Biochemical Genetics
https://doi.org/10.1007/s10528-024-10874-7

ORIGINAL ARTICLE

Repetitive Sequences, Codon Usage Bias and Phylogenetic Analysis of the Plastome of *Miliusa glochidioides*

Yangying Gan[1] · Jingyao Ping[2] · Xiaojing Liu[1] · Caixia Peng[3]

Received: 4 December 2023 / Accepted: 18 June 2024

Abstract

Annonaceae is the largest family in Magnoliales, exhibiting the greatest diversity among and within genera. In this study, we conducted an analysis of repetitive sequences and codon usage bias in the previously acquired plastome of *Miliusa glochidioides.* Using a concatenated dataset of shared genes, we constructed the phylogenetic relationships among 27 Annonaceae species. The results showed that the size of the plastomes in the Annonaceae ranged from 159 to 202 kb, with the size of the inverted repeat region ranging from 40 to 65 kb. Within the plastome of *M. glochidioides*, we identified 42 SSRs, 36 tandem repeats, and 9 dispersed repeats. These SSRs consist of three nucleotide types and eight motif types, with a preference for A/T bases, primarily located in the large single-copy regions and intergenic spacers. Tandem and dispersed repeat sequences were predominantly detected in the IR region. Through codon usage bias analysis, we identified 30 high-frequency codons and 11 optimal codons. The plastome of *M. glochidioides* demonstrated relatively weak codon usage bias, favoring codons with A/T endings, primarily influenced by natural selection. Phylogenetic analysis revealed that all four subfamilies formed monophyletic groups, with *Cananga odorata* (Ambavioideae) and *Anaxagorea javanica* (Anaxagoreoideae) successively nested outside Annonoideae + Malmeoideae. These findings improve our understanding of the plastome of *M. glochidioides* and provide additional insights for studying plastome evolution in Annonaceae.

Keywords *Miliusa glochidioides* plastome · SSRs · Codon usage bias · Phylogenetics

Extended author information available on the last page of the article

Published online: 02 July 2024

Introduction

In recent years, plastid genomes (plastomes) have been used more frequently in phylogenetic evolution studies. Compared with the nuclear genome, they have a small genome size, a large copy number, and are easy to obtain. Additionally, compared with plastid DNA markers, they provide more informative (Li et al. 2021a; Raubeson and Jansen 2005; Tonti-Filippini et al. 2017). Moreover, it has been found that the evolution of plastome is often accompanied by gene loss or pseudogenization (Chaw et al. 2018; Li et al. 2021b; Ruhlman and Jansen 2014), inversion of sequence fragments (Ping et al. 2021a; Raubeson and Jansen 1992), or expansion and contraction of IR regions (Cauz-Santos et al. 2020; Mower and Vickrey 2018; Ping et al. 2021b; Zhu et al. 2016). By comparing features such as genome size, gene capacity, boundary displacement, and structural rearrangement, it is possible to shed new light on phylogenetic studies and provide rich information on the evolution process of plastomes.

The plastome typically contains certain repeat sequences, including direct repeat sequences, inverted repeat sequences, dispersed repeat sequences, and simple sequence repeats (SSRs), which may play a significant role in the structure and function of plastomes (Wu et al. 2021). Recently, Zhou et al. (2022) discovered that short repeat sequences and intergenic spacers mediated the formation of six new configurations of plastomes in the *Selaginella*. Tandem repeat sequences refer to identical or similar DNA sequences that are continuously repeated in the genome. Li et al. (2023) found that the number of tandem repeat sequences in the Alismatidae plastomes is significantly correlated with genome size. Dispersed repeats refer to sequences that are scattered and repeatedly present in the genome. SSRs are one of the features of plastomes. Studying repetitive sequences in plastomes is important for understanding evolutionary origins and relationships, identifying functional elements, analyzing mutations and diversity, molecular markers, genetic analysis, and plastome engineering (Li et al. 2015; Chmielewski et al. 2015). It has been observed that SSRs are positively correlated with recombination frequency (Weng et al. 2014; Amenu et al. 2022). In Liverwort plastomes, SSRs contribute to the expansion of the inverted repeat region through the incorporation of *rps12* and *rps7* genes (Sawicki et al. 2020).

The phenomenon of preferential or non-random usage of synonymous codons is widely observed in bacteria, plants, and animals, and this phenomenon is known as codon usage bias (Parvathy et al. 2022). Codon usage bias influences a range of cellular processes, such as transcription and translation (Quax et al. 2015; Liu 2020). Several factors, including natural selection, mutation pressure, and genetic drift in translation efficiency, contribute to codon usage bias (Ingvarsson 2008; Liu 2012). Analyzing codon usage patterns can help us better understand gene evolution, environmental adaptation, and evolutionary relationships among species (Athey et al. 2017; Mazumdar et al. 2017). In the *Chlamydomonas reinhardtii* plastome, it has been discovered that highly expressed genes strongly prefer optimized codons for translation, while genes of lower functional importance are influenced by biased directional mutations (Fages-Lartaud et al. 2022).

Recently, the codon usage bias in the plastomes of several species, such as *Trivalvaria costata* (Ping et al. 2023), *Oryza* (Chakraborty et al. 2020), and Boraginales (Li and Wei 2022), has been deciphered.

Annonaceae is the largest pantropical family of trees and lianas in the early-divergent order of Magnoliales, with 107 genera and about 2300 species (Chatrou et al. 2012; Guo et al. 2017). It is the most diverse family and contributes significantly to tree diversity in rain forests around the world (Punyasena et al. 2008; Couvreur et al. 2011; Gan et al. 2015; Gan and Xu 2018; Lei et al. 2022). However, its remarkable diversity has also brought great difficulty and controversy to its phylogenetic research (Chaowasku et al. 2015, 2018; Maas et al. 2015; Ortiz-Rodriguez et al. 2018; Saunders et al. 2020; Xue and Tan 2016; Xue et al. 2021, 2020). Currently, there is limited research on the repetitive sequences and codon usage bias of plastomes in Annonaceae. To understand the genetic characteristics of the Annonaceae plastome, this study analyzed the repetitive sequences and codon usage bias of the previously obtained plastome of *Miliusa glochidioides*. Additionally, the study constructed the phylogenetic relationships among the published plastomes of Annonaceae species.

Materials and Methods

Preparation of Sequence Data

The complete plastome sequences were downloaded from the NCBI database [https://www.ncbi.nlm.nih.gov/nuccore/?term=, (accessed on 1 June 2023)], including 27 Annonaceae and 2 Magnoliaceae (as an out-group) (Table 1). Among them, *M. glochidioides* (Gan et al. 2022) and *T. costata* (Ping et al. 2023) were obtained from previous studies conducted by our research group. The sequences were imported into Genious Prime 2022.0.1 (Kearse et al. 2012) software for statistical genome structure information and screening of common genes. All sampled species were used for the comparison of basic information of plastomes and the construction of phylogenetic relationships.

Detection of Repeat Sequences

MISA (Microsatellite Identification Tool) online website (https://webblast.ipk-gatersleben.de/misa/index.php?action=1, accessed on 10 June 2023) predicts SSRs, with minimal iterations of ten repeat motifs for mononucleotides, six for dinucleotide repeats, and five for tri-, etra-, penta- and hexanucleotides. When the distance between two SSRs is less than 100 bp, it can be used as a compound SSR (Beier et al. 2017).

Tandem repeat sequences were identified using the online software Tandem Repeats Finder v4.09 (http://tandem.bu.edu/trf/trf.html, accessed on 11 June 2023) (Benson 1999). Default parameters were used in the advanced module, with match, mismatch, and gap parameters set to 2, 7, and 7, respectively. The minimum

Table 1 The information of sample species

Species name		Size (bp)				GC content (%)	NCBI
		LSC	IR	SSC	Genome		Accession number
Annonoideae							
Annoneae	*Annona squamosa*	69,649	63,860	2966	200,335	39.6	MN241494
	Annona atemoya	69,685	65,026	2966	202,703	39.5	MN241495
	Annona cherimola	69,771	64,493	2966	201,723	39.6	KU563738
	Annona reticulata	69,650	64,621	3014	196,038	39.9	MT742547
	Rollinia mucosa	70,842	63,938	2948	201,666	39.3	MN241491
	Annona montana	75,172	58,609	3105	195,495	39.9	MK087989
	Annona muricata	75,339	58,797	3105	201,906	39.6	MT742546
	Goniothalamus tamirensis	87,019	51,468	3047	193,002	38.8	MN241496
Uvarieae	*Desmos chinensis*	84,251	51,646	3538	191,081	38.4	MN253545
	Monanthotaxis ambrensis	83,130	51,717	3558	190,122	38.6	MN241488
	Sphaerocoryne affinis	83,281	51,343	3630	189,597	38.7	MN241489
	Uvaria macrophylla	83,581	52,730	3741	192,782	38.7	MH992130
	Anomianthus dulcis	83,727	51,620	3715	190,682	38.9	MN241490
	Fissistigma oldhamii	82,584	51,062	3074	187,782	38.9	MW136266
	Fissistigma polyanthum	83,000	51,936	3273	189,920	38.7	MW829282
Xylopieae	*Artabotrys hexapetalus*	90,803	42,294	3066	178,457	38.8	MZ936420
	Artabotrys pilosus	90,797	42,150	3098	178,195	38.8	OK216144
Malmeoideae							
Miliuseae	*Alphonsea hainanensis*	88,771	25,861	18,548	159,041	39.2	MN253543
	Stelechocarpus burahol	88,218	25,943	18,733	158,837	39.2	MN253544
	Monoon laui	89,555	26,324	18,975	161,178	39.1	OL979152
	Miliusa glochidioides	88,782	26,029	18,949	159,789	39.2	OM047203

Table 1 (continued)

Species name		Size (bp)				GC content (%)	NCBI
		LSC	IR	SSC	Genome		Accession number
	Trivalvaria costata	87,143	28,021	18,817	162,002	39.0	OM914484
	Polyalthia suberosa	88,566	25,921	19,000	159,408	39.2	OM937139
	Polyalthiopsis verrucipes	89,030	25,974	18,987	159,965	39.0	MW018366
	Chieniodendron hainanense	89,424	26,062	18,949	160,497	39.1	MK035708
Ambavioideae							
	Cananga odorata	83,620	32,008	20,310	167,946	39.0	MN016933
Anaxagoreoideae							
	Anaxagorea javanica	89,887	40,202	4354	174,645	38.7	MK087990
Magnoliaceae (outgroup)							
	Michelia alba	88,137	26,596	18,777	160,106	39.2	NC_037005
	Liriodendron chinense	87,766	26,333	18,997	159,429	39.2	NC_030504

alignment score was set to 50, the maximum period size was set to 500, and the maximum tandem repeat array size (in bp, millions) was set to 2.

The online tool REPuter (http://bibiserv.techfak.uni-bielefeld.de/reputer, accessed on 11 June 2023) was used to search for dispersed repeat sequences of forward and palindromic repeats between genomes (Kurtz and Schleiermacher 1999). The parameters used for detecting dispersed repeat sequences were set as follows: a Hamming distance of 3, a maximum of 500 computed repeats, and a minimal repeat size of 30.

Codon Bias Analysis

The protein-coding genes (PCGs) were extracted using Geneious Prime 2020.0.1 software. Sequences shorter than 300 bp were removed, resulting in a total of 53 PCGs. CodonW 1.4.2 software was used to calculate codon usage parameters for the 53 PCGs, including the number of occurrences (N) of each codon, the effective number of codons (ENC), and the relative synonymous codon usage (RSCU). MEGA X (Kumar et al. 2018) software was used to compute the overall GC content (GC_{all}) and the GC content at each position of the codons: the first position (GC_1), the second position (GC_2), and the third position (GC_3).

Codons with RSCU > 1 were considered high-frequency codons. The ENC values of the 53 PCGs were ranked from high to low, and genes representing the top and bottom 10% of the total were selected to establish high- and low-expression gene pools, respectively. The parameter ΔRSCU was calculated as the difference between RSCU values of high-expression genes and low-expression genes. Codons that satisfied RSCU > 1 and ΔRSCU ≥ 0.08 were identified as optimal codons.

A neutrality plot was generated with GC_3 on the *x*-axis and the average of GC_1 and GC_2 (GC_{12}) on the *y*-axis. An ENC plot was created with GC_3 on the *x*-axis and ENC on the *y*-axis, and the expected ENC curve was inserted: $ENC = 2 + GC_3 + 29/[GC_3^2 + (1 - GC_3)^2]$. The ENC ratio, calculated as $(ENC_{expected} - ENC_{observed})/ENC_{expected}$, provided insights into the deviation between the observed and expected ENC values.

A PR2 plot was constructed by plotting the ratio of G content at the third position of codons (G_3) to the sum of G and C content at the third position ($G_3 + C_3$) on the *x*-axis, and the ratio of A content at the third position (A_3) to the sum of A and T content at the third position ($A_3 + T_3$) on the *y*-axis. This scatter plot allowed for analysis of the relationship between the A, C, G, and T content at the third position of codons.

Construction of Phylogenetic Relationships

Based on the tandem dataset of 75 common genes, we used MEGA X software to predict the best phylogenetic tree model (Best Model: GTR+G). Subsequently, we constructed maximum-likelihood (ML) trees with 1000 bootstrap and GTR+GAMMA model using RaxmlGUI2 (Stamatakis 2014).

Results

Structural Characteristics of Sampled Species Plastomes

The nucleotide sequences of the 27 Annonaceae plastomes vary in size from 158,837 bp (*Stelechocarpus burahol*) to 202,703 bp (*Annona atemoya*). The overall GC content ranges from 38.4 to 39.9% (Table 1). Compared to other species, the plastomes of Annonoideae and Anaxagoreoideae are larger, particularly in the IR region, ranging from 40,202 bp (*Anaxagorea javanica*) to 65,026 bp (*A. atemoya*), accounting for 46–65.9% of the total genome size (Fig. 1). In contrast, their SSC region has undergone significant reduction, ranging from 2948 to 4354 bp, which accounts for 1.5–2.5% of the total genome size. Among them, *Annona* has the largest plastomes, with five species having genome sizes exceeding 200 kb. The non-parametric rank sum test results reveal that significant differences exist among the four subfamilies in terms of the complete genome, inverted repeat (IR) region, small single-copy (SSC), and large single-copy (SSC) size. Pairwise comparisons further demonstrated significant differences between Annonoideae and Malmeoideae (Table S1).

Distribution Pattern of Repeat Sequences in *M. glochidioides*

The study detected a total of 42 SSRs in the *M. glochidioides* plastome, including 8 types of repeat motifs (Fig. 2a), which can be categorized into 3 nucleotide types

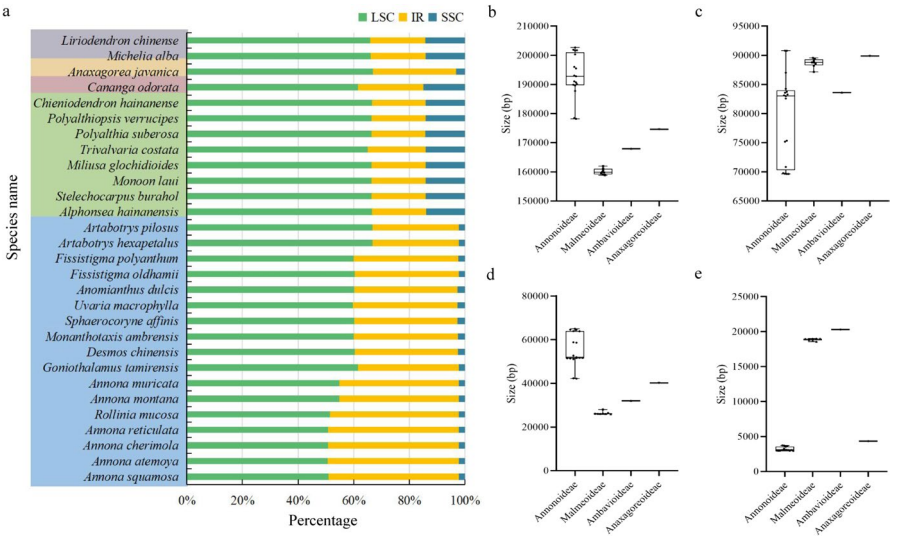

Fig. 1 Size distribution of plastomes in the sampled species. **a** Proportion of three regions in the genome; **b** size distribution of complete genomes in four subfamilies; **c** size distribution of the LSC region in four subfamilies; **d** size distribution of the IR region in four subfamilies; **e** size distribution of the SSC region in four subfamilies

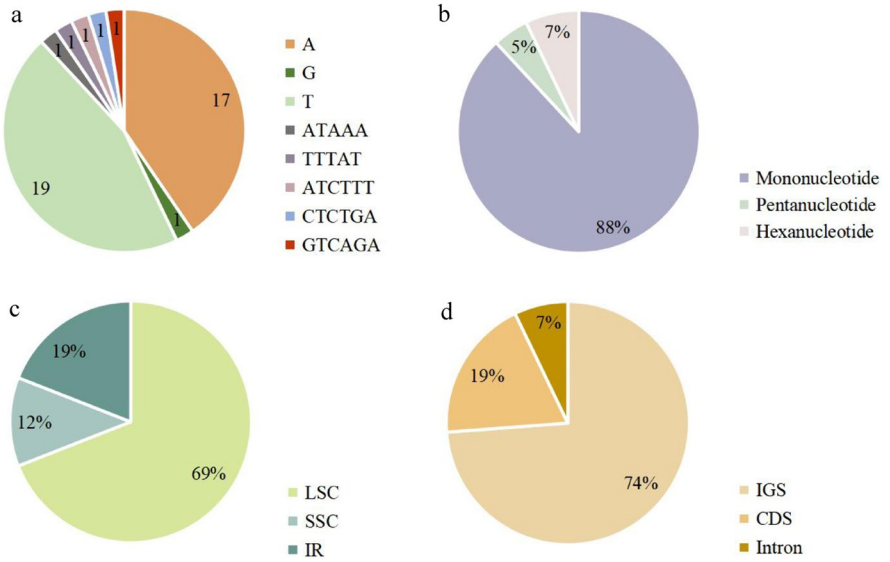

Fig. 2 Distribution characteristics of SSRs in *M. glochidioides* plastome. **a** Type and number of repetitive motifs. **b** Nucleotide types of SSRs. **c** and **d** Location distribution of repetitive motifs

(Fig. 2b). The SSRs, in terms of base composition, are predominantly composed of A/T bases. In terms of genomic distribution, they are mainly located in the LSC region (69%, Fig. 2c) and intergenic spacer regions (IGS) (74%, Fig. 2d). Five SSRs of the A/T repeat type were detected in the coding sequences of four genes (*rpoC2*, *rpoB*, *rpoA*, and *ycf2*). There is only one AAAGAT located in the *infA* gene, and two ACTCTG located in the *ycf2* gene. Additionally, one A motif type SSR was detected in the introns of *atpF*, *clpP*, and *rpl16*, respectively. There was only one G motif type, which was located between *trnS-GGA* and *rps4* (Table S2).

Additionally, a total of 36 tandem repeat sequences were detected in the study, with 16, 19, and 1 located in the LSC region, IR region, and SSC region, respectively. Among these, five forward and four palindromic dispersed repeat sequences were identified, with only one forward dispersed repeat sequence found in the LSC region, while the rest were located in the IR region.

Composition of Codons in the Protein Coding Genes

The analysis of 53 PCGs reveals a total of 21,484 codons. The average GC content is 39.59%. The average GC content for each position of the codon is as follows: $GC_1 = 47.5\%$, $GC_2 = 40.4\%$, $GC_3 = 30.9\%$. The values of ENC range from 35.8 (*rps14*) to 59.9 (*rpl22*), with an average of 48.99. Using ENC = 45 as the criterion for evaluating codon bias strength, there are 45 genes (84.9%) with ENC values greater than 45 and 8 genes (15.1%) with ENC values less than 45 (Table S3).

Pearson correlation analysis results indicate that GC_{all} is highly correlated with GC_1, GC_2, and GC_3, with correlation coefficients of 0.765, 0.698, and 0.433, respectively, and all show significant correlations. The correlation coefficient between GC_1 and GC_2 is 0.282, and there is a significant correlation. GC_3 is significantly positively correlated with ENC ($p = 0.14$), with a correlation coefficient of 0.334 (Table 2).

Analysis of RSCU and Optimal Codons

CodonW calculated the RSCU values of 53 PCGs of *M. glochidioides*. Among them, 30 codons have RSCU values greater than 1, with 16 ending in U and 13 ending in A (Fig. 3, Table S4). In the determination of optimal codons, we ranked the ENC values of the 53 PCGs from high to low and selected genes from both ends that accounted for a total of 10% of the genes to establish a high-expression gene library (*clpP*, *ccsA*, *rpl2*, *rpl22*, and *rps4*) and a low-expression gene library (*rps14*, *rps18*, *rpl16, petB*, and *psbA*). Based on $\Delta RSCU = RSCU_{high\text{-}expression} - RSCU_{low\text{-}expression}$, codons with $RSCU > 1$ and $\Delta RSCU \geq 0.08$ were considered as optimal codons. A total of 11 optimal codons were identified: UUU (F), UUA (L), GUU (V), CAU (H), CAA (Q), AAU (N), UCA (S), ACA (T), GCA (A), AGA (R), and GGA (G).

Plot Analysis of Codon Bias

The analysis results from the neutrality plot (Fig. 4a) indicate that most genes are positioned above the diagonal line. The GC_{12} range is 35.95–54%, while the GC_3 range is 22.5–40.5%. The Pearson test results show a correlation coefficient of 0.041 between GC_{12} and GC_3, which is not statistically significant ($p = 0.772$).

The analysis results from the ENC plot demonstrate that a majority of genes exhibit deviation from the standard curve (Fig. 4b, Table S3). The statistical analysis of the ENC ratio reveals that 17 genes (47.2%) have an ENC ratio ranging from −0.05 to 0.05, whereas 23 genes (52.8%) have an ENC ratio ranging from 0.05 to 0.15 (Fig. 5).

The PR2 plot analysis (Fig. 4c) presents the $A_3/(A_3 + T_3)$ range as 0.307–0.596, with 40 genes (75.5%) exhibiting values below 0.5. The $G_3/(G_3 + C_3)$ range is 0.366–0.751, with 37 genes (69.8%) exhibiting values above 0.5.

Table 2 Correlation analysis of GC content and ENC

	GC_1	GC_2	GC_3	GC_{all}
GC_2	0.268			
GC_3	0.115	−0.049		
GC_{all}	0.765**	0.698**	0.433**	
ENC	−0.112	−0.056	0.334*	0.036

*Indicates significant correlation at the 0.05 level

**Indicates significant correlation at the 0.01 level

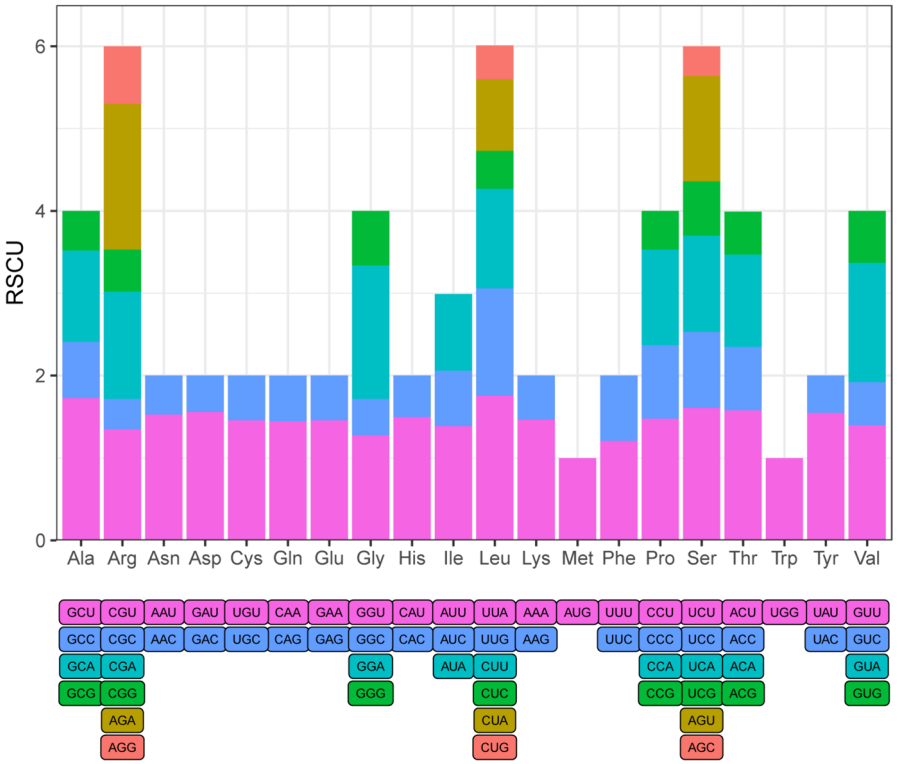

Fig. 3 RSCU values of the amino acids in the 53 PCGs

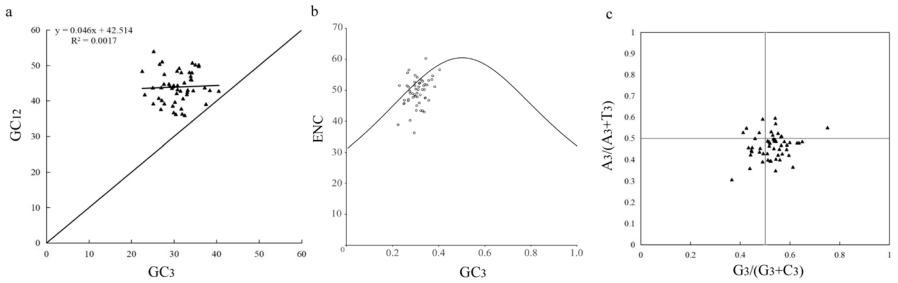

Fig. 4 Plot analysis. **a** Neutral plot. **b** ENC plot. **c** PR2 plot

Phylogenetic Relationships of Sampled Species

Based on the tandem dataset of shared genes, a maximum likelihood tree constructed using *M. alba* and *L. chinense* as outgroups reveals highly supported phylogenetic relationships between different taxa at the subfamily and tribe levels (Fig. 6). Annonoideae (17 species) and Malmeoideae (8 species) form monophyletic clades and are sister groups. *Cananga odorata* and *A. javanica* nest subsequently

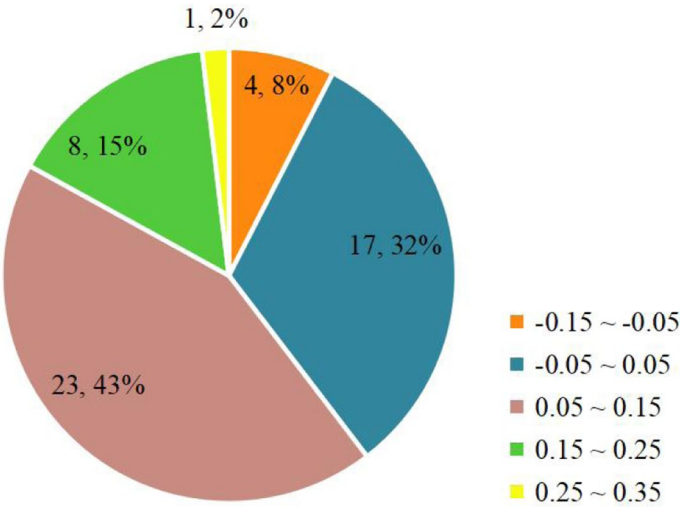

Fig. 5 Frequency distribution of ENC ratio

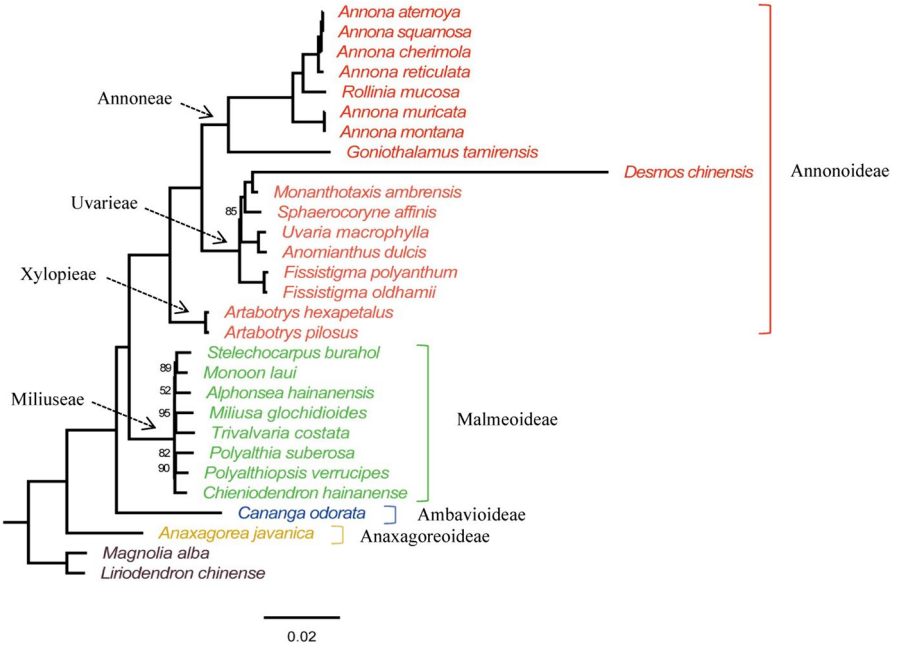

Fig. 6 The maximum likelihood (ML) tree was constructed using the RaxML software with the GTR-GAMMA substitution model based on a concatenated dataset of 75 conserved genes. The tree was generated by running 1000 bootstraps. The numerical values on the branches represent the bootstrap support for each branch, with a support value of 100 for all other branches

to Annonoideae + Malmeoideae, with *A. javanica* serving as the basal genus of the Annonaceae.

Discussion

We observe the gradually enlarged plastomes of Annonaceae, from 159 to 201.9 kb. The plastomes size of Annonoideae exceeds 160 kb, and the size of the IR region is also between 42 and 64 kb. This is well beyond the size of the IR region of most land plants (15 kb to 30 kb) (Zhu et al. 2016). The enlargement of Annonoideae plastomes is mainly related to the expansion of the IR region, a single IR region expanded by about 20–30 kb. In addition, we note that the expansion of the IR region in the Annonaceae is dynamic. In the basal group (*Cananga*), only 6 kb was extended to LSC. In Annonoideae, they expanded 20 kb, 14 kb or 6 kb to the LSC region, and 16 kb to the SSC region. These suggest that the expansion of the IR region occurred in the early evolution of Annonaceae, but the mechanism of expansion might be different. Large expansions of IR regions have been observed in multiple lineages (Chumley et al. 2006; Dugas et al. 2015; Zhu et al. 2016). Some studies suggested that IR region expansion was related to the Poly A region (Goulding et al. 1996; Dugas et al. 2015). Unlike the massive expansion of IR regions, IR regions are lost in the plastomes of some plant lineages and algae (Cai et al. 2017; Cauz-Santos et al. 2020; Jin et al. 2020; Karnkowska et al. 2018; Ruhlman et al. 2017).

Studies of the distribution, variation, and evolution of plastid repetitive sequences can be applied to fields such as population genetics, species identification, and genome comparisons. SSRs are often used to study polymorphisms or as molecular markers (Chmielewski et al. 2015; Fasanella et al. 2020; Huang et al. 2019). Consistent with previous studies (Gui et al. 2020; Li et al. 2021b; Ping et al. 2021a, 2023), the SSRs of *M. glochidioides* are predominantly composed of A/T motifs, with the majority located in the IGS region. However, in some other plants such as Polypodiaceae, the SSRs are dominated by C/G mononucleotides, which are presumed to be an adaptation to the environment (Gao et al. 2018; Liu et al. 2021; Schneider et al. 2004). In a previous study, the distribution of SSRs was reported to be lineage specific. Ping et al. (2021a) found that the distributions of SSRs in *Cupressus* and *Hesperocyparis* were highly uniform, and suggested that *Callitropsis funebris* was closer to *Cupressus* than other *Callitropsis* species. Recently, Zhu et al. (2021) found that the number, type, and localization of SSRs in the Cyatheaceae plastomes were genus specific, providing valid evidence for the phylogenetic analysis of Cyatheaceae. In Annonaceae, we observed lineage-specific expansion of the IR region, and the distribution pattern of SSRs may also exhibit lineage specificity. However, further exploration is needed to determine whether SSRs play a role in the construction of phylogenetic relationships in the Annonaceae family.

The codon usage bias in genes reflects the evolutionary mechanism under specific natural selection pressures and mutation pressures, representing the adaptability of species to genomic environment and natural evolutionary pressures. In the study of codon usage bias in the *M. glochidioides* plastome, a total of 30 high-frequency codons were detected, among which 29 ended with A/T, indicating a

bias for using codons ending with A/T. The ENC values of 45 genes were greater than 45, revealing relatively weak codon usage bias. Eleven optimal codons were identified, all of which ended with A/U. Neutrality plot analysis showed differences in nucleotide composition at the five positions of codons, suggesting that the impact of mutations on codon usage bias was weak and mainly influenced by natural selection. ENC plot results indicated that most genes did not fall on the expected curve, indicating significant differences between the actual ENC values and the expected ENC values. The ENC ratio of 17 genes ranged from − 0.05 to 0.05, indicating a significant impact of mutations on this group of genes. The other 36 genes were primarily influenced by natural selection. Results from the PR2 plot demonstrated a higher usage rate of thymine (T) compared to adenine (A), and a higher usage rate of guanine (G) compared to cytosine (C). Consistent with many studies, the plastome favored the usage of codons ending with A/T (Ping et al. 2023; Wang et al. 2022). The formation of codon usage bias is influenced by multiple factors, with natural selection typically playing a major role. This has been observed in various species such as *T. costata* (Ping et al. 2023), *Mesona chinensis* (Tang et al. 2021), *Panicum* (Li et al. 2021c), and Theaceae (Wang et al. 2022).

does this mean inconsistency

Due to the great diversity of Annonaceae, their phylogenetic studies have always been controversial (Chatrou et al. 2018; Maas et al. 2015; Mols et al. 2004; Ortiz-Rodriguez et al. 2018; Sauquet et al. 2003). In recent years, molecular evidence has been successfully used to solve some fuzzy classification problems (Guo et al. 2017; Xue et al. 2020; Thomas et al. 2012). In this study, we constructed the phylogenetic relationship of 27 species from Annonaceae based on chloroplast common PCGs. Our results support that Annonoideae form a monophyletic branch, and *C. hainanense* locate inside Malmeoideae. Among them, *C. hainanense* was classified as a separate category in the NCBI database. This is consistent with the results of Thomas et al. (2012), who incorporated *Chieniodendron* into *Meiogyne* (Malmeoideae). However, to date, the reported complete plastome sequences of Annonaceae are too limited. More data needs to be added and combined with morphological features to clarify the relationship and evolution process of Annonaceae.

Conclusions

Our research findings reveal the repetitive sequences and codon usage bias of the plastome of *M glochidioides*, providing valuable information for further genetic engineering studies. Additionally, we identified significant variation in the plastomes of Annonaceae. However, to date, the reported complete plastomes of Annonaceae are still limited, and it is necessary to acquire more plastomes to gain a better understanding of the evolutionary process of Annonaceae plastomes.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10528-024-10874-7.

Acknowledgements The authors thank Dr. Xiao Chunfen for her assistance with material collection.

Author Contributions Y.G. designed the study and wrote the manuscript. J.P. analyzed the data and revised the manuscript. X.L. and C.P. analyzed some of the data and revised the manuscript. All authors read and contributed to the final version of the manuscript.

Funding This work was funded by the National Natural Science Foundation of China (31800184).

Data Availability The data source is the NCBI database: https://www.ncbi.nlm.nih.gov/nuccore/MN241494,MN241495,KU563738,MT742547,MN241491,MK087989,MT742546,MN241496,MN253545,MN241488,MN241489,MH992130,MN241490,MW136266,MW829282,MZ936420,OK216144,MN253543,MN253544,OL979152,OM047203,OM914484,OM937139,MW018366,MK035708,MN016933,MK087990,NC_037005,NC_030504, (accessed on 1 June 2023). Data is provided within the manuscript or supplementary information files.

Declarations

Conflict of interest The authors declare no conflict of interest.

Ethical Approval This article does not contain any studies with human participants or animals performed by any of the authors.

Informed Consent All authors have read the manuscript and approved the submission.

References

Amenu SG, Wei N, Wu L, Oyebanji O, Hu GW, Zhou YD, Wang QF (2022) Phylogenomic and comparative analyses of *Coffeeae alliance* (Rubiaceae): deep insights into phylogenetic relationships and plastome evolution. BMC Plant Biol 22:88

Athey J, Alexaki A, Osipova E, Rostovtsev A, Santana-Quintero LV, Katneni U, Simonyan V, Kimchi-Sarfaty C (2017) A new and updated resource for codon usage tables. BMC Bioinform 18:391

Beier S, Thiel T, Münch T, Scholz U, Mascher M (2017) MISA–web: a web server for microsatellite prediction. Bioinformatics 33(16):2583–2585

Benson G (1999) Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res 27:573–580

Cai C, Wang LK, Jiang T, Zhou LJ, He PM, Jiao BH (2017) Complete chloroplast genome of green tide algae *Ulva flexuosa* (Ulvophyceae, Chlorophyta) with comparative analysis. PLoS ONE 12:e0184196

Cauz-Santos LA, da Costa ZP, Callot C, Cauet S, Zucchi MI, Bergès H, van den Berg C, Vieira MLC (2020) A repertory of rearrangements and the loss of an inverted repeat region in *Passiflora* chloroplast genomes. Genome Biol Evol 12:1841–1857

Chakraborty S, Yengkhom S, Uddin A (2020) Analysis of codon usage bias of chloroplast genes in *Oryza* species. Planta 252:67

Chaowasku T, Johnson DM, van der Ham RWJM, Chatrou LW (2015) *Huberantha*, a substitution name for *Hubera* (Annonaceae—Malmeoideae—Miliuseae). Kew Bull 70:23

Chaowasku T, Damthongdee A, Jongsook H, Nuraliev MS, Ngo D, Le HT, Lithanatudom P, Osathanunkul M, Deroin T, Xue B et al (2018) Genus *Huberantha* (Annonaceae) revisited: erection of *Polyalthiopsis*, a new genus for *H. floribunda*, with a new combination *H. luensis*. Ann Bot Fenn 55:121–136

Chatrou LW, Pirie MD, Erkens RHJ, Couvreur TLP, Neubig KM, Abbott JR, Mols JB, Maas JW, Saunders MK, Chase MW (2012) A new subfamilial and tribal classification of the pantropical plant family Annonaceae informed by molecular phylogenetics. Bot J Linn Soc 169:5–40

Chatrou LW, Turner IM, Klitgaard BB, Maas PJ, Utteridge T (2018) A linear sequence to facilitate curation of herbarium specimens of Annonaceae. Kew Bull 73:1–10

Chaw SM, Wu CS, Sudianto E (2018) Evolution of gymnosperm plastid genomes. Adv Bot Res 85:195–222

Chmielewski M, Meyza K, Chybicki IJ, Dzialuk A, Litkowiec M, Burczyk J (2015) Chloroplast microsatellites as a tool for phylogeographic studies: the case of white oaks in poland. Iforest 8:765

Chumley TW, Palmer JD, Mower JP, Fourcade HM, Calie PJ, Boore JL, Jansen RK (2006) The complete chloroplast genome sequence of *Pelargonium* × *hortorum*: organization and evolution of the largest and most highly rearranged chloroplast genome of land plants. Mol Biol Evol 23:2175–2190

Couvreur TL, Pirie MD, Chatrou LW, Saunders RM, Su YC, Richardson JE, Erkens RH (2011) Early evolutionary history of the flowering plant family Annonaceae: Steady diversification and boreotropical geodispersal. J Biogeogr 38:664–680

Dugas DV, Hernandez D, Koenen EJ, Schwarz E, Straub S, Hughes CE, Jansen RK, Nageswara-Rao M, Staats M, Trujillo JT et al (2015) *Mimosoid legume* plastome evolution: IR expansion, tandem repeat expansions and accelerated rate of evolution in *clpP*. Sci Rep 5:1–13

Fages-Lartaud M, Hundvin K, Hohmann-Marriott MF (2022) Mechanisms governing codon usage bias and the implications for protein expression in the chloroplast of *Chlamydomonas reinhardtii*. Plant J 112:919–945

Fasanella M, Souto CP, Premoli AC (2020) Preliminary cross–genera transferability of SSRs among threatened South American Cupressaceae. N Z J Bot 58:153–166

Gan YY, Xu FX (2018) The coexistence of bicellular and tricellular pollen. Grana 58:129–132

Gan YY, Liu Y, Xu FX (2015) Pollen morphology of selected species from Annonaceae. Grana 54:271–281

Gan Y, Zhang Q, Ping J (2022) The complete chloroplast genome of *Miliusa glochidioides* (Annonaceae) and phylogenetic analysis. Mitochondrial DNA B 7:1604–1605

Gao R, Wang WZ, Huang QY, Fan RF, Wang X, Feng P, Zhao GM, Bian S, Ren HL, Chang Y (2018) Complete chloroplast genome sequence of *Dryopteris fragrans* (L.) Schott and the repeat structures against the thermal environment. Sci Rep 8:1–11

Goulding SE, Wolfe KH, Olmstead RG, Morden CW (1996) Ebb and flow of the chloroplast inverted repeat. Mol Genet Genomics 252:195–206

Gui L, Jiang S, Xie D, Yu L, Huang Y, Zhang Z, Liu Y (2020) Analysis of complete chloroplast genomes of *Curcuma* and the contribution to phylogeny and adaptive evolution. Gene 732:144355

Guo X, Tang CC, Thomas DC, Couvreur TL, Saunders RM (2017) A mega–phylogeny of the Annonaceae: taxonomic placement of five enigmatic genera and support for a new tribe, Phoenicantheae. Sci Rep 7:1–11

Huang CY, Yin QY, Khadka D, Meng KK, Fan Q, Chen SF, Liao WB (2019) Identification and development of microsatellite (SSRs) makers of *Exbucklandia* (Hamamelidaceae) by high–throughput sequencing. Mol Biol Rep 46:3381–3386

Ingvarsson PK (2008) Molecular evolution of synonymous codon usage in *Populus*. BMC Evol Biol 8:307

Jin DM, Wicke S, Gan L, Yang JB, Jin JJ, Yi TS (2020) The loss of the inverted repeat in the putranjivoid clade of Malpighiales. Front Plant Sci. https://doi.org/10.3389/fpls.2020.00942

Karnkowska A, Bennett MS, Triemer RE (2018) Dynamic evolution of inverted repeats in Euglenophyta plastid genomes. Sci Rep 8:1–10

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A (2012) Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28:1647–1649

Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol 35:1547

Kurtz S, Schleiermacher C (1999) REPuter: fast computation of maximal repeats in complete genomes. Bioinformatics 15:426–427

Lei JM, Liang ZR, Zhang H, Lim YQ, Xue BE (2022) Advances in exploitation and utilization of wild fruit resources of Annonaceae in China. J Fruit Sci 39:121–130 (**in Chinese**)

Li Q, Wei R (2022) Comparison of boraginales plastomes: insights into codon usage bias, adaptive evolution, and phylogenetic relationships. Diversity 14:1104

Li X, Yang Y, Henry RJ, Rossetto M, Wang Y, Chen S (2015) Plant DNA barcoding: from gene to genome. Biol Rev 90:157–166

Li L, Hu YF, He M, Zhang B, Wu W, Cai PM, Huo D, Hong YC (2021a) Comparative chloroplast genomes: Insights into the evolution of the chloroplast genome of *Camellia sinensis* and the phylogeny of *Camellia*. BMC Genomics 22:138

Li XP, Zhao YM, Tu XD, Li CR, Zhu YT, Zhong ZJ, Wu SS, Zhai JW (2021b) Comparative analysis of plastomes in Oxalidaceae: phylogenetic relationships and potential molecular markers. Plant Divers 43:281–291

Li G, Zhang L, Xue P (2021c) Codon usage pattern and genetic diversity in chloroplast genomes of *Panicum* species. Gene 802:145866

Li ZZ, Lehtonen S, Chen JM (2023) The dynamic history of plastome structure across aquatic subclass Alismatidae. BMC Plant Biol 23:125

Liu Q (2012) Mutational bias and translational selection shaping the codon usage pattern of tissue–specific genes in rice. PLoS ONE 7:e48295

Liu Y (2020) A code within the genetic code: codon usage regulates co–translational protein folding. Cell Commun Signal 18:145

Liu SS, Wang Z, Su YJ, Wang T (2021) Comparative genomic analysis of Polypodiaceae chloroplasts reveals fine structural features and dynamic insertion sequences. BMC Plant Biol 21:1–15

Maas PJM, Westra LYT, Guerrero SA, Lobão AQ, Scharf U, Zamora NA, Erkens RHJ (2015) Confronting a morphological nightmare: revision of the neotropical genus *Guatteria* (Annonaceae). Blumea-Bio Evol Biogeogra Plants 60:1–219

Mazumdar P, Othman RB, Mebus K, Ramakrishnan N, Harikrishna JA (2017) Codon usage and codon pair patterns in non–grass monocot genomes. Ann Bot-London 120:893909

Mols JB, Gravendeel B, Chatrou LW, Pirie MD, Bygrave PC, Chase MW, Keßler PJ (2004) Identifying clades in Asian Annonaceae: monophyletic genera in the polyphyletic *Miliuseae*. Am J Bot 91:590–600

Mower JP, Vickrey TL (2018) Structural diversity among plastid genomes of land plants. Adv Bot Res 85:263–292

Ortiz-Rodriguez AE, Ornelas JF, Ruiz-Sanchez E (2018) A jungle tale: molecular phylogeny and divergence time estimates of the *Desmopsis-Stenanona* clade (Annonaceae) in Mesoamerica. Mol Phylogenet Evol 122:80–94

Parvathy ST, Udayasuriyan V, Bhadana V (2022) Codon usage bias. Mol Biol Rep 49:539–565

Ping JY, Feng PP, Li JY, Zhang RJ, Su YJ, Wang T (2021a) Molecular evolution and SSRs analysis based on the chloroplast genome of *Callitropsis funebris*. Ecol Evol 11:4786–4802

Ping JY, Feng PP, Hao J, Li JY, Su YJ, Wang T (2021b) The molecular evolution pattern of *rps12* gene in gymnosperms. Chin Sci Bull 66:3182–3193 (**in Chinese**)

Ping JY, Zhong XN, Wang T, Su YJ (2023) Structural characterization of *Trivalvaria costata* chloroplast genome and molecular evolution of *rps12* gene in Magnoliids. Forests 14:1101

Punyasena SW, Eshel G, McElwain JC (2008) The influence of climate on the spatial patterning of Neotropical plant families. J Biogeogr 35:117–130

Quax TE, Claassens NJ, Söll D, van der Oost J (2015) Codon bias as a means to fine–tune gene expression. Mol Cell 59:149–161

Raubeson LA, Jansen RK (1992) Chloroplast DNA evidence on the ancient evolutionary split in vascular land plants. Science 255:1697–1699

Raubeson LA, Jansen RK (2005) Chloroplast genomes of plants. In: Henry RJ (ed) Plant diversity and evolution: genotypic and phenotypic variation in higher plants. CAB International, London, pp 45–68

Ruhlman TA, Jansen RK (2014) The plastid genomes of flowering plants. In: Maliga P (ed) Chloroplast biotechnology: methods and protocols. Springer, New York, pp 3–38

Ruhlman TA, Zhang J, Blazier JC, Sabir JS, Jansen RK (2017) Recombination–dependent replication and gene conversion homogenize repeat sequences and diversify plastid genome structure. Am J Bot 104:559–572

Saunders RMK, Guo X, Tang CC (2020) *Friesodielsia subaequalis* (Annonaceae) a new nomenclatural combination following conservation of the generic name against *Schefferomitra*. Phytotaxa 464:183–184

Sauquet H, Doyle JA, Scharaschkin T, Borsch T, Hilu KW, Chatrou LW, Le Thomas A (2003) Phylogenetic analysis of Magnoliales and Myristicaceae based on multiple data sets: Implications for character evolution. Bot J Linn Soc 142:125–186

Sawicki J, Bączkiewicz A, Buczkowska K, Górski P, Krawczyk K, Mizia P, Myszczyński K, Ślipiko M, Szczecińska M (2020) The increase of simple sequence repeats during diversification of marchantiidae, an early land plant lineage, leads to the first known expansion of inverted repeats in the evolutionarily–stable structure of liverwort plastomes. Genes 11:299

Schneider H, Schuettpelz E, Pryer KM, Cranfill R, Magallón S, Lupia R (2004) Ferns diversified in the shadow of angiosperms. Nature 428(6982):553–557
Stamatakis A (2014) RaxML version 8: a tool for phylogenetic analysis and post–analysis of large phylogeies. Bioinformatics 30:1312–1313
Tang D, Wei F, Cai Z, Wei Y, Khan A, Miao J, Wei K (2021) Analysis of codon usage bias and evolution in the chloroplast genome of *Mesona chinensis* Benth. Dev Genes Evol 231:1–9
Thomas DC, Surveswaran S, Xue B, Sankowsky G, Mols JB, Keßler PJ, Saunders RM (2012) Molecular phylogenetics and historical biogeography of the *Meiogyne-Fitzalania* clade (Annonaceae): generic paraphyly and late Miocene-Pliocene diversification in Australasia and the Pacific. Taxon 61:559–575
Tonti-Filippini J, Nevill PG, Dixon K, Small I (2017) What can we do with 1000 plastid genomes? Plant J 90:808–818
Wang Z, Cai Q, Wang Y, Li M, Wang C, Wang Z, Jiao C, Xu C, Wang H, Zhang Z (2022) Comparative analysis of codon bias in the chloroplast genomes of theaceae species. Fronteir Genet 13:824610
Weng ML, Blazier JC, Govindu M, Jansen RK (2014) Reconstruction of the ancestral plastid genome in geraniaceae reveals a correlation between genome rearrangements, repeats, and nucleotide substitution rates. Mol Biol Evol 31:645–659
Wu S, Chen J, Li Y, Liu A, Li A, Yin M, Shrestha N, Liu J, Ren G (2021) Extensive genomic rearrangements mediated by repetitive sequences in plastomes of *Medicago* and its relatives. BMC Plant Biol 21:1–16
Xue B, Tan YH (2016) Excluding *Miliusa velutina* (Annonaceae) from Flora of China. Phytotaxa 282:166–169
Xue B, Ding HB, Yao G, Shao YY, Fan XJ, Tan YH (2020) From *Polyalthia* to *Polyalthiopsis* (Annonaceae): transfer of species enlarges a previously monotypic genus. PhytoKeys 148:71
Xue B, Nurmawati S, Xu Y, Li YQ (2021) Excluding the species *Monoon fragrans* (Annonaceae) from the flora of China. Phytotaxa 487:91–96
Zhou XM, Zhao J, Yang JJ, Le Péchon T, Zhang L, He ZR, Zhang LB (2022) Plastome structure, evolution, and phylogeny of *Selaginella*. Mol Phylogenet Evol 169:107410
Zhu AD, Guo WH, Gupta S, Fan WS, Mower JP (2016) Evolutionary dynamics of the plastid inverted repeat: the effects of expansion, contraction, and loss on substitution rates. New Phytol 209:1747–1756
Zhu M, Feng PP, Ping JY, Li JY, Su YJ, Wang T (2021) Phylogenetic significance of the characteristics of simple sequence repeats at the genus level based on the complete chloroplast genome sequences of Cyatheaceae. Ecol Evol 11:14327–14340

Authors and Affiliations

Yangying Gan[1] · Jingyao Ping[2] · Xiaojing Liu[1] · Caixia Peng[3]

✉ Yangying Gan
sharelot@163.com

Jingyao Ping
pingjy@mail.sysu.edu.cn

Xiaojing Liu
xiaojing.e.liu@foxmail.com

Caixia Peng
pengcx@scbg.ac.cn

[1] Key Laboratory of Urban Agriculture in South China, Ministry of Agriculture and Rural Affairs, Institute of Agricultural Economics and Information, Guangdong Academy of Agricultural Sciences, Guangzhou 510640, China

[2] College of Life Sciences, Sun Yet-sen University, Guangzhou 510275, China

[3] Horticulture Center, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou 510650, China