Assembly and analysis of the *Populus deltoides* mitochondrial genome: the first report of a multicircular mitochondrial conformation for the genus *Populus*

Yanshu Qu[1] · Pengyan Zhou[1] · Chunfa Tong[1] ·
Changwei Bi[2] · Li'an Xu[1]

Abstract Genomics research of *Populus deltoides*, an important timber species that is widely planted worldwide, is an important part of poplar breeding. Currently, the nuclear and chloroplast genome of *P. deltoides* have been sequenced, but its mitochondrial genome (mitogenome) has not been reported. To further explore the evolution and phylogeny of *P. deltoides*, the mitogenome of *P. deltoides* I-69 was assembled using reads from Nanopore and Illumina sequencing platforms and found to consist of 802,637 bp and three circular chromosomes (336,205, 280,841, and 185,591 bp) containing 58 genes (34 protein-coding genes, 21 tRNA genes, and 3 rRNA genes). RNA analysis in combination with several species showed significantly fewer RNA editing sites in the mitogenomes of poplar and other angiosperms than in gymnosperms. Sequence transfer analysis showed extensive mitogenome rearrangements in *Populus* species, and with evolution from lower to higher plants, tRNA transfer from chloroplasts to mitochondria became increasingly frequent. In a phylogenetic analysis, the evolutionary status of *P. deltoides* was determined, and the section *Populus* was supported. Our results based on the first report of a multicircular conformation of the *Populus* mitogenome provide a basis for further study of the evolution and genetics of *P. deltoides* and other *Populus* species and for breeding programs.

The online version is available at https://www.springerlink.com.

Corresponding editor: Yu Lei.

✉ Changwei Bi
bichwei@163.com

✉ Li'an Xu
laxu@njfu.edu.cn

[1] Key Laboratory of Forestry Genetics & Biotechnology of Ministry of Education, Co-Innovation Center for Sustainable Forestry in Southern China, Nanjing Forestry University, Nanjing 210037, People's Republic of China

[2] College of Information Science and Technology, Nanjing Forestry University, Nanjing 210037, People's Republic of China

Introduction

The endosymbiosis theory, currently the most commonly accepted hypothesis to explain the origin of organelles in eukaryotic cells, suggests that mitochondria were formed by the invasion of host cells resembling archaea by α-proteobacteria, the ancestor of mitochondria, approximately 1.5 billion years ago and slightly before chloroplasts (Gupta and Golding 1996). Since the first sequencing of a mitochondrial genome in a land plant (*Marchantia polymorpha*) in 1992, the National Center for Biotechnology Information (NCBI) has collected 327 mitogenomes from land plants (as of March 2022), far fewer than the 6104 chloroplast genomes that have been sequenced (Oda et al. 1992). To some extent, this lag reflects the difficulty and complexity of extracting plant mitochondrial DNA and of genome assembly, but as the cost of sequencing decreases and assembly techniques improve, the number of plant mitogenomes should soon increase rapidly.

Plant mitogenomes are typically species specific, with large differences in size and structure among different species (Alverson et al. 2010). The mitogenomes of some plants contain a high number of introns and undergo trans-splicing after transcription (Handa 2003; Malek et al. 1997). In addition, plant mitogenomes are characterized by a wide range of RNA editing sites and universal genetic code usage (Gray and Covello 1993; Kubo et al. 2000). In contrast to the relatively stable genome size of animal mitochondria and plant chloroplasts, the size of mitogenomes in higher plants varies greatly, even among species of the same genus, from 66 kb in *Viscum scuruloideum* (Santalaceae) to 11.3 Mb in *Silene conica* (Caryophyllaceae) (Skippington et al. 2015; Sloan et al. 2012) and from 735.17 kb in *Salix cardiophylla* (Salicaceae) to 562.92 kb in *Salix polaris* (Chen et al. 2020a, b). This difference in mitogenome size is due in large part to the insertion of foreign DNA and extensive replication of noncoding regions (Bergthorsson et al. 2003; Wynn and Christensen 2019). The number of genes in plant mitogenomes also varies, ranging from 19 (*Viscum album*) to 221 (*Capsicum annuum*, Solanaceae) (Petersen et al. 2015; Jo et al. 2011). Although plant mitogenomes vary in size and number of genes, the genes encoding ATP synthase, NADH dehydrogenase, coenzyme Q-cytochrome *c* reductase, and cytochrome *c* oxidase are relatively conserved (Bi et al. 2020). Other genes, such as those encoding succinate dehydrogenase and ribosomal proteins, are not conserved and have been lost on occasion during plant evolution. For example, *sdh3* has been lost in all studied species of Saliaceae, and *rps11* has also been frequently lost in many angiosperms (Bi et al. 2016). Although the content of these genes in different species is not conserved, the gene sequence is relatively conserved (Allen et al. 2007; Chaw et al. 2008; Goremykin et al. 2009). Structurally, in contrast to that of plant chloroplasts and animal mitochondria, the structure of plant mitochondria has a variety of conformations, with not only the presence of a common single circular structure, but also linear structures and multiple chromosomal conformations. For example, the mitogenome of *Lactuca saligna* forms a branched linear structure (Kozik et al. 2019); the *Cucumis sativus* mitogenome has three independent chromosomal structures (Alverson et al. 2011); the mitogenome of *Scutellaria tsinyunensis* has a double-ring crossover structure (Li et al. 2021); and that of *Lactuca sativa* appears to undergo a dynamic change as evidenced by cryoelectron microscopy (Kozik et al. 2019). These features also indicate the instability of the plant mitogenome structure and make the assembly of mitogenomes in higher plants more difficult.

Populus deltoides (section *Aigerios*) is an important timber species with a natural distribution from southern Canada to the southeastern United States (Fahrenkrog et al. 2017). The trees are tall, leafy and geographically adaptable, and their wood is an important raw material for wood-based panels and papermaking. Cultivar I-69 is one of the most important cultivars of *P. deltoides*, with characteristics such as rapid growth and resistance to brown spot disease, and has been promoted worldwide and widely used for afforestation in China (Bai et al. 2021). Therefore, the study of *P. deltoides* I-69 at the genomic level is of both theoretical and practical importance.

At present, the nuclear and chloroplast genomes of *P. deltoides* have been assembled and annotated, but its mitogenome has not been studied. In this study, the first *P. deltoides* mitogenome was assembled and annotated by combining Nanopore and Illumina sequencing data using *P. deltoides* I-69 as the material. The gene content, codon usage bias, repeat sequences, RNA editing sites, genome collinearity and gene transfer between different genomes were analyzed, and phylogenetic analyses were performed based on conserved mitochondrial genes. The results of this study further support the existence of multiple conformations of the plant mitogenome and provide valuable information for the further study of the phylogenetic status of *P. deltoides*.

Materials and methods

Plant materials, DNA extraction and sequencing

P. deltoides I-69 branches were collected at Siyang Farm in Siyang County, Jiangsu Province, China (33 °40′ N, 118 °41′ E), and propagated in soil at the Xiashu Forestry Farm, Jurong City, Jiangsu Province, China (32 °7′ N, 119 °12′ E). Fresh leaves were collected and kept at −80 °C in a freezer. Genomic DNA was extracted from frozen leaves using the CTAB method (Arseneau et al. 2017). The purity and concentration of DNA were determined using Qubit and NanoDrop (Thermo Fisher Scientific), respectively. Size selection (> 20 kb) was performed using 15 μg of genomic DNA and the Blue Pippin System (Sage Science, Beverly, MA, USA), and PCR-free libraries were constructed using the standard protocol of Nanopore (Oxford Nanopore, Oxford, UK). Libraries prepared in two flow cells were sequenced at Shanghai Biozeron Biotechnology Co., Ltd and Nextomics Biosciences Co., Ltd for Nanopore sequencing. The original signal data were based on ONT MinKNOW software for base detection, and the statistics for the generated Nanopore sequencing data were calculated using Nanostat software (v1.4.2) (De Coster et al. 2018).

In addition, for leaf samples from the same individual, the whole genome was sequenced using the Illumina HiSeq 2000 sequencing platform by Biomarker Technologies Corporation, Beijing, China. To correct contigs assembled from Nanopore sequencing data, these sequencing data were processed using the NGS QC Toolkit (v2.3.3) with

default parameters to obtain high-quality reads (Patel and Jain 2012).

Mitogenome assembly and annotation

De novo assembly of the *P. deltoides* mitogenome was performed using the Oxford Nanopore data obtained by sequencing in this study and the wtdbg2 software (Ruan and Li 2020). The contigs obtained after preliminary assembly were polished by Pilon software (v1.18) in combination with Illumina sequencing data for error correction (Walker et al. 2014).

The complete mitogenome of *P. deltoides* was annotated using GE-Seq (Tillich et al. 2017), and the mitogenome of *Populus tremula* (NC_028096.1) was used as the reference genome. Protein-coding genes (PCGs) of the *P. deltoides* mitogenome were manually compared with those in the mitogenomes of other *Populus* species, then manually corrected using MacVector (Bernt et al. 2013). The tRNAscan-SE and default parameters were used to predict tRNA genes (Chan and Lowe 2019; Chan et al. 2021). rRNA genes were identified using RNAmmer (v1.2) and manually corrected with local BLASTn (Camacho et al. 2009; Lagesen et al. 2007). The OGDRAW web server was used to visualize the complete multi-circular *P. deltoides* mitogenome (Greiner et al. 2019).

Sequence repeat analysis

Simple sequence repeats (SSRs) were analyzed using the MISA program (Beier et al. 2017), and the SSR motif size thresholds for $1-6$ nucleotides were set to 8, 4, 4, 3, 3, and 3. Tandem repeat sequences in *P. deltoides* mitogenome were analyzed by Tandem Repeats Finder (v4.09; Benson 1999) with default parameters. Finally, the size and position of the dispersed repeats, including forward repeats, palindromic repeats, reverse repeats and complement repeats, were detected using Reputer (Kurtz et al. 2001); the minimum repeat length was set to 30 bp, the Hamming distance was set to 3, and the repeats containing the overlap regions were manually verified and merged.

Codon usage analysis and RNA editing site prediction

The relative synonymous codon usage (RSCU) of *P. deltoides* was calculated using MEGA X (Kumar et al. 2018). The amino acid compositions of PCGs of *P. deltoides*, *P. tremula*, *Salix suchowensis*, *Gossypium raimondii*, *Arabidopsis thaliana*, *Ginkgo biloba* and *Cycas taitungensis* were analyzed using a local Perl language script (Kumar et al. 2016).

RNA editing sites in the mitogenome of *P. deltoides* and four other species (*S. suchowensis*, *P. tremula*, *G. biloba* and

Pinus taeda) were predicted by PREP-Mt online software (Mower 2009) with a *C* value set to 0.2.

Multigenomic alignment

Nucleotides of the *P. deltoides* mitogenome and chloroplast genomes (NC_040929.1) were aligned using BLASTn (version 2.9.0) with E-value $\leq 1e-10$, matching rate $\geq 80\%$, and length ≥ 40 bp.

To explore the mitogenome collinearity relationship between different poplars, we aligned the nucleotides of the *P. deltoides* and *P. tremula* mitogenomes using BLASTn with E-value $\leq 1e-10$, matching rate $\geq 80\%$, and length ≥ 1000 bp. The synteny circle plot for the above two analyses was plotted with Tb tools (Chen et al. 2020a, b).

Phylogenetic analysis

The genes for 43 species (see Supplementary Table S1 for all NCBI accessions) were first extracted using a local Perl script to build a matrix from which a total of 20 conserved genes were selected, then integrated and linked by another local Perl language program. Then, multisequence alignment was performed with the default parameters of Muscle (Edgar 2004). Finally, IQ-TREE (Minh et al. 2020) was used to construct the maximum likelihood-based phylogenetic tree, and the best model was determined by Model Finder (Kalyaanamoorthy et al. 2017; Minh et al. 2020) as GTR + F + R10 with bootstrap set to 1000 times, using two gymnosperms (*G. biloba* and *P. taeda*) as outgroups.

Results and discussion

Assembly and annotation of the P. deltoides mitogenome

A total of 44.3 Gb of Nano pore sequencing data were generated, containing 2.7 million reads with an average length of 16 kb and N50 of 25 kb, which is approximately $100\times$ of the *P. deltoides* genome. A total of 235.4 million Illumina reads with 23.8 Gb bases were generated. All sequencing-generated Nano pore data were used for de novo assembly, and $\sim 55\times$ Illumina sequencing reads were used for error correction to ensure the accuracy of the results. Finally, the complete *P. deltoides* mitogenome was harvested.

The total length of the *P. deltoides* mitogenome is 802637 bp, which is similar to that of *P. tremula* (783,442 bp) and *P. alba* (838,420 bp) (Kersten et al. 2016; Brenner et al. 2019). However, unlike the common plant monocyclic mitogenome structure, the *P. deltoides* mitogenome was assembled into three circular chromosomal structures with lengths of 336,205 bp (Chr1), 280,841 bp (Chr2),

and 185,591 bp (Chr3) (Fig. 1), and the three chromosome sequences were submitted to NCBI Genome Database (MZ951174.1, MZ951175.1, MZ951176.1). The nucleotide composition of the complete *P. deltoides* mitogenome was 27.67% A, 22.36% C, 22.42% G, and 27.56% T, with a GC content of 44.77% (Table 1). The GC contents of the three independent chromosomes were approximately the same: 44.89% (Chr1), 44.65% (Chr2), and 44.75% (Chr3).

The 58 genes that were identified in the mitogenome of *P. deltoides* included 34 PCGs, 21 tRNA genes and 3 rRNA genes (Table 2). There were 28 genes on Chr1 (13 complete PCGs, 11 tRNA genes and all 3 rRNA genes), 22 genes on Chr2 (13 complete PCGs, 7 tRNA genes), and 10 on Chr3 (6 complete PCGs and 3 tRNA genes). Interestingly, we found that the exons of the gene *nad1* were distributed on Chr1 and Chr2, and the exons of *nad5* were distributed on Chr2 and Chr3. In addition, three copies of the *trnM-CAU* gene were found, all on Chr1, and two copies of the *trnP-UGG* gene were found, all on Chr2.

As shown in Table 1, the total length of the PCGs in the *P. deltoides* mitogenome was 30,483 bp, accounting for 3.80% of the entire mitogenome; the total length of tRNA genes and
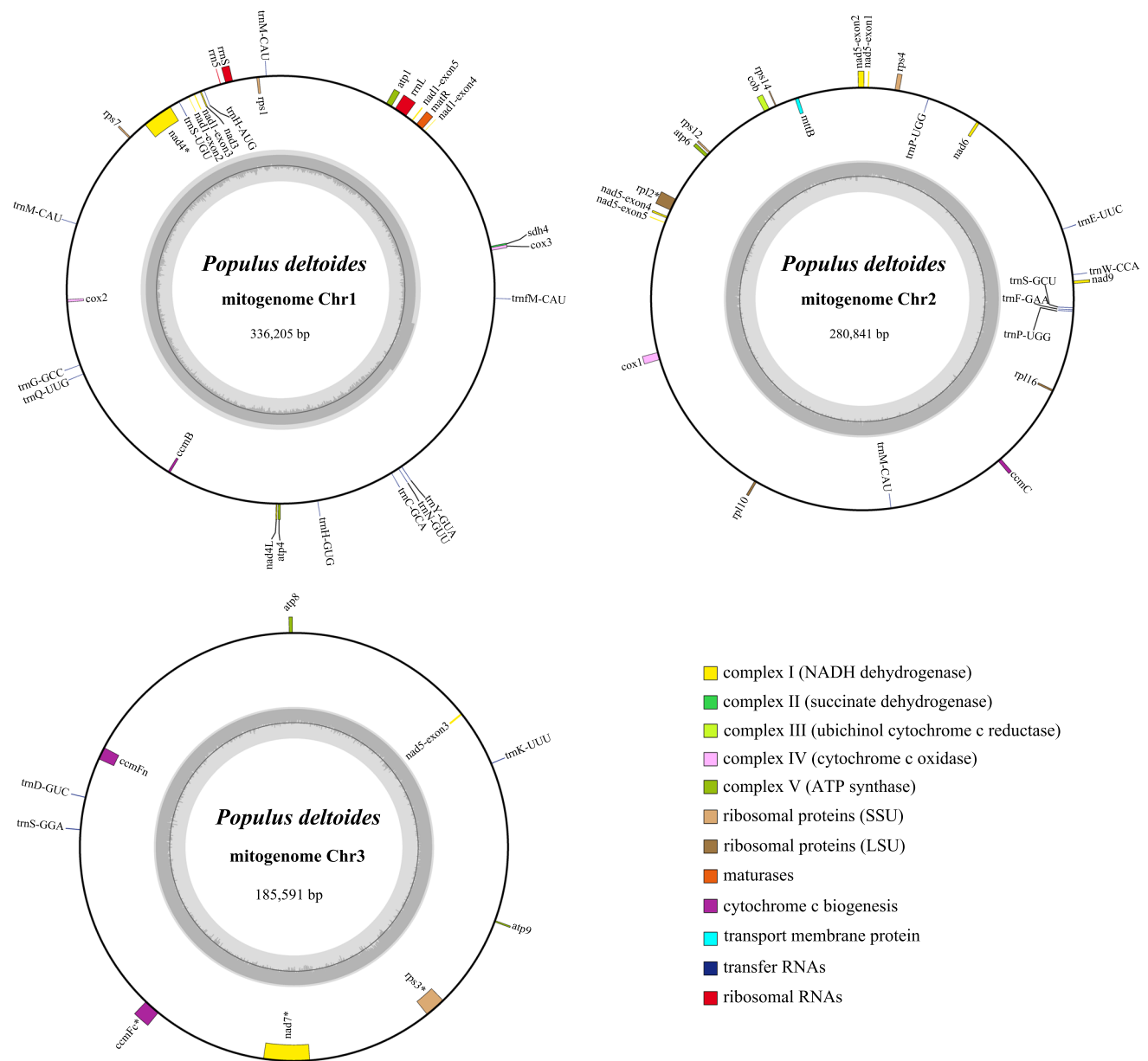

Fig. 1 Circular map of *Populus deltoides* mitogenome Chr1, Chr2 and Chr3. The genes on the outside are transcribed in the clockwise direction and those inside the circle in counterclockwise direction. GC content is represented by the dark gray plot in the inner circle. Genes with an asterisk (*) have introns

Table 1 Genomic features of the *Populus deltoides* mitogenome

Feature	A (%)	C (%)	G (%)	T (%)	GC (%)	Size (bp)	Percentage in genome (%)
Whole genome	27.67	22.36	22.42	27.56	44.77	802,637	100
Chr1	27.72	22.41	22.48	27.39	44.89	336,205	41.89
Chr2	27.57	22.26	22.4	27.78	44.65	280,841	34.99
Chr3	27.72	22.4	22.34	27.53	44.75	185,591	23.12
Protein-coding genes	26.46	20.62	21.62	31.3	42.24	30,483	3.8
cis-Spliced introns	25.43	24.6	27.06	22.91	51.66	27,077	3.37
tRNA genes	24.01	23.88	26.53	25.58	50.41	1583	0.2
rRNA genes	25.89	22.76	29.41	21.94	52.17	5369	0.67
Noncoding regions	27.82	22.34	22.22	27.62	44.56	738,125	91.96

Table 2 Gene annotation of the *Populus deltoides* mitogenome

Group	Chr1	Chr2	Chr3
Complex I	*nad1*, nad2, nad3, nad4, nad4L*	*nad1*, nad5*, nad6, nad9*	*nad5*, nad7*
Complex II	*sdh4*		
Complex III		*cob*	
Complex IV	*cox2, cox3*	*cox1*	
Complex V	*atp1, atp4*	*apt6*	*apt8, atp9*
Cytochrome c biogenesis	*ccmB*	*ccmC*	*ccmFc, ccmFn*
Ribosomal proteins (SSU)	*rps1, rps7*	*rps4, rps12, rps14*	*rps3*
Ribosomal proteins (LSU)		*rpl2, rpl10, rpl16*	
Maturases	*matR*		
Methyltransferase		*mttB*	
Ribosomal RNAs	*rrnL, rrn5, rrnS*		
Transfer RNAs	*trnM-CAU* (3 copies)	*trnE-UUC*	*trnK-UUU*
	trnH-AUG	*trnP-UGG* (2 copies)	*trnD-GUC*
	trnS-UGU	*trnfM-CAU*	*trnS-GGA*
	trnG-GCC	*trnS-GCU*	
	trnQ-UUG	*trnW-CCA*	
	trnH-GUG	*trnF-GAA*	
	trnC-GCA		
	trnN-GUU		
	trnY-GUA		

*Genes with introns

rRNA genes accounted for 0.20 and 0.67%, respectively; *cis*-spliced introns accounted for 3.37%, and the total length of noncoding regions was 738,125 bp, accounting for 91.96%. Noncoding sequences occupied more than 80% of the mitogenome in most plants. These noncoding region sequences consisted of repetitive fragments, sequences transferred from chloroplast and nuclear genomes, and even sequences from other species obtained by horizontal gene transfer, such as the mitogenome of the oldest relict angiosperm *Amborella trichopoda*, in which a large number of sequence fragments from mosses, green algae, and other plants are present (Rice et al. 2013). Many of these sequences were relatively short, repetitive fragments, which is an important reason for the very easy recombination and rearrangement of plant mitogenomes (Alverson et al. 2011; Rice et al. 2013).

Codon usage analysis of PCGs

ATG is a typical start codon in plant mitochondria, but plant mitochondria may also use several other different start codons (Bock et al. 1994; Dong et al. 1998). The full length of PCGs of *P. deltoides* is 30483 bp. Most of these PCGs have a typical ATG start codon, but the start codon of *mttB* is ATT and GGG for *rpl16*, which may be the result of RNA editing. In addition, three different stop codons were identified in the PCGs, namely, TAA (*atp1*, *atp6*, *cox1*,

cox2, *nad1*, *nad2*, *nad3*, *nad4L*, *nad5*, *nad9*, *rps1*, *rps3*, *rps4*, *rps7*, *rpl2*, *rpl10*, *rpl16*, and *sdh4*), TAG (*atp4*, *atp8*, *atp9*, *ccmFc*, *matR*, *mttB*, *nad6*, *nad7*, and *rps14*) and TGA (*ccmB*, *ccmC*, *ccmFn*, *cob*, *cox3*, *nad4*, and *rps12*) (Supplementary Table S2). As shown in Fig. 2, we analyzed the codon use of *P. deltoides*, *P. tremula*, *S. suchowensis*, *G. raimondii*, *A. thaliana*, *G. biloba*, and *C. taitungensis*. The results showed that the three most commonly used amino acids in plant mitochondria were leucine (Leu), serine (Ser), and arginine (Arg), while methionine (Met) and tryptophan (Trp) were rarely used (Handa 2003; Chaw et al. 2008; Bi et al. 2016; Guo et al. 2016; Kersten et al. 2016; Ye et al. 2017).

We found that the amino acid composition of mitochondrial proteins in plants of Salicaceae was very similar. In comparison with *Arabidopsis thaliana* and *Gossypium*, *Populus deltoides* had a similar amino acid composition. In the gymnosperms *G. biloba* and *C. taitungensis*, the use frequency of Gly, Pro and Arg was significantly higher than that of the angiosperms examined. Proline and arginine are involved in many stages of plant growth and development, especially in physiological and biochemical processes such as stress resistance. Glycine can also improve plant stress resistance, which may explain the longer lifespan of gymnosperms. The use frequency of lysine (Lys) and leucine (Leu), which promote photosynthesis and regulation in *G. biloba* and *C. taitungensis*, is significantly lower than that of the above five angiosperms, which may be why *G. biloba* and *C. taitungensis*, as gymnosperms, grow slower than

angiosperms (Chaw et al. 2008; Guo et al. 2016;). Except for the above amino acids, the distribution of other amino acids between angiosperms (*P. deltoides*, *P. tremula*, *S. suchowensis*, *G. raimondii*, and *A. thaliana*) and gymnosperms (*G. biloba* and *C. taitungensis*) is relatively conserved.

For the relative probability of the occurrence of specific codons between synonymous codons that encode corresponding amino acids (RSCU), RSCU = 1 indicates that there is no preference for codon use, while RSCU > 1 indicates that the codon is a relatively frequently used codon (Yang et al. 2021; Sau et al. 2006). Figure 3 shows the RSCU analysis of the mitogenome of *P. deltoides*, and we found that all codons are present in PCGs.

As shown in Fig. 3, the RSCU values of almost all NNA and NNT codons were greater than or equal to 1.00 except for Leu (CUA, 0.99), Ile (AUA, 0.92), Ser (UCA, 0.97; AGU, 0.94) and Arg (CGU, 0.88). This phenomenon indicates high A/T content at the third codon position in the mitogenome of *P. deltoides*, which is very similar to what has been reported in the mitogenomes of other land plants. The strong AT bias of the third codon is considered to be a universal phenomenon in higher plants (Bi et al. 2020; Wang et al. 2021; Yang et al. 2021).

Analysis of repeat regions

Compared with chloroplasts, plant mitochondria are characterized by a large distribution of repetitive sequences, and most of the size differences and rearrangements among plant
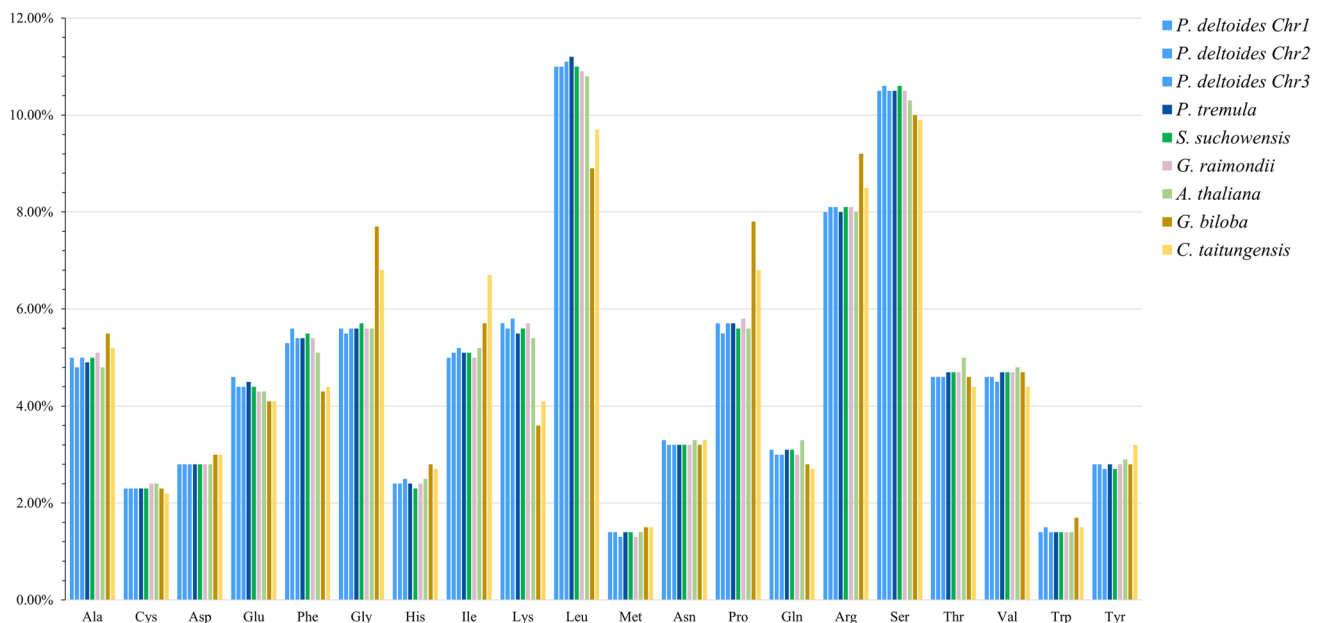

Fig. 2 Codon usage pattern in the *Populus deltoides* mitogenome compared with those of *P. tremula*, *Salix suchowensis*, *Gossypium raimondii*, *Arabidopsis thaliana*, *Ginkgo biloba* and *Cycas taitungensis*. The fraction of each amino acid residue in all mitochondrial proteins is given by the y-axis

Fig. 3 RSCU values of the *Populus deltoides* mitogenome. The *x*-axis indicates the codon families

mitogenomes can be explained by the size and arrangement of repetitive sequences, which play an important role in the evolution of plant mitochondria (Tanaka et al. 2014; Wynn and Christensen 2019; Yang et al. 2021).

SSRs are highly variable tandem repeat fragments of 1–6 bases, which are more frequent in plant genomes and are widely used as DNA markers for species identification and genetic diversity research (Ma et al. 2017; Li and Ye 2020). A total of 714 SSRs were detected in the mitogenome of *P. deltoides* (Table 3), including 302 mononucleotide repeats (42.3%), 277 dinucleotide repeats (38.80%), 27 trinucleotide

repeats (3.78%), 87 tetranucleotide repeats (12.18%), 18 pentanucleotide repeats (2.52%) and 3 hexanucleotide repeats (0.42%). Among them, more than 80% were mononucleotide and dinucleotide repeats, comparable to the findings for a variety of species such as *S. suchowensis*, *G. raimondii*, and *Gleditsia sinensis* (Xu et al. 2017). Further analysis showed that 90.06% of the mononucleotide SSRs were A or T. A higher A/T content in SSRs may lead to higher AT abundance in the mitogenome of *P. deltoides*, which is consistent with the results for many other angiosperms (Bi et al. 2020; Yang et al. 2021). Our results provide valuable reference

Table 3 Frequency of identified SSRs in the *Populus deltoides* mitogenome

Motif type	Number of repeats														Total	Proportion (%)
	3	4	5	6	7	8	9	10	11	12	13	14	15	16		
Monomer	0	0	0	0	0	133	82	52	17	10	4	3	0	1	302	42.3
Dimer	0	236	35	5	1	0	0	0	0	0	0	0	0	0	277	38.8
Trimer	0	27	0	0	0	0	0	0	0	0	0	0	0	0	27	3.78
Tetramer	82	4	1	0	0	0	0	0	0	0	0	0	0	0	87	12.18
Pentamer	16	2	0	0	0	0	0	0	0	0	0	0	0	0	18	2.52
Hexamer	3	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0.42
Total	101	269	36	5	1	133	82	52	17	10	4	3	0	1	714	100

Table 4 Distribution of tandem repeats in the *P. deltoides* mitogenome

Code	Chr	Size (bp)	Start	End	Repeat (bp)×Copy number	Location
1	Chr1	20	34,197	34,234	(CCCTTGTCTTGGCGTTCTTC)×2	IGS (*sdh4*, *nad1-exon4*)
2	Chr1	18	52,587	52,622	(CATAGTCGCGAGCTGTTT)×2	*rrnL*
3	Chr1	33	71,579	71,646	(CTAATACAGACGACGAGAAAAAGAAAC CGAAAC)×2	IGS (*atp1*, *trnM-CAU*)
4	Chr1	13	153,010	153,035	(CTTGCCTTATTAG)×2	IGS (*trnM-CAU*, *cox2*)
5	Chr1	22	195,695	195,737	(ATAAGGAAGAAACAGGTACAAT)×2	IGS (*trnQ-UUG*, *ccmB*)
6	Chr1	19	202,785	202,823	(TAAAAAAGAAATTCCCCTA)×2	IGS (*trnQ-UUG*, *ccmB*)
7	Chr1	25	237,622	237,669	(TCGACTGTTAAGGACACAGAGGGGA)×2	IGS (*ccmB*, *nad4L*)
8	Chr2	21	23,165	23,206	(GCAATCAAGTTAAAATCACCA)×2	IGS (*trnE-UUC*, *nad1-exon1*)
9	Chr2	17	77,802	77,835	(AGGAAATCTGTATTTAT)×2	IGS (*nad5-exon2*, *mttB*)
10	Chr2	24	276,996	277,043	(CATAACCAGGCAGTGAGGAATCTT)×2	IGS (*rpl16*, *trnP-UGG*)
11	Chr3	18	12,559	12,594	(TTCCCATCTATACCTTCT)×2	IGS (*trnK-UUU*, *nad5-exon3*)
12	Chr3	20	46,230	46,268	(TTCTGTTCTGTCATACTCTA)×2	IGS (*nad5-exon3*, *atp8*)
13	Chr3	19	109,978	110,015	(AAACAGAAGCAAAAAGAAA)×2	IGS (*trnS-GGA*, *ccmFc*)
14	Chr3	11	168,256	168,277	(AAGGAAAAGGG)×2	IGS (*rps3*, *atp9*)
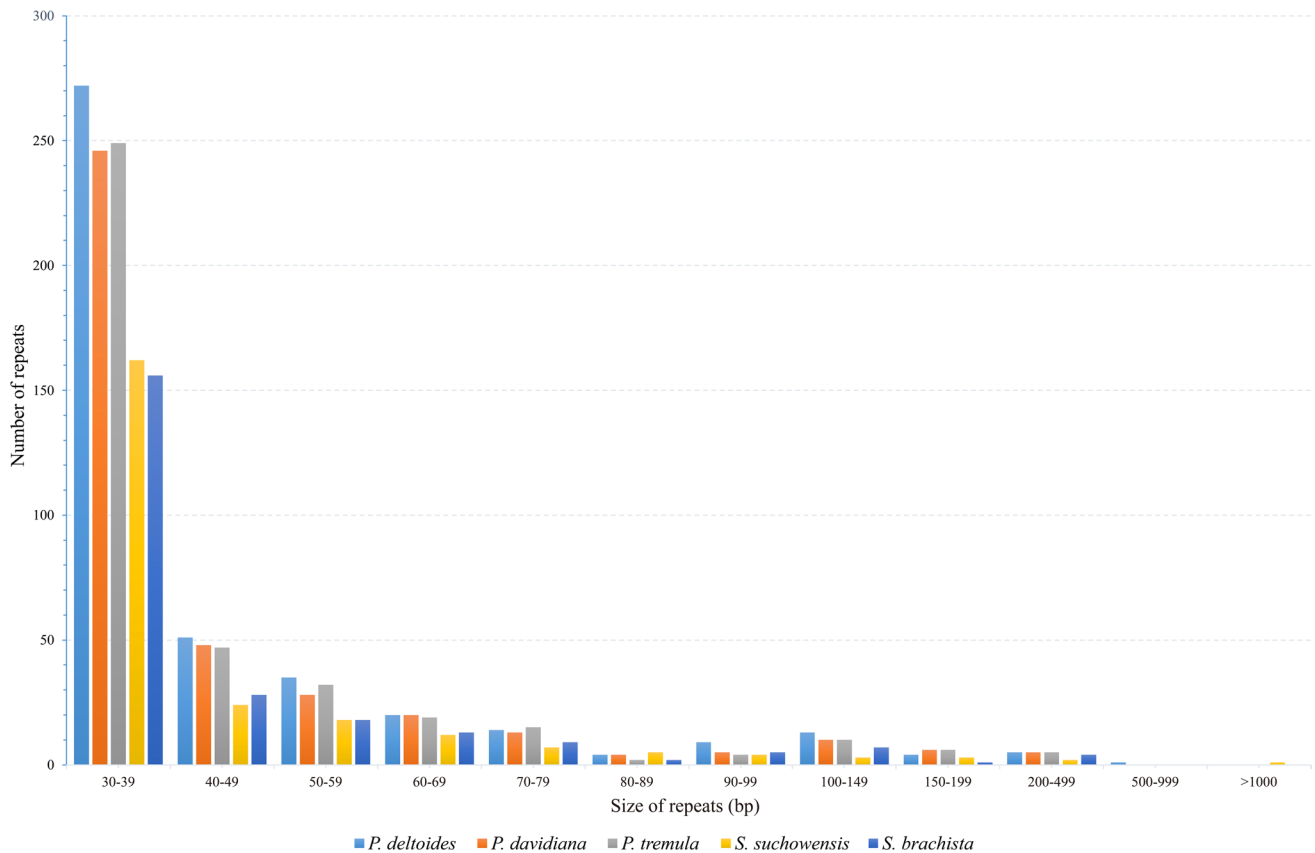
Fig. 4 Frequency distribution of dispersed repeats in the *Populus deltoides* mitogenome compared with four other Salicaceae plants

information for molecular marker development and population variation studies of *P. deltoides*.

As shown in Table 4, 14 tandem repeats were identified in the mitogenome of *P. deltoides*, with unit lengths ranging from 11 to 33 bp, of which 7 were distributed on Chr1, 3 on Chr2, and 4 on Chr3. Further analysis showed that 13 of them were located in the intergenic region, and only one tandem repeat sequence on Chr1 was located in a coding
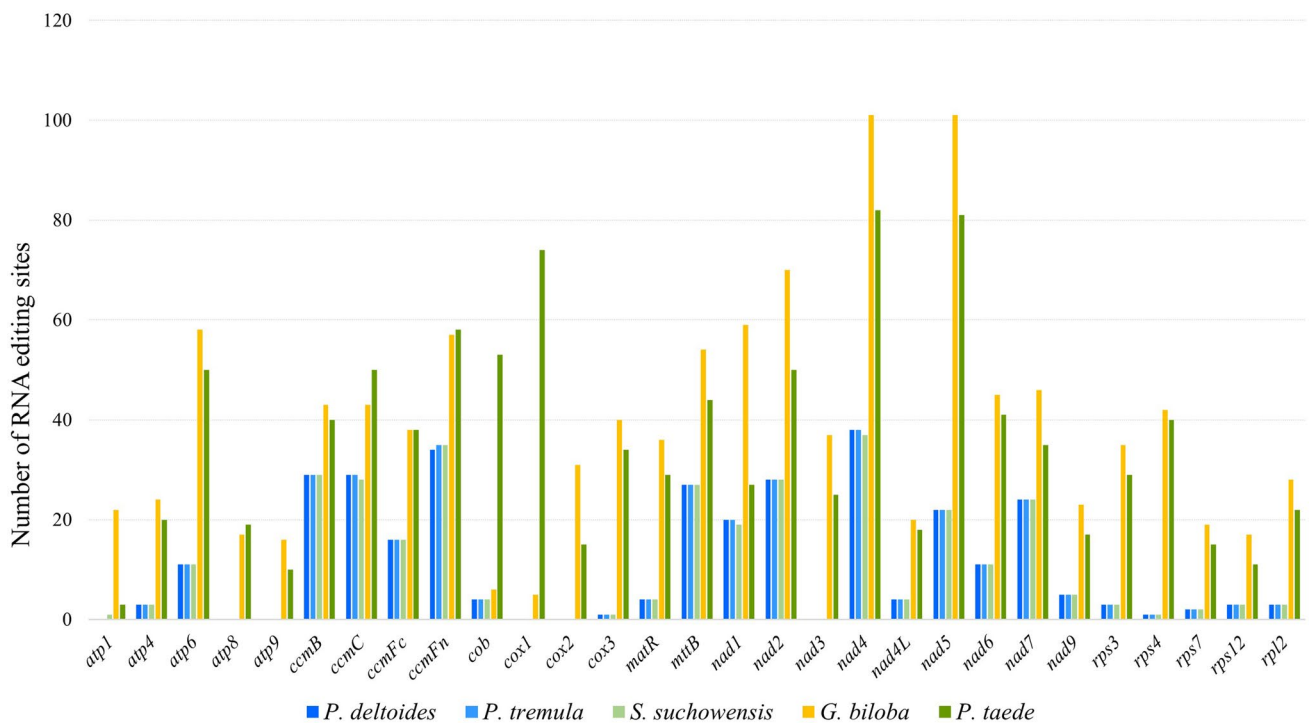
Fig. 5 The distribution of RNA editing sites in the mitogenomes of three Salicaceae plants and two gymnosperms (*Ginkgo biloba* and *Pinus taeda*)

region (*rrnL*, [CATAGTCGCGAGCTGTTT] × 2). The large distribution of SSRs and tandem repeats in intergenic regions indicates that repetitive sequences do not affect the transcription, translation and functional expression of coding genes (Bi et al. 2016).

According to previous descriptions, in addition to SSRs and tandem repeats, there are a large number of dispersed repeat sequences with repeat units larger than 30 bp in the plant mitogenome, which are considered to be long repeat sequences, including forward repeats (F), palindromic repeats (P), reverse repeats (R) and complement repeats (C) (Bi et al. 2016). In this study, 428 dispersed repeats larger than 30 bp were found in the mitogenome of *P. deltoides*, with a total length of 18,863 bp, accounting for 2.35% of the total length of the mitogenome. Among the 428 repeat sequences were 212 forward repeats, 215 palindromic repeats and 1 reverse repeat. No complement repeats were found.

As shown in Fig. 4, most of the dispersed repeat sequences in the mitogenome of *P. deltoides* are between 30 and 39 bp (272, 63.55%). The proportion of repeat sequences in regions of this length was significantly higher than in regions of other lengths, consistent with the findings in mitochondria of many plants that have been studied (Bi et al. 2016, 2020; Yang et al. 2021). Twenty-three repetitive sequences were longer than 100 bp, and only one was longer than 500 bp (550 bp, Chr1). Compared with other species

of Salicaceae (*P. davidiana*, *P. tremula*, *Salix suchowensis*, *S. brachista*), the number of dispersed repeat sequences in *Populus* was significantly higher than in *Salix*, and the number of repeat sequences was well conserved within the same genus. Further analysis showed that the proportion of repeat sequences of most different length segments was very similar among Salicaceae species. The dispersed repetitive sequences in Salicaceae are generally less than 500 bp long, and the only ones larger than 500 bp are a 550-bp repetitive sequence in *P. deltoides* and a 15,592-bp sequence in *S. suchowensis* (Ye et al. 2017). Such long repeat sequences have been found in several species. These long repeats may be associated with subgenomic conformations or isomers in plant mitogenomes (Dong et al. 2018; Bi et al. 2020) and deserve attention for their impact on the structure of the plant mitogenome and on expansion of the mitogenome size in plants (Chang et al. 2013; Dong et al. 2018).

Prediction of RNA editing sites

Previous studies have shown that RNA editing events are among the necessary steps to regulate gene expression in plant growth and development and are also a common phenomenon in chloroplasts and mitochondria (Bock and Khan 2004; Chen et al. 2011; Raman and Park 2015). In this study, 330 RNA editing sites of 34 PCGs in *P. deltoides* were predicted by the PREP-MT program, and analysis showed that
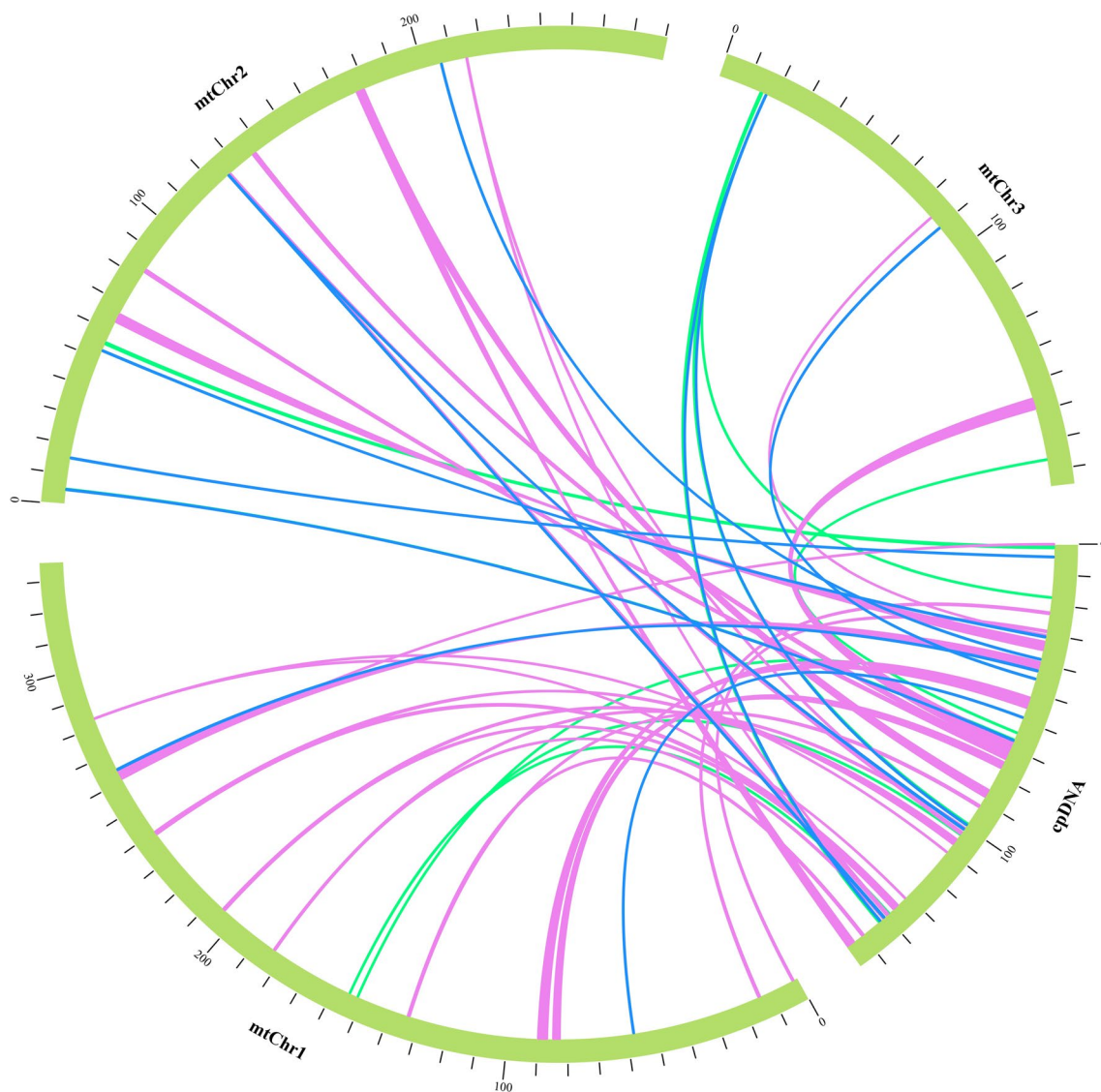
Fig. 6 Collinearity of the mitogenome and chloroplast genome of *Populus deltoides*. In the purple region, identity is greater than 95%, in the green between 90 and 95%, and in the blue region between 85 and 90%

all were C-U edits and occurred in the first (119, 36.06%) and second (207, 62.73%) positions of the codon, whereas no RNA editing was found in the third position, similar to the case in most angiosperms (Grewe et al. 2014; Kovar et al. 2018; Bi et al. 2020; Li et al. 2021; Yang et al. 2021). Interestingly, four codons in the PCGs of *P. deltoides* mitochondria underwent C-U editing (CCT→TTT) at both the first and second positions, which has not been observed in other Salicaceae species. RNA editing can change the start and stop codons of PCGs; for example, the start codon of *mttB* in *P. deltoides* is ATT, which may be caused by RNA editing, but this was not predicted by the PREP-MT program. Some studies have pointed out that the failure of PREP-MT to predict the third codon for RNA editing may result from the algorithmic rules of PREP-MT. The main

limitation of PREP-MT is that it is unable to recognize silent editing sites. Because most of the editing of the third codon does not change the amino acids encoded by the protein, PREP-MT cannot always predict the outcome (Mower 2009). Therefore, the identification of this particular case requires relevant biological experiments to obtain accurate results.

RNA editing sites were predicted for three Salicaceae species (*P. deltoides*, *P. tremula*, *S. suchowensis*) and two gymnosperms (*G. biloba*, *P. taeda*) with 330, 327, 328, 1291 and 1159 sites, respectively. As shown in Fig. 5, the number of RNA editing sites varied greatly among different genes in the *P. deltoides* mitogenome, with the most RNA editing sites occurring in the Complex I genes and

the cytochrome *c* biosynthesis gene, while no RNA editing was found in *atp1*, *atp8*, *atp9*, *cox1*, *cox2*, or *rps1*.

According to the predicted number of RNA editing sites, the number of RNA editing sites in Salicaceae is extremely conserved, while the number of RNA editing sites in *G. biloba* and *P. taeda* is much larger than in *Populus* and *Salix* species. Analysis of RNA editing of PCGs common to the five species revealed that genes (*atp1*, *atp8*, *cox1*, *cox2*, etc.) with no RNA editing events in *Populus* and *Salix* had a large number of RNA editing sites in both *G. biloba* and *P. taeda*. Further studies revealed that the number of RNA editing loci in angiosperms was mostly within 500 and concentrated between 300 and 500, while in gymnosperms, it was significantly

greater than that in angiosperms. In addition to 1291 RNA editing sites in *G. biloba* and 1159 in *P. taeda*, past studies also predicted 1084 RNA editing sites in the mitochondrial PCGs of *C. taitungensis* (Chaw et al. 2008; Bi et al. 2020), The reason for this phenomenon may be that as land plants diverged from early stages and more-advanced angiosperms evolved, angiosperms lost editing sites due to random mutations and reverse transcription of genes, generating a more stable genome structure to provide intrinsic conditions for adaptation and resistance to various environmental changes (Rudinger et al. 2012).

In addition, we found that most of the RNA-edited amino acids (251, 76.06%) were hydrophobic, and this hydrophobicity is the main force driving protein-folding; thus, such
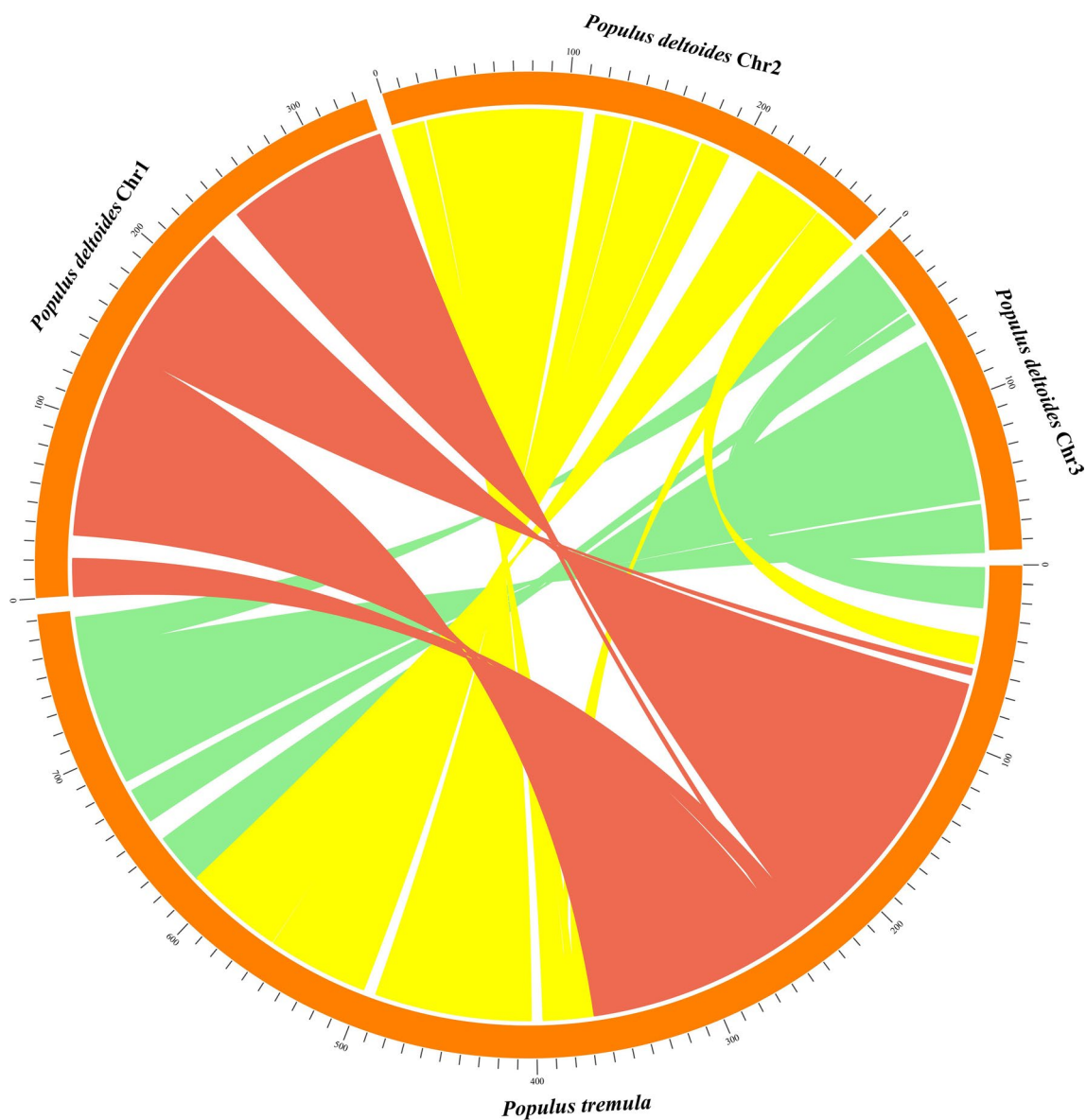
Fig. 7 Collinearity of the mitogenomes of *Populus deltoides* and *P. tremula*
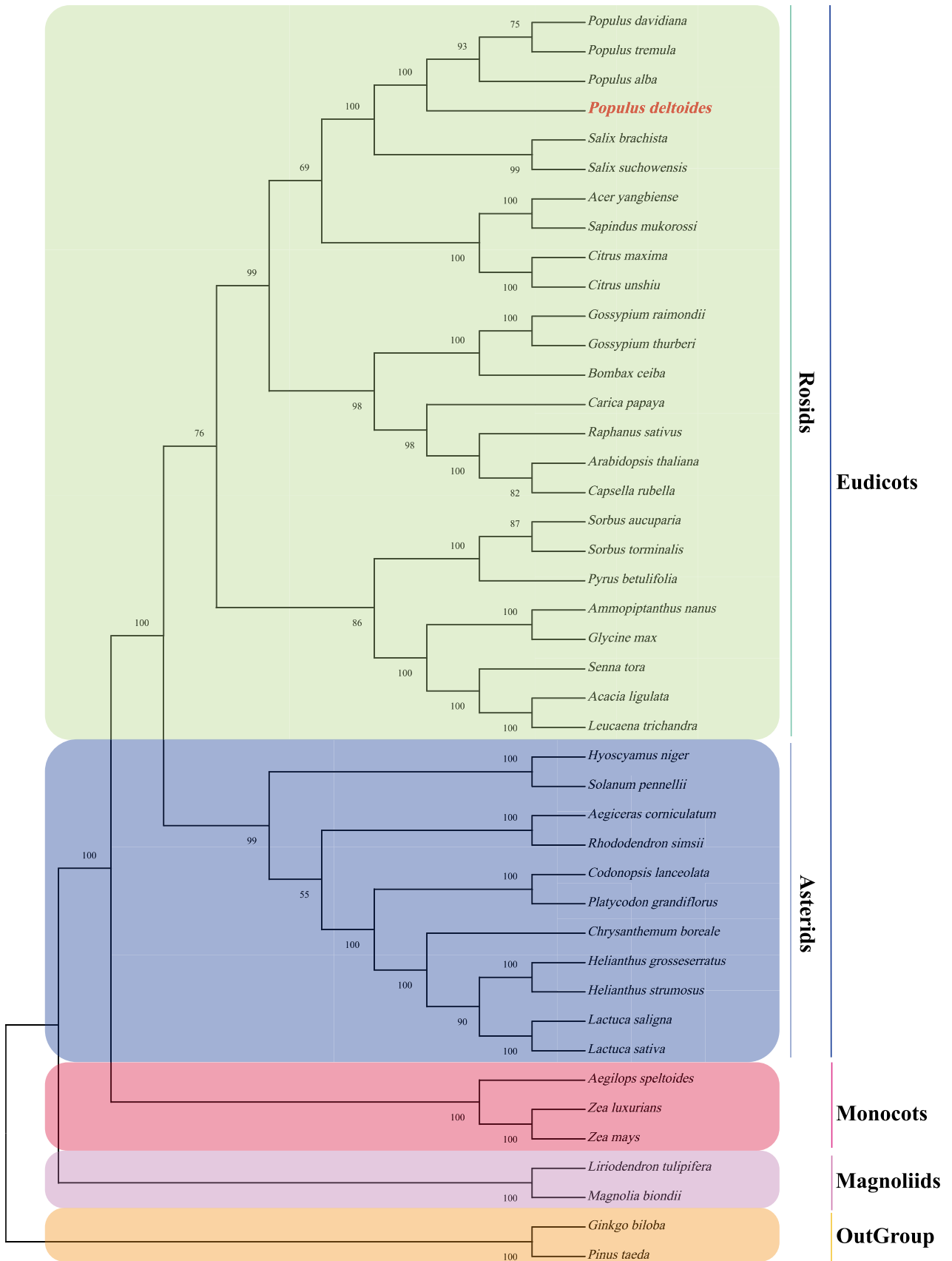
RNA editing enables better protein function and thus better gene expression (Galtier 2011; Brenner et al. 2019).

Sequence transfer and collinearity between different genomes

The horizontal transfer of DNA fragments between organellar genomes and the nuclear genome is an important event in plant evolution. Current genome data and experimental studies show that the main direction of sequence transfer between the three genomes is from the organellar genomes to the nuclear genome and from the nuclear genome and chloroplast genome to the mitogenome. In higher plants, sequences of the nuclear genome and mitogenome have not been found to be transferred to the chloroplast genome (Bergthorsson et al. 2003; Leister 2005; Martin et al. 1998).

To better understand the horizontal sequence transfer events in *P. deltoides*, we used the mitogenome as a query and compared it to the nuclear genome and the chloroplast genome separately. The results showed that 830 fragments, totaling 471 kb of nuclear genome sequences, were transferred to the mitogenome in *P. deltoides*. Further analysis showed that each chromosome in the *P. deltoides* nuclear genome had sequences transferred to the mitogenome, with chromosome 18 having the longest transfer sequence (77,364 bp), much longer than those of the other chromosomes; the smallest transfer sequence was on chromosome 14, only 6241 bp. In addition, statistical analysis of the lengths of the transferred fragments showed that the fragments transferred from the nuclear genome to the mitochondria were mainly in the range of 200–500 bp (674, 81.20%), similar to the findings for other angiosperm mitogenomes (Bi et al. 2016). In addition, four fragments were greater than 10,000 bp, with the longest being 20,520 bp. The gene sequences were aligned with those of the *P. deltoides* nuclear genome, which shares 59,788 bp of gene sequences with the mitogenome, including one complete gene sequence (*trnN-GUU*) and partial sequences of 25 genes (*atp1, atp4, atp6, atp8, nad1, nad2, nad4, nad4L, nad5, nad7, cox1, cox2, cox3, cob, ccmFc, sdh4, matR, rps1, rps3, rps4, rrnL, trnC-GCA, trnY-GUA, trnE-UUC*, and *trnD-GUC*).

As shown in Fig. 6, comparison of the mitochondrial and chloroplast genomes of *P. deltoides* showed that a total of 51 fragments (32.97 kb) were transferred from the chloroplast to the mitogenome, accounting for 21.01% and 4.11% of the chloroplast and mitogenomes, respectively, as detailed in Supplementary Table S3. According to the annotation results of chloroplasts and mitochondria of *P. deltoides*, 27 intact genes were transferred to mitochondria in the chloroplast genome. In addition to 14 PCGs, 13 tRNA genes in the chloroplast of *P. deltoides* had migrated into the mitogenome. Past studies have shown that tRNAs in plant mitochondria have a dual origin; i.e., one part of tRNA is inherited from the ancestor of the mitochondria, and the other part is derived from horizontal gene transfer from chloroplasts (Sprinzl and Vassilenko 2005). There are 18 different tRNAs in the mitogenome of *P. deltoides*, and eight are from chloroplast gene transfer, accounting for 44.44%, similar to the values in *G. raimondii*, *Boea hygrometrica* and some other species (Zhang et al. 2011; Bi et al. 2016). Further comparison with the gymnosperms *G. biloba* and *C. taitungensis* showed that less than 20% of tRNAs in the chloroplast genomes of *G. biloba* and *C. taitungensis* migrated to mitochondria, significantly less than for many angiosperms that have been studied. No chloroplast-to-mitochondria tRNA gene transfer was found in *Marchantia polymorpha*, the first plant to assemble mitochondria (Oda et al. 1992; Chaw et al. 2008; Guo et al. 2016). This phenomenon suggests that the transfer of tRNAs from chloroplasts to mitochondria is becoming more frequent as plants evolve from lower to higher levels, probably because lower plants have a complete set of tRNA genes, but as plants evolve, their own tRNAs in the mitochondria of higher plants are being lost and replaced by homologous tRNAs of chloroplast origin to meet the needs of both chloroplast and mitochondrial protein synthesis (Glover et al. 2001).

To further explore the similarities and differences between the multiring structure and single-ring structure of the mitogenome of *Populus*, the mitogenomes of *P. deltoides* and *P. tremula* (783.44 kb) were compared and subjected to collinear analysis. Figure 7 shows that almost all fragments of the mitogenome of *P. tremula* mapped to the three circular chromosomes of *P. deltoides*. However, the distribution order and direction of many large collinear fragments are different. Interestingly, the collinear analysis of *P. tremula* and *P. davidiana* showed that the two species had a good collinear relationship, and the sequence fragments were basically the same in order and direction (Supplementary Fig. S1), perhaps because *P. deltoides* and *P. tremula* do not belong to the same section and the mitogenome of poplar appears to be extensively rearranged between sections (Yang et al. 2016; Cole et al. 2018).

Phylogenetic analysis

Because the plant mitogenome also carries relevant genetic information, it is increasingly used for the study of phylogeny. With the improvement of sequencing and assembly technology, increasing attention has been given to the phylogenetic analysis of plants using mitochondrial information (Bi et al. 2016; Choi et al. 2019; Yang et al. 2021).

To determine the phylogenetic location of *P. deltoides*, 42 mitogenome data sets from among eudicots, monocots, magnoliids and gymnosperms, were obtained from the NCBI genome database (Supplementary Table S2). Based on 20 single-copy homologous genes shared by 43 species (*atp1, atp4, atp6, atp8, atp9, ccmB, ccmC, ccmFc, ccmFn, cob, cox1, cox2, cox3, nad2, nad3, nad4, nad4L, nad6, nad7,* and *nad9*), a maximum likelihood (ML) tree was constructed with gymnosperms *G. biloba* and *P. taeda* as outgroups. As shown in Fig. 8, most of the 40 branch nodes in the phylogenetic tree have bootstrap values above 90% (34, 85%), and 26 of them have bootstrap values of 100%. It can be seen from the ML tree that the mitochondrial PCGs can perfectly distinguish rosids, asterids, monocots, magnoliids and gymnosperms and classify each order, family and genus clearly, which is consistent with the results of the APG IV classification system (Byng et al. 2016) (APG 2016). Among the phylogenetic trees, the genera *Populus* and *Salix* under the family Salicaceae are distinguished by a 100% bootstrapping rate. Further study found that within the genus *Populus*, *P. davidiana,* and *P. tremula* clustered together but are more closely related to *P. alba*, while *P. deltoides* is in a separate branch. The reason for this is that *Populus* is widely distributed and has many species. *P. davidiana*, *P. tremula,* and *P. alba* all belong to *Populus* sect. *Populus*, so the genetic relationship is relatively close, while *P. deltoides* belongs to *Populus* sect. *Aegiros*, so it is more distantly related to the other three species, which is well reflected in the evolutionary tree. The above results also confirm the reliability of the ML tree based on plant mitochondrial PCGs, and with the rapid development of plant mitogenome sequencing and assembly technologies, extensive evolutionary research on plants through mitochondria will become possible.

Conclusion

We assembled and annotated the mitogenome of *P. deltoides* I-69 and systematically analyzed it based on the obtained genome and annotation information. The *P. deltoides* I-69 mitogenome consists of three circular chromosome structures with a genome size of 802,637 bp, which is similar to the mitogenomes of other poplars that have been studied. The 58 genes were distributed on three chromosomes, and the number of genes on each chromosome was positively correlated with the size of the chromosome. RNA editing and codon usage analysis showed that the mitogenome of poplar was relatively conserved. Further analysis of *Arabidopsis thaliana*, *Gossypium raimondii* and other angiosperms and gymnosperms, such as *Ginkgo biloba* and *Cycas taitungensis*, showed significantly fewer RNA editing sites in the mitogenomes of poplar and other angiosperms than in those of gymnosperms, such as *Ginkgo biloba*. Phylogenetic

analysis of the mitogenome of *P. deltoides* and 42 other species based on 20 common PCGs showed that the evolutionary status of this species could be resolved well, and the section of *Populus* was verified at the genome level. Sequence transfer and collinear analyses showed extensive rearrangements among different sections in *Populus*, which provided new clues and evidence for the study of *Populus* phylogeny. In addition, tRNA translocation analysis revealed that as plants evolved from lower to higher levels, tRNAs in chloroplasts were transferred to mitochondria more frequently. This study is the first to reveal the multichromosomal conformation of mitochondria in *Populus*, elucidating more details on the phylogenetic relationships of *Populus* species and on genetic information for the study of *Populus* species.

References

Allen JO, Fauron CM, Minx P, Roark L, Oddiraju S, Lin GN, Meyer L, Sun H, Kim K, Wang C, Du F, Xu D, Gibson M, Cifrese J, Clifton SW, Newton KJ (2007) Comparisons among two fertile and three male-sterile mitochondrial genomes of maize. Genetics 177(2):1173–1192. https://doi.org/10.1534/genetics.107.073312

Alverson AJ, Wei X, Rice DW, Stern DB, Barry K, Palmer JD (2010) Insights into the evolution of mitochondrial genome size from complete sequences of *Citrullus lanatus* and *Cucurbita pepo* (Cucurbitaceae). Mol Biol Evol 27(6):1436–1448. https://doi.org/10.1093/molbev/msq029

Alverson AJ, Rice DW, Dickinson S, Barry K, Palmer JD (2011) Origins and recombination of the bacterial-sized multichromosomal mitochondrial genome of cucumber. Plant Cell 23(7):2499–2513. https://doi.org/10.1105/tpc.111.087189

Arseneau JR, Steeves R, Laflamme M (2017) Modified low-salt CTAB extraction of high-quality DNA from contaminant-rich tissues. Mol Ecol Resour 17(4):686–693. https://doi.org/10.1111/1755-0998.12616

Bai S, Wu H, Zhang J, Pan Z, Zhao W, Li Z, Tong C (2021) Genome assembly of Salicaceae *Populus deltoides* (Eastern Cottonwood) I-69 based on nanopore sequencing and Hi-C technologies. J Hered 112(3):303–310. https://doi.org/10.1093/jhered/esab010

Beier S, Thiel T, Munch T, Scholz U, Mascher M (2017) MISA-web: a web server for microsatellite prediction. Bioinformatics 33(16):2583–2585. https://doi.org/10.1093/bioinformatics/btx198

Bergthorsson U, Adams KL, Thomason B, Palmer JD (2003) Widespread horizontal transfer of mitochondrial genes in flowering plants. Nature 424(6945):197–201. https://doi.org/10.1038/nature01743

Bernt M, Donath A, Juhling F, Externbrink F, Florentz C, Fritzsch G, Putz J, Middendorf M, Stadler PF (2013) MITOS: improved de novo metazoan mitochondrial genome annotation. Mol Phylogenet Evol 69(2):313–319. https://doi.org/10.1016/j.ympev.2012.08.023

Bi C, Paterson AH, Wang X, Xu Y, Wu D, Qu Y, Jiang A, Ye Q, Ye N (2016) Analysis of the complete mitochondrial genome sequence of the diploid cotton *Gossypium raimondii* by comparative genomics approaches. Biomed Res Int 2016:5040598. https://doi.org/10.1155/2016/5040598

Bi C, Lu N, Xu Y, He C, Lu Z (2020) Characterization and analysis of the mitochondrial genome of common bean (*Phaseolus vulgaris*) by comparative genomic approaches. Int J Mol Sci. https://doi.org/10.3390/ijms21113778

Bock R, Khan MS (2004) Taming plastids for a green future. Trends Biotechnol 22(6):311–318. https://doi.org/10.1016/j.tibtech.2004.03.005

Bock H, Brennicke A, Schuster W (1994) Rps3 and rpl16 genes do not overlap in Oenothera mitochondria: GTG as a potential translation initiation codon in plant mitochondria? Plant Mol Biol 24(5):811–818. https://doi.org/10.1007/BF00029863

Brenner WG, Mader M, Muller NA, Hoenicka H, Schroeder H, Zorn I, Fladung M, Kersten B (2019) High level of conservation of mitochondrial RNA editing sites among four *Populus* species. G3 Bethesda 9(3):709–717. https://doi.org/10.1534/g3.118.200763

Byng JW, Chase MW, Christenhusz MJM, Fay MF, Judd WS, Mabberley DJ, Sennikov AN, Soltis DE, Soltis PS, Stevens PF, Briggs B, Brockington S, Chautems A, Clark JC, Conran J, Haston E, Moller M, Moore M, Olmstead R, Perret M, Skog L, Smith J, Tank D, Vorontsova M, Weber A, Grp AP (2016) An update of the angiosperm phylogeny group classification for the orders and families of flowering plants: APG IV. Bot J Linn Soc 181(1):1–20. https://doi.org/10.1111/boj.12385

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL (2009) BLAST+: architecture and applications. BMC Bioinform 10:421. https://doi.org/10.1186/1471-2105-10-421

Chan PP, Lowe TM (2019) tRNAscan-SE: Searching for tRNA genes in genomic sequences. Methods Mol Biol 1962:1–14. https://doi.org/10.1007/978-1-4939-9173-0_1

Chan PP, Lin BY, Mak AJ, Lowe TM (2021) tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. Nucleic Acids Res 49(16):9077–9096. https://doi.org/10.1093/nar/gkab688

Chang S, Wang Y, Lu J, Gai J, Li J, Chu P, Guan R, Zhao T (2013) The mitochondrial genome of soybean reveals complex genome structures and gene evolution at intercellular and phylogenetic levels. PLoS ONE 8(2):e56502. https://doi.org/10.1371/journal.pone.0056502

Chaw SM, Shih AC, Wang D, Wu YW, Liu SM, Chou TY (2008) The mitochondrial genome of the gymnosperm *Cycas taitungensis* contains a novel family of short interspersed elements, Bpu sequences, and abundant RNA editing sites. Mol Biol Evol 25(3):603–615. https://doi.org/10.1093/molbev/msn009

Chen H, Deng L, Jiang Y, Lu P, Yu J (2011) RNA editing sites exist in protein-coding genes in the chloroplast genome of *Cycas taitungensis*. J Integr Plant Biol 53(12):961–970. https://doi.org/10.1111/j.1744-7909.2011.01082.x

Chen C, Chen H, Zhang Y, Thomas HR, Frank MH, He Y, Xia R (2020a) TBtools: An integrative toolkit developed for interactive analyses of big biological data. Mol Plant 13(8):1194–1202. https://doi.org/10.1016/j.molp.2020.06.009

Chen X, Zhang L, Huang Y, Zhao F (2020b) Mitochondrial genome of *Salix cardiophylla* and its implications for infrageneric division of the genus of *Salix*. Mitochondrial DNA B Resour 5(3):3485–3486. https://doi.org/10.1080/23802359.2020.1827065

Choi IS, Schwarz EN, Ruhlman TA, Khiyami MA, Sabir JSM, Hajarah NH, Sabir MJ, Rabah SO, Jansen RK (2019) Fluctuations in Fabaceae mitochondrial genome size and content are both ancient and recent. BMC Plant Biol 19(1):448. https://doi.org/10.1186/s12870-019-2064-8

Cole LW, Guo W, Mower JP, Palmer JD (2018) High and variable rates of repeat-mediated mitochondrial genome rearrangement in a genus of plants. Mol Biol Evol 35(11):2773–2785. https://doi.org/10.1093/molbev/msy176

De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C (2018) NanoPack: visualizing and processing long-read sequencing data. Bioinformatics 34(15):2666–2669. https://doi.org/10.1093/bioinformatics/bty149

Dong FG, Wilson KG, Makaroff CA (1998) The radish (Raphanus sativus L.) mitochondrial cox2 gene contains an ACG at the predicted translation initiation site. Curr Genet 34(2):79–87. https://doi.org/10.1007/s002940050369

Dong S, Zhao C, Chen F, Liu Y, Zhang S, Wu H, Zhang L, Liu Y (2018) The complete mitochondrial genome of the early flowering plant *Nymphaea colorata* is highly repetitive with low recombination. BMC Genomics 19(1):614. https://doi.org/10.1186/s12864-018-4991-4

Edgar RC (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5:113. https://doi.org/10.1186/1471-2105-5-113

Fahrenkrog AM, Neves LG, Resende MFR Jr, Dervinis C, Davenport R, Barbazuk WB, Kirst M (2017) Population genomics of the eastern cottonwood (*Populus deltoides*). Ecol Evol 7(22):9426–9440. https://doi.org/10.1002/ece3.3466

Galtier N (2011) The intriguing evolutionary dynamics of plant mitochondrial DNA. BMC Biol 9:61. https://doi.org/10.1186/1741-7007-9-61

Glover KE, Spencer DF, Gray MW (2001) Identification and structural characterization of nucleus-encoded transfer RNAs imported into wheat mitochondria. J Biol Chem 276(1):639–648. https://doi.org/10.1074/jbc.M007708200

Goremykin VV, Salamini F, Velasco R, Viola R (2009) Mitochondrial DNA of *Vitis vinifera* and the issue of rampant horizontal gene transfer. Mol Biol Evol 26(1):99–110. https://doi.org/10.1093/molbev/msn226

Gray MW, Covello PS (1993) RNA editing in plant mitochondria and chloroplasts. FASEB J 7(1):64–71. https://doi.org/10.1096/fasebj.7.1.8422976

Greiner S, Lehwark P, Bock R (2019) OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. Nucleic Acids Res 47(W1):W59–W64. https://doi.org/10.1093/nar/gkz238

Grewe F, Edger PP, Keren I, Sultan L, Pires JC, Ostersetzer-Biran O, Mower JP (2014) Comparative analysis of 11 brassicales mitochondrial genomes and the mitochondrial transcriptome of

Brassica oleracea. Mitochondrion 19:135–143. https://doi.org/10.1016/j.mito.2014.05.008

Guo W, Grewe F, Fan W, Young GJ, Knoop V, Palmer JD, Mower JP (2016) Ginkgo and welwitschia mitogenomes reveal extreme contrasts in gymnosperm mitochondrial evolution. Mol Biol Evol 33(6):1448–1460. https://doi.org/10.1093/molbev/msw024

Gupta RS, Golding GB (1996) The origin of the eukaryotic cell. Trends Biochem Sci 21(5):166–171

Handa H (2003) The complete nucleotide sequence and RNA editing content of the mitochondrial genome of rapeseed (Brassica napus L.): comparative analysis of the mitochondrial genomes of rapeseed and Arabidopsis thaliana. Nucleic Acids Res 31(20):5907–5916. https://doi.org/10.1093/nar/gkg795

Jo YD, Park J, Kim J, Song W, Hur CG, Lee YH, Kang BC (2011) Complete sequencing and comparative analyses of the pepper (Capsicum annuum L.) plastome revealed high frequency of tandem repeats and large insertion/deletions on pepper plastome. Plant Cell Rep 30(2):217–229. https://doi.org/10.1007/s00299-010-0929-2

Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods 14(6):587–589. https://doi.org/10.1038/nmeth.4285

Kersten B, Faivre Rampant P, Mader M, Le Paslier MC, Bounon R, Berard A, Vettori C, Schroeder H, Leple JC, Fladung M (2016) Genome sequences of *Populus tremula* chloroplast and mitochondrion: implications for holistic poplar breeding. PLoS ONE 11(1):e0147209. https://doi.org/10.1371/journal.pone.0147209

Kovar L, Nageswara-Rao M, Ortega-Rodriguez S, Dugas DV, Straub S, Cronn R, Strickler SR, Hughes CE, Hanley KA, Rodriguez DN, Langhorst BW, Dimalanta ET, Bailey CD (2018) PacBio-based mitochondrial genome assembly of *Leucaena trichandra* (Leguminosae) and an intrageneric assessment of mitochondrial RNA editing. Genome Biol Evol 10(9):2501–2517. https://doi.org/10.1093/gbe/evy179

Kozik A, Rowan BA, Lavelle D, Berke L, Schranz ME, Michelmore RW, Christensen AC (2019) The alternative reality of plant mitochondrial DNA: One ring does not rule them all. PLoS Genet 15(8):e1008373. https://doi.org/10.1371/journal.pgen.1008373

Kubo T, Nishizawa S, Sugawara A, Itchoda N, Estiati A, Mikami T (2000) The complete nucleotide sequence of the mitochondrial genome of sugar beet (Beta vulgaris L.) reveals a novel gene for tRNA(Cys)(GCA). Nucleic Acids Res 28(13):2571–2576. https://doi.org/10.1093/nar/28.13.2571

Kumar S, Stecher G, Tamura K (2016) MEGA7: Molecular evolutionary genetics analysis Version 70 for bigger datasets. Mol Biol Evol 33(7):1870–1874. https://doi.org/10.1093/molbev/msw054

Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: Molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol 35(6):1547–1549. https://doi.org/10.1093/molbev/msy096

Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R (2001) REPuter: the manifold applications of repeat analysis on a genomic scale. Nucleic Acids Res 29(22):4633–4642. https://doi.org/10.1093/nar/29.22.4633

Lagesen K, Hallin P, Rodland EA, Staerfeldt HH, Rognes T, Ussery DW (2007) RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res 35(9):3100–3108. https://doi.org/10.1093/nar/gkm160

Leister D (2005) Origin, evolution and genetic effects of nuclear insertions of organelle DNA. Trends Genet 21(12):655–663. https://doi.org/10.1016/j.tig.2005.09.004

Li J, Ye C (2020) Genome-wide analysis of microsatellite and sex-linked marker identification in Gleditsia sinensis. BMC Plant Biol 20(1):338. https://doi.org/10.1186/s12870-020-02551-9

Li J, Xu Y, Shan Y, Pei X, Yong S, Liu C, Yu J (2021) Assembly of the complete mitochondrial genome of an endemic plant, *Scutellaria tsinyunensis*, revealed the existence of two conformations generated by a repeat-mediated recombination. Planta 254(2):36. https://doi.org/10.1007/s00425-021-03684-3

Ma Q, Li S, Bi C, Hao Z, Sun C, Ye N (2017) Complete chloroplast genome sequence of a major economic species, *Ziziphus jujuba* (Rhamnaceae). Curr Genet 63(1):117–129. https://doi.org/10.1007/s00294-016-0612-4

Malek O, Brennicke A, Knoop V (1997) Evolution of trans-splicing plant mitochondrial introns in pre-Permian times. Proc Natl Acad Sci U S A 94(2):553–558. https://doi.org/10.1073/pnas.94.2.553

Martin W, Stoebe B, Goremykin V, Hapsmann S, Hasegawa M, Kowallik KV (1998) Gene transfer to the nucleus and the evolution of chloroplasts. Nature 393(6681):162–165. https://doi.org/10.1038/30234

Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R (2020) IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol 37(5):1530–1534. https://doi.org/10.1093/molbev/msaa015

Mower JP (2009) The PREP suite: predictive RNA editors for plant mitochondrial genes, chloroplast genes and user-defined alignments. Nucleic Acids Res 37:W253-259. https://doi.org/10.1093/nar/gkp337

Oda K, Kohchi T, Ohyama K (1992) Mitochondrial DNA of *Marchantia polymorpha* as a single circular form with no incorporation of foreign DNA. Biosci Biotechnol Biochem 56(1):132–135. https://doi.org/10.1271/bbb.56.132

Patel RK, Jain M (2012) NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. PLoS ONE 7(2):e30619. https://doi.org/10.1371/journal.pone.0030619

Petersen G, Cuenca A, Moller IM, Seberg O (2015) Massive gene loss in mistletoe (Viscum, Viscaceae) mitochondria. Sci Rep 5:17588. https://doi.org/10.1038/srep17588

Raman G, Park S (2015) Analysis of the complete chloroplast genome of a medicinal plant, Dianthus superbus var. longicalyncinus, from a comparative genomics perspective. PLoS ONE 10(10):e0141329. https://doi.org/10.1371/journal.pone.0141329

Rice DW, Alverson AJ, Richardson AO, Young GJ, Sanchez-Puerta MV, Munzinger J, Barry K, Boore JL, Zhang Y, dePamphilis CW, Knox EB, Palmer JD (2013) Horizontal transfer of entire genomes via mitochondrial fusion in the angiosperm Amborella. Science 342(6165):1468–1473. https://doi.org/10.1126/science.1246275

Ruan J, Li H (2020) Fast and accurate long-read assembly with wtdbg2. Nat Methods 17(2):155–158. https://doi.org/10.1038/s41592-019-0669-3

Rudinger M, Volkmar U, Lenz H, Groth-Malonek M, Knoop V (2012) Nuclear DYW-type PPR gene families diversify with increasing RNA editing frequencies in liverwort and moss mitochondria. J Mol Evol 74(1–2):37–51. https://doi.org/10.1007/s00239-012-9486-3

Sau K, Gupta SK, Sau S, Mandal SC, Ghosh TC (2006) Factors influencing synonymous codon and amino acid usage biases in Mimivirus. Biosystems 85(2):107–113. https://doi.org/10.1016/j.biosystems.2005.12.004

Skippington E, Barkman TJ, Rice DW, Palmer JD (2015) Miniaturized mitogenome of the parasitic plant Viscum scurruloideum is extremely divergent and dynamic and has lost all nad genes. Proc Natl Acad Sci U S A 112(27):E3515-3524. https://doi.org/10.1073/pnas.1504491112

Sloan DB, Alverson AJ, Chuckalovcak JP, Wu M, McCauley DE, Palmer JD, Taylor DR (2012) Rapid evolution of enormous, multichromosomal genomes in flowering plant mitochondria with

exceptionally high mutation rates. PLoS Biol 10(1):e1001241. https://doi.org/10.1371/journal.pbio.1001241

Sprinzl M, Vassilenko KS (2005) Compilation of tRNA sequences and sequences of tRNA genes. Nucleic Acids Res 33:D139-140. https://doi.org/10.1093/nar/gki012

Tanaka Y, Tsuda M, Yasumoto K, Terachi T, Yamagishi H (2014) The complete mitochondrial genome sequence of *Brassica oleracea* and analysis of coexisting mitotypes. Curr Genet 60(4):277–284. https://doi.org/10.1007/s00294-014-0433-2

Tillich M, Lehwark P, Pellizzer T, Ulbricht-Jones ES, Fischer A, Bock R, Greiner S (2017) GeSeq - versatile and accurate annotation of organelle genomes. Nucleic Acids Res 45(W1):W6-W11. https://doi.org/10.1093/nar/gkx391

Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM (2014) Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS ONE 9(11):e112963. https://doi.org/10.1371/journal.pone.0112963

Wang X, Zhang RG, Yun QZ, Xu YY, Zhao GC, Liu JM, Shi SL, Chen Z, Jia LM (2021) Comprehensive analysis of complete mitochondrial genome of Sapindus mukorossi Gaertn.: an important industrial oil tree species in China. Ind Crops Prod. https://doi.org/10.1016/j.indcrop.2021.114210

Wynn EL, Christensen AC (2019) Repeats of unusual size in plant mitochondrial genomes: Identification, incidence and evolution. G3 (bethesda) 9(2):549–559. https://doi.org/10.1534/g3.118.200948

Xu C, Dong W, Li W, Lu Y, Xie X, Jin X, Shi J, He K, Suo Z (2017) Comparative analysis of six *lagerstroemia* complete chloroplast genomes. Front Plant Sci 8:15. https://doi.org/10.3389/fpls.2017.00015

Yang J, Liu G, Zhao N, Chen S, Liu D, Ma W, Hu Z, Zhang M (2016) Comparative mitochondrial genome analysis reveals the evolutionary rearrangement mechanism in Brassica. Plant Biol (stuttg) 18(3):527–536. https://doi.org/10.1111/plb.12414

Yang H, Li W, Yu X, Zhang X, Zhang Z, Liu Y, Wang W, Tian X (2021) Insights into molecular structure, genome evolution and phylogenetic implication through mitochondrial genome sequence of *Gleditsia sinensis*. Sci Rep 11(1):14850. https://doi.org/10.1038/s41598-021-93480-6

Ye N, Wang X, Li J, Bi C, Xu Y, Wu D, Ye Q (2017) Assembly and comparative analysis of complete mitochondrial genome sequence of an economic plant *Salix suchowensis*. PeerJ 5:e3148. https://doi.org/10.7717/peerj.3148

Zhang T, Zhang X, Hu S, Yu J (2011) An efficient procedure for plant organellar genome assembly, based on whole genome data from the 454 GS FLX sequencing platform. Plant Methods 7:38. https://doi.org/10.1186/1746-4811-7-38

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