RESEARCH ARTICLE

Assembly and comparative analysis of the complete mitochondrial genome of three *Macadamia* species (*M. integrifolia, M. ternifolia* and *M. tetraphylla*)

Yingfeng Niu[1☯], Yongjie Lu[2☯], Weicai Song[2☯], Xiyong He[1], Ziyan Liu[1], Cheng Zheng[1], Shuo Wang[2], Chao Shi[2]*, Jin Liu[1]*

1 Yunnan Institute of Tropical Crops, Xishuangbanna, China, 2 Qingdao University of Science & Technology, Qingdao, China

☯ These authors contributed equally to this work.
* chsh1111@aliyun.com (CS); liujin2416@163.com (JL)

Abstract

Background

Macadamia is a true dicotyledonous plant that thrives in a mild, humid, low wind environment. It is cultivated and traded internationally due to its high-quality nuts thus, has significant development prospects and scientific research value. However, information on the genetic resources of *Macadamia* spp. remains scanty.

Results

The mitochondria (mt) genomes of three economically important *Macadamia* species, *Macadamia integrifolia*, *M. ternifolia* and *M. tetraphylla*, were assembled through the Illumina sequencing platform. The results showed that each species has 71 genes, including 42 protein-coding genes, 26 tRNAs, and 3 rRNAs. Repeated sequence analysis, RNA editing site prediction, and analysis of genes migrating from chloroplast (cp) to mt were performed in the mt genomes of the three *Macadamia* species. Phylogenetic analysis based on the mt genome of the three *Macadamia* species and 35 other species was conducted to reveal the evolution and taxonomic status of *Macadamia*. Furthermore, the characteristics of the plant mt genome, including genome size and GC content, were studied through comparison with 36 other plant species. The final non-synonymous (Ka) and synonymous (Ks) substitution analysis showed that most of the protein-coding genes in the mt genome underwent negative selections, indicating their importance in the mt genome.

Conclusion

The findings of this study provide a better understanding of the *Macadamia* genome and will inform future research on the genus.

1. Introduction

Macadamia spp belongs in the family Proteaceae, class *Magnoliopsida*, and order *Proteales*. The Proteaceae family has five subfamilies, 80 genera, and over 1600 species [1, 2]. Most of them are distributed in Oceania and South Africa, while a few are produced in East Asia and South America. Notably, more than 100 species in the Proteaceae family produce flowers that are traded internationally [3]. Besides, the species grown in the northeastern part of Oceania are also rich in nuts. The genus *Macadamia* comprises four species: *Macadamia integrifolia*, *M. jansenii*, *M. ternifolia*, and *M. tetraphylla*. These species are naturally distributed in the subtropical rain forests from southeastern Queensland in Australia to northeastern New South Wales [4, 5]. Among them, *M. integrifolia* and *M. tetraphylla* produce edible nuts; thus, most commercial cultivars are either these two species or their hybrids. The other two species, *M. Jansenii* and *M. ternifolia* produce non-edible nuts containing high levels of bitter cyanide glycosides, thus has not been used to guide the breeding [6, 7]. *Macadamia* seeds are sweet with high nutritional and medicinal value. Therefore, they have enjoyed the reputation of "King of Thousand Fruits". They are also used in international transactions due to their high economic value [8].

Mitochondria (mt) are organelles that primarily convert biomass energy in living cells into chemical energy to fuel biological activities [9]. Additionally, they participate in other biological processes, including cell differentiation, cell apoptosis, cell growth, and cell division [10–13]. Therefore, mt are central to life activities within individual cells and the entire living body [14]. Both plastids and mt harbor genetic information and are thought to have evolved through endosymbiosis of freely living bacteria [15–17]. In most seed plants, nuclear genetic information is inherited from both parents, while cp and mt are derived from maternal genes [18]. Thus, we can temporarily ignore the influence of paternal genes, thereby reducing the difficulty of genetic research and promoting the research of genetic mechanisms [19].

Studies have shown that the size of the mt genome varies significantly between different species. For example, plants have a larger mt genome than animals [20]. Furthermore, mt genome size in seed plants can vary by at least one order of magnitude ranging from ~ 222 bp in *Brassica napus* [21] and ~ 316 Kb in *Allium cepa* [22] to ~ 3.9 Mb in *Amborella trichopoda* [23] and a striking ~ 11.3 Mb in *Silene conica* [24]. This phenomenon may be caused by the abundance of non-coding regions and repeated elements in the plant mt genome [25]. DNA recombination between homologous sequences produces small circular sub-genomic DNA. The circular genomic DNA coexists with the complete "master" genome in the cell. These genomes typically have several kb repeats, leading to multiple heterogeneous forms of the genome [26–31]. The mutation rate of plant mt genomes is very low; however, their rearrangement rate is so high that there is almost no conservation of synteny [32–34].

The development of cost-effective and more efficient DNA sequencing methods like high-throughput sequencing has accelerated mt genome sequencing. So far (until June 2021), the mt genomes of 618 green plant species have been released in the NCBI (https://www.ncbi.nlm.nih.gov/) database. Long-term mutually beneficial symbiosis caused the mt to lose some of the original DNA, possibly by transfer, leaving only the DNA encoding it [35, 36]. Mt DNA integrates DNA from various sources by intracellular and horizontal transfer [37]. Therefore, regardless of the length, gene sequence and content, mt genome varies remarkably among different plant species [33]. The mt genome length of the smallest terrestrial plant is about 66 Kb, and that of the largest terrestrial plant is 11.3 Mb [24, 38, 39]; the number of genes is usually between 32 and 67 [40]. In this study, the mt genomes of three *Macadamia* species were sequenced, assembled, and annotated. Also, their genomic and structural features were analyzed and compared with other angiosperms (and gymnosperms). This study improves our

understanding of *Macadamia* genetics and provides crucial data to inform future research on the evolution of mt genomes of land plants.

2. Materials and methods

2.1 Genome sequencing

The three *Macadamia* species examined in this study were collected from Yunnan Institute of Tropical Crops (Xishuangbanna, China; 101˚28' E, 21˚92' N). Total genomic DNA was extracted from fresh leaves using modified CTAB [41]. Meanwhile, the quantity and quality of extracted DNA was assessed by spectrophotometry and the integrity was evaluated using a 1% (w/v) agarose gel electrophoresis. The qualified DNA samples were used for Illumian DNA library construction, according to the standard procedure. Subsequently, a paired-end sequencing library with an insert size of 350 bp was constructed. The Illumina Hiseq 4000 high-throughput sequencing platform was used for sequencing. The sequencing strategy involved PE150 (Pair-End 150) and the sequencing data volume of not less than 1 Gb. Illumina high-throughput sequencing results initially existing as original image data files were con-verted into Raw Reads. CASAVA software was used for Base Calling.

2.2 Genome assembly and annotation

SPAdes v.3.5.0 [42] software was used to splice and assemble mt genome sequences. To correct the splicing results, the raw sequencing data were mapped to mitochondrial sequences using Geneious software [43]. DOGMA [44] and NCBI were used to annotate the mt genome. The Blastn and Blastp method was used to compare mt gene-encoding protein and rRNAs among related species. TRNA scan-SE2.0 [45] and ARWEN [46] were used to annotate tRNA. The tRNAs with unreasonable length and incomplete structure were eliminated. Subsequently, a tRNA secondary structure diagram was generated. The final mt genomes of *M. integrifolia*, *M. ternifolia*, and *M. tetraphylla* have been deposited in the GenBank (Accession number: MW566570/MW566571/MW566572).

2.3 Analysis of repeat structure and sequence

Microsatellites within the mt genomes of the three *Macadamia* species were identified using MISA [47, 48]. The minimum number of repeats for the motif length of 1, 2, 3, 4, 5, and 6 were 10, 6, 5, 4, 3, and 3, respectively, were identified in this analysis. The tandem repeats were detected using Tandem Repeats Finder v4.09 software [49] with default parameters.

2.4 DNA transformation from cp to mt and RNA editing analyses

The cp genome of *M. integrifolia* (NC_025288) was downloaded from the NCBI database. Chloroplast-like sequences were identified and the genome was mapped using TBtools [50]. The online program Predictive RNA Editor for Plants (PREP) suite [51] was adopted to iden-tify the possible RNA editing sites in the protein-coding genes of the three *Macadamia* species. The cutoff value was set as 0.2 to ensure accurate prediction. The protein-coding genes from other plant mt genomes were used as references to reveal the RNA editing sites in the mt genomes of the three *Macadamia* species.

2.5 Phylogenetic tree construction and Ka/Ks analysis

The genome sequences of the three *Macadamia* species were compared with those of 35 (S1 Table) other plant species to further verify their phylogenetic position. Notably, the complete mt genome sequences of these species were available in the NCBI database. Phylogenetic

analyses were performed on 23 conserved protein-coding genes (*atp1*, *atp4*, *atp6*, *atp8*, *atp9*, *ccmB*, *ccmC*, *ccmFc*, *ccmFn*, *cob*, *cox1*, *cox2*, *cox3*, *matR*, *nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5*, *nad6*, *nad7* and *nad9*) that were extracted from the mt genomes of the 35 plant species using TBtools [51]. These conserved genes were then aligned using Muscle [52] implemented in MEGA X [53]; the alignment was modified manually to eliminate gaps and missing data. The GTR + G + I model was determined to be the best model based on the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) calculated by ModelFinder [54]. The Maximum Likelihood (ML) algorithm in MEGA X [53] was used to construct a phylogenetic tree. The bootstrap consensus tree was inferred from 1000 replications. *Cycas taitungensis* and *Ginkgo biloba* were designated as the outgroup in this analysis.

The Ka and Ks replacement rates of protein-coding genes in mitochondrial genomes of the three *Macadamia* species and other higher plants were analyzed. blastn in TBtools was used to extract the sequences of corresponding protein-coding genes in *Macadamia* and *N. nucifera* genomes. The Ka and Ks replacement rates of each protein-coding gene were estimated using *N. nucifera* genome as a reference.

3. Results and discussion

3.1 Genomic features of the mt genomes of the three *Macadamia* species

The mt genomes of *M. integrifolia*, *M. ternifolia* and *M. tetraphylla* have a typical terrestrial plant genome ring structure (Fig 1). A total of 71 unique genes were identified in the mt genomes of the three *Macadamia* species, including 42 protein-coding, 26 tRNA, and 3 rRNA genes (Table 1). In addition, two copies of *rRNA26*, *ccmB*, *rps19*, *trnN*-GTT, and *trnH*-GTG, and seven copies of *trnM*-CAT were identified. It has been established that the mt genomes of land plants contain a variable number of introns [55]. In the present study, the three mt genomes had ten genes with introns, length ranging from 13 bp (*rps3*) to 31,841 bp (*cox2*) where *ccmFC*, *rpl2*, *rps3*, and *rps10* had two introns, *cox2* had three, *nad1*, *nad4*, and *nad5* had four and *nad2* and *nad7* had five introns. Besides, in all protein-coding genes, except *atp6*, *cox1*, *nad1*, *nad4L*, *rps4*, and *rps10*, which had ACG as the start codon, all the others had ATG as their start codon. In addition, the stop codons in all the protein-coding genes were: TAA 45.2%, TGA 28.6%, TAG 14.3%, CAA 9.5%, and CGA 2.4%.

The size and GC content of mt genome are the primary characteristics. Here, we compared the size and GC content of mt genomes between three *Macadamia* species and 36 other green plants, including four phorophytes, three bryophytes, two gymnosperms, four monocots, and 23 dicots (S1 Table). The size of the mt genomes ranged from 22,897 bp (*Chlamydomonas moewusii*) to 2,709,526 bp (*Cucumis melo*) (Fig 2). Compared to phorophytes and bryophytes, the mt genomes of the three *Macadamia* species are larger. The GC content in the mt genomes was also highly variable, ranging from 32.24% in *Sphagnum palustric* to 50.36% in *Ginkgo biloba*. Overall, the GC content of angiosperm mt genome (including monocots and dicots) is higher than that in bryophytes but less than in gymnosperms [56, 57], implying that the GC contents fluctuated following the angiosperms divergence from bryophytes and gymnosperms. Interestingly, the GC content significantly fluctuated in algae and was mostly conserved in angiosperms, although their genome sizes vary significantly.

3.2 Repeat sequences analysis

Microsatellites or simple sequence repetitions (SSRs) are DNA fragments composed of short sequence repeating units of 1–6 base pairs [58]. Their unique value is created by their polymorphism, relative abundance, codominant inheritance, large-scale genome coverage, and PCR detection simplicity [59]. Based on the SSRs analysis, we identified 87 SSRs with SSRs
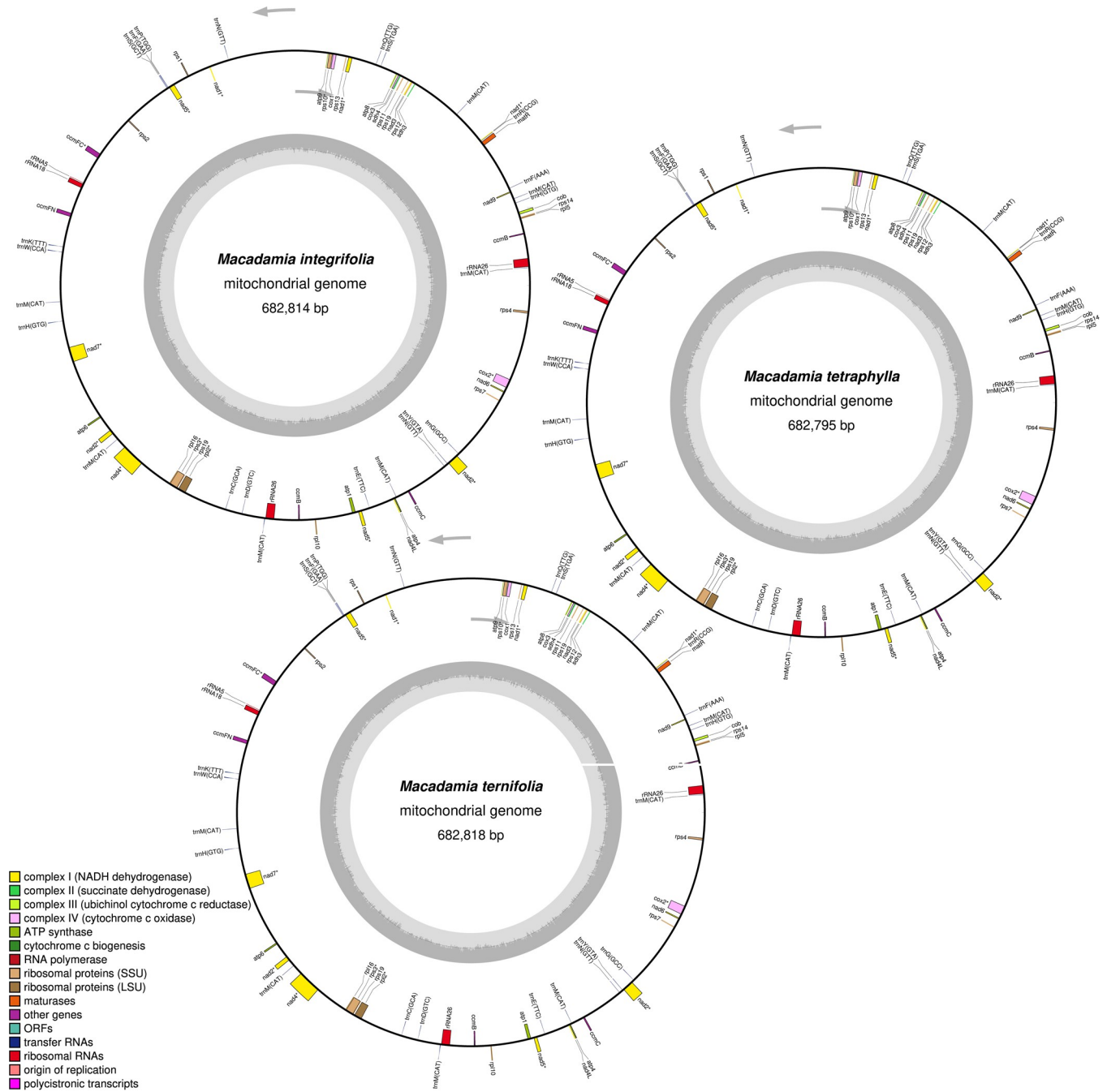
Fig 1. The circular map of three *Macadamia species* mitochondrial genome. Gene map showing 71 annotated genes of different functional groups.

https://doi.org/10.1371/journal.pone.0263545.g001

monomers and dimers accounting for 70.11% of the total SSRs. Adenine (A) was the most repeated monomer with 19 (38%) out of the 50 identified monomer SSRs. The AT repeat was the most common dimer SSR, accounting for 66.67% of all the identified dimers. However, one hexamer [ATTAGG(X3)] was present in the mt genomes of three *Macadamia* species.

Among the reference mt genome, only *Nelumbo nucifera* has been published in the NCBI database. *N. nucifera* belongs to the family Nelumbonaceae and the same order (Proteales) with *Macadamia*. Therefore, the mt genome of *N. nucifera* was used as a reference for

Table 1. Gene profile and organization of three *Macadamia* species (*M. integrifolia*, *M. ternifolia* and *M. tetraphylla*).

Group of genes	Gene/element	Size(bp)	GC_Percent	AminoAcids (aa)	InferredInitiation Codon	Inferred TerminationCodon
ATP synthase	atp1	1530	45.29%	509	ATG	TGA
	atp4	597	43.05%	198	ATG	TAG
	atp6	783	39.08%	260	ACG	TAA
	atp8	480	40.63%	159	ATG	TAA
	atp9	225	46.67%	74	ATG	CAA
Cytochrome c biogenesis	ccmB(2)	621,621	42.83%	206	ATG	TGA
	ccmC	771	44.23%	256	ATG	TAA
	ccmFCa	1356	46.53%	451	ATG	TAA
	ccmFN	1734	47.58%	577	ATG	TGA
Ubichinol cytochrome c reductase	cob	1182	42.39%	393	ATG	TGA
Cytochrome c oxidase	cox1	1584	44.26%	527	ACG	TAA
	cox2a	822	42.34%	273	ATG	TAG
	cox3	798	45.11%	265	ATG	TGA
Maturases	matR	1968	52.64%	655	ATG	TAG
NADH dehydrogenase	nad1a	978	44.99%	325	ACG	TAA
	nad2a	1467	40.90%	488	ATG	TAA
	nad3	357	41.74%	118	ATG	TAA
	nad4a	1488	42.67%	495	ATG	TGA
	nad4L	303	37.29%	100	ACG	TAA
	nad5a	1989	41.78%	662	ATG	TAA
	nad6	630	40.95%	209	ATG	TGA
	nad7a	1185	45.23%	394	ATG	TAG
	nad9	573	42.93%	190	ATG	TAA
Ribosomal proteins (LSU)	rpl2a	999	52.15%	332	ATG	CAA
	rpl5	561	44.74%	186	ATG	TAA
	rpl10	516	46.32%	171	ATG	TAA
	rpl16	492	43.09%	163	ATG	TAA
Ribosomal proteins (SSU)	rps1	606	43.56%	201	ATG	TAA
	rps2	648	39.20%	215	ATG	CAA
	rps3a	1692	43.91%	563	ATG	TAG
	rps4	1059	40.51%	352	ACG	TAA
	rps7	447	43.18%	148	ATG	TAA
	rps10a	333	39.04%	110	ACG	CGA
	rps11	444	45.27%	147	ATG	CAA
	rps12	378	45.50%	125	ATG	TGA
	rps13	351	39.60%	116	ATG	TGA
	rps14	303	40.92%	100	ATG	TAG
	rps19(2)	285,285	40.00%	94	ATG	TAA
Transport membrane protein	sdh3	336	37.20%	111	ATG	TGA
	sdh4	450	41.33%	149	ATG	TGA
Ribosomal RNAs	rrn5	119	52.94%			
	rrn18	2061	55.12%			
	rrn26(2)	3989,3989	53.02%			

(Continued)

Table 1. (Continued)

Group of genes	Gene/element	Size(bp)	GC_Percent	AminoAcids (aa)	InferredInitiation Codon	Inferred TerminationCodon
Transfer RNAs	trnR-CCG	75	57.33%			
	trnN-GTTb(2)	75,72	**49.33%**			
	trnD-GTCb	74	63.51%			
	trnC-GCA	76	52.63%			
	trnQ-TTG	72	47.22%			
	trnE-TTC	72	50.00%			
	trnG-GCC	74	54.05%			
	trnH-GTGb(2)	*75,75*	54.67%			
	trnK-TTT	75	46.67%			
	trnM-CATb(7)	72,75,73,72,77,72,72	59.72%,46.67%,43.84%, 59.72%,44.16%,59.72%,59.72%			
	trnF-AAA	75	49.33%			
	trnF-GAA	74	*47.30%*			
	trnP-TGG	75	54.67%			
	trnS-TGA	*88*	*51.14%*			
	trnS-GCT	91	46.15%			
	trnW-CCAb	74	51.35%			
	trnY-GTA	84	51.19%			

Notes: The numbers after the gene names indicate the duplication number. Lowercase a indicates the genes containing introns, and lowercase b indicates the chloroplast-derived genes.

https://doi.org/10.1371/journal.pone.0263545.t001

comparative analysis in the present study. The monomers in *N. nucifera* were lower than in the three *Macadamia* species, while pentamers and hexamers in *N. nucifera* were significantly higher than in the three *Macadamia* species (Fig 3A). Moreover, the SSRs in mt genomes of *M. integrifolia*, *M. ternifolia*, *M. tetraphylla*, and *N. nucifera* were mainly single-nucleotide A/T motifs, and dimer AT/TA motifs. Within the *Macadamia* genus, the mt SSRs among the
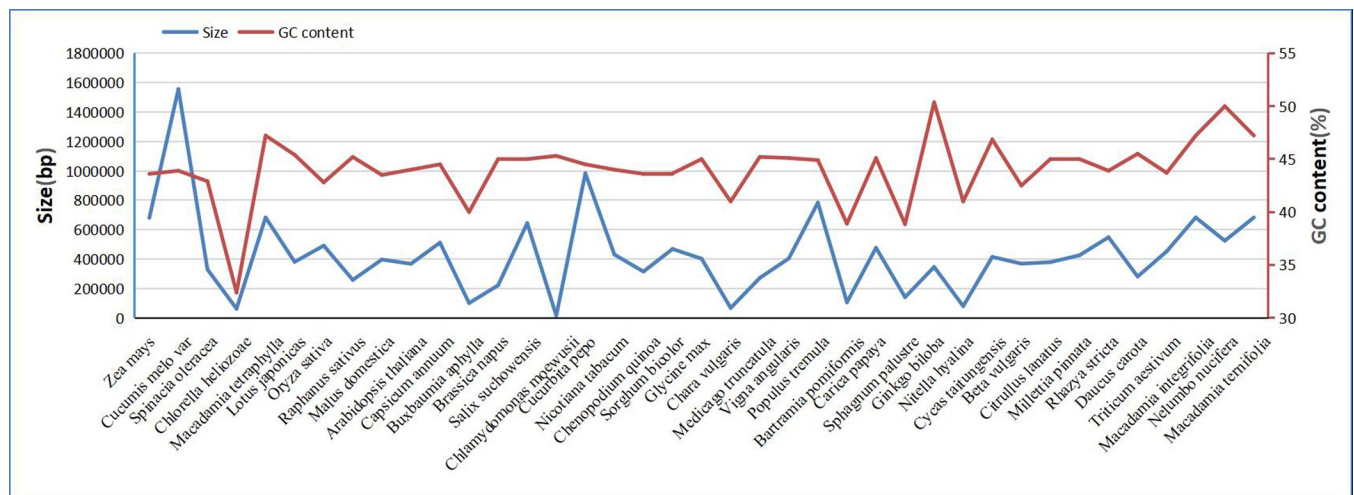

Fig 2. The sizes and GC contents of 39 plant mitochondrial genomes. The blue dots represent the genome size and the orange trend line shows the variation of GC content across the different taxa.
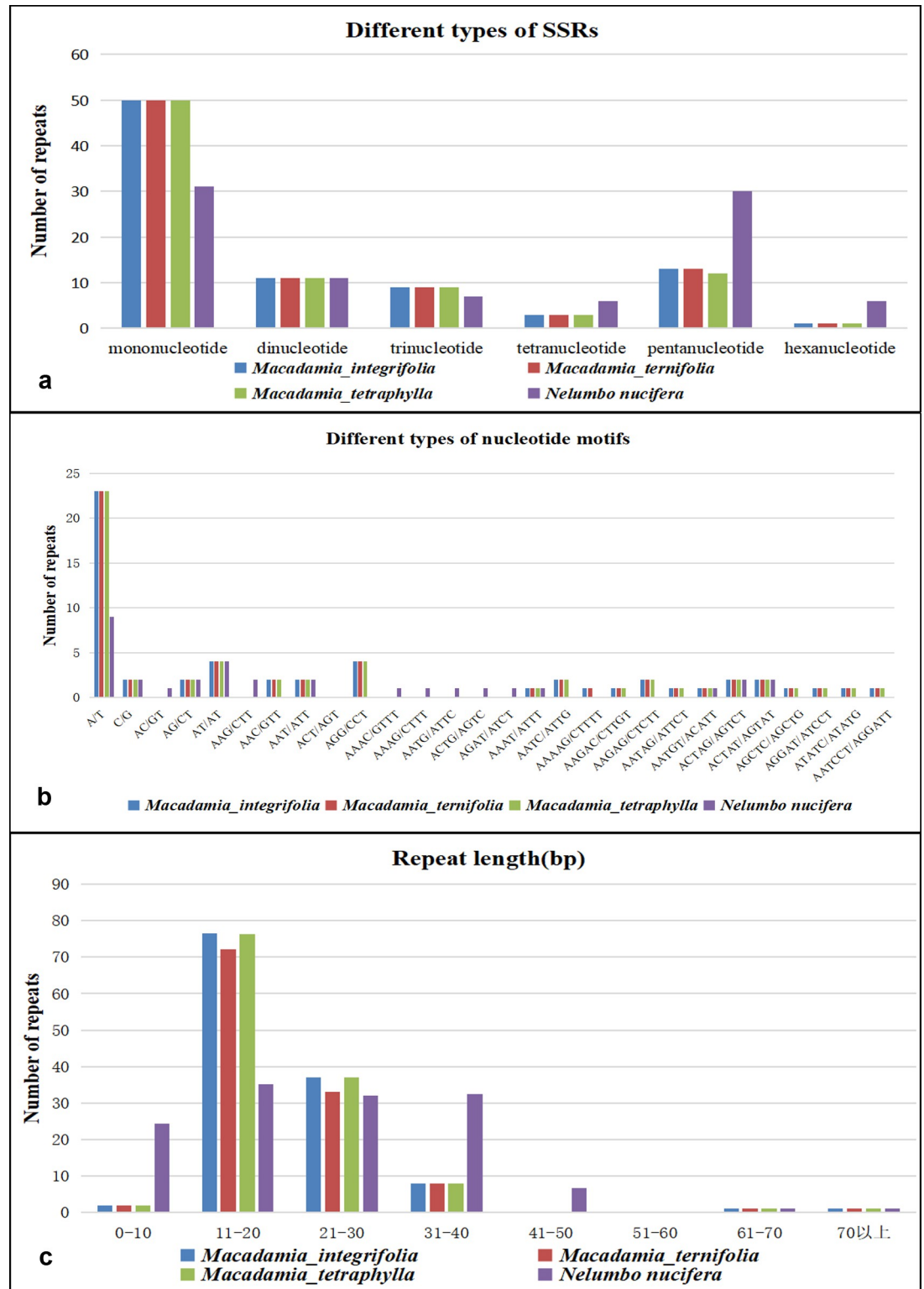
https://doi.org/10.1371/journal.pone.0263545.g002

Fig 3. The comparison of microsatellites and oligonucleotide repeats in three *Macadamia* species and *N. nucifera* mitochondrial genomes.

https://doi.org/10.1371/journal.pone.0263545.g003

different species are highly similar (Fig 3B). However, compared with *N. nucifera*, there were both differences and similarities. For example, the single nucleotide A/T in the three *Macadamia* species has 23-unit repeats, while *N. nucifera* has only nine. Nevertheless, their single-nucleotide C/G numbers were the same (two-unit repeats) (Fig 3B). In addition, the AG/CT and AT/AT motifs unit repetitions are the same, although *N. nucifera* also has an AC/GT motif, lacking in the three *Macadamia* species. Interestingly, the pentanucleotide AATGT/ACATT, ACTAG/AGTCT, and ACATT/AGTAT also had the same number of repetitions in the three *Macadamia* species and *N. nucifera*. Overall, the greater the nucleotide motif, the greater the difference between the three *Macadamia* species and *N. nucifera*.

Core repeating units ranging from 1 to 200 bases (tandem repeats) are widely present in eukaryotes and some prokaryotes genomes [60]. In the present study, 25, 21, and 20 tandem repeats (10–33 bp) were identified in the *M. integrifolia*, *M. ternifolia*, and *M. tetraphylla* with a match greater than 95% (S2–S4 Tables). The tandem repeats (11–20 bp and 21–30 bp) significantly varied among the three *Macadamia* species (Fig 3C), where *M. ternifolia* had the least number of repetitions, while *M. integrifolia* and *M. tetraphylla* had a very similar number of repetitions. However, *N. nucifera* had the least (11–20 bp and 21–30 bp) and had the highest (0–10 bp, 31–40 bp, 41–50 bp) tandem repeated compared to the three *Macadamia* species. Besides, no repetitions ranged from 51–60 bp among the four genomes, while the number of repetitions was the same for 60–70 bp and above.

3.3 The prediction of RNA editing

RNA editing is a post-transcriptional process entailing the addition, deletion, or conversion of bases in the coding region of a transcribed RNA. The conversion of cytosine to uridine is common in cp and mt genomes of plants [61–65], which improves protein preservation in plants. The accurate detection of ribonucleic acid editing is inseparable from the proteomics data. In the present study, we predicted 42 protein-coding genes (including two multi-copy genes: *ccmB* and *rps19*) in the mt genomes of the three *Macadamia* species using the PREP-mt program [51]. The findings revealed that the RNA editing sites were 688, 689, and 688 (Fig 4).
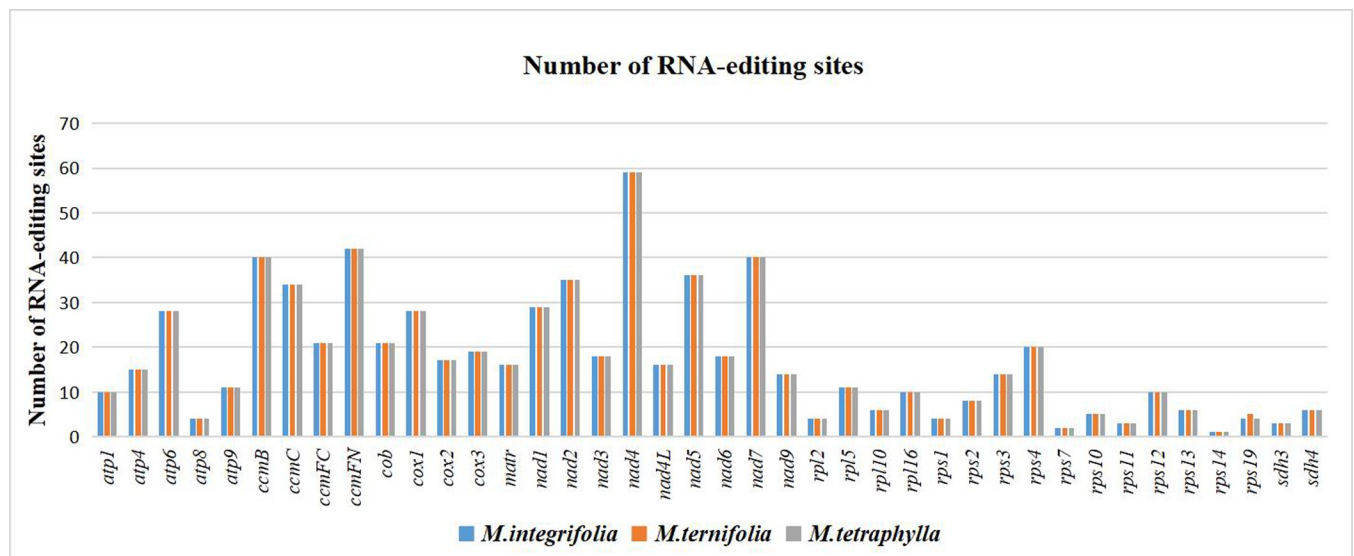

Fig 4. The distribution of RNA-editing sites in the mt protein-coding genes of three species of *Macadamia*. The bars of different colors represent the number of RNA-editing sites of each gene.

Among the protein-coding genes, *nad4* had the most RNA editing sites (59 sites), while *atp8*, *rpl2*, *rpl10*, *rps1*, *rps2*, *rps7*, *rps10*, *rps11*, *rps13*, *rps14*, *rps19*, *sdh3*, and *sdh4* had less than 10 RNA editing sites. 236 RNA editing sites occurred in the first base position of the codon, 472 sites appeared in the second base position, and there was no RNA editing in the third base position. *M. ternifolia* had more than one RNA editing site, unlike the other two *Macadamia* species.

The RNA editing increases the diversity at the start and stop codons in protein-coding genes. However, even with RNA editing, 30.2% (208 positions) of amino acid hydrophobicity and 12.5% (86 positions) of amino acid hydrophilicity remained unchanged in the *M. integrifolia* and *M. tetraphylla* mt genomes. However, 6.7% (46 positions) of amino acids were converted from hydrophobic to hydrophilic, and 47.9% (330 positions) from hydrophilic to hydrophobic. In addition, five amino acids were converted from glutamine to stop codons and two from arginine to stop codons (Table 2). The findings in this study revealed that most

Table 2. Prediction of RNA editing sites.

Type	RNA-editing	Number	Percentage
hydrophobic	GCA (A) = > GTA (V)	1	30.23%
	GCG (A) = > GTG (V)	6	
	GCT (A) = > GTT (V)	4	
	CTC (L) = > TTC (F)	7	
	CTT (L) = > TTT (F)	16	
	CCC (P) = > TTC (F)	6	
	CCT (P) = > TTT (F)	14	
	CCA (P) = > CTA (L)	61	
	CCC (P) = > CTC (L)	14	
	CCG (P) = > CTG (L)	44	
	CCT (P) = > CTT (L)	35	
hydrophilic	CAT (H) = > TAT (Y)	24	12.50%
	CAC (H) = > TAC (Y)	11	
	CGC (R) = > TGC (C)	15	
	CGT (R) = > TGT (C)	36	
hydrophobic-hydrophilic	CCA (P) = > TCA (S)	16	8.28%
	CCC (P) = > TCC (S)	13	
	CCG (P) = > TCG (S)	6	
	CCT (P) = > TCT (S)	22	
hydrophilic-hydrophobic	CGG (R) = > TGG (W)	43	47.97%
	TCC (S) = > TTC (F)	47	
	TCT (S) = > TTT (F)	58	
	TCA (S) = > TTA (L)	101	
	TCG (S) = > TTG (L)	59	
	ACA (T) = > ATA (I)	7	
	ACC (T) = > ATC (I)	1	
	ACG (T) = > ATG (M)	8	
	ACT (T) = > ATT (I)	6	
hydrophilic-stop	CGA (R) = > TGA (X)	2	1.02%
	CAG (Q) = > TAG (X)	1	
	CAA (Q) = > TAA (X)	4	

Notes: Compared with the other two species of *Macadamia*, *M. ternifolia* had only one more RNA-editing site (CTT (L) = >TTT (F)).

amino acids were converted from serine to leucine (23.3%, 160 sites), proline to leucine (22.4%), and serine to phenylalanine (15.3%). The remaining 269 RNA editing sites included other RNA editing types, such as Ala-Val, His-Tyr, Leu-Phe, Pro-Phe, Pro-Ser, Arg-Cys, Arg-Trp, Thr-Ile, Thr- Met, Gln-X, and Arg-X (X = stop codon). Compared to *M. integrifolia* and *M. tetraphylla*, *M. ternifolia* only had one more RNA-edited site (Leu-Phe).

3.4 DNA migration from cp to mt

The cp-like sequences in the mt genome were detected by comparing against the complete cp genome sequence of *M. integrifolia* obtained from the NCBI database (Fig 5). We detected 28 fragments in the mt genome of *M. integrifolia*, ranging in size from 32 bp to 5,210 bp. The cp-like sequence had 36,902 bp, accounting for 5.4% of the mt genome. Five complete annotated tRNA genes were detected, namely *trnH*-GTG, *trnM*-CAT, *trnW*-CCA, *trnD*-GTC, and *trnN*-GTT, with some fragments of *rrn18* genes. The findings also revealed that 28 insertion regions accounted for 23.2% of the cp genome, including seven complete protein-coding genes (*petL*, *petG*, *ndhE*, *rps15*, *rpl23(X2)*, *rpl2*) and eight complete tRNA genes (*trnH*-GUG, *trnD*-GUC, *trnM*-CAU, *trnW*-CCA, *trnP*-UGG, *trnP*-GGG, *trnI*-CAU, *trnN*-GUU). Besides, several protein-coding genes were also identified, including *psbA*, *rpoB*, *psbD*, *psbC*, *ndhC*, *rpl2*, *ycf2(X2)*, *ndhB*, *rps7(X2)*, *ndhD*, *ndhB* and *ycf1*, and some tRNA genes (t*rnI*-GAU, *trnA*-UGC, *trnN*-GUU), which migrated from the cp genome into the mt genome. But, most of these genes lost their integrity during the evolution process, and only their partial sequences were found in the mt genome. Furthermore, most cp-like sequences were located in the spacer region of the mt genome. These findings are consistent with previous research, where during evolution, tRNA genes were more conserved than the protein-coding genes and rRNA genes since they play an important role in mt genome [66].

3.5 Phylogenetic analysis within higher plant mt genomes

Australia is the origin and center of diversity of the Proteaceae, and this is distributed across remnant landmasses of the southern supercontinent Gondwana [67]. The order Proteales inclusive of Proteaceae, Platanaceae and Nelumbonaceae was established relatively recently, on the basis of molecular data, and morphological synapomorphies for the order are yet to be identified [68, 69]. Phylogenetic analysis was performed to understand the evolution of the three *Macadamia* species compared to 29 dicots, four monocots, and two gymnosperms (out-groups). The phylogenetic tree was constructed based on the comparisons in the data matrix of 23 conserved protein-coding genes (Fig 6). The findings revealed that the phylogenetic tree strongly supports the separation of Proteales from rosids and asterids, the separation of eudicots from monocots and angiosperms from gymnosperms. The evolutionary relationships among all the taxa separated into 20 families (Leguminosae, Cucurbitaceae, Apiaceae, Apocynaceae, Solanaceae, Rosaceae, Caricaceae, Brassicaceae, Salicaceae, Bataceae, Malvaceae, Vitaceae, Lamiaceae, Nelumbonaceae, Proteaceae, Butomaceae, Arecaceae, Poaceae, Cycadaceae, and Ginkgoaceae) were efficiently deduced in the phylogenetic tree (Fig 6). The *Macadamia* chloroplast genome confirms the placement of this family with the morphologically divergent Plantanaceae (plane tree family) and Nelumbonaceae (sacred lotus family) in the basal eudicot order Proteales [70]. In addition, Phylogenetic analysis of chloroplast genomic variation revealed a latitudinal population structure of wild *M. integrifolia* germplasm, suggesting long-term regional isolation of maternal lineages [71]. Overall, evolutionary analyses of organelle genomes suggest that Proteaceae are most closely related to Nelumbonaceae.
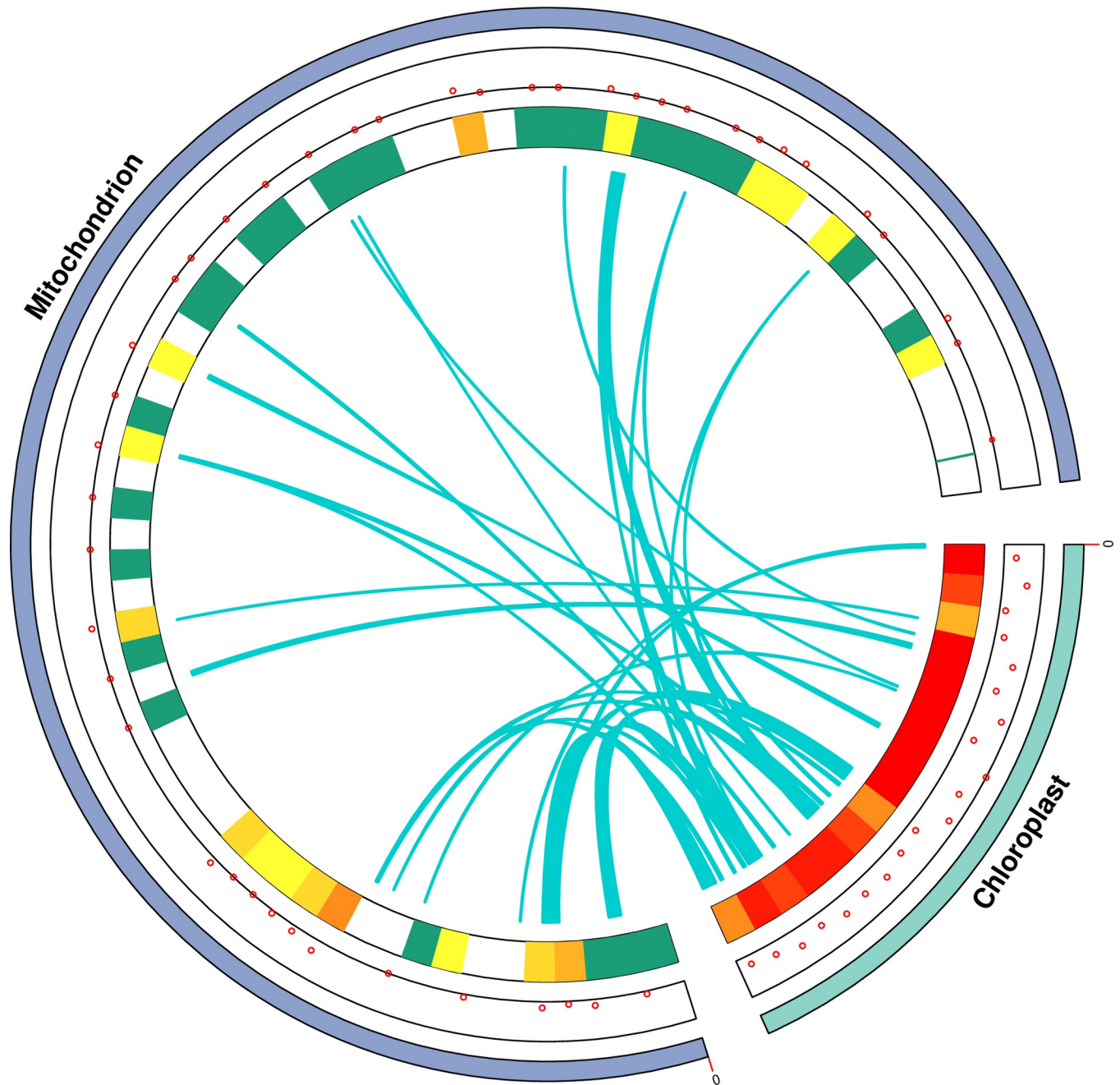
Fig 5. Schematic representation of mitochondrial genome, chloroplast genome and chloroplast-like sequence of *M. integrifolia*. Dots and heat maps inside the two chromosomes show where genes are located. The green lines in the circle show the regions of chloroplast-like sequences inserted from the chloroplast genome into the mt genome.

https://doi.org/10.1371/journal.pone.0263545.g005

3.6 The substitution rates of protein-coding genes

In genetics, non-synonymous (Ka) and synonymous (Ks) substitution rates help understand the evolutionary dynamics of protein-coding genes among similar species since the Ka to Ks ratio indicates gene selection [72, 73]. In the present study, *N. nucifera* was used as a reference
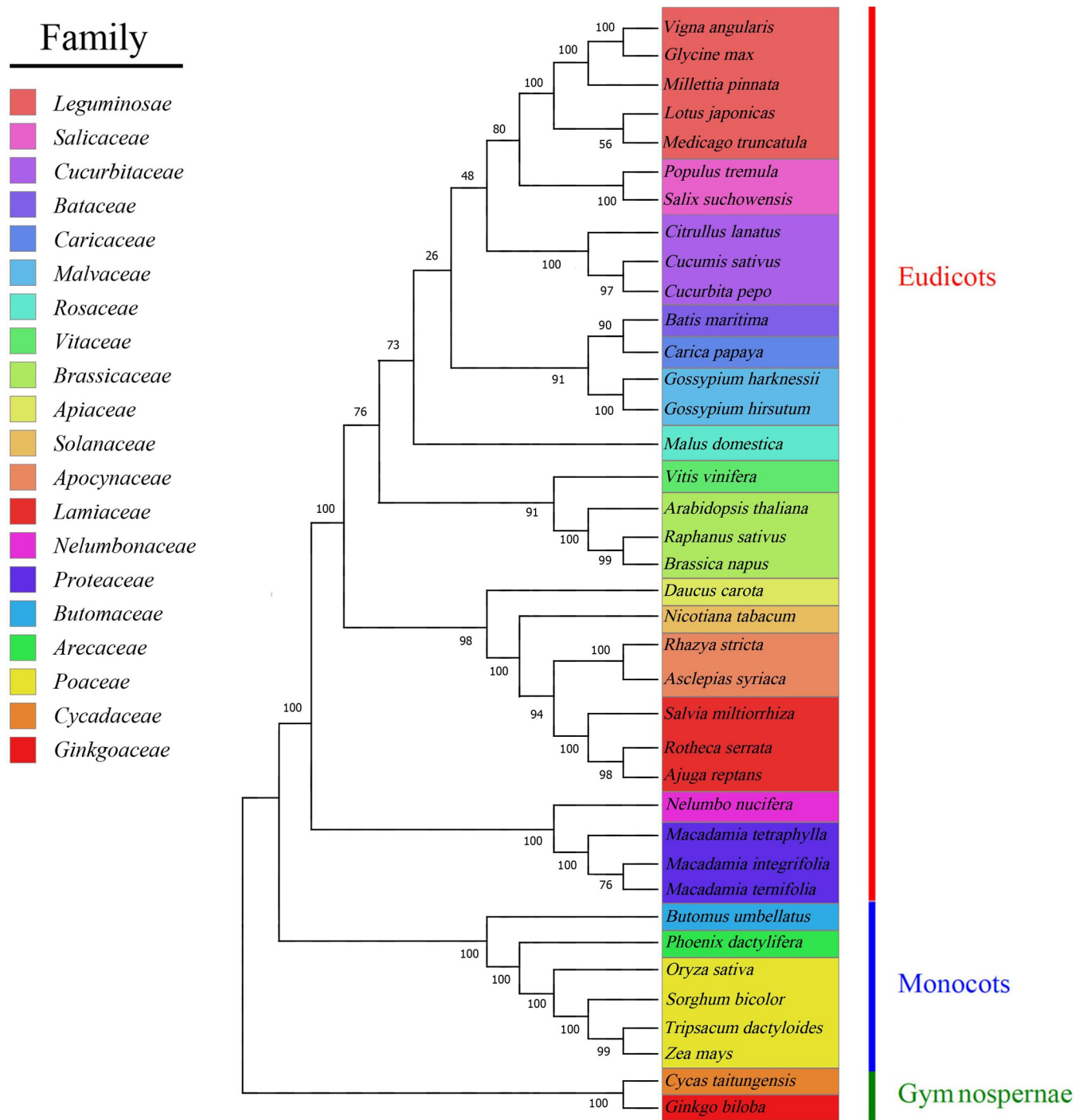
Fig 6. The phylogenetic relationships of three species of *Macadamia* with other 35 plant species. The Maximum Likelihood tree was constructed based on the sequences of 23 conserved protein-coding genes. Colors indicate the families that the specific species belongs.

species to calculate the Ka/Ks ratio of 40 protein-coding genes present in the mt genome of three *Macadamia* species. The Ks of *atp9* and *rps14*, and the Ka of *rps12* was 0. Besides, in most protein-coding genes, the Ka/Ks ratio was significantly less than 1 (Fig 7). However, the
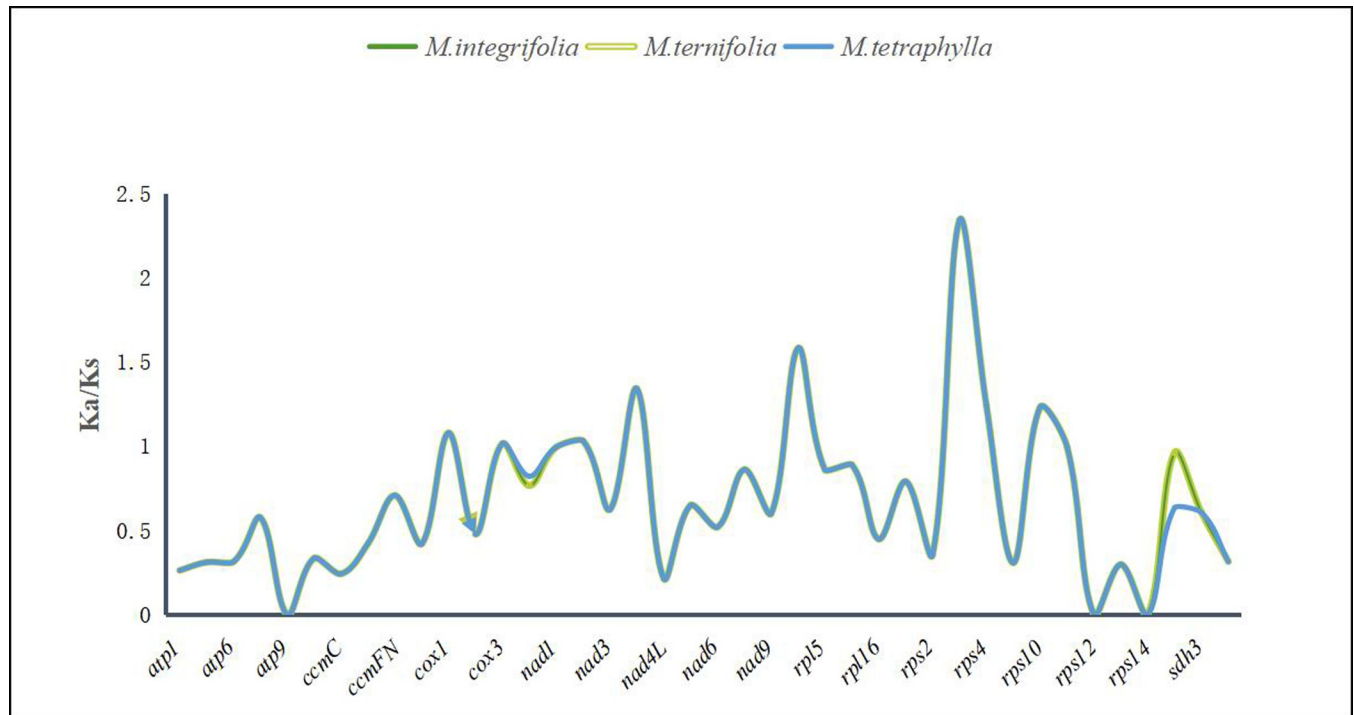
Fig 7. The Ka/Ks values of 40 protein-coding genes of three *Macadamia* species.

https://doi.org/10.1371/journal.pone.0263545.g007

Ka/Ks ratio of *nad4*, *rpl2*, *rps3*, *rps4*, and *rps10* was greater than 1, with the *rps3* ratio being 2.34, implying that these genes might have undergone mutation related positive selection following *Macadamia* and *N. nucifera* differentiation from their last common ancestor [74]. Besides, the ATP synthase, Cytochrome C biogenesis, Ubiquinol Cytochrome C reductase, and Maturases of Ka/Ks ratios were below 1, implying that the negative selection acted on these genes (Table 2). Therefore, these genes may be highly conserved during the evolution of higher plants [75].

4. Conclusions

The complete mt genomes of *M. integrifolia*, *M. ternifolia* and *M. tetraphylla* share many common features with angiosperm mt genomes. In this study, we found that the mt genomes of the three *Macadamia* species were circular like most mt genomes. Compared them with the GC content of the mt genome of 36 other green plants, the results supported the conclusion that the GC content in the *Macadamia* species and angiosperms are highly conserved. In addition, we conducted studies on SSRs and longer tandem repeats in the three sets of data. Besides, 688 RNA editing sites were identified in 42 protein-coding genes, providing important clues for predicting gene function with new codons. By detecting gene migration, we observed 28 fragments (with five complete tRNA genes) were transferred from the cp genome to mt genome. The subsequent phylogenetic analysis results also showed their accuracy in plant classification. Moreover, based on the Ka/Ks substitution of protein-coding genes, most coding genes have undergone negative selection, indicating that the protein-coding genes in the mt genome are conserved in *Macadamia* species. The findings of this study provide information on the mt genome of *Macadamia* species, which is key in understanding the evolutionary history of the family Proteaceae.

Supporting information

S1 Table. The abbreviations and NCBI accession numbers of mt genomes used in this study.
(XLSX)

S2 Table. Perfect tandem repeats in the *Macadamia integrifolia* mitochondrial genome.
(XLSX)

S3 Table. Perfect tandem repeats in the *Macadamia ternifolia* mitochondrial gemone.
(XLSX)

S4 Table. Perfect tandem repeats in the *Macadamia tetraphylla* mitochondrial gemone.
(XLSX)

Author Contributions

Conceptualization: Shuo Wang, Chao Shi, Jin Liu.

Data curation: Yingfeng Niu, Yongjie Lu, Weicai Song, Xiyong He.

Formal analysis: Yingfeng Niu, Yongjie Lu, Weicai Song.

Funding acquisition: Ziyan Liu, Cheng Zheng, Shuo Wang.

Investigation: Yingfeng Niu, Yongjie Lu.

Methodology: Yingfeng Niu, Yongjie Lu, Xiyong He, Ziyan Liu.

Project administration: Xiyong He, Cheng Zheng, Chao Shi, Jin Liu.

Resources: Shuo Wang, Jin Liu.

Software: Yongjie Lu, Weicai Song, Xiyong He, Cheng Zheng, Shuo Wang.

Supervision: Ziyan Liu, Chao Shi, Jin Liu.

Validation: Yingfeng Niu, Xiyong He, Ziyan Liu, Cheng Zheng, Chao Shi.

Visualization: Weicai Song.

Writing – original draft: Yingfeng Niu, Yongjie Lu.

Writing – review & editing: Yingfeng Niu, Yongjie Lu.

References

1. Bai SH, Brooks P, Gama R, Nevenimo T, Hannet G, Hannet D, et al. Nutritional quality of almond, canarium, cashew and pistachio and their oil photooxidative stability. J Food Sci Technol. 2019; 56. https://doi.org/10.1007/s13197-018-3539-6 PMID: 30906037

2. De Souza RGM, Schincaglia RM, Pimente GD, Mota JF. Nuts and human health outcomes: A systematic review. Nutrients. 2017. https://doi.org/10.3390/nu9121311 PMID: 29207471

3. Lin J, Zhang W, Zhang X, Ma X, Zhang S, Chen S, et al. Signatures of selection in recently domesticated *Macadamia*. Nat Commun. 2022; 13. https://doi.org/10.1038/s41467-021-27937-7 PMID: 35017544

4. Hardner CM, Peace C, Lowe AJ, Neal J, Pisanu P, Powell M, et al. Genetic Resources and Domestication of *Macadamia*. Horticultural Reviews. 2009. https://doi.org/10.1002/9780470593776.ch1

5. Topp BL, Nock CJ, Hardner CM, Alam M, O'Connor KM. *Macadamia* (*Macadamia* spp.) breeding. Advances in Plant Breeding Strategies: Nut and Beverage Crops. 2020. https://doi.org/10.1007/978-3-030-23112-5_7

6. Dahler JM, McConchie CA, Turnbull CGN. Quantification of cyanogenic glycosides in seedlings of three *Macadamia* (Proteaceae) species. Aust J Bot. 1995; 43. https://doi.org/10.1071/BT9950619

7. Gross CL, Weston PH. *Macadamia jansenii* (Proteaceae), a new species from central queensland. Aust Syst Bot. 1992; 5. https://doi.org/10.1071/SB9920725

8. Taylor PJ, Grass I, Alberts AJ, Joubert E, Tscharntke T. Economic value of bat predation services–A review and new estimates from *Macadamia* orchards. Ecosyst Serv. 2018; 30. https://doi.org/10.1016/j.ecoser.2017.11.015

9. Newton KJ. Plant Mitochondrial Genomes: Organization, Expression and Variation. Annu Rev Plant Physiol Plant Mol Biol. 1988; 39. https://doi.org/10.1146/annurev.pp.39.060188.002443

10. Bonora M, De Marchi E, Patergnani S, Suski JM, Celsi F, Bononi A, et al. Tumor necrosis factor-α impairs oligodendroglial differentiation through a mitochondria-dependent process. Cell Death Differ. 2014; 21. https://doi.org/10.1038/cdd.2014.35 PMID: 24658399

11. Van Loo G, Saelens X, Van Gurp M, MacFarlane M, Martin SJ, Vandenabeele P. The role of mitochondrial factors in apoptosis: A Russian roulette with more than one bullet. Cell Death and Differentiation. 2002. https://doi.org/10.1038/sj.cdd.4401088 PMID: 12232790

12. Kroemer G, Reed JC. Mitochondrial control of cell death. Nature Medicine. 2000. https://doi.org/10.1038/74994 PMID: 10802706

13. Rehman J, Zhang HJ, Toth PT, Zhang Y, Marsboom G, Hong Z, et al. Inhibition of mitochondrial fission prevents cell cycle progression in lung cancer. FASEB J. 2012; 26. https://doi.org/10.1096/fj.11-196543 PMID: 22321727

14. Ogihara Y, Yamazaki Y, Murai K, Kanno A, Terachi T, Shiina T, et al. Structural dynamics of cereal mitochondrial genomes as revealed by complete nucleotide sequencing of the wheat mitochondrial genome. Nucleic Acids Res. 2005; 33. https://doi.org/10.1093/nar/gki925 PMID: 16260473

15. Greiner S, Bock R. Tuning a ménage à trois: Co-evolution and co-adaptation of nuclear and organellar genomes in plants. BioEssays. 2013; 35. https://doi.org/10.1002/bies.201200137 PMID: 23361615

16. Keeling PJ. The endosymbiotic origin, diversification and fate of plastids. Philosophical Transactions of the Royal Society B: Biological Sciences. 2010. https://doi.org/10.1098/rstb.2009.0103 PMID: 20124341

17. Sagan L. On the origin of mitosing cells. J Theor Biol. 1967; 14. https://doi.org/10.1016/0022-5193(67)90079-3 PMID: 11541392

18. Birky CW. Uniparental inheritance of mitochondrial and chloroplast genes: Mechanisms and evolution. Proceedings of the National Academy of Sciences of the United States of America. 1995. https://doi.org/10.1073/pnas.92.25.11331 PMID: 8524780

19. Wallace DC, Singh G, Lott MT, Hodge JA, Schurr TG, Lezza AMS, et al. Mitochondrial DNA mutation associated with Leber's hereditary optic neuropathy. Science (80-). 1988; 242. https://doi.org/10.1126/science.3201231 PMID: 3201231

20. Ward BL, Anderson RS, Bendich AJ. The mitochondrial genome is large and variable in a family of plants (Cucurbitaceae). Cell. 1981; 25. https://doi.org/10.1016/0092-8674(81)90187-2

21. Handa H. 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. 2003; 31. https://doi.org/10.1093/nar/gkg795 PMID: 14530439

22. Kim B, Kim K, Yang TJ, Kim S. Completion of the mitochondrial genome sequence of onion (*Allium cepa* L.) containing the CMS-S male-sterile cytoplasm and identification of an independent event of the ccmF N gene split. Curr Genet. 2016; 62. https://doi.org/10.1007/s00294-016-0595-1 PMID: 27016941

23. 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 (80-). 2013; 342. https://doi.org/10.1126/science.1246275 PMID: 24357311

24. Sloan DB, Alverson AJ, Chuckalovcak JP, Wu M, McCauley DE, Palmer JD, et al. Rapid evolution of enormous, multichromosomal genomes in flowering plant mitochondria with exceptionally high mutation rates. PLoS Biol. 2012; 10. https://doi.org/10.1371/journal.pbio.1001241 PMID: 22272183

25. Smith DR, Keeling PJ. Mitochondrial and plastid genome architecture: Reoccurring themes, but significant differences at the extremes. Proceedings of the National Academy of Sciences of the United States of America. 2015. https://doi.org/10.1073/pnas.1422049112 PMID: 25814499

26. Folkerts O, Hanson MR. Three copies of a single recombination repeat occur on the 443 kb mastercircle of the *Petunia hybrida* 3704 mitochondrial genome. Nucleic Acids Res. 1989; 17. https://doi.org/10.1093/nar/17.18.7345 PMID: 2798096

27. Klein M, Eckert-Ossenkopp U, Schmiedeberg I, Brandt P, Unseld M, Brennicke A, et al. Physical mapping of the mitochondrial genome of *Arabidopsis thaliana* by cosmid and YAC clones. Plant J. 1994; 6. https://doi.org/10.1046/j.1365-313x.1994.06030447.x PMID: 7920724

28. Palmer JD, Shields CR. Tripartite structure of the *Brassica campestris* mitochondrial genome. Nature. 1984; 307. https://doi.org/10.1038/307437a0

29. Siculella L, Damiano F, Cortese MR, Dassisti E, Rainaldi G, Gallerani R, et al. Gene content and organization of the oat mitochondrial genome. Theor Appl Genet. 2001; 103. https://doi.org/10.1007/s001220100568

30. Sloan DB, Alverson AJ, Štorchová H, Palmer JD, Taylor DR. Extensive loss of translational genes in the structurally dynamic mitochondrial genome of the angiosperm *Silene latifolia*. BMC Evol Biol. 2010; 10. https://doi.org/10.1186/1471-2148-10-274 PMID: 20831793

31. Stren DB, Palmer JD. Tripartite mitochondrial genome of spinach: Physical structure, mitochondrial gene mapping, and locations of transposed chloroplast DNA sequences. Nucleic Acids Res. 1986; 14. https://doi.org/10.1093/nar/14.14.5651 PMID: 3016660

32. Drouin G, Daoud H, Xia J. Relative rates of synonymous substitutions in the mitochondrial, chloroplast and nuclear genomes of seed plants. Mol Phylogenet Evol. 2008; 49. https://doi.org/10.1016/j.ympev.2008.09.009 PMID: 18838124

33. 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. https://doi.org/10.1186/1741-7007-11-29 PMID: 23587068

34. 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. https://doi.org/10.1073/pnas.84.24.9054 PMID: 3480529

35. Simon C, Frati F, Beckenbach A, Crespi B, Liu H, Flook P. Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. Ann Entomol Soc Am. 1994; 87. https://doi.org/10.1093/aesa/87.6.651

36. Knoop V. The mitochondrial DNA of land plants: Peculiarities in phylogenetic perspective. Current Genetics. 2004. https://doi.org/10.1007/s00294-004-0522-8 PMID: 15300404

37. Bergthorsson U, Richardson AO, Young GJ, Goertzen LR, Palmer JD. Massive horizontal transfer of mitochondrial genes from diverse land plant donors to the basal angiosperm *Amborella*. Proc Natl Acad Sci U S A. 2004; 101. https://doi.org/10.1073/pnas.0408336102 PMID: 15598737

38. Skippingtona E, Barkmanb TJ, Ricea DW, Palmera 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. https://doi.org/10.1073/pnas.1504491112 PMID: 26100885

39. Song W, Feng Q, Zhang Y, Wu X, Shi C, Wang S. The complete chloroplast genome sequence of *Duranta erecta* (Verbenaceae). Mitochondrial DNA Part B Resour. 2021; 6: 1832–1833. https://doi.org/10.1080/23802359.2021.1934164 PMID: 34124359

40. Hsu CL, Mullin BC. Physical characterization of mitochondrial DNA from cotton. Plant Mol Biol. 1989; 13. https://doi.org/10.1007/BF00015558 PMID: 2562383

41. Song W, Ji C, Chen Z, Cai H, Wu X, Shi C, et al. Comparative Analysis the Complete Chloroplast Genomes of Nine *Musa* Species: Genomic Features, Comparative Analysis, and Phylogenetic Implications. Front Plant Sci. 2022; 13: 1–15. https://doi.org/10.3389/fpls.2022.832884 PMID: 35222490

42. 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. https://doi.org/10.1089/cmb.2012.0021 PMID: 22506599

43. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 2012; 28. https://doi.org/10.1093/bioinformatics/bts199 PMID: 22543367

44. Wyman SK, Jansen RK, Boore JL. Automatic annotation of organellar genomes with DOGMA. Bioinformatics. 2004; 20. https://doi.org/10.1093/bioinformatics/bth352 PMID: 15180927

45. Lowe TM, Chan PP. tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. Nucleic Acids Res. 2016; 44. https://doi.org/10.1093/nar/gkw413 PMID: 27174935

46. Laslett D, Canbäck B. ARWEN: A program to detect tRNA genes in metazoan mitochondrial nucleotide sequences. Bioinformatics. 2008; 24. https://doi.org/10.1093/bioinformatics/btm573 PMID: 18033792

47. Beier S, Thiel T, Münch T, Scholz U, Mascher M. MISA-web: A web server for microsatellite prediction. Bioinformatics. 2017; 33: 2583–2585. https://doi.org/10.1093/bioinformatics/btx198 PMID: 28398459

48. Song W, Chen Z, He L, Feng Q, Zhang H, Du G, et al. Comparative Chloroplast Genome Analysis of Wax Gourd (*Benincasa hispida*) with Three Benincaseae Species, Revealing. Genes (Basel). 2022; 13: 461. https://doi.org/10.3390/genes13030461 PMID: 35328015

49. Benson G. Tandem repeats finder: A program to analyze DNA sequences. Nucleic Acids Res. 1999; 27. https://doi.org/10.1093/nar/27.2.573 PMID: 9862982

50. Chen C, Xia R, Chen H, He Y. TBtools, a Toolkit for Biologists integrating various HTS-data handling tools with a user-friendly interface. TBtools, a Toolkit Biol Integr Var HTS-data Handl tools with a user-friendly interface. 2018. https://doi.org/10.1101/289660

51. Mower JP. The PREP suite: Predictive RNA editors for plant mitochondrial genes, chloroplast genes and user-defined alignments. Nucleic Acids Res. 2009; 37. https://doi.org/10.1093/nar/gkp337 PMID: 19433507

52. Edgar RC. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004; 32. https://doi.org/10.1093/nar/gkh340 PMID: 15034147

53. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol. 2018; 35: 1547–1549. https://doi.org/10.1093/molbev/msy096 PMID: 29722887

54. Kalyaanamoorthy S, Minh BQ, Wong TKF, Von Haeseler A, Jermiin LS. ModelFinder: Fast model selection for accurate phylogenetic estimates. Nat Methods. 2017; 14. https://doi.org/10.1038/nmeth.4285 PMID: 28481363

55. Liao X, Zhao Y, Kong X, Khan A, Zhou B, Liu D, et al. Complete sequence of kenaf (*Hibiscus cannabinus*) mitochondrial genome and comparative analysis with the mitochondrial genomes of other plants. Sci Rep. 2018; 8. https://doi.org/10.1038/s41598-018-30297-w PMID: 30143661

56. Shearman JR, Sonthirod C, Naktang C, Pootakham W, Yoocha T, Sangsrakru D, et al. The two chromosomes of the mitochondrial genome of a sugarcane cultivar: Assembly and recombination analysis using long PacBio reads. Sci Rep. 2016; 6. https://doi.org/10.1038/srep31533 PMID: 27530092

57. Adams KL, Qiu YL, Stoutemyer M, Palmer JD. Punctuated evolution of mitochondrial gene content: High and variable rates of mitochondrial gene loss and transfer to the nucleus during angiosperm evolution. Proc Natl Acad Sci U S A. 2002; 99. https://doi.org/10.1073/pnas.042694899 PMID: 12119382

58. Liu Y chun, Liu S, Liu D cheng, Wei Y xiang, Liu C, Yang Y min, et al. Exploiting EST databases for the development and characterization of EST-SSR markers in blueberry (*Vaccinium*) and their cross-species transferability in *Vaccinium* spp. Sci Hortic (Amsterdam). 2014; 176. https://doi.org/10.1016/j.scienta.2014.07.026

59. Powell W, Machray GC, Proven J. Polymorphism revealed by simple sequence repeats. Trends in Plant Science. 1996. https://doi.org/10.1016/1360-1385(96)10028-5 PMID: 11539828

60. Gao H, Kong J. Distribution characteristics and biological function of tandem repeat sequences in the genomes of different organisms. Zool Res. 2005; 26.

61. Bock R, Khan MS. Taming plastids for a green future. Trends in Biotechnology. 2004. https://doi.org/10.1016/j.tibtech.2004.03.005 PMID: 15158061

62. Chen H, Deng L, Jiang Y, Lu P, Yu J. RNA editing sites exist in protein-coding genes in the chloroplast genome of *Cycas taitungensis*. J Integr Plant Biol. 2011;53. https://doi.org/10.1111/j.1744-7909.2011.01082.x PMID: 22044752

63. Raman G, Park SJ. Analysis of the complete chloroplast genome of a medicinal plant, Dianthus superbus var. longicalyncinus, from a comparative genomics perspective. PLoS One. 2015; 10. https://doi.org/10.1371/journal.pone.0141329 PMID: 26513163

64. Wakasugi T, Hirose T, Horihata M, Tsudzuki T, Kössel H, Sugiura M. Creation of a novel protein-coding region at the RNA level in black pine chloroplasts: The pattern of RNA editing in the gymnosperm chloroplast is different from that in angiosperms. Proc Natl Acad Sci U S A. 1996; 93. https://doi.org/10.1073/pnas.93.16.8766 PMID: 8710946

65. Zandueta-Criado A, Bock R. Surprising features of plastid *ndhD* transcripts: Addition of non-encoded nucleotides and polysome association of *mRNAs* with an unedited start codon. Nucleic Acids Res. 2004; 32. https://doi.org/10.1093/nar/gkh217 PMID: 14744979

66. Cheng Y, He X, Priyadarshani SVGN, Wang Y, Ye L, Shi C, et al. Assembly and comparative analysis of the complete mitochondrial genome of *Suaeda glauca*. BMC Genomics. 2021; 22: 1–15. https://doi.org/10.1186/s12864-020-07350-y PMID: 33388042

67. Sauquet H, Weston PH, Anderson CL, Barker NP, Cantrill DJ, Mast AR, et al. Contrasted patterns of hyperdiversification in Mediterranean hotspots. Proc Natl Acad Sci U S A. 2009;106. https://doi.org/10.1073/pnas.0805607106 PMID: 19116275

68. Bremer K, Chase MW, Stevens PF. An ordinal classification for the families of flowering plants. Ann Missouri Bot Gard. 1998; 85. https://doi.org/10.2307/2992015

69. Soltis DE, Soltis PS, Chase MW, Mort ME, Albach DC, Zanis M, et al. Angiosperm phylogeny inferred from *18S rDNA*, *rbcL*, and *atpB* sequences. Bot J Linn Soc. 2000; 133. https://doi.org/10.1006/bojl.2000.0380

70. Nock CJ, Baten A, King GJ. Complete chloroplast genome of *Macadamia integrifolia* confirms the position of the Gondwanan early-diverging eudicot family Proteaceae. BMC Genomics. 2014; 15: 1–10. https://doi.org/10.1186/1471-2164-15-1 PMID: 24382143

71. Nock CJ, Hardner CM, Montenegro JD, Ahmad Termizi AA, Hayashi S, Playford J, et al. Wild origins of *Macadamia* domestication identified through intraspecific chloroplast genome sequencing. Front Plant Sci. 2019; 10: 1–15. https://doi.org/10.3389/fpls.2019.00001 PMID: 30723482

72. Fay JC, Wu CI. Sequence Divergence, Functional Constraint, and Selection in Protein Evolution. Annual Review of Genomics and Human Genetics. 2003. https://doi.org/10.1146/annurev.genom.4.020303.162528 PMID: 14527302

73. Wang D, Zhang Y, Zhang Z, Zhu J, Yu J. KaKs_Calculator 2.0: A Toolkit Incorporating Gamma-Series Methods and Sliding Window Strategies. Genomics, Proteomics Bioinforma. 2010; 8. https://doi.org/10.1016/S1672-0229(10)60008-3 PMID: 20451164

74. Bi C, Paterson AH, Wang X, Xu Y, Wu D, Qu Y, et al. Analysis of the Complete Mitochondrial Genome Sequence of the Diploid Cotton Gossypium raimondii by Comparative Genomics Approaches. Biomed Res Int. 2016; 2016. https://doi.org/10.1155/2016/5040598 PMID: 27847816

75. Wendel JF, Greilhuber J, Doležel J, Leitch IJ. Plant genome diversity volume 1: Plant genomes, their residents, and their evolutionary dynamics. Plant Genome Diversity Volume 1: Plant Genomes, their Residents, and their Evolutionary Dynamics. 2012. https://doi.org/10.1007/978-3-7091-1130-7