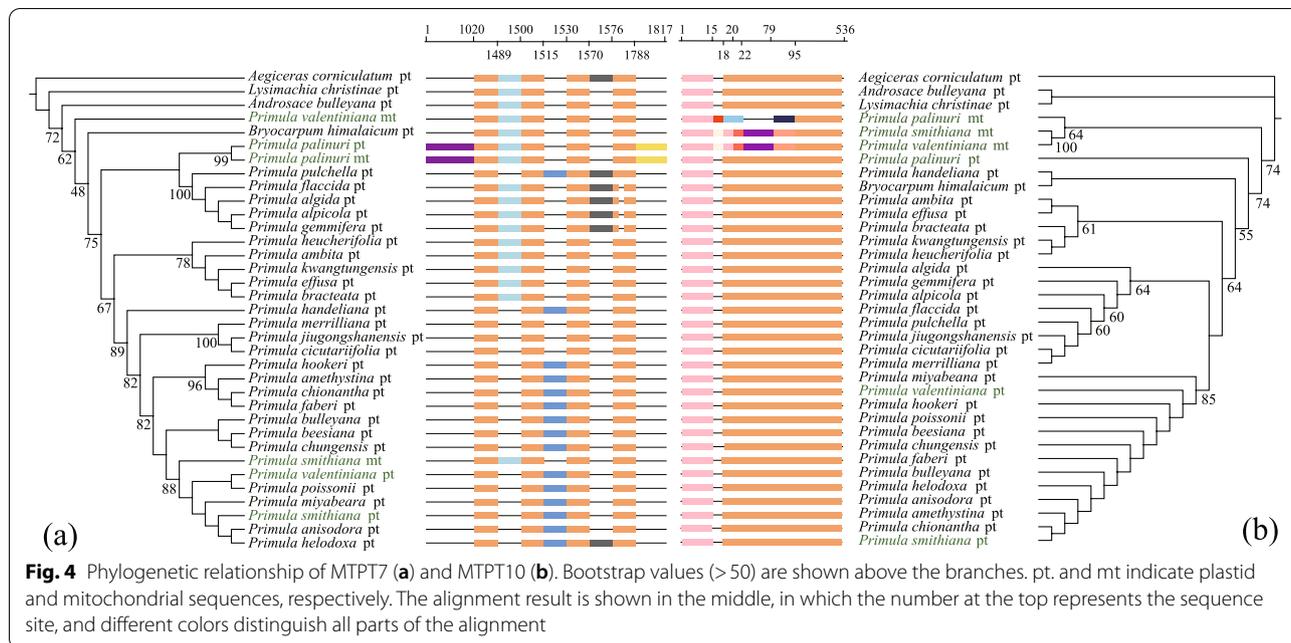
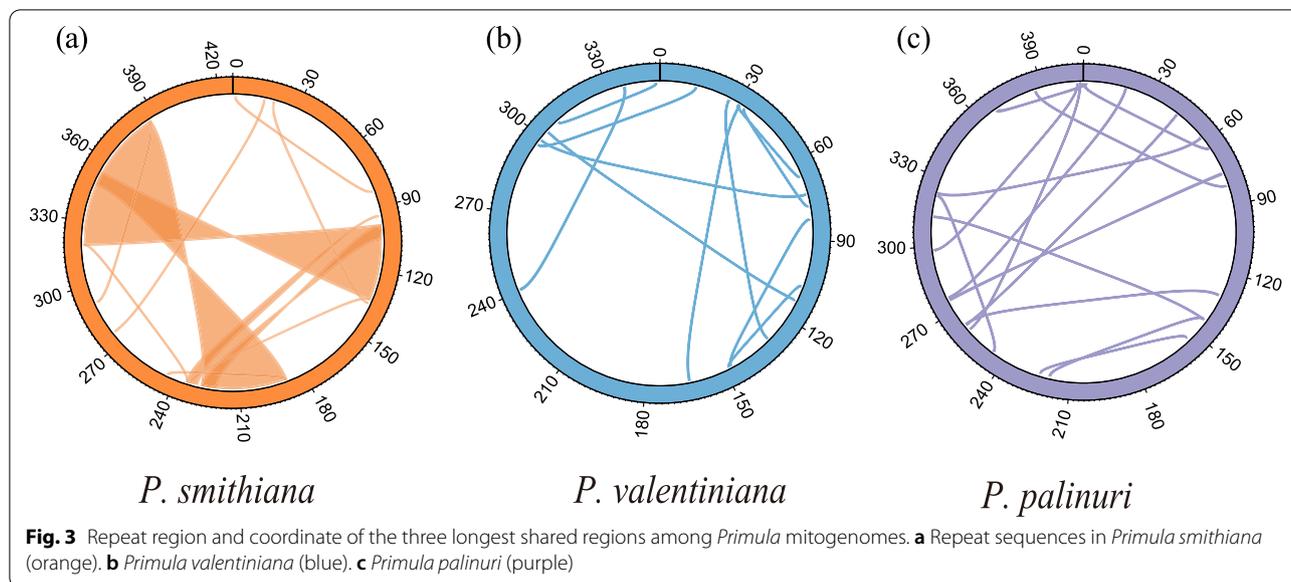
	<i>P. palinuri</i>	<i>P. smithiana</i>	<i>P. valentiniana</i>
Size	407,597 bp	426,527 bp	349,360 bp
GC content	45.5%	45.2%	45.5%
Gene numbers	61	68	64
protein coding	39	41	39
rRNA	3	3	3
tRNA	19	23	20
Introns	18	22	18
Repeat region	8105 bp	181,003 bp	4036 bp
Plastid-derived region	3073 bp	2491 bp	2857 bp
Tandem repeat patterns	8	16	7
Total length of tandem repeats	161 bp	367 bp	198 bp

Locally collinear blocks and gene clusters in *Primula* mitogenomes

The locally collinear blocks (LCBs) are the conserved genomic sequences among all genomes, which is useful for mitogenome structural analysis and for constructing a robust phylogenetic topology [36]. The LCBs among the three *Primula* mitogenomes were scattered in the mitogenomes, and no specific conserved region was longer than 50 kb (Fig. 1; Fig. S5). Despite the relatively short shared regions, the total length of the shared regions was extremely long. The shared sequences between *P. smithiana* and *P. valentiniana* were over 364 kb in length, which was markedly longer than in the other species pairs [*P. smithiana* and *P. palinuri* (288 kb), and *P. valentiniana* and *P. palinuri* (243 kb), Fig. 1].

Among the 29 conserved gene clusters identified in angiosperms [25], the three *Primula* mitogenomes possessed

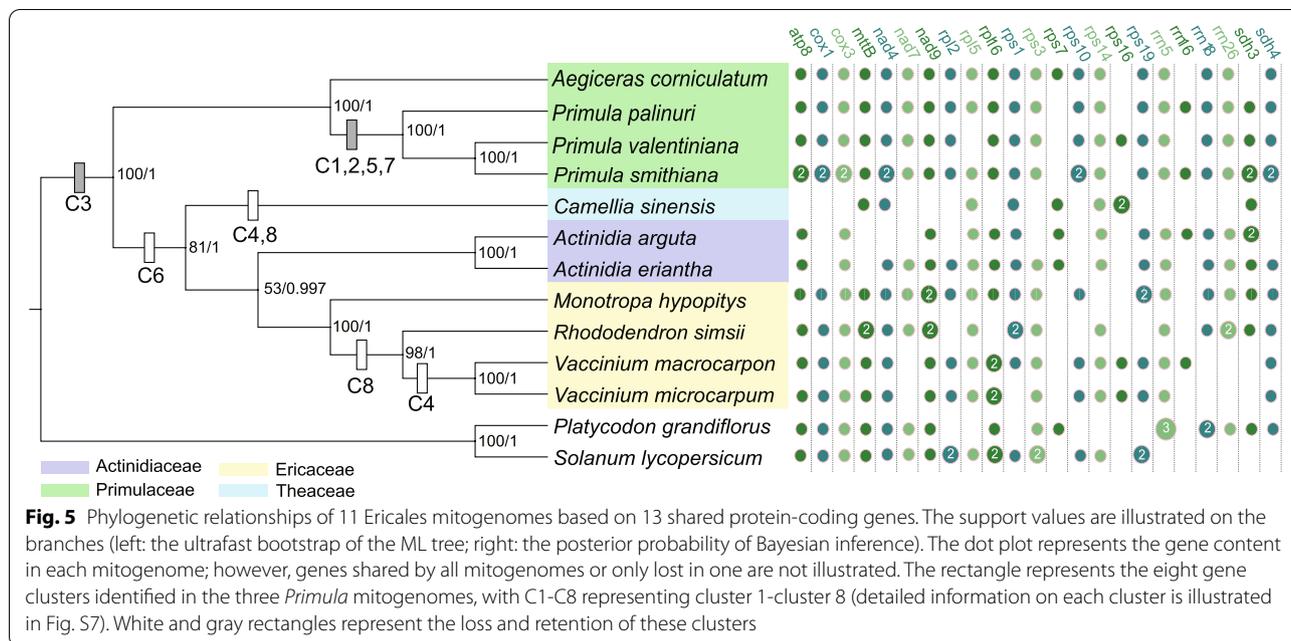




13 of these shared gene clusters, whereas *P. palinuri* exhibited one additional gene cluster (>*trnI-CAT*><*trnD-GTC*<) (Fig. S6). Eight additional gene clusters were also identified, four of which were specific to *Primula* compared to the other Ericales mitogenomes (Fig. 5; Fig. S7). Among the four specific gene clusters, each of the three gene clusters [*nad9-trnW* (*CGA*)-*trnA* (*UGC*) and *rps12-nad3-trnM* (*CAU*)] contained two (*trnW-CGA* and *trnA-UGC*) and one (*trnM-CAU*) plastid-derived gene, respectively.

Nucleotide diversity based on locally collinear blocks of mitogenomes

The nucleotide diversity (π) of LCBs was calculated for different plant lineages (Table S8). First, *Primula* mitogenomes exhibited relatively high nucleotide diversity (π : 0.009) among all examined plant lineages, which was 10-fold higher than the lowest diversity, found in *Zea* (π : 0.0009) (Table S8). The π value (0.005) among the CDS regions of *Primula* was also higher than that in most of



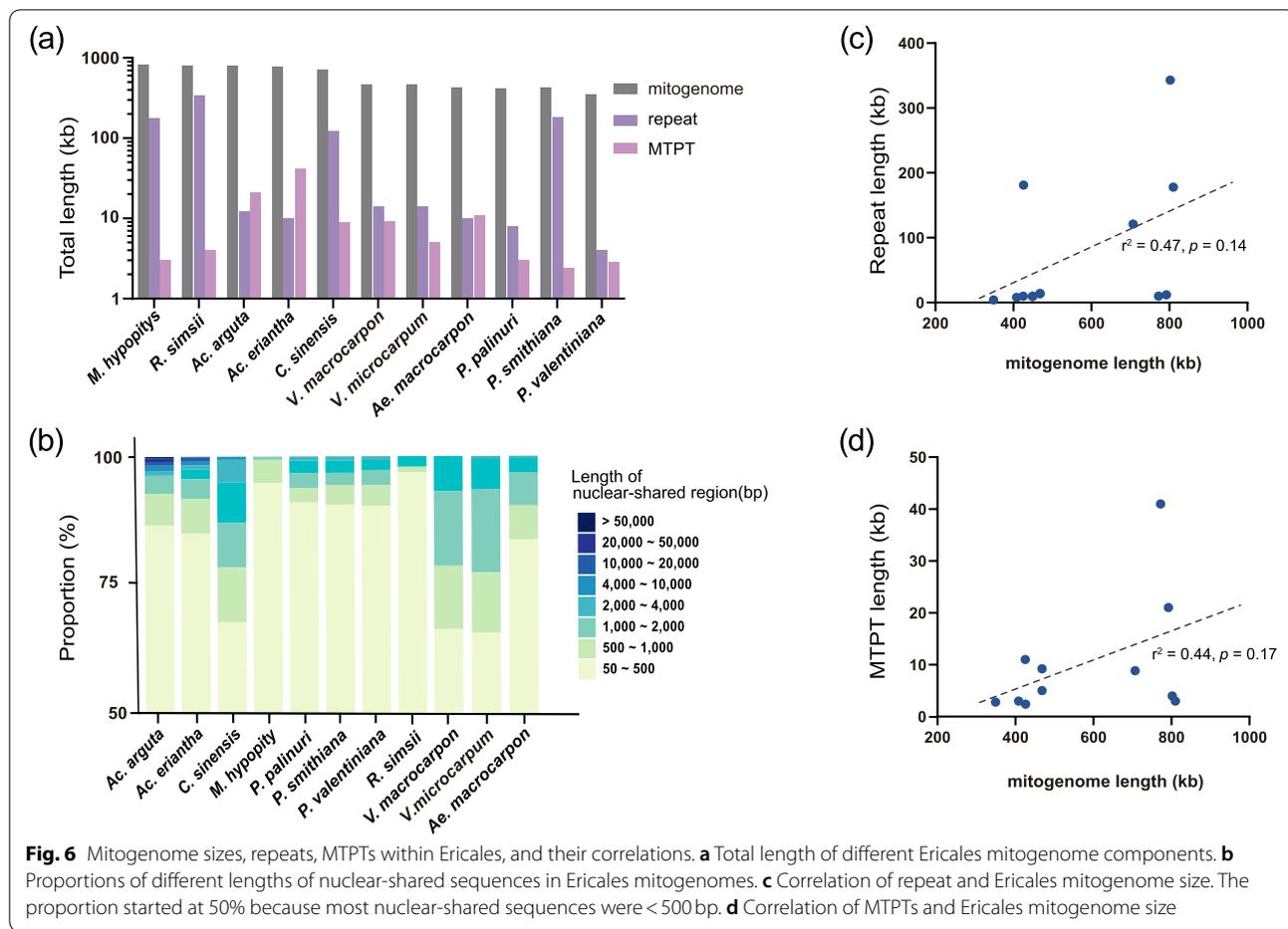
the analyzed taxa (Table S8). Second, the high nucleotide diversity at the genus level mainly contributed to the striking variation between *P. palinuri* (from the Mediterranean region) and the two alpine *Primula* taxa (from the Himalaya–Hengduan Mountains). The π value between the two alpine *Primula* taxa was 0.004, while the π value between the taxa from the two regions (i.e., the Himalaya–Hengduan Mountains and the Mediterranean regions) was 0.013 (Table S9).

Size variations and phylogenetic analyses among Ericales mitogenomes

The present study included eight published Ericales mitogenomes (from Primulaceae, Ericaceae, Actinidiaceae, and Theaceae) from NCBI, with the aim of revealing the potential causes of the observed mitogenome size variation at the order level (Table S10). The *Monotropa hypopitys* Crantz, *Rhododendron simsii* Planch., *Actinidia arguta* Miq, *Actinidia eriantha* Benth., and *Camellia sinensis* (L.) Kuntze mitogenome sizes were 810 kb, 802 kb, 792 kb, 772 kb, and 707 kb, respectively, which were much larger than those of *Vaccinium macrocarpon* Aiton (459 kb), *V. microcarpum* Miyabe & T. Miyake (468 kb), *Ae. corniculatum* (425 kb), and the three *Primula* taxa (349–426 kb) (Fig. 6a). The total repeat lengths of *R. simsii* (343 kb), *P. smithiana* (181 kb), *M. hypopitys* (178), and *C. sinensis* (121 kb) were significantly larger than those of the remaining mitogenomes (each < 20 kb) (Fig. 6a). Most Ericales mitogenomes lacked large MTPTs, whereas total lengths of 21 kb and 41 kb were identified in *Ac. arguta* and *Ac. eriantha*, respectively (Fig. 6a).

Most Ericales NUMTs were < 500 bp, and the longest NUMT was identified in *Ac. arguta* (53 kb) (Fig. 6b). Additionally, there was a positive correlation between the NUMTs and Ericales mitogenome length, although the *p* value was slightly higher than 0.05 (Fig. S8). The relatively small mitogenome of *Ae. corniculatum* contained a large number of NUMTs (Table S10), which indicated that this species served as an outlier in this analysis. There was a very strong positive relationship between NUMTs and mitogenome sizes when *Ae. corniculatum* was removed ($r^2 = 0.55$, $P = 0.01$, Fig. S9a). Pearson’s correlation test also indicated that repeats (Fig. 6c) were not significantly correlated with Ericales mitogenome size variation. However, a significant correlation was observed between repeats and mitogenome size when the kiwifruit mitogenomes were removed (Fig. S9b), since the large kiwifruit mitogenomes contained a small number of repeats [3]. There was also no significant relationship between MTPTs and mitogenome sizes in Ericales (Fig. 6d). *R. simsii* and *M. hypopitys* both showed relatively large mitogenomes but fewer MTPTs (Table S10). Therefore, the two species act as outliers in this analysis. There was a very strong positive relationship between MTPTs and mitogenome sizes when *R. simsii* and *M. hypopitys* were removed ($r^2 = 0.6$, $P = 0.01$, Fig. S9c).

Protein-coding genes also varied within Ericales, and only 13 of these genes were shared among all eight Ericales taxa (Fig. 5). 38 protein-coding genes were shared between *Primula* taxa and *Ae. corniculatum*, whereas only 28 and 27 genes were observed in *C. sinensis* and *Ac. arguta*, respectively.



Both maximum likelihood (ML) and Bayesian inference trees were reconstructed based on the 13 shared protein-coding genes (with a total aligned length of 11,360bp). Both trees suggested that *P. smithiana* (Sect. *Proliferae*) and *P. valentiniana* (Sect. *Amethyatina*) exhibited the closest relationship (Fig. 5). The interfamilial relationships (i.e., Primulaceae, Ericaceae, Actinidiaceae, and Theaceae) within Ericales species with available mitogenome data to date were resolved robustly. Primulaceae is sister to other families. Among the three other families, Theaceae was sister to a clade that contained Ericaceae and Actinidiaceae.

Discussion

Complete mitogenomes of three *Primula* taxa

The plant mitogenome is known for its considerable variation in both structure and content [1, 5, 12]. Most mitogenomes are reconstructed as a master circle containing the complete mitochondrial gene set [1, 7, 19–21]. A previously assembled pipeline [37–39] was followed to extend and merge the *Primula* mitochondrial contigs into a master circle. These master circles

indicated highly homologous recombination (HR) frequencies, as no long collinear sequences were shared within *Primula* (Fig. S5) [3]. Most potential HR events occurred in the noncoding region and seemed to occur randomly among *Primula* mitogenomes, as no synteny could be detected based on the coordinates of each shared region (Fig. S5). Large repeats (> 500bp) are often involved in reversible reciprocal HR that modulates plant mitogenome plasticity [25, 40, 41]. However, only the *P. smithiana* mitogenome exhibited a high proportion of repeats. As mitogenome reconstruction for *P. palinuri* and *P. valentiniana* involved the merging of several contigs, long repeats (over 30kb in *P. smithiana*, Fig. 3a) might have been consolidated and eventually “vanished” from the final assemblies. This scenario could be examined using long-read sequencing technologies in the future.

Studies have discussed the subgenomic molecular forms of plant mitogenomes from linear branches to several circular sequences [42, 43]. The master circle reconstructed in the present study might not fully represent the only state of the *Primula* mitogenomes. However,

the master circle contained all the mitochondrial genetic information [20, 23, 44]. The present study implemented two methods for confirming the reliability of our master circles. First, the clean reads were remapped against the master circle to identify potential assembly mistakes. Second, several PCR primer pairs were designed to verify the accuracy of the merging of contigs. The consecutive mapping and PCR results indicated the reliability of the master circles of the *Primula* mitogenomes (Fig. S1).

Contrasting patterns of nucleotide diversity across plant lineages

Compared with plant nuclear and plastid genomes, high rates of rearrangements in plant mitogenome frequently occur among closely related species [12]. Therefore, the conserved coding region and collinear segments might be relatively good resources for comparative analyses among different plant lineages [12, 20]. The study of mitochondrial collinear regions could further help us understand the evolution of plant mitogenomes [3, 36]. However, the lack of shared collinear fragments among plant lineages is due to frequent repeat-mediated rearrangement [12] and a lack of large collinear fragments, as shown in *Picea* [45].

In this study, we calculated and compared the nucleotide diversity of three collinear fragments of the mitogenomes within each genus (including at least three mitogenomes each). *Primula* mitogenomes exhibited the highest nucleotide diversity (Table S8), which was 2–3 times lower than the nucleotide diversity of the plastid genomes (complete sequence: 0.026; CDS: 0.019). This scenario is consistent with the synonymous-site divergence levels observed within the plastome and mitogenome [46].

There are several possible explanations for the contrasting patterns of nucleotide diversity within *Primula* and among plant lineages. First, *Primula* probably presents an elevated mutation rate in the mitogenome, as shown in several plant lineages [47]. Several functional genes might be positively selected by severe environmental conditions (such as strong ultraviolet radiation and extremely low temperatures in alpine environments) [48, 49], with far-reaching impacts on noncoding regions because of the hitchhiking effect. However, there is no clear evidence of elevated rates of mutation in these core genes. Second, a plausible alternative is that the LCB evolved under an assumed constant mutation rate, although substantial rate variation occurs among plant lineages. The specific evolutionary history of each plant genus should result in striking mitogenomic diversity. For example, the most recent common ancestor (MRCA) of the *Primula* taxa in the two regions (Europe vs. Asia) originated 20–40 Mya [30, 50], which is the oldest age among the taxa examined in this study, contributing to the accumulation of the greatest amount of nucleotide polymorphism. However,

the MRCA of the two alpine *Primula* taxa in the Himalaya–Hengduan Mountains probably originated in the Late Miocene (approx. 9 Mya, [50]), resulting in approximately three times less nucleotide polymorphism than was observed between the two regions (Europe vs. Asia). At the other extreme, the genus *Zea*, which has diversified since approx. 0.18 Mya [51], presented the lowest inter-specific polymorphism (Table S8). Third, the role of sampling error should be taken into account in considering these contrasting patterns of nucleotide polymorphism among plant lineages. Given that the LCB examined here accounted for a varying fraction of the whole mitogenome (4–64%), more LCB data and more comprehensive taxon sampling are required to test this in the future.

RNA editing is essential for mitochondrial gene expression

RNA editing is a common phenomenon within land plant mitogenomes involving the conversion of cytidines to uridines (C-U editing) or uridines to cytidines (U-C editing) in some plants [11, 52]. The most predictable RNA editing sites were located on the second codon in *Primula*, similar to most angiosperm mitogenomes [53].

Most of the predicted RNA editing sites could result in the conversion of the encoded amino acid from neutral to hydrophobic, ultimately increasing the hydrophobic character of the coding protein [53]. Hydrophobicity is conducive to protein folding and secondary structure formation [54]. The most frequent RNA editing events were found in *ccmB* and *nad4L* in *P. smithiana*. The secondary and tertiary structures of these two genes in *P. smithiana* showed increased stability of the protein.

In this study, only two RNA editing sites (in the *atp6* and *ccmFc* genes) that could result in an early stop codon were identified in *Primula*, which is lower than the number found in kiwifruit mitogenomes [3]. Although both RNA editing sites could also be detected in other Ericales taxa (Figs. S2–3), only the termination in the *atp6* gene was effective. The regulation of *atp6* is essential for plant fertilization [55] and is related to cytoplasmic male sterility (CMS) in pepper [56], sunflower [57], rice [58], and *Sorghum* [59]. RNA editing in the *atp6* gene also alters seed formation in maize [60]. Therefore, this RNA editing site in the *atp6* gene may potentially impact the reproduction of Ericales species, but further studies are needed to verify this. Furthermore, our results highlight the importance of considering RNA editing sites during annotation.

Plastid-derived sequences in *Primula* mitogenomes

Transfer events from plastid to mitochondrial genomes frequently occur in angiosperm plants [61, 62]. Studies have revealed plastid-derived backgrounds of 0.1–11.5% of plant mitogenomes [63]. The present study identified 0.5–0.7% of *Primula* mitogenomes as MTPTs (Table 1).

MTPT transfer occurs independently among species [15], and plastid-derived fragments ultimately become nonfunctional pseudogenes [63]. Sequence transfer events from plastids to mitochondria inferred based on the oldest MTPT (*trnV (uac)-trnM (cau)-atpE-atpB-rbcL*) occurred at least 300 million years ago (Mya), before the divergence of extant gymnosperms and angiosperms [64]. In the present study, the oldest MTPT transfer events (MTPT7 and MTPT10) might date back to the divergence time of *Primula* and its relatives (approximately 23–40 Mya) (Fig. 4) [30]. However, a specific transfer event (such as MTPT7) probably occurred multiple times, as reported in several plant lineages [65].

Two hypotheses may explain the multiple transfer events observed. First, these transfer events could have been acquired directly via intracellular gene transfer (IGT). An ancient plastome-to-mitogenome transfer event occurred in ancestral species of *Primula* and its relatives but was partially lost in several species during the subsequent speciation process, after which IGT reoccurred recently. Second, the transfer event was acquired by plant-to-plant horizontal gene transfer (HGT). This means that the “ancient” MTPT was initially acquired via mitochondrion-to-mitochondrion HGT among plant taxa, rather than via an ancient plastid-to-mitochondrial IGT event [66]. Regardless, the underlying mechanism of MTPT deserves in-depth study in the future.

Phylogenetic implications from mitogenomes

Unlike plastid and nuclear genes (such as low copy nuclear genes or ITS), the mitogenome is not frequently used to reconstruct phylogenies or phylogeographies in higher plants due to its slow mutation rate [46], frequent genomic rearrangement [12], and incorporation of foreign DNA from the nuclear and plastid genomes [23]. However, several mitogenomic genes (such as *atp1*, *matR*, *nad5*, and *rps3*) have been widely used for plant phylogenetic studies at different levels since they may offer different insights than plastid and nuclear genes [67], see references therein). Theoretically, the coding genes of mitogenomes are more suitable for elucidating ancient diversification patterns in plants because of their generally slow mutation rates compared to plastid and nuclear genes [46].

In this study, the interfamilial relationships (i.e., Primulaceae, Ericaceae, Actinidiaceae, and Theaceae) within Ericales were resolved robustly based on the available mitogenome data (Fig. 5). Notably, these relationships are broadly consistent with previous studies with multiple fragment combinations (containing plastid, mitogenome, and nuclear genes [50]);, plastomes [68], and transcriptomes and genomes [69, 70].

The three studied *Primula* taxa belong to two subgenera (*Aleuritia* and *Auriculastrum*) [28, 29]. Subgenus

Aleuritia contains *P. smithiana* (Sect. *Proliferae*) and *P. valentiniana* (Sect. *Amethyatina*), whereas subgenus *Auriculastrum* includes *P. palinuri* (Sect. *Auricula*). Here, *P. palinuri* is sister to a clade that contains *P. smithiana* and *P. valentiniana*, consistent with the relationships based on plastid genes [30, 71].

The identified collinear gene clusters and LCB verified the above relationships in *Primula*. First, among 13 conserved mitochondrial gene clusters in *Primula* taxa (Fig. S6), one unique gene cluster (*trnI-CAT-trnD-GTC*) occurred only in *P. palinuri*. Second, the collinear fragments of the mitogenomes between *P. smithiana* and *P. valentiniana* exhibited lower nucleotide diversity than the other pairs containing *P. palinuri* (Table S9). The above two lines of evidence indicated the close relationship of *P. smithiana* and *P. valentiniana*. These results show that some conserved gene clusters and LCBs of plant mitogenomes present phylogenetic signals [36, 72].

Mitogenome size expansion in Ericales

Several hypotheses have been proposed to explain the considerable variation in mitogenome size observed in land plants [3, 13, 22, 23]. However, further specific studies are required [73]. The variation in noncoding sequences probably results in size variation [74], as most angiosperm mitogenomes generally contain a core set of 24 protein-coding genes [47], and 100-fold variation is observed among total plant mitogenome sizes [5].

The uptake of foreign sequences is a vital source of noncoding sequences [17, 23, 61, 75] and might contribute to angiosperm mitogenome size expansion. A large plant mitogenome often contains longer nuclear-shared sequences [76]. NUMT seems to contribute to mitogenome size in Ericales, despite the lack of a strong overall NUMT-mitogenome size relationship (Fig. S8). MTPT is also sometimes linked to the size variation of angiosperm mitogenomes [25, 77]. However, no significant relationship between MTPTs and mitogenome sizes was found in Ericales (Fig. S6d). The relatively large mitogenomes of *R. simsii* and *M. hypopitys* do not possess large MTPTs, probably because of the lack of an inverted repeat region (IR) in *R. simsii* [78] or because all genes encode products with photosynthetic functions and RNA polymerase subunits in the mycoheterotrophic plant *M. hypopitys* [79]. Collectively, the results indicated that the uptake of foreign sequences is a vital source of mitogenome size expansion within Ericales, albeit with some exceptions.

Repeats usually cause plant mitogenome size variation [20]. Approximately 42.7% of the *R. simsii* mitogenome consisted of repeat sequences, probably resulting in the largest mitogenome in Ericales. However, an exception was observed in the kiwifruit mitogenomes (Fig. 6c), which possessed large mitogenomes but lacked repeats

[3], suggesting diverse reasons for mitogenome size variation in the Ericales order. Further studies are needed to examine plant mitogenome size expansion and variation.

Conclusions

This study successfully assembled the complete mitogenomes of three *Primula* species in the form of master circles. These mitogenomes shared similar gene contents but varied in structure. Relatively high nucleotide diversity was found in the *Primula* mitogenomes among all examined plant lineages. The RNA editing of these *Primula* mitogenomes could increase protein stability and alleviate the influence of amino acid mutation. MTPT events that co-occurred in the mitogenome and plastome were identified using the phylogenetic method, indicating that multiple transfer events probably occurred in the evolutionary history of *Primula*. In addition, mitogenomic genes, collinear gene clusters, and LCB all showed phylogenetic signals. Although the size of IGT events and repeats might drive mitogenome size variation in Ericales, the diverse reasons for mitogenome size variation in the order must be studied further.

Methods

Plant materials, DNA extraction, and sequencing

Fresh young leaves of *P. smithiana* (voucher no. Xu et al. 150,170) were collected from the South China Botanical Garden greenhouse, Chinese Academy of Sciences (Guangzhou, China), and cultivated from Yadong County of Tibet, China. Mitochondria were extracted from the leaves using density gradient centrifugation as described in a previous study [80]. mtDNA was extracted using the modified cetyltrimethylammonium bromide (CTAB) method [81]. Silica gel-dried leaves of *P. valentiniana* (voucher no. Hao et al. 120,274) were used for DNA extraction via the CTAB method [81]. The Kew DNA Bank provided the total DNA of *P. palinuri* (voucher no. Chase M.W. 16,567, Kew) (<https://dnabank.science.kew.org>). These voucher specimens were formally identified or checked by the third author (Prof. Gang Hao, South China Agricultural University) and deposited in the herbarium of the South China Botanical Garden, Chinese Academy of Sciences (IBSC). All experimental research complied with relevant institutional, national, and international guidelines and legislation. No specific permissions or licenses were required for our collection activities and experiments. A 250-bp paired-end library was prepared and sequenced on the Illumina HiSeq 2500 platform (Illumina, San Diego, CA, USA).

Genome assembly and validation

The short reads were checked using FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) and trimmed

using Trimmomatic [82] for the *P. smithiana* assembly. The clean reads were assembled de novo using Spades v3.13.0 [83], and putative contigs were then selected and extended using PRICE [84] with the following parameters: `600 95 -nc 50 -dbmax 72 -mol 30 -mpi 90 -target 90 2 1 1`.

Genome skimming data for *P. palinuri* and *P. valentiniana* were used after quality control for de novo assembly by MEGAHIT v1.0 [85] and Spades v3.13.0 [83] to maximize the utilization of paired-end reads. Furthermore, the potential mitochondrial contigs were selected based on coverage and rechecked manually by BLASTN [86] against the NCBI nonredundant nucleotide database. The whole trimmed reads were then mapped to the putative mitogenome contigs with Geneious (<http://www.geneious.com/>). Contigs were extended based on the mapping coverage as described by Mower et al. [87] and Zhang et al. [39]. Mapped sequences were excluded if their sequencing coverage depth was $<5\times$ and $>500\times$ or the overlap was <50 bp. In the continuous extension cycle, contigs showing overlap of greater than 80 bp and identity over 99% were merged.

The complete reads were aligned against these draft mitogenomes using BWA v0.7 [67] with the default parameters to evaluate the coverage consistency and read connectivity. The sites that differed from 75% of the other aligned reads were corrected.

Although several isoforms of angiosperm mitochondrial sequences have been reported, the assembled form of a master circle containing all mitochondrial genes is commonly found in various plant groups [1, 3, 5, 42]. To further check the reliability of our assembly, five PCR primer pairs were designed between two initial contigs for each species to confirm whether these assembled regions could be successfully amplified (Fig. S1; Table S11).

Annotation of *Primula* mitogenomes

Protein and rRNA genes were annotated using the GeSeq online server [88] and local BLASTN [86], with eight Ericales mitogenomes [*Ae. Corniculatum* (MT130509.1), *R. simsii* (MW030508.1), *C. sinensis* (NC_043914.1), *V. macrocarpon* (NC_023338.1), *V. microcarpum* (MK715445.1), *M. hypopitys*, (MK990822, MK990823), *Ac. arguta* (MH559343), and *A. eriantha* (MH645952.1)] as references. tRNAs were identified by using tRNAscan-SE v2.0.7 [89]. The start and stop codons in exons were adjusted manually. Group I and II introns were detected with the RNAweasel tool [90]. ORF-Finder was used for predicting open reading frames longer than 300 bp with the standard genetic code (<https://www.ncbi.nlm.nih.gov/orffinder/>). The GC count per 500 bp was calculated with an in-house Python3 script. The depth of sequencing coverage per locus was calculated using the `genomcov` command in bbMap (<https://www.sourceforge.net/projects/bbmap/>) with the parameter `-d`.

Repeat regions were identified by alignment against each assembly using BLASTN [86] with a minimum identity of 85% and a minimum length of 100bp. Tandem repeat sequences were detected by using Tandem Repeats Finder [91] with the default parameters. The detailed information of the three *Primula* mitogenomes was visualized using Circos v0.69 [92].

Identification of RNA editing sites and gene clusters

RNA editing sites in the three mitogenomes were predicted using the PRET-Mt server [35] with a cutoff value of 0.2. The hydrophobicity of *ccmB* and *nad4L* in *P. smithiana* were calculated using ProtScale [93]. In contrast, the average hydrophobic character of *P. smithiana* protein-coding genes was calculated based on the method described by Kyte and Doolittle [94].

The *atp6* and *ccmFc* gene sequences were extracted from the eight publicly available Ericales and *Primula* mitogenomes and were aligned using MAFFT v7.4 [95]. The secondary structures of *atp6* and *ccmFc* were inferred using PSIPRED [96]. To validate the predicted editing sites, the RNA-seq data of *Ac. arguta* (SRR3823655), *Ae. Corniculatum* (SRR1688722), *C. sinensis* (SRR20083852), *M. hypopitys* (SRR10159707), *R. simsii* (SRR10415549), and *V. macrocarpon* (SRR18449568) were downloaded from the NCBI SRA database and mapped to the *atp6* and *ccmFc* gene sequences using BWA v0.7 [67]. Editing efficiency was estimated by calculating the proportion of cDNA reads that contained the edited nucleotide.

The 29 gene clusters in the plant mitogenomes described by Richardson et al. [97] were identified in the Ericales mitogenomes by manual inspection. Additional gene clusters were searched among *Primula* taxa through visual checking, as described by Kan et al. [25].

Identification of conserved sequences and foreign-shared sequences

Each pair of *Primula* mitogenomes was aligned against the others to identify LCBs using the nucmer command in MUMMER4 [98] under the default parameters. The aligned areas with lengths < 2000bp were removed. To calculate the interspecific nucleotide diversity of mitogenomes in plant genera, more than three mitogenomes (from different species) within each genus were considered and retrieved from GenBank (Table S8). The top three LCB and their CDSs within each genus were aligned using MAFFT v7.4 [95] and then concentrated using SequenceMatrix [99]. Nucleotide diversity was calculated using DnaSP v.6 [100].

Three *Primula* plastid genomes were assembled using GetOrganelle [101] and annotated using the GeSeq server [88]. Transfer events were identified by aligning each *Primula* mitogenome against the *P. palinuri* plastid genome (after removing one inverted repeat) using

local BLASTN with a minimum length of 100bp and an e-value of 1e-20. Putative MTPTs, including *atp1/atpA*, *rrn26/rrn23*, and *rrn18/rrn16*, were excluded because they simultaneously occurred in the mitogenome and plastid genome [20, 25, 66].

Two MTPTs longer than 500bp (containing enough polymorphic variation for phylogenetic analysis) were extracted and validated using read mapping and PCR amplification. Each clean read was mapped against the 100bp up- and downstream regions of MTPT in the mitogenome and its plastid genome counterpart for read mapping. MTPT7 and MTPT10 were aligned with the other 30 publicly available plastid genomes (Table S12) after validation within Primulaceae using MAFFT v7.4 [95]. Each aligned region was trimmed using Gblocks [102]. IQ-TREE v1.6.12 [103] was used to construct an ML tree with 5000 ultrafast bootstrapping replicates under the best model detected using ModelFinder [104]. Then, the GC content of the two sequences was calculated using an in-house Python3 script.

To account for gene transfer events between the mitogenome and nuclear genome, the nuclear-shared sequences were identified by aligning each *Primula* mitogenome against the *P. veris* nuclear genome (GCA_000788445.1) using BLASTN with a minimum length of 50bp and an e-value of 1e-20. The nuclear-shared regions in other Ericales mitogenomes were also identified against their corresponding nuclear genomes. This reference nuclear genome information can be found in Table S13. The size information of mitogenomes, repeats, MTPTs, and NUMTs can be found in Table S10.

Repeat regions and MTPTs were also identified among Ericales mitogenomes. Pearson's correlation was calculated between repeats, NUMTs, MTPTs, and mitogenome size in Ericales to reveal the underlying causes of mitogenome size variation [3].

Phylogenetic analysis and identification of the contents of Ericales mitogenomes

The above eight publicly available Ericales mitogenomes and three *Primula* mitogenomes (representing Primulaceae, Ericaceae, Actinidiaceae, and Theaceae) were included in the phylogenetic analysis, with *Platycodon grandiflorus* A. DC. (NC_035958.1; Campanulaceae) and *Solanum lycopersicum* L. (NC_035963.1; Solanaceae) as outgroups. A total of 13 coding sequences (CDSs) among the mitogenomes were aligned with MAFFT v7.4 [95]. Conserved aligned regions were extracted with Gblocks [102] and concatenated with SequenceMatrix [99]. ML phylogenetic trees were built using IQ-TREE v1.6.12 [103] with 5000 ultrafast bootstrapping replicates under the GTR+G model. Bayesian inference was conducted using MrBayes v3.2 [105] under the GTR+G+F model.

The Markov chain Monte Carlo (MCMC) algorithm was run for 2.0×10^7 generations with four incrementally heated chains, starting from random trees and sampling one out of every 1000 generations. The stability of the Markov chain was ascertained by plotting likelihood values against the number of generations (effective sample size > 200) using Tracer v1.7 [106] and by splitting variances < 0.01. The burn-in fraction was set to 0.25, and the remaining trees were used to construct the 50%-majority rule consensus tree. Bayesian posterior probabilities were used to estimate support for each branch in the consensus tree.

Abbreviations

bp: base pair; CDS: Coding sequence; DNA: Deoxyribo Nucleic acid; GC: Guanine- cytosine; HGT: Horizontal gene transfer; HR: Homologous rearrangement; kb: kilobase; LCB: Locally collinear blocks; MTPT: Mitochondrial plastid DNA; ML: Maximum likelihood; mtDNA: mitochondrial DNA; NUMT: Nuclear mitochondrial DNA sequences; IGT: Intracellular gene transfer; ORF: Open reading frame; PCR: Polymerase chain reaction; RNA: Ribonucleic acid; atp1, atp4, atp6, atp8 and atp9: Genes for ATP synthase subunits 1, 4, 6, 8 and 9; ccmC, ccmC, ccmFc and ccmFn: Genes for cytochrome c biogenesis proteins B, C, Fc and Fn; cob: Gene for cytochrome b; cox1, cox2 and cox3: Genes for cytochrome c oxidase subunits 1, 2 and 3; matR: Gene for maturase; mttB: transport membrane protein B; nad1, nad2, nad3, nad4, nad4L, nad5, nad6, nad7 and nad9: Mitochondrial genes for NADH dehydrogenase subunits 1–7, 9 and 4L; rpl2, rpl5 and rpl16: Genes for ribosomal proteins L2, L5 and L16; rps1, rps2, rps3, rps4, rps7, rps10, rps11, rps12, rps13, rps14 and rps19: Genes for ribosomal proteins S1, S2, S3, S4, S7, S10, S11, S12, S13, S14, and S19; sdh3 and sdh4: Genes for succinate dehydrogenase cytochrome subunits 3 and 4; trnL: tRNA gene for leucine; trnT: tRNA gene for threonine; trnS: tRNA gene for serine; Arg: Arginine; Cys: L-cysteine; His: Histidine; Pro: Proline; Ser: Serine; Tyr: Tyrosine.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-022-08983-x>.

Additional file 1: Figure S1. Electrophoretic gel visualization of the amplified fragments of the three draft mitogenome assemblies and MTPTs. M1 is the DL2000 DNA marker, whereas M2 is the DL1000 DNA marker. The first two wells in each gel represent the corresponding primer pairs for MTPT7 and MTPT10 in each *Primula* mitogenome, whereas the other primer pairs were used for assembly validation in each draft mitogenome (primer details can be found in Table S11).

Additional file 2: Figure S2. The inferred secondary structure of the *atp6* gene among the Ericales mitogenomes. *Primula* contains three *Primula* mitogenomes; *Actinidia* contains two kiwifruit mitogenomes.

Additional file 3: Figure S3. The inferred secondary structure of the *ccmFc* gene among the Ericales mitogenomes. *Primula* contains three *Primula* mitogenomes; *Actinidia* contains two kiwifruit mitogenomes.

Additional file 4: Figure S4. Average coverage of MTPTs and their plastid counterparts.

Additional file 5: Figure S5. The three longest LCBs and their distribution in *Primula* taxa.

Additional file 6: Figure S6. Gene clusters of each Ericales mitogenome. The red cell indicates the existence of one specific gene cluster.

Additional file 7: Figure S7. The newly identified gene clusters of each Ericales mitogenome. The red cell indicates the existence of one specific gene cluster.

Additional file 8: Figure S8. Correlation of NUMT length and Ericales mitogenome size.

Additional file 9: Figure S9. Correlation of NUMTs, MTPTs, and repeat length and Ericales mitogenome size after removing outlier mitogenomes. (a) NUMT length and Ericales mitogenome size after removing the *Aegiceras corniculatum* mitogenome; (b) repeat length and Ericales mitogenome size after removing the kiwifruit mitogenomes; (c) MTPT length and Ericales mitogenome size after removing the *Rhododendron simsii* and *Monotropa hypopitys* mitogenomes.

Additional file 10: Table S1. Assembly information of three *Primula* taxa. **Table S2.** The number of protein-coding genes in the three *Primula* mitogenomes. **Table S3.** Tandem repeats in the *Primula smithiana* mitogenome. **Table S4.** Tandem repeats in the *Primula palinuri* mitogenome. **Table S5.** Tandem repeats in the *Primula valentiniana* mitogenome. **Table S6.** Information on MTPTs within the three *Primula* mitogenomes. **Table S7.** The shared nuclear sequences within the three *Primula* mitogenomes. **Table S8.** The nucleotide diversity of mitogenomes within different plant lineages. **Table S9.** The nucleotide diversity of mitogenomes among *Primula* taxa. **Table S10.** The general features of the 11 mitogenomes used for this study. **Table S11.** The primers used for *Primula palinuri*, *Primula smithiana*, and *Primula valentiniana*. **Table S12.** The 30 publicly available plastid genomes used for MTPT analysis. **Table S13.** The publicly available nuclear genomes used for NUMT analysis.

Acknowledgements

We would like to thank Drs. Bao-Sheng Wang, Zheng Li, and Lin-Feng Li for their valuable comments on the early version of the manuscript. We also thank Yu-Yin Zhou and Qiong Dong for helping with the PCR validation. Special thanks to Kew DNA Bank of the Royal Botanic Gardens (Richmond, London, UK) for kindly providing *Primula palinuri* DNA. We thank American Journal Experts (AJE) for its linguistic assistance during the preparation of this manuscript.

Authors' contributions

TJL and HFY conceived this study. LW and TJL performed the analysis. LW conducted the experimental study. LW and TJL wrote the manuscript. HFY, XJG, and GH contributed feedback and edits to the manuscript. All authors read and approved the final manuscript.

Funding

This work was financially supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant No. XDB31000000) through Xue-Jun Ge and the National Natural Foundation of China (Grant No. 31870192) through Hai-Fei Yan.

Availability of data and materials

The data of this study have been deposited in the NCBI with BioProject accession number PRJNA794031. The genome skimming sequencing reads can be found under the number SRR17422315-SRR17422317. Mitogenome assembly of *Primula* has been deposited to GenBank with accession numbers OM971881-OM971883. Plastid genomes of three *Primula* taxa have been deposited to GenBank with accession numbers OM313289-OM313291.

Declarations

Ethics approval and consent to participate

This study's material collections and experimental research complied with relevant institutional, national, and international guidelines and legislation. No specific permissions or licenses were required.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Key Laboratory of Plant Resources Conservation and Sustainable Utilization, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou,

China. ²University of Chinese Academy of Sciences, Beijing, China. ³College of Life Sciences, South China Agricultural University, Guangzhou, China.

Received: 7 March 2022 Accepted: 1 November 2022

Published online: 24 November 2022

References

- Gualberto JM, Newton KJ. Plant mitochondrial genomes: dynamics and mechanisms of mutation. *Annu Rev Plant Biol.* 2017;68(1):225–52.
- Schleicher S, Binder S. In *Arabidopsis thaliana* mitochondria 5' end polymorphisms of *nad4L-atp4* and *nad3-rps12* transcripts are linked to RNA PROCESSING FACTORS 1 and 8. *Plant Mol Biol.* 2021;106(4–5):335–48.
- Wang S, Li D, Yao X, Song Q, Wang Z, Zhang Q, et al. Evolution and diversification of kiwifruit mitogenomes through extensive whole-genome rearrangement and mosaic loss of intergenic sequences in a highly variable region. *Genome Biol Evol.* 2019;11(4):1192–206.
- Diaz-Garcia L, Rodriguez-Bonilla L, Rohde J, Smith T, Zalapa J. Pacbio sequencing reveals identical organelle genomes between American cranberry (*Vaccinium macrocarpon* Ait.) and a wild relative. *Genes.* 2019;10(4):291.
- Zardoya R. Recent advances in understanding mitochondrial genome diversity. *F1000Res.* 2020;9:270.
- Skippington E, Barkman TJ, Rice DW, Palmer JD. Miniaturized mitogenome of the parasitic plant *Viscum scurruloideum* is extremely divergent and dynamic and has lost all nad genes. *Proc Natl Acad Sci U S A.* 2015;112(27):E3515–E24.
- Putintseva YA, Bondar EI, Simonov EP, Sharov VV, Oreshkova NV, Kuzmin DA, et al. Siberian larch (*Larix sibirica* Ledeb.) mitochondrial genome assembled using both short and long nucleotide sequence reads is currently the largest known mitogenome. *BMC Genomics.* 2020;21:654.
- Covello PS, Gray MW. RNA editing in plant mitochondria. *Nature.* 1989;341(6243):662–6.
- Castandet B, Araya A. RNA editing in plant organelles. Why make it easy? *Biochem Mosc.* 2011;76(8):924–31.
- Small ID, Schallenberg-Rüdinger M, Takenaka M, Mireau H, Ostersetzer-Biran O. Plant organellar RNA editing: what 30 years of research has revealed. *Plant J.* 2020;101(5):1040–56.
- Binder S, Marchfelder A, Brennicke A, Wissinger B. RNA editing in trans-splicing intron sequences of *nad2* mRNAs in *Oenothera* mitochondria. *J Biol Chem.* 1992;267(11):7615–23.
- Mower JP. Variation in protein gene and intron content among land plant mitogenomes. *Mitochondrion.* 2020;53:203–13.
- Smith DR. The mutational hazard hypothesis of organelle genome evolution: 10 years on. *Mol Ecol.* 2016;25(16):3769–75.
- Brown GG, Colas des Francs-Small C, Ostersetzer-Biran O. Group II intron splicing factors in plant mitochondria. *Front Plant Sci.* 2014;5:35.
- Wang X, Chen H, Yang D, Liu C. Diversity of mitochondrial plastid DNAs (MTPDs) in seed plants. *Mitochondrial DNA A DNA Mapp Seq Anal.* 2017;29(4):635–42.
- Wang D, Rousseau-Gueutin M, Timmis JN. Plastid sequences contribute to some plant mitochondrial genes. *Mol Biol Evol.* 2012;29(7):1707–11.
- Goremykin VV, Salamini F, Velasco R, Viola R. Mitochondrial DNA of *Vitis vinifera* and the issue of rampant horizontal gene transfer. *Mol Biol Evol.* 2008;26(1):99–110.
- Cui H, Ding Z, Zhu Q, Wu Y, Qiu B, Gao P. Comparative analysis of nuclear, chloroplast, and mitochondrial genomes of watermelon and melon provides evidence of gene transfer. *Sci Rpt.* 2021;11:1595.
- Goremykin VV, Lockhart PJ, Viola R, Velasco R. The mitochondrial genome of *Malus domestica* and the import-driven hypothesis of mitochondrial genome expansion in seed plants. *Plant J.* 2012;71(4):615–26.
- Guo W, Grewe F, Fan W, Young GJ, Knoop V, Palmer JD, et al. *Ginkgo* and *Welwitschia* mitogenomes reveal extreme contrasts in gymnosperm mitochondrial evolution. *Mol Biol Evol.* 2016;33(6):1448–60.
- Kovar L, Nageswara-Rao M, Ortega-Rodriguez S, Dugas DV, Straub S, Cronn R, et al. PacBio-based mitochondrial genome assembly of *Leucaena trichandra* (Leguminosae) and an Intrageneric assessment of mitochondrial RNA editing. *Genome Biol Evol.* 2018;10(9):2501–17.
- Christensen AC. Plant mitochondrial genome evolution can be explained by DNA repair mechanisms. *Genome Biol Evol.* 2013;5(6):1079–86.
- Alverson AJ, Wei X, Rice DW, Stern DB, Barry K, Palmer JD. Insights into the evolution of mitochondrial genome size from complete sequences of *Citrullus lanatus* and *Cucurbita pepo* (Cucurbitaceae). *Mol Biol Evol.* 2010;27(6):1436–48.
- Choi I, Ruhlman TA, Jansen RK. Comparative mitogenome analysis of the genus *Trifolium* reveals independent gene fission of *ccmFn* and intracellular gene transfers in Fabaceae. *Int J Mol Sci.* 2020;21(6):1959.
- Kan SL, Shen TT, Gong P, Ran JH, Wang XQ. The complete mitochondrial genome of *Taxus cuspidata* (Taxaceae): eight protein-coding genes have transferred to the nuclear genome. *BMC Evol Biol.* 2020;20:10.
- Makarenko MS, Omelchenko DO, Usatov AV, Gavrilova VA. The insights into mitochondrial genomes of sunflowers. *Plants.* 2021;10(9):1774.
- Basak SK, Maiti G, Hajra PK. The genus *Primula* L. in India (a taxonomic revision). Dehradun: Bishen Singh Mahendra Pal Singh; 2014.
- Hu CM, Kelso S. Primulaceae. In: Wu ZY, Raven PH, editors. *Flora of China* (Vol 15). St. Louis: Science Press, Beijing & Missouri Botanical Garden Press; 1996.
- Richards J. *Primula*. Portland: Timber Press; 2003.
- De Vos JM, Hughes CE, Schneeweiss GM, Moore BR, Conti E. Heterostyly accelerates diversification via reduced extinction in primroses. *Proc Biol Sci.* 2014;281(1784):20140075.
- Cocker JM, Wright J, Li J, Swarbreck D, Dyer S, Caccamo M, et al. *Primula vulgaris* (primrose) genome assembly, annotation and gene expression, with comparative genomics on the heterostyly supergene. *Sci Rpt.* 2018;8(1):17942.
- Nowak MD, Russo G, Schlapbach R, Huu CN, Lenhard M, Conti E. The draft genome of *Primula veris* yields insights into the molecular basis of heterostyly. *Genome Biol.* 2015;16(1):12.
- Ren T, Yang Y, Zhou T, Liu Z. Comparative plastid genomes of *Primula* species: sequence divergence and phylogenetic relationships. *Int J Mol Sci.* 2018;19(4):1050.
- Zhang J, Zhang S, Zhang Y. The complete mitochondrial genome of a mangrove plant: *Aegiceras corniculatum* and its phylogenetic implications. *Mitochondrial DNA B Resour.* 2020;5(2):1502–3.
- Mower JP. The PREP suite: predictive RNA editors for plant mitochondrial genes, chloroplast genes and user-defined alignments. *Nucleic Acids Res.* 2009;37:W253–W59.
- Wang S, Song Q, Li S, Hu Z, Dong G, Song C, et al. Assembly of a complete mitogenome of *Chrysanthemum nankingense* using Oxford Nanopore long reads and the diversity and evolution of Asteraceae mitogenomes. *Genes.* 2018;9(11):547.
- Wang X, Bi C, Xu Y, Wei S, Dai X, Yin T, et al. The whole genome assembly and comparative genomic research of *Thellungiella parvula* (extremophile crucifer) mitochondrion. *Int J Genomics.* 2016;2016:1–13.
- Wang X, Cheng F, Rohlsen D, Bi C, Wang C, Xu Y, et al. Organellar genome assembly methods and comparative analysis of horticultural plants. *Hortic Res.* 2018;5(1):3.
- Zhang R, Jin J, Moore MJ, Yi TS. Assembly and comparative analyses of the mitochondrial genome of *Castanospermum australe* (Papilionoideae, Leguminosae). *Aust Syst Bot.* 2019;32:484–94.
- Chevigny N, Schatz-Daas D, Lotfi F, Gualberto JM. DNA repair and the stability of the plant mitochondrial genome. *Int J Mol Sci.* 2020;21(1):328.
- Štorchová H, Stone JD, Sloan DB, Abeyawardana OAJ, Müller K, Walterová J, et al. Homologous recombination changes the context of cytochrome b transcription in the mitochondrial genome of *Silene vulgaris* Kra. *BMC Genomics.* 2018;19:874.
- Kozik A, Rowan BA, Lavelle D, Berke L, Schranz ME, Michelmore RW, et al. The alternative reality of plant mitochondrial DNA: one ring does not rule them all. *PLoS Genet.* 2019;15(8):e1008373.
- Yu R, Sun C, Zhong Y, Liu Y, Sanchez Puerta MV, Mower JP, et al. The minicircular and extremely heteroplasmic mitogenome of the holoparasitic plant *Rhopalocnemis phalloides*. *Curr Biol.* 2022;32(2):470–79.e5.
- Palmer JD, Shields CR. Tripartite structure of the *Brassica campestris* mitochondrial genome. *Nature.* 1984;307(5950):437–40.
- Sullivan AR, Eldfjell Y, Schiffthaler B, Delhomme N, Asp T, Hebelstrup KH, et al. The mitogenome of Norway spruce and a reappraisal of mitochondrial recombination in plants. *Genome Biol Evol.* 2020;12(1):3586–98.

46. Wolfe KH, Li WH, Sharp PM. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proc Natl Acad Sci U S A*. 1987;84(24):9054–8.
47. Sloan DB, Müller K, McCauley DE, Taylor DR, Štorchová H. Intraspecific variation in mitochondrial genome sequence, structure, and gene content in *Silene vulgaris*, an angiosperm with pervasive cytoplasmic male sterility. *New Phytol*. 2012;196(4):1228–39.
48. Sinha RP, Hader DP. UV-induced DNA damage and repair: a review. *Photochem Photobiol Sci*. 2002;1(4):225–36.
49. He X, Burgess KS, Gao LM, Li DZ. Distributional responses to climate change for alpine species of *Cyananthus* and *Primula* endemic to the Himalaya-Hengduan Mountains. *Plant Divers*. 2019;41(1):26–32.
50. Rose JP, Kleist TJ, Lofstrand SD, Drew BT, Schonenberger J, Sytsma KJ. Phylogeny, historical biogeography, and diversification of angiosperm order Ericales suggest ancient Neotropical and east Asian connections. *Mol Phylogeny Evol*. 2018;122:59–79.
51. Orton LM, Burke SV, Wysocki WP, Duvall MR. Plastid phylogenomic study of species within the genus *Zea*: rates and patterns of three classes of microstructural changes. *Curr Genet*. 2017;63(2):311–23.
52. Fan W, Guo W, Funk L, Mower JP, Zhu A. Complete loss of RNA editing from the plastid genome and most highly expressed mitochondrial genes of *Welwitschia mirabilis*. *Sci China Life Sci*. 2019;62(4):498–506.
53. He ZS, Zhu AD, Yang JB, Fan WS, Li DZ. Organellar genomes and transcriptomes of *Nymphaea* reveal the interplay between intron splicing and RNA editing. *Int J Mol Sci*. 2021;22(18):9842.
54. Li Z, Liu Z, Zhong W, Huang M, Wu N, Xie Y, et al. Large-scale identification of human protein function using topological features of interaction network. *Sci Rpt*. 2016;6(1):37179.
55. Zancani M, Braidot E, Filippi A, Lippe G. Structural and functional properties of plant mitochondrial F-ATP synthase. *Mitochondrion*. 2020;53:178–93.
56. Kim DH, Kim BD. The organization of mitochondrial *atp6* gene region in male fertile and CMS lines of pepper (*Capsicum annuum* L.). *Curr Genet*. 2006;49(1):59–67.
57. Makarenko MS, Usatov AV, Tatarinova TV, Azarin KV, Logacheva MD, Gavrilova VA, et al. Characterization of the mitochondrial genome of the MAX1 type of cytoplasmic male-sterile sunflower. *BMC Plant Biol*. 2019;19(1):41–7.
58. Hu J, Yi R, Zhang H, Ding Y. Nucleo-cytoplasmic interactions affect RNA editing of *cox2*, *atp6* and *atp9* in alloplasmic male-sterile rice (*Oryza sativa* L.) lines. *Mitochondrion*. 2013;13(2):87–95.
59. Howard W, Kempken F. Cell type-specific loss of *atp6* RNA editing in cytoplasmic male sterile *Sorghum bicolor*. *Proc Natl Acad Sci U S A*. 1997;94(20):11090–5.
60. Li X, Huang W, Yang H, Jiang R, Sun F, Wang H, et al. EMP 18 functions in mitochondrial *atp6* and *cox2* transcript editing and is essential to seed development in maize. *New Phytol*. 2019;221(2):896–907.
61. Rice DW, Alverson AJ, Richardson AO, Young GJ, Sanchez-Puerta MV, Munzinger J, et al. Horizontal transfer of entire genomes via mitochondrial fusion in the angiosperm *Amborella*. *Science*. 2013;342(6165):1468–73.
62. Sanchez-Puerta MV, Cho Y, Mower JP, Alverson AJ, Palmer JD. Frequent, phylogenetically local horizontal transfer of the *cox1* group I intron in flowering plant mitochondria. *Mol Biol Evol*. 2008;25(8):1762–77.
63. Warren JM, Sloan DB. Interchangeable parts: the evolutionarily dynamic tRNA population in plant mitochondria. *Mitochondrion*. 2020;52:144–56.
64. Wang D, Wu YW, Shih ACC, Wu CS, Wang YN, Chaw SM. Transfer of chloroplast genomic DNA to mitochondrial genome occurred at least 300 Mya. *Mol Biol Evol*. 2007;24(9):2040–8.
65. Cummings MP, Nugent JM, Olmstead RG, Palmer JD. Phylogenetic analysis reveals five independent transfers of the chloroplast gene *rbcl* to the mitochondrial genome in angiosperms. *Curr Genet*. 2003;43(2):131–8.
66. Gandini CL, Sanchez-Puerta MV. Foreign plastid sequences in plant mitochondria are frequently acquired via mitochondrion-to-mitochondrion horizontal transfer. *Sci Rpt*. 2017;7(1):43402.
67. Qiu YL, Li LB, Wang B, Xue JY, Hendry TA, Li RQ, et al. Angiosperm phylogeny inferred from sequences of four mitochondrial genes. *J Syst Evol*. 2010;48(6):391–425.
68. Yan MH, Fritsch PW, Moore MJ, Feng T, Meng AP, Yang J, et al. Plastid phylogenomics resolves infrafamilial relationships of the Styracaceae and sheds light on the backbone relationships of the Ericales. *Mol Phylogeny Evol*. 2018;121:198–211.
69. Larson DA, Walker JF, Vargas OM, Smith SA. A consensus phylogenomic approach highlights paleopolyploid and rapid radiation in the history of Ericales. *Am J Bot*. 2020;107(5):773–89.
70. Zhang L, Wu W, Yan HF, Ge XJ. Phylotranscriptomic analysis based on coalescence was less influenced by the evolving rates and the number of genes: a case study in Ericales. *Evol Bioinforma*. 2015;11(Suppl 1):81.
71. Mast AR, Kelso S, Conti E. Are any primroses (*Primula*) primitively monomorphic? *New Phytol*. 2006;171(3):605–16.
72. Dong S, Zhao C, Chen F, Liu Y, Zhang S, Wu H, et al. The complete mitochondrial genome of the early flowering plant *Nymphaea colorata* is highly repetitive with low recombination. *BMC Genomics*. 2018;19:614.
73. Alverson AJ, Zhuo S, Rice DW, Sloan DB, Palmer JD. The mitochondrial genome of the legume *Vigna radiata* and the analysis of recombination across short mitochondrial repeats. *PLoS One*. 2011;6(1):e16404.
74. Wynn EL, Christensen AC. Repeats of unusual size in plant mitochondrial genomes: identification, incidence and evolution. *G3: Genes, Genom, Genet*. 2019;9(2):549–59.
75. Sloan DB, Wu Z. History of plastid DNA insertions reveals weak deletion and AT mutation biases in angiosperm mitochondrial genomes. *Genome Biol Evol*. 2014;6(12):3210–21.
76. Ko YJ, Kim S. Analysis of nuclear mitochondrial DNA segments of nine plant species: size, distribution, and insertion loci. *Genomics Inform*. 2016;14(3):90.
77. Petersen G, Cuenca A, Zervas A, Ross GT, Graham SW, Barrett CF, et al. Mitochondrial genome evolution in *Alismatales*: size reduction and extensive loss of ribosomal protein genes. *PLoS One*. 2017;12(5):e0177606.
78. Xu J, Luo H, Nie S, Zhang RG, Mao JF. The complete mitochondrial and plastid genomes of *Rhododendron simsii*, an important parent of widely cultivated azaleas. *Mitochondrial DNA Part B-Resour*. 2021;6(3):1197–9.
79. Liu XD, Liao XY, Chen DQ, Zheng Y, Yu X, Xu XY, et al. The complete chloroplast genome sequence of *Monotropa uniflora* (Ericaceae). *Mitochondrial DNA Part B-Resour*. 2020;5(3):3186–7.
80. Tanaka Y, Tsuda M, Yasumoto K, Yamagishi H, Terachi T. A complete mitochondrial genome sequence of Ogura-type male-sterile cytoplasm and its comparative analysis with that of normal cytoplasm in radish (*Raphanus sativus* L.). *BMC Genomics*. 2012;13(1):352.
81. Allen GC, Flores Vergara MA, Krasynanski S, Kumar S, Thompson WF. A modified protocol for rapid DNA isolation from plant tissues using cetyltrimethylammonium bromide. *Nat Protoc*. 2006;1(5):2320–5.
82. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30(15):2114–20.
83. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol*. 2012;19(5):455–77.
84. Ruby JG, Bellare P, Derisi JL. PRICE: software for the targeted assembly of components of (Meta) genomic sequence data. *G3: Genes, Genom, Genet*. 2013;3(5):865–80.
85. Li DH, Luo RB, Liu CM, Leung CM, Ting HF, Sadakane K, et al. MEGAHIT v1.0: a fast and scalable metagenome assembler driven by advanced methodologies and community practices. *Methods*. 2016;102:3–11.
86. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. *BMC Bioinform*. 2009;10(1):421.
87. Mower JP, Sloan DB, Alverson AJ. Plant mitochondrial genome diversity: the genomics revolution. In: Springer Vienna; 2012. p. 123–144.
88. Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, et al. GeSeq-versatile and accurate annotation of organelle genomes. *Nucleic Acids Res*. 2017;45(W1):W6–W11.
89. Chan PP, Lowe TM. tRNAscan-SE: searching for tRNA genes in genomic sequences. *Methods Mol Biol*. 2019;1962:1–14.
90. Lang BF, Laforest MJ, Burger G. Mitochondrial introns: a critical view. *Trends Genet*. 2007;23(3):119–25.
91. Benson G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res*. 1999;27(2):573–80.

92. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, et al. Circos: an information aesthetic for comparative genomics. *Genome Res.* 2009;19(9):1639–45.
93. Wilkins MR, Gasteiger E, Bairoch A, Jean-Charles S, Williams KL, Appel RD, Hochstrasser DF. Protein identification and analysis tools in the ExPASy server. In: *The proteomics protocols handbook*: Humana Press; 2005. p. 531–52.
94. Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. *J Mol Biol.* 1982;157(1):105–32.
95. Katoh K. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 2002;30(14):3059–66.
96. McGuffin LJ, Bryson K, Jones DT. The PSIPRED protein structure prediction server. *Bioinformatics.* 2000;16(4):404–5.
97. Richardson AO, Rice DW, Young GJ, Alverson AJ, Palmer JD. The “fossilized” mitochondrial genome of *Liriodendron tulipifera*: ancestral gene content and order, ancestral editing sites, and extraordinarily low mutation rate. *BMC Biol.* 2013;11(1):29.
98. Marçais G, Delcher AL, Phillippy AM, Coston R, Salzberg SL, Zimin A. MUMmer4: a fast and versatile genome alignment system. *PLoS Comp Biol.* 2018;14(1):e1005944.
99. Vaidya G, Lohman DJ, Meier R. SequenceMatrix: concatenation software for the fast assembly of multi-gene datasets with character set and codon information. *Cladistics.* 2011;27(2):171–80.
100. Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, et al. DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol Biol Evol.* 2017;34(12):3299–302.
101. Jin JJ, Yu WB, Yang JB, Song Y, Depamphilis CW, Yi TS, et al. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biol.* 2020;21:241.
102. Talavera G, Castresana J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol.* 2007;56(4):564–77.
103. Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol.* 2015;32(1):268–74.
104. Kalyaanamoorthy S, Minh BQ, Wong TKF, Von Haeseler A, Jermini LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods.* 2017;14(6):587–9.
105. Ronquist F, Teslenko M, Van Der Mark P, Ayres DL, Darling A, Höhna S, et al. MrBayes 3.2: efficient bayesian phylogenetic inference and model choice across a large model space. *Syst Biol.* 2012;61(3):539–42.
106. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior summarization in bayesian phylogenetics using tracer 1.7. *Syst Biol.* 2018;67(5):901–4.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

