Codon usage bias in chloroplast genes implicate adaptive evolution of four ginger species

Qian Yang[1], Cheng Xin[2], Qing-Song Xiao[1], Ya-Ting Lin[1], Li Li[1]* and Jian-Li Zhao[1]*

[1]Ministry of Education Key Laboratory for Transboundary Ecosecurity of Southwest China, Yunnan Key Laboratory of Plant Reproductive Adaptation and Evolutionary Ecology and Centre for Invasion Biology, Institute of Biodiversity, School of Ecology and Environmental Science, Yunnan University, Kunming, Yunnan, China, [2]School of Life Sciences, Zhengzhou University, Zhengzhou, China

Codon usage bias (CUB) refers to different codons exhibiting varying frequencies of usage in the genome. Studying CUB is crucial for understanding genome structure, function, and evolutionary processes. Herein, we investigated the codon usage patterns and influencing factors of protein-coding genes in the chloroplast genomes of four sister genera (monophyletic *Roscoea* and *Cautleya*, and monophyletic *Pommereschea* and *Rhynchanthus*) from the Zingiberaceae family with contrasting habitats in southwestern China. These genera exhibit distinct habitats, providing a unique opportunity to explore the adaptive evolution of codon usage. We conducted a comprehensive analysis of nucleotide composition and codon usage on protein-coding genes in the chloroplast genomes. The study focused on understanding the relationship between codon usage and environmental adaptation, with a particular emphasis on genes associated with photosynthesis. Nucleotide composition analysis revealed that the overall G/C content of the coding genes was ‹ 48%, indicating an enrichment of A/T bases. Additionally, synonymous and optimal codons were biased toward ending with A/U bases. Natural selection is the primary factor influencing CUB characteristics, particularly photosynthesis-associated genes. We observed differential gene expressions related to light adaptation among sister genera inhabiting different environments. Certain codons were favored under specific conditions, possibly contributing to gene expression regulation in particular environments. This study provides insights into the adaptive evolution of these sister genera by analyzing CUB and offers theoretical assistance for understanding gene expression and regulation. In addition, the data support the relationship between RNA editing and CUB, and the findings shed light on potential research directions for investigating adaptive evolution.

1 Introduction

Plant codons are three-base (nucleotide) units in DNA sequences. Codons act as links between nucleic acids (DNA and RNA) and proteins, providing instructions for translating the genetic information in DNA from RNA into proteins, and play a crucial role in the process of genetic information transfer (Radwan et al., 2010). Plants contain 64 codons, with each codon corresponding to an amino acid or termination signal. Codon usage bias (CUB) refers to the unequal use of synonymous codons during gene transcription (Jia and Xue, 2009; Wei et al., 2014). This phenomenon is widely observed in numerous organisms in a non-random manner and believed to reflect certain evolutionary events in cellular and environmental adaptations (Botzman and Margalit, 2011). CUB can be influenced by various factors, including genetic and environmental factors, as well as evolutionary pressures (Sharp et al., 1988; Shackelton et al., 2006; Tao et al., 2009; Pandey et al., 2020). Current research suggests that CUB is primarily the result of a balance between mutational bias and natural selection pressure (Gu et al., 2004; Tyagi et al., 2023). Therefore, studying CUB is essential for understanding the functionality and evolutionary mechanisms of plant genomes and can contribute to understanding the evolutionary and environmental adaptations of related species.

The chloroplast (cp) genome of typical angiosperms comprises a circular, double-stranded DNA molecule ranging from 120 to 160 kb (Yang et al., 2020). It contains approximately 130 genes distributed within two identical inverted repeat regions, referred to as the large and small single-copy regions (Xin and Yang, 2020; Yang et al., 2021). Due to its stable structure and uniparental inheritance, the cp genome has been used for comparative analyses to understand the evolution of closely related species. CUB plays a significant role in gene expression, therefore, research on codons in the cp genome is highly important (Song et al., 2017; Yang et al., 2018). By studying the codon usage patterns of plants from different habitats, researchers can identify specific codons or amino acids that are preferred or avoided in response to environmental factors, such as temperature, light and nutrient availabilities, and water stress. Analyzing codon usage patterns in stress-related genes, of which CUB may be associated, can provide deeper insights into how plants respond to biotic and abiotic stresses, such as drought, salinity, pathogens, or extreme temperatures. For example, studies have shown correlations between gene expression levels and synonymous codon usage in rice under biotic or abiotic stresses, such as drought stress (Chamani Mohasses et al., 2020; Tyagi et al., 2023). However, from the perspective of codon usage patterns, research on the adaptive evolution of plants to different habitats is relatively limited at present.

Recent research has revealed the potential relationship between plant RNA editing and CUB. RNA editing is a post-transcriptional modification process that alters the coding sequence of mRNA, consequently influencing protein translation and function (Wang et al., 2015; Wang et al., 2016). This phenomenon plays a critical role in gene expression regulation and adaptation in plants (Ichinose and Sugita, 2016). Similarly, CUB involves the preferential usage of certain synonymous codons over others and can affect gene regulation. Although the relationship between RNA editing and CUB is not direct, both processes may contribute to the overall regulation of plant genes. Theoretically, CUB could influence RNA editing in specific regions of the genome, and the different frequencies of codon usage may affect the recognition and action of RNA-editing enzymes at specific sites (Monian et al., 2022). As a result, the preference for certain codons can affect the frequency and location of RNA editing events. Therefore, the relationship between plant RNA editing and CUB is a promising area of investigation. Both processes have been implicated in gene expression and may contribute to plant adaptation. Understanding these interactions may provide crucial insights into the regulatory mechanisms underlying plant gene expression and evolution. Further research efforts encompassing diverse experimental approaches and sophisticated analyses are required to unravel the intricate relationship between RNA editing and CUB in the context of plant genetics.

Zingiberaceae plants exhibit rich species diversity and habitat variability from the equator to the mountains (Zhao et al., 2022). Certain species tolerate strong light, whereas others thrive in the shade, for instance, sister genera with distinct habitat preferences, such as *Roscoea* versus *Cautleya* and *Pommereschea* versus *Rhynchanthus*. The phylogenetic relationships of these four sister genera are (((*Roscoea*, *Cautleya*), (*Pommereschea*, *Rhynchanthus*)), outgroups) (Supplementary Figure S1) (Kress et al., 2002; Yang et al., 2021). *Roscoea* and *Pommereschea* grow primarily in open forests, meadows, and mountaintops and endure a certain degree of intense light. In contrast, *Cautleya* and *Rhynchanthus* grow epiphytically on shaded tree trunks or rocky streams, displaying shade adaptation. This study aimed to explore their adaptation to different habitats by investigating their codon preferences. In this study, we conducted a comprehensive comparative analysis of CUB in the cp genomes of four sister genera. Several CUB parameters, including the effective number of codons (ENc) and relative synonymous codon usage (RSCU), were estimated (Sharp et al., 2005; Angellotti et al., 2007). The primary objectives of this study encompass the following: 1) To investigate whether mutation pressure or selection pressure predominantly influences CUB in four taxonomic groups. 2) To determine if the optimal codon usage is correlated with habitat adaptation. 3) To explore the potential relationship between RNA editing and CUB.

2 Materials and methods

2.1 Sequence acquisition and filtering

The complete cp genome sequences of *Roscoea tibetica* (NC047420) were downloaded from GenBank (https://www.ncbi.nlm.nih.gov), and the complete cp genomes of *Cautleya gracilis*, *Pommereschea lackneri*, and *Rhynchanthus beesianus* (MW769781–MW769783) were assembled. We collected fresh leaves from three species in the wild (99.70°E, 24.18°N; 101.23°E, 21.99°N; 99.50°E, 22.48°N) and generated sequencing data for each

species using the Illumina Hiseq 2500 platform. The sequencing data produced 45 gigabases of data for each species, with read counts of 277,483,161, 631,731,352, and 691,955,913, respectively. Subsequently, we performed *de novo* assembly of the raw data using the GetOrganelle tool with parameters (-R 15 -k 105,121; Jin et al., 2020). We annotated the coding sequences (CDS) using two different methods. Firstly, we selected complete and high-quality chloroplast genomes from closely related species within the Zingiberaceae family as references. The automatic annotation was performed using the GeSeq and DOGMA annotation tools (Wyman et al., 2004; Tillich et al., 2017). Secondly, we used chloroplast genomes of closely related species reported in the literature as reference sequences. The assembled chloroplast genomes were aligned with the reference sequences in Geneious software (Biomatters Ltd., Auckland, New Zealand) using the MAFFT tool (Katoh and Standley, 2013). Open reading frames were identified, and manual annotation for known genes was conducted based on similarity and positional information. To ensure accuracy, the consistency of coding sequences obtained from both annotation methods was further confirmed through BLAST searches. The accurately annotated CDS were then utilized for subsequent codon usage bias studies.

2.2 Content calculation and RSCU analysis

Perl programming scripts were used to calculate the GC content of the filtered CDSs, as well as the GC content at the first, second, and third positions of the codons (GC1, GC2, and GC3, respectively). Additionally, CodonW software (version 1.4.2, https://codonw.sourceforge.net/) was employed to calculate the RSCU, which represents the relative probability of specific codon usage among synonymous codons encoding the same amino acid (Sharp and Li, 1986). RSCU was determined by comparing the frequency of a codon with the average frequency of all codons encoding the corresponding amino acid. The RSCU value of a codon equals 1 if there is no codon preference. An RSCU value that exceeds 1 indicates a significant positive bias in codon usage. Conversely, an RSCU value less than 1 indicates a notable negative bias in codon usage (Sharp et al., 1993).

2.3 ENc and optimal codon analysis

The ENc refers to the number of codons effectively used in a gene (Wright, 1990). This reflects the deviation of codon usage from random selection and serves as an important indicator of the degree of synonymous CUB. The ENc values ranged from 20 (each amino acid used only one synonymous codon) to 61 (all synonymous codons were used equally). ENc and CAI values were calculated using the CodonW software (version 1.4.2) for the four species. Based on the ENc and CAI values, genes were rearranged, and the top and bottom 5% of genes were selected as high- and low-expression datasets, respectively. The optimal codons were determined using ΔRSCU with a threshold of 0.08. If the ΔRSCU

value was > 0.08 and the absolute RSCU value was > 1, the codon was defined as optimal (Zhang et al., 2011).

2.4 ENc plot analysis

ENc-plot analysis was used to examine the codon usage patterns of genes. The expected ENc value was calculated using the following formula (Wright, 1990):

$$ENc = 2 + GC3_S + 29/[GC3_S^2 + (1 - GC3_S)^2]$$

To explore the factors influencing CUB, an ENc-GC3s plot was generated, with GC3s and ENc on the x- and y-axes, respectively. GC3s are important nucleotide composition indicators that refer to the content of G and C at the third position of codons (excluding Trp, Met, and stop codons). If the CUB is primarily influenced by mutations, the genes will be located on or close to the standard curve. Conversely, if the CUB is influenced by natural selection, the genes will be located below the standard curve (Lü et al., 2005).

2.5 PR2-plot analysis

PR2-plot analysis is commonly used to estimate the effects of mutations and natural selection on CUB, specifically by analyzing the composition of the four bases at the third position of codons. Theoretically, the bases used at the third position of the codon should follow the PR2 rule (A=T and C=G). A 2D plot was constructed with $A_3/(A_3+T_3)$ and $G_3/(G_3+C_3)$ on the y- and x-axes, respectively (Sueoka, 1999). The vectors emanating from the center point (A=T, G=C) reflect the extent and direction of the base deviations (Wang et al., 2021). If the vectors are evenly distributed around the center point, mutations may entirely cause codon bias.

2.6 Neutrality plot analysis

Neutrality plot analysis was used to estimate the degree to which mutations and natural selection influenced CUB (Sueoka, 1988). A regression curve was plotted with GC12 (the average of GC1 and GC2) on the y-axis and GC3 on the x-axis. If GC12 is significantly correlated with GC3, and the slope of the regression curve is close to or equal to 1, the mutation is the primary factor affecting CUB. Conversely, if the slope is close to or equal to 0, then natural selection is the primary factor influencing CUB (Huo et al., 2021).

2.7 Determination of gene expression and RNA editing sites

Before the flowering period, in the afternoon under clear weather conditions, we collected leaves from plants in four locations (102.91°E, 24.76°N; 99.70°E, 24.18°N; 101.23°E, 21.99°N; and 99.50°E, 22.48°N) for these four species. We selected leaves from plants exhibiting normal growth and development. The leaves

were rinsed with RNase-free water and dried with absorbent paper to remove any contaminants. Subsequently, the samples were quickly placed in liquid nitrogen for 30 minutes and then transferred for long-term storage at -80°C. For each plant population, we selected no fewer than three individuals with consistent growth as the source of biological replicate samples, aiming to minimize potential variations in results due to individual differences. The frozen samples were sent to Novogene Corporation for RNA extraction and transcriptome sequencing. These data are stored in NCBI under the accession number PRJNA997822, with the specific samples labeled as SRR25433125 to SRR25433136. The high-quality clean reads obtained were aligned to the reference sequences using the HISAT2 tool (Kim et al., 2019), and the resulting alignment results were visualized using the IGV tool (Robinson et al., 2023). The expression levels are typically quantified in Fragments Per Kilobase of transcript per Million fragments mapped (FPKM) values. The main process involved using the featureCounts function from the Rsubread library to process specified BAM files and GTF annotation files (Liao et al., 2014). It conducted read counting based on the mapping to genomic features, such as genes. The resulting read count information was stored in the fCountsList variable. Subsequently, the script utilized the rpkm function to calculate FPKM values from the obtained count data. To explore the potential relationship between RNA editing and CUB, the PREP-cp algorithm was employed to predict the RNA editing sites (Mower, 2009). A parameter threshold (cutoff value) of 0.8 was used to ensure prediction accuracy. The predicted editing sites were further validated by integrating transcriptomic data to confirm their authenticity.

3 Results

3.1 Nucleotide composition and RSCU analysis of the cp genes in four genera of sister plants

The overall nucleotide composition and codon usage at the third position of the codons were determined for the cp genes of the four sister plant species (*R. tibetica*, *C. gracilis*, *P. lackneri*, and *R. beesianus*). Among the four sister genera, the nucleotide T showed

the highest usage, followed by A, G, and C. This uneven distribution of A, T, G, and C nucleotides indicated an A/T bias in the cp genomes. Based on their biological functions, the encoded genes were categorized as those related to photosynthesis, transcription, translation, or other miscellaneous functions. For photosynthesis-related genes, the nucleotide T had the highest usage, followed by A, C, and G. For transcription- and translation-related genes, the nucleotide A had the highest usage, followed by T, G, and C. Nucleotide usage of the other genes was consistent with the overall trend (Supplementary Figure S2). Additionally, compared with codons ending in G/C, cp genes showed a greater preference for codons ending in A/T, with a higher enrichment of A/T bases at the third codon position. The GC content (%) of the cp genes in the four sister species did not exceed 48%, and the GC content at different positions varied. Notably, both the overall GC content and GC content for the three types of genes followed the trend GC1 > GC2 > GC3, further indicating a bias toward A and T at the three codon positions (Supplementary Figure S3). RSCU (relative synonymous codon usage) analysis of the codons revealed that all four sister genera had 29 synonymous codons with relative usage rates > 1 (Figure 1), of which seven codons were overrepresented (RSCU > 1.6). Both photosynthesis-related and transcription- and translation-related genes had nine overrepresented codons, while four codons were overrepresented in the other gene category (Figure 2). The average ENc (effective number of codons) values for both the four sister genera and photosynthesis-related genes ranged between 46 and 47. In contrast, the average ENc values for transcription- and translation-related genes ranged between 45 and 46, whereas for other genes, the average ENc values were between 49 and 50. These results indicated the presence of certain codon preferences.

3.2 ENc-plot analysis

To further investigate the relationship between nucleotide composition and codon usage preference, ENc-plot analysis was conducted to study the factors influencing codon usage, including mutation pressure and natural selection. If mutation pressure is the primary factor influencing codon usage preference, the real ENc values should align closely with the ENcexp curve. Conversely, if natural selection is the primary factor affecting the codon usage
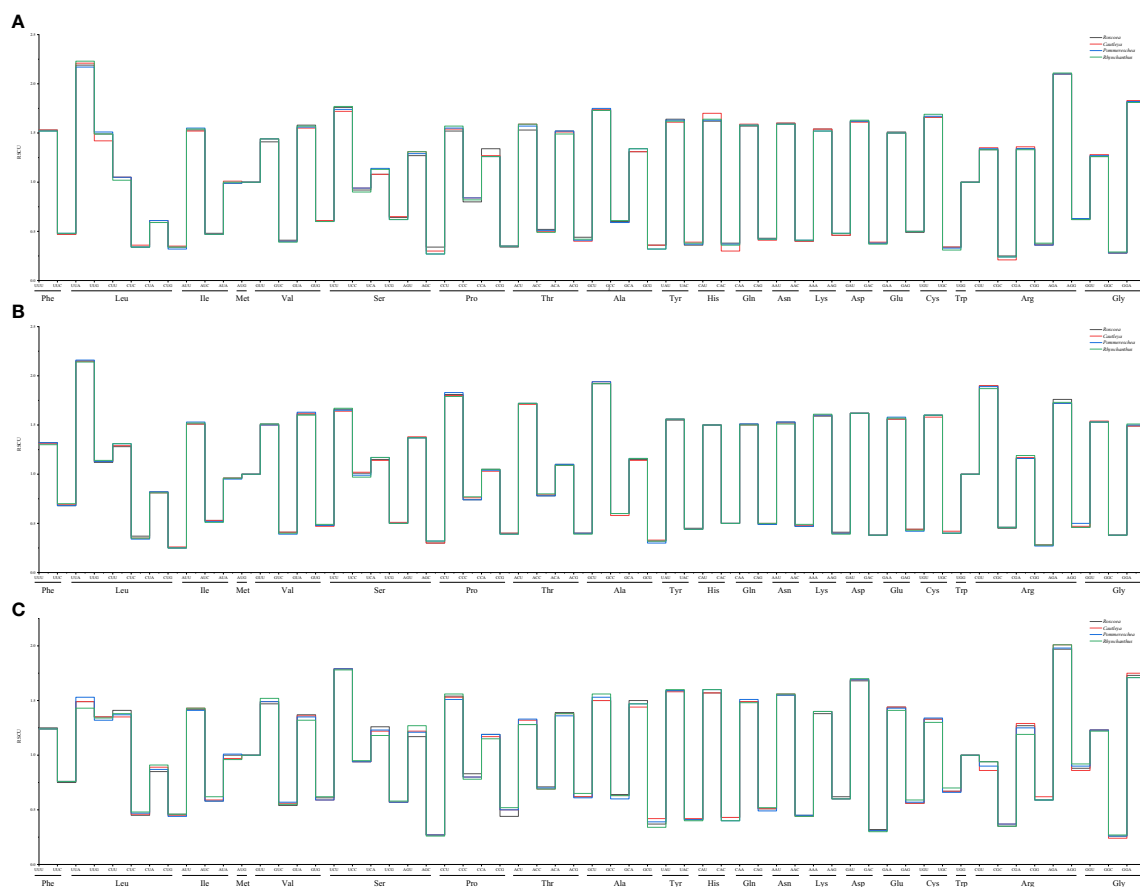

FIGURE 1
Relative synonymous codon use (RSCU) of chloroplast (cp) genes in four sister genera.
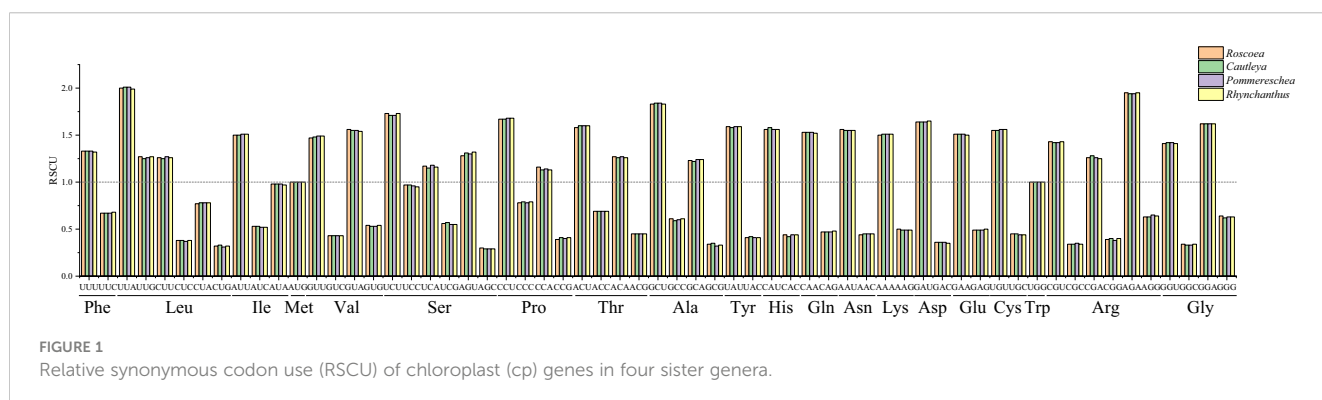
FIGURE 2
RSCU of photosynthesis-related genes in four sister genera **(A)**; RSCU of transcription- and translation-related genes in four sister genera **(B)**; RSCU of other genes in four sister genera **(C)**.

preference, the real ENc values will deviate from the ENcexp curve, thus positioning below the curve (Gao et al., 2022). Among both sister genera groups, certain genes related to photosynthesis, transcription, and translation were distributed below the standard curve, with ENc values between 35 and 57, thus indicating a certain level of CUB (Figure 3; Supplementary Figure S4). In addition to mutations, natural selection may influence the variability in codon usage.

3.3 PR2-plot analysis

To understand the impact of mutations and selection pressures on codon usage in cp genes, a parity rule 2 (PR2) bias plot analysis was conducted to explore the relationship between the third base of codons. Equal nucleotide usage frequencies indicate that only mutation pressure influences CUB (Kawabe and Miyashita, 2003). Conversely, natural selection may enable the unequal usage of A and T with G and C bases (Sueoka, 1995). Herein, for both pairs of sister genera, the $A_3/(A_3+T_3)$ ratio of genes related to photosynthesis primarily showed a distribution < 0.5, suggesting a higher frequency of T bases than that of A bases. Additionally, the $G_3/(G_3+C_3)$ ratio showed a distribution of 0.5, indicating a higher

frequency of G bases than that of C bases. Regarding genes related to transcription and translation, the $A_3/(A_3+T_3)$ ratio for both pairs of sister genera primarily showed a distribution of 0.5, indicating a higher frequency of A bases than that of T bases. The $G_3/(G_3+C_3)$ ratio, in contrast, predominantly showed a distribution > 0.5, indicating a higher frequency of G bases than that of C bases (Figure 4; Supplementary Figure S5). These results further demonstrate that both base mutations and natural selection jointly influence codon usage preferences in the cp genomes of the two pairs of sister genera, with natural selection primarily affecting codon usage preference.

3.4 Neutrality plot

A separate neutral plot analysis of the cp genes of the two pairs of sister genera was conducted to investigate the relationships between the bases at the three codon positions. No significant correlations between the GC12 and GC3 genes related to photosynthesis were found in *Roscoea* and *Cautleya*, indicating that natural selection may influence their CUB (Nie et al., 2012). A similar phenomenon was observed for *Pommereschea* and *Rhynchanthus*. Neutral plot analysis further explained the roles of
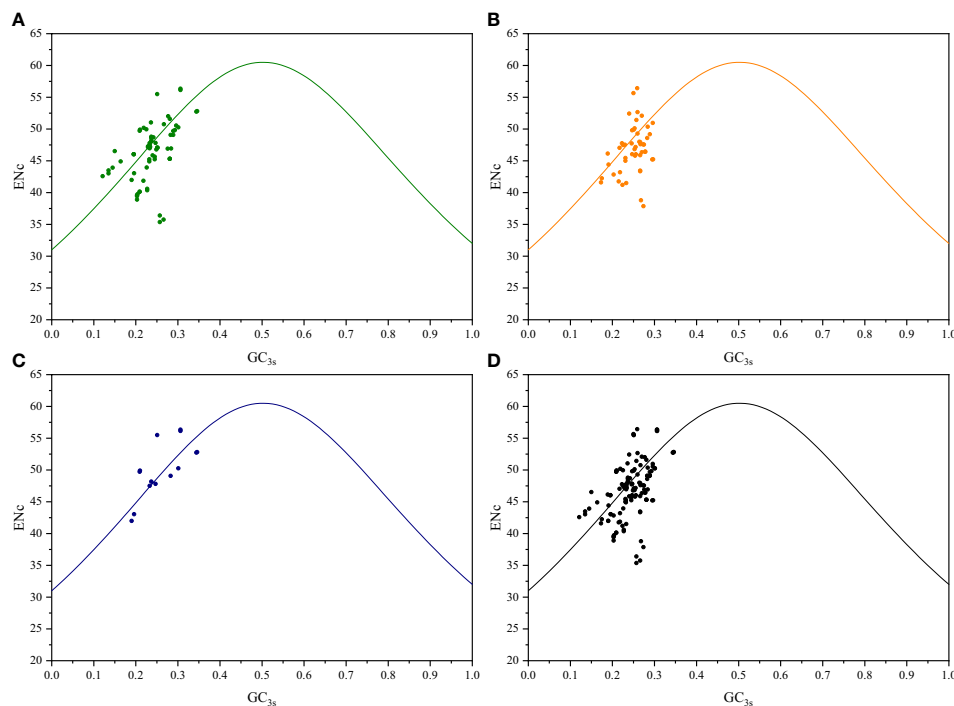
FIGURE 3
ENc-plot analysis of *Roscoea* and *Cautleya* cp genes. **(A)** represents genes related to photosynthesis, **(B)** represents genes related to transcription and translation, **(C)** represents other genes, and **(D)** represents all genes.

mutation pressure and natural selection. When the slope of the regression curve is approximately 0, natural selection plays a dominant role in the CUB, whereas a slope of approximately 1 indicates the dominance of mutation pressure (Butt et al., 2016; Zhao et al., 2016). For genes related to photosynthesis, transcription and translation, other genes, and all genes in *Roscoea* and *Cautleya*, the absolute values of the slopes of the regression curves were 0.07, 0.67, 0.55, and 0.14, respectively. Similar results were obtained for *Pommereschea* and *Rhynchanthus* with absolute values of the regression curve slopes of 0.09, 0.62, 0.50, and 0.10, respectively (Figure 5; Supplementary Figure S6). These results further indicate that natural selection primarily influences codon usage patterns in the cp genomes of the two pairs of sister genera, particularly in genes related to photosynthesis, whereas base mutations play a secondary role.

3.5 Optimal codon determination

To determine the optimal codons in the protein-coding sequences of the cp genomes of the four sister genera, genes were sorted based on their ENc and CAI values. The top and bottom 5% of the sequences were selected within their corresponding ENc and CAI value ranges as high- and low-expression gene samples, respectively. Using previously calculated RSCU values for all synonymous codons, 10, 11, 11, and 9 optimal codons were identified for *Roscoea*, *Cautleya*, *Pommereschea*, and *Rhynchanthus*, respectively. The majority of optimal codons ended with A/U (except for UUG). Six optimal codons were

shared among genera: CUU (Leu), GUU (Val), GUA (Val), UCA (Ser), ACA (Thr), and GCA (Ala). These findings suggest a preference for A/U at the third position of their codons. Notably, the two optimal codons (AGU and GAU) were present only in *Roscoea* and *Pommereschea* (Table 1), which inhabit similar environments. Therefore, the presence of these codons may be related to their adaptation to similar habitats.

3.6 Relationship between codon usage and gene expression

To gain a better understanding of the relationship between codon usage and gene expression, the expression levels of cp genes were obtained from the section 2.7, as detailed in the Materials and Methods section. Through a comparison of gene expression levels across the four taxa, we identified eight coding genes with differential expression in different environments. Among the seven genes analyzed, *Cautleya* and *Rhynchanthus* exhibited higher expression levels of *psbA*, *psaB*, *ndhI*, *matK*, *rps3*, *rps7*, and *rps14* than those of *Roscoea* and *Pommereschea*, whereas *Roscoea* and *Pommereschea* showed higher expression levels of *rbcL* than those for *Cautleya* and *Rhynchanthus* (Figure 6). The ENc values for these eight genes were < 50, indicating a relatively strong CUB. In addition, the third base of these codons typically ended in A/U (Supplementary Figure S7 and Supplementary Table S1). Notably, the *psbA* gene contained no codon for Leu.

Furthermore, the RNA editing of these eight genes in the four sister genera were investigated. PREP-cp prediction was used and
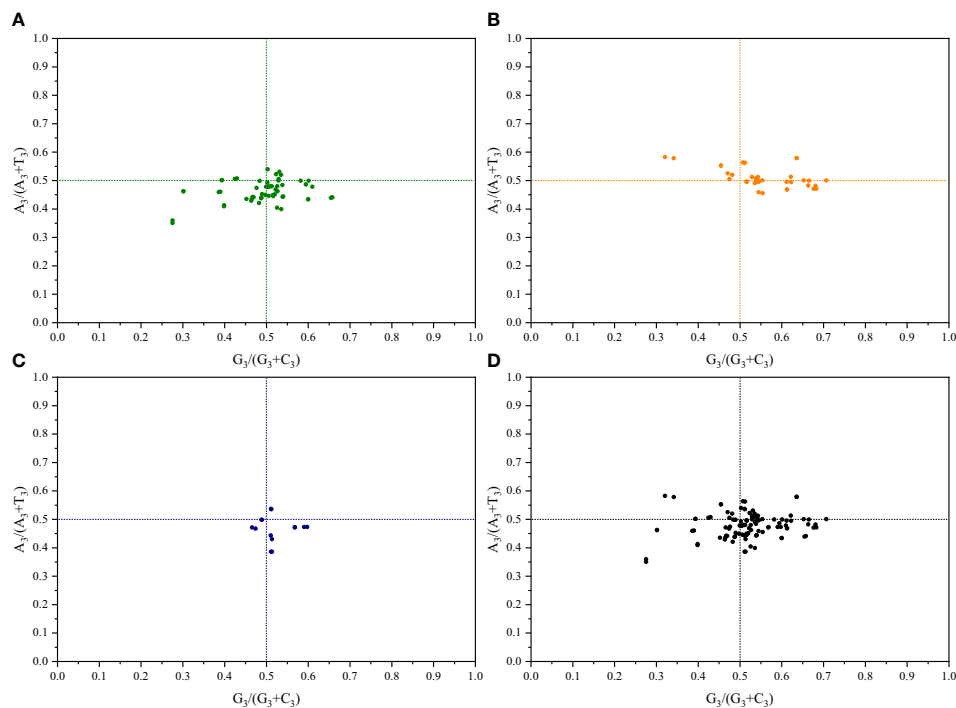
FIGURE 4

PR2-plot analysis of *Roscoea* and *Cautleya* cp genes. **(A)** represents genes related to photosynthesis, **(B)** represents genes related to transcription and translation, **(C)** represents other genes, and **(D)** represents all genes.
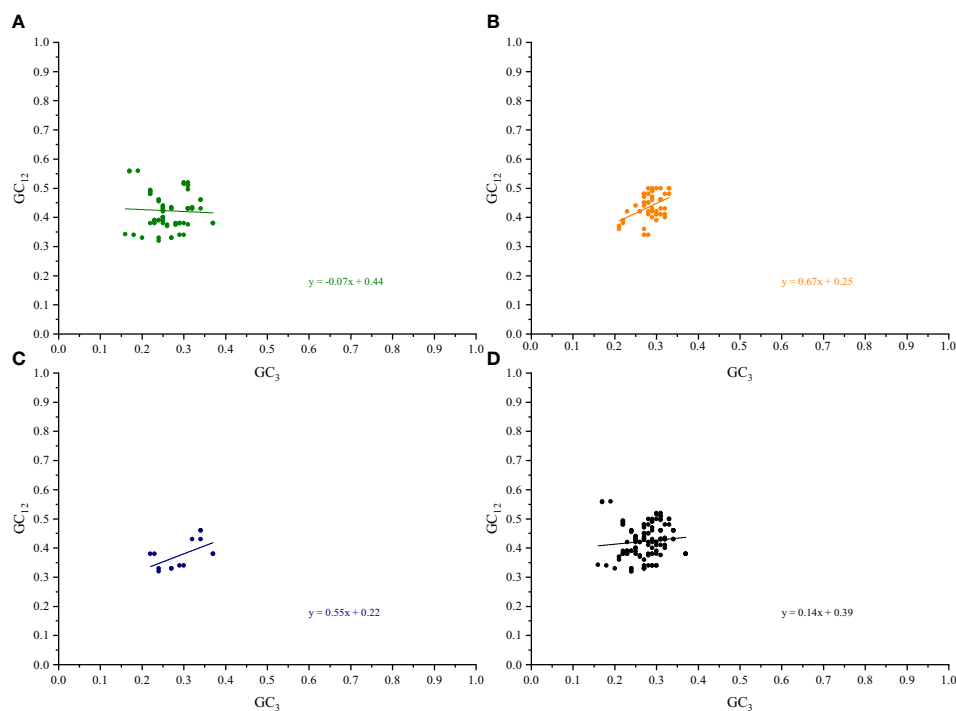
FIGURE 5

Neutrality plot analysis of *Roscoea* and *Cautleya* cp genes. **(A)** represents genes related to photosynthesis, **(B)** represents genes related to transcription and translation, **(C)** represents other genes, and **(D)** represents all genes.

TABLE 1 Optimal codons of chloroplast genes in four sisters genera.

		Roscoea		Cautleya		Pommereschea		Rhynchanthus	
		RSCU	ΔRSCU	RSCU	ΔRSCU	RSCU	ΔRSCU	RSCU	ΔRSCU
Leu	UUG			1.25	0.25				
	CUU	1.26	0.51	1.25	0.38	1.27	0.62	1.26	0.70
Val	GUU	1.47	0.43	1.48	0.40	1.49	0.37	1.49	0.53
	GUA	1.56	0.36	1.55	0.35	1.55	0.43	1.54	0.16
Ser	UCA	1.17	0.77	1.15	1.03	1.18	0.59	1.16	0.67
	AGU	1.28	0.22			1.30	0.23		
Pro	CCU	1.67	0.32			1.68	0.30	1.68	0.09
	ACA					1.27	0.58		
Thr	ACA	1.27	0.83	1.26	0.56			1.26	0.60
Ala	GCA	1.23	0.66	1.22	0.76	1.24	0.68	1.24	0.63
Tyr	UAU			1.58	0.11	1.59	0.11	1.59	0.14
Asp	GAU	1.64	0.17			1.64	0.09		
Cys	UGU	1.55	0.24						
Arg	AGA			1.94	0.38				
Gly	GGU			1.42	0.26	1.42	0.26	1.41	0.32
	GGA			1.62	0.23				

validated using transcriptome data (Mower, 2009), resulting in the identification of 20 authentic RNA editing sites among the five genes, all involving C-to-U transitions. Among them, 20% (4 sites), 65% (13 sites), and 15% (3 sites) occurred at the first, second, and third codon positions, respectively. Moreover, 40% (8 sites) of the RNA editing events resulted in the conversion of neutral amino acids to hydrophobic amino acids, whereas another 40% (8 sites) led to the conversion of hydrophobic amino acids to other hydrophobic amino acids. Additionally, 10% (2 sites) of the RNA editing events caused basic amino acids to change into neutral amino acids, and the remaining 10% (2 sites) involved basic amino acids being converted into other basic amino acids or hydrophobic amino acids turning into neutral amino acids (Supplementary Table S2). We predicted RNA editing sites for these eight genes and observed a preference for RNA editing to occur in codons with high RSCU values in these genes.

# 4 Discussion

4.1 Base composition

In the study of CUB in plant cp genomes, base composition analysis is an essential aspect that involves investigating the frequency of AGCT nucleotides, the GC content, and their relationship with codon usage preference. Valuable information regarding gene expression and evolution can be obtained by analyzing the codons and bases in cp genomes (Southworth et al., 2018; Chakraborty et al.,

2019; Shen et al., 2020). Moreover, certain studies have suggested that biases in the base composition of the genome may cause mutation pressure and influence codon usage preferences via an appropriate balance of G + C nucleotide pairs (Li et al., 2002). In this study, the cp genes of two pairs of sister genera that exhibited significant habitat differentiation were analyzed, and their codon usage preferences and patterns influencing CUB evolution were explored.

Herein, the overall frequency of codon usage for A and T was significantly higher than that for G and C. All gene-coding sequences were enriched with AT bases. In the cp genes of the four sister genera, codons ending with A and T were preferred, with T being the most common nucleotide in genes associated with photosynthesis. Among the cp genes of the sister genera, the GC content was highest in the first position of the codon and lowest in the third. Previous studies reported that codons preferentially ended with A or T in angiosperm nuclear genomes, whereas they typically ended with G or C in monocot genes (Kawabe and Miyashita, 2003). A highly conserved GC content in the coding sequences of cp genes in these two pairs of sister genera were observed, suggesting that evolutionary equilibrium was attained over time. The usage frequency of pyrimidines (C and T) was higher than that of purines (A and G) in the third position of their codons. Similar results have been reported in previous studies (Zhang et al., 2013). Additionally, certain studies have proposed that codon usage preferences are primarily influenced by differences in G and C content, rather than differences in A and T (Zhang et al., 2007).

Generally, the term "optimal codon" refers to the codons that efficiently facilitate translation and protein synthesis under specific

FIGURE 6
Expression of cp genes in four sister genera.
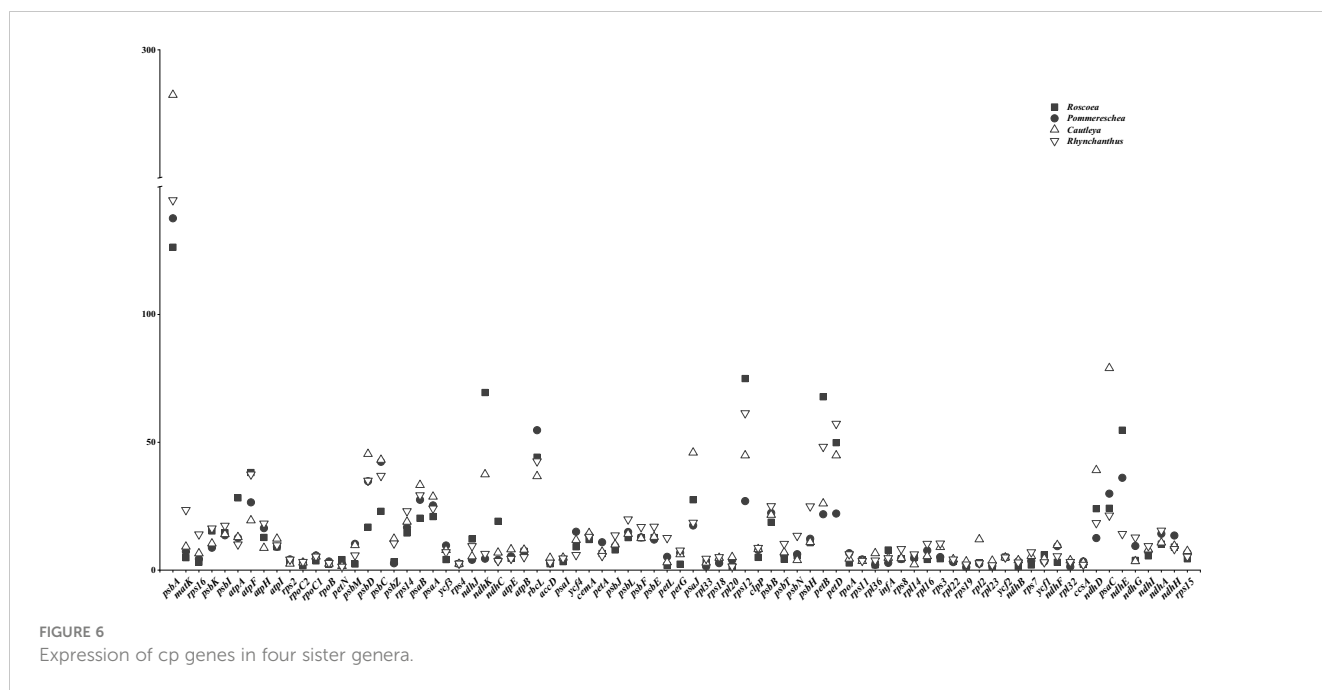
environmental conditions. Research findings suggest that CUB can influence the speed and accuracy of gene expression in cps, thereby affecting cellular energy utilization and genome complexity (Lane and Martin, 2010). Furthermore, codon selection may have played a role in the evolution of cp genomes. A preference for A/U as the ending nucleotide in optimal codons was observed, and shared optimal codons in similar habitats were identified, thus indicating a potential association with their adaptation to the environment. In addition, optimal codons can be used to design degenerate primers and introduce site mutations in evolutionary studies.

4.2 Factors influencing CUB

The formation of codon bias is a result of mutations, and the bias in mutations is shaped by the underlying mutational mechanisms. Mutational selection is neutral (it does not affect normal mechanisms) and acts equally on all DNA sequences of the organism. In most organisms, the frequency of TA dinucleotides is lower than expected based on random nucleotide usage. This phenomenon is a consequence of avoiding the use of termination codons, such as TAA and TAG, just as mRNA avoids using UA because it is susceptible to ribonuclease activity (Bulmer, 1991). Numerous base mutations are caused by random mutations, mispairing during repair, replication errors, or methylation.

Unlike pure mutational mechanisms, selective pressure influences the use of synonymous codons. Codon bias that results from selective pressure may exist in specific genes or positions within codons to enhance translation efficiency, accuracy, or protein-folding dynamics. At the RNA level, selective pressure for efficient transcription has been reported, where rich nucleotides improve mRNA translation efficiency (Xia, 2005). Selection also occurs at the mRNA level and affects mRNA folding and decay through preferential selection. This suggests the presence of overall

optimization to reduce the time of action of the ribosome during mRNA translation, thereby enhancing its efficiency. The use of optimal codons can influence translation speed and elongation rate, thus affecting protein-folding co-translational processes (Kawabe and Miyashita, 2003). In large populations of effective codons, organism selection plays a significant role in translational optimization (Bulmer, 1987).

Herein, both natural selection and mutations influenced the CUB of cp genes in the two pairs of sister genera, with natural selection primarily shaping codon usage patterns, particularly for genes related to photosynthesis, while base mutations played a secondary role. This is consistent with previous research on *Panicum*, *Malus*, and Euphorbiaceae (Wang et al., 2020; Li et al., 2021; Li et al., 2023). Notably, the factors influencing CUB are highly complex, and other factors, such as gene length, gene expression levels, recombination rates, and genetic code repair, may also play a role (Marais et al., 2001; Barbhuiya et al., 2021; Huang et al., 2022).

4.3 CUB and gene expression and RNA editing

Codon usage patterns in cp genomes are associated with environmental adaptation. Chloroplast genomes are the plant organelles responsible for photosynthesis and energy conversion, therefore, the proteins encoded by the cp genome are closely related to these functions. Research indicates that CUB in the cp genome is also linked to the adaptability of cps to their environment and efficiency of gene expression (Zhang et al., 2007).

Significant differences in the expression levels of eight genes were found between the two sister genera in similar habitats, and the ENc values showed strong preferences. Among them, the *psbA*, *psaB*, *rbcL*, and *ndhI* genes are directly involved in the regulation of

photosynthesis, including light capture, photoprotection, stress adaptation, and redox reactions (Jensen et al., 2000; Adams et al., 2001; Munekage et al., 2002; Heinemeyer et al., 2004; Nishizawa et al., 2006; Takahashi et al., 2007; Whitney and Sharwood, 2008; Sage and Sage, 2009). In certain studies, the *rps14*, *rps3*, and *rps7* genes have been suggested to participate in translation. In addition, their expression levels may be regulated by light, temperature, and other environmental factors, implying their potential roles in photosynthesis and plant growth and development regulation (Barkan, 2011; Shanmugabalaji et al., 2013; Dent et al., 2015; Wang et al., 2017). However, the role of the *matK* gene in plant light adaptation has not been extensively studied. To date, most studies have focused on the phylogeny and species identification of *matK* gene. Certain studies have used cp genome data from multiple plant species, including the *matK* gene, to analyze the phylogenetic relationships and evolutionary patterns of plant species, which may be related to their photosynthetic adaptation ability (Jansen et al., 2007; Timme et al., 2007). Therefore, the differential expression of these genes may be associated with their adaptation to different habitats, particularly in response to light conditions.

RNA editing in the cp genome of plants is a post-transcriptional modification process that involves changing the bases in the transcribed RNA sequence to other bases, thereby introducing diversity between the transcript and genomic sequences (Ichinose and Sugita, 2016). This type of RNA editing is widespread in plants and plays crucial roles in regulating gene expression, increasing functional diversity, and adapting to environmental changes. Studies have suggested that plants utilize RNA editing to adapt to different environmental pressures, such as temperature, light, and salinity. Adaptive RNA editing can confer survival advantages to plants in different habitats (Wang et al., 2015). In addition, RNA editing plays a significant role in regulating the expression of genes involved in photosynthesis. The present study found that in differentially expressed genes, RNA editing resulted in more hydrophobic amino acid changes, thus significantly increasing the hydrophobicity of the edited proteins. This pattern is similar to the changes observed in RNA editing sites in cotton cp genes, which may favor the formation of stable protein structures. When hydrophilic mutations occur in the core of a protein structure, they can cause severe instability or even loss of folding and function (De Vos et al., 2001). RNA editing appears to be more likely to occur in codons with high RSCU values. However, further data analysis and experimental validation are required to investigate whether there is a mutual influence between RNA editing and CUB and to understand the mechanisms underlying this potential interaction.

5 Conclusions

This study investigated the codon usage patterns of the cp genomes in two pairs of sister genera of the Zingiberaceae family that showed clear habitat differentiation. The relationship between CUB and gene expression was also explored. High CUBs were observed, with protein-coding genes and the third position of

codons enriched in A/T bases. Different patterns of codon usage were observed among the different gene types. Natural selection likely plays a considerable role in shaping the CUB in the cp genes of species from different habitats, whereas mutations play a secondary role. Nine to 11 optimal codons were identified in the cp genome, with two showing similarities between species from the same habitat. Genes related to light adaptation exhibited differential expression in species from different habitats, and RNA editing is likely to occur in codons with higher RSCU values. Future research into the potential associations between CUB and RNA editing, as well as multidimensional analyses of factors influencing CUB, is necessary.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/genbank/, PRJNA997822.

Author contributions

QY: Conceptualization, Software, Writing – original draft. CX: Software, Writing – review & editing. Q-SX: Resources, Writing – review & editing. Y-TL: Data curation, Writing – review & editing. LL: Resources, Writing – review & editing. J-LZ: Conceptualization, Funding acquisition, Writing – review & editing.

Funding

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1304264/full#supplementary-material

SUPPLEMENTARY FIGURE 1

The evolutionary relationship of four sisters genera (*Roscoea*, *Cautleya*, *Pommereschea*, and *Rhynchanthus*).

SUPPLEMENTARY FIGURE 2

Comparison of the base composition of four sister genera with their cp genes. **(A)** is all genes, **(B)** is photosynthesis-related genes, **(C)** is transcription- and translation-related genes, and **(D)** is other genes. The y-coordinate represents the frequency of base usage.

SUPPLEMENTARY FIGURE 3

Content of the cp genes GC, GC1, GC2, and GC3 in four sister genera. **(A)** is all genes, **(B)** is photosynthesis-related genes, **(C)** is transcription- and translation-related genes, and **(D)** is other genes. The y-coordinate represents the GC content.

SUPPLEMENTARY FIGURE 4

ENc-plot analysis of Pommereschea and Rhynchanthus cp genes. A represents genes related to photosynthesis, B represents genes related to transcription and translation, C represents other genes, and D represents all genes.

SUPPLEMENTARY FIGURE 5

PR2-plot analysis of Pommereschea and Rhynchanthus cp genes. A represents genes related to photosynthesis, B represents genes related to transcription and translation, C represents other genes, and D represents all genes.

SUPPLEMENTARY FIGURE 6

Neutrality plot analysis of Pommereschea and Rhynchanthus cp genes. A represents genes related to photosynthesis, B represents genes related to transcription and translation, C represents other genes, and D represents all genes.

SUPPLEMENTARY FIGURE 7

Relative synonymous codon use (RSCU) of differentially expressed genes. **(A)** is *psbA*, **(B)** is *psaB*, **(C)** is *rbcL*, **(D)** is *ndhI*, **(E)** is *matK*, **(F)** is *rps14*, **(G)** is *rps3*, and **(H)** is *rps7*.

References

Adams, W. W. R., Demmig-Adams, B., Rosenstiel, T. N., and Ebbert, V. (2001). Dependence of photosynthesis and energy dissipation activity upon growth form and light environment during the winter. *Photosynthesis Res.* 67, 51–62. doi: 10.1023/A:1010688528773

Angellotti, M. C., Bhuiyan, S. B., Chen, G., and Wan, X. (2007). CodonO: codon usage bias analysis within and across genomes. *Nucleic Acids Res.* 35, W132–W136. doi: 10.1093/nar/gkm392

Barbhuiya, P. A., Uddin, A., and Chakraborty, S. (2021). Understanding the codon usage patterns of mitochondrial CO genes among *Amphibians*. *Gene* 777, 145462. doi: 10.1016/j.gene.2021.145462

Barkan, A. (2011). Expression of plastid genes: organelle-specific elaborations on a prokaryotic scaffold. *Plant Physiol.* 155, 1520–1532. doi: 10.1104/pp.110.171231

Botzman, M., and Margalit, H. (2011). Variation in global codon usage bias among prokaryotic organisms is associated with their lifestyles. *Genome Biol.* 12, R109. doi: 10.1186/gb-2011-12-10-r109

Bulmer, M. (1987). Coevolution of codon usage and transfer RNA abundance. *Nature* 325, 728–730. doi: 10.1038/325728a0

Bulmer, M. (1991). The selection-mutation-drift theory of synonymous codon usage. *Genetics* 129, 897–907. doi: 10.1093/genetics/129.3.897

Butt, A. M., Nasrullah, I., Qamar, R., and Tong, Y. (2016). Evolution of codon usage in Zika virus genomes is host and vector specific. *Emerging Microbes Infections* 5, e107. doi: 10.1038/emi.2016.106

Chakraborty, S., Mazumder, T. H., and Uddin, A. (2019). Compositional dynamics and codon usage pattern of *BRCA1* gene across nine mammalian species. *Genomics* 111, 167–176. doi: 10.1016/j.ygeno.2018.01.013

Chamani Mohasses, F., Solouki, M., Ghareyazie, B., Fahmideh, L., and Mohsenpour, M. (2020). Correlation between gene expression levels under drought stress and synonymous codon usage in rice plant by in-silico study. *PloS One* 15, e237334. doi: 10.1371/journal.pone.0237334

De Vos, S., Backmann, J., Prévost, M., Steyaert, J., and Loris, R. (2001). Hydrophobic core manipulations in ribonuclease T1. *Biochemistry* 40, 10140–10149. doi: 10.1021/bi010565n

Dent, R. M., Sharifi, M. N., Malnoë, A., Haglund, C., Calderon, R. H., Wakao, S., et al. (2015). Large-scale insertional mutagenesis of Chlamydomonas supports phylogenomic functional prediction of photosynthetic genes and analysis of classical acetate-requiring mutants. *Plant Journal: For Cell Mol. Biol.* 82, 337–351. doi: 10.1111/tpj.12806

Gao, Y., Lu, Y., Song, Y., and Jing, L. (2022). Analysis of codon usage bias of WRKY transcription factors in *Helianthus annuus*. *BMC Genomic Data* 23, 46. doi: 10.1186/s12863-022-01064-8

Gu, W., Zhou, T., Ma, J., Sun, X., and Lu, Z. (2004). The relationship between synonymous codon usage and protein structure in *Escherichia coli* and *Homo sapiens*. *Biosystems* 73, 89–97. doi: 10.1016/j.biosystems.2003.10.001

Heinemeyer, J., Eubel, H., Wehmhöner, D., Jänsch, L., and Braun, H. (2004). Proteomic approach to characterize the supramolecular organization of photosystems in higher plants. *Phytochemistry* 65, 1683–1692. doi: 10.1016/j.phytochem.2004.04.022

Huang, X., Jiao, Y., Guo, J., Wang, Y., Chu, G., and Wang, M. (2022). Analysis of codon usage patterns in *Haloxylon ammodendron* based on genomic and transcriptomic data. *Gene* 845, 146842. doi: 10.1016/j.gene.2022.146842

Huo, X., Liu, S., Li, Y., Wei, H., Gao, J., Yan, Y., et al. (2021). Analysis of synonymous codon usage of transcriptome database in *Rheum palmatum*. *Peerj* 9, e10450. doi: 10.7717/peerj.10450

Ichinose, M., and Sugita, M. (2016). RNA editing and its molecular mechanism in plant organelles. *Genes* 8, 5. doi: 10.3390/genes8010005

Jansen, R. K., Cai, Z., Raubeson, L. A., Daniell, H., Depamphilis, C. W., Leebens-Mack, J., et al. (2007). Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. *Proc. Natl. Acad. Sci. United States America* 104, 19369–19374. doi: 10.1073/pnas.0709121104

Jensen, P. E., Gilpin, M., Knoetzel, J., and Scheller, H. V. (2000). The PSI-K subunit of photosystem I is involved in the interaction between light-harvesting complex I and the photosystem I reaction center core. *J. Biol. Chem.* 275, 24701–24708. doi: 10.1074/jbc.M000550200

Jia, J., and Xue, Q. (2009). Codon usage biases of transposable elements and host nuclear genes in *Arabidopsis thaliana* and *Oryza sativa*. *Genomics Proteomics Bioinf.* 7, 175–184. doi: 10.1016/S1672-0229(08)60047-9

Jin, J. J., Yu, W. B., Yang, J. B., Song, Y., and Li, D. Z. (2020). Getorganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biol.* 21, 241. doi: 10.1186/s13059-020-02154-5

Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010

Kawabe, A., and Miyashita, N. T. (2003). Patterns of codon usage bias in three dicot and four monocot plant species. *Genes Genet. Syst.* 78, 343–352. doi: 10.1266/ggs.78.343

Kim, D., Paggi, J. M., Park, C., Bennett, C., and Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* 37, 907–915. doi: 10.1038/s41587-019-0201-4

Kress, W. J., Prince, L. M., and Williams, K. J. (2002). The phylogeny and a new classification of the gingers (Zingiberaceae): evidence from molecular data. *Am. J. Bot.* 89, 1682–1696. doi: 10.3732/ajb.89.10.1682

Lane, N., and Martin, W. (2010). The energetics of genome complexity. *Nature* 467, 929–934. doi: 10.1038/nature09486

Li, Y., Korol, A. B., Fahima, T., Beiles, A., and Nevo, E. (2002). Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. *Mol. Ecol.* 11, 2453–2465. doi: 10.1046/j.1365-294x.2002.01643.x

Li, G., Zhang, L., and Xue, P. (2021). Codon usage pattern and genetic diversity in chloroplast genomes of *Panicum* species. *Gene* 802, 145866. doi: 10.1016/j.gene.2021.145866

Li, G., Zhang, L., Xue, P., and Zhu, M. (2023). Comparative analysis on the codon usage pattern of the chloroplast genomes in *Malus* species. *Biochem. Genet.* 61, 1050–1064. doi: 10.1007/s10528-022-10302-8

Liao, Y., Smyth, G. K., and Shi, W. (2014). Featurecounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30, 923–930. doi: 10.1093/bioinformatics/btt656

Lü, H., Zhao, W., Zheng, Y., Wang, H., Qi, M., and Yu, X. (2005). Analysis of synonymous codon usage bias in Chlamydia. *Acta Biochim. Et Biophys. Sin.* 37, 1–10. doi: 10.1093/abbs/37.1.1

Marais, G., Mouchiroud, D., and Duret, L. (2001). Does recombination improve selection on codon usage? Lessons from nematode and fly complete genomes. *Proc. Natl. Acad. Sci. United States America* 98, 5688–5692. doi: 10.1073/pnas.091427698

Monian, P., Shivalila, C., Lu, G., Shimizu, M., Boulay, D., Bussow, K., et al. (2022). Endogenous ADAR-mediated RNA editing in non-human primates using stereopure chemically modified oligonucleotides. *Nat. Biotechnol.* 40, 1093–1102. doi: 10.1038/s41587-022-01225-1

Mower, J. P. (2009). The PREP suite: predictive RNA editors for plant mitochondrial genes, chloroplast genes and user-defined alignments. *Nucleic Acids Res.* 37, W253–W259. doi: 10.1093/nar/gkp337

Munekage, Y., Hojo, M., Meurer, J., Endo, T., Tasaka, M., and Shikanai, T. (2002). *PGR5* is involved in cyclic electron flow around photosystem I and is essential for photoprotection in *Arabidopsis. Cell* 110, 361–371. doi: 10.1016/s0092-8674(02)00867-x

Nie, X., Lv, S., Zhang, Y., Du, X., Wang, L., Biradar, S. S., et al. (2012). Complete chloroplast genome sequence of a major invasive species, crofton weed (*Ageratina adenophora*). *PloS One* 7, e36869. doi: 10.1371/journal.pone.0036869

Nishizawa, A., Yabuta, Y., Yoshida, E., Maruta, T., Yoshimura, K., and Shigeoka, S. (2006). *Arabidopsis* heat shock transcription factor A2 as a key regulator in response to several types of environmental stress. *Plant Journal: For Cell Mol. Biol.* 48, 535–547. doi: 10.1111/j.1365-313X.2006.02889.x

Pandey, S., Prasad, A., Sharma, N., and Prasad, M. (2020). Linking the plant stress responses with RNA helicases. *Plant Science: Int. J. Exp. Plant Biol.* 299, 110607. doi: 10.1016/j.plantsci.2020.110607

Radwan, O. E., Ahmed, T. A., and Knapp, S. J. (2010). Phylogenetic analyses of peanut resistance gene candidates and screening of different genotypes for polymorphic markers. *Saudi J. Biol. Sci.* 17, 43–49. doi: 10.1016/j.sjbs.2009.12.007

Robinson, J. T., Thorvaldsdottir, H., Turner, D., and Mesirov, J. P. (2023). igv.js: an embeddable JavaScript implementation of the Integrative Genomics Viewer (IGV). *Bioinformatics* 39, btac830. doi: 10.1093/bioinformatics/btac830

Sage, T. L., and Sage, R. F. (2009). The functional anatomy of rice leaves: implications for refixation of photorespiratory CO2 and efforts to engineer C4 photosynthesis into rice. *Plant Cell Physiol.* 50, 756–772. doi: 10.1093/pcp/pcp033

Shackelton, L. A., Parrish, C. R., and Holmes, E. C. (2006). Evolutionary basis of codon usage and nucleotide composition bias in vertebrate DNA viruses. *J. Mol. Evol.* 62, 551–563. doi: 10.1007/s00239-005-0221-1

Shanmugabalaji, V., Besagni, C., Piller, L. E., Douet, V., Ruf, S., Bock, R., et al. (2013). Dual targeting of a mature plastoglobulin/fibrillin fusion protein to chloroplast plastoglobules and thylakoids in transplastomic tobacco plants. *Plant Mol. Biol.* 81, 13–25. doi: 10.1007/s11103-012-9977-z

Sharp, P. M., Bailes, E., Grocock, R. J., Peden, J. F., and Sockett, R. E. (2005). Variation in the strength of selected codon usage bias among bacteria. *Nucleic Acids Res.* 33, 1141–1153. doi: 10.1093/nar/gki242

Sharp, P. M., Cowe, E., Higgins, D. G., Shields, D. C., Wolfe, K. H., and Wright, F. (1988). Codon usage patterns in *Escherichia coli, Bacillus subtilis, Saccharomyces cerevisiae, Schizosaccharomyces pombe, Drosophila melanogaster* and *Homo sapiens*; a review of the considerable within-species diversity. *Nucleic Acids Res.* 16, 8207–8211. doi: 10.1093/nar/16.17.8207

Sharp, P. M., and Li, W. H. (1986). An evolutionary perspective on synonymous codon usage in unicellular organisms. *J. Mol. Evol.* 24, 28–38. doi: 10.1007/BF02099948

Sharp, P. M., Stenico, M., Peden, J. F., and Lloyd, A. T. (1993). Codon usage: mutational bias, translational selection, or both? *Biochem. Soc. Trans.* 21, 835–841. doi: 10.1042/bst0210835

Shen, Z., Gan, Z., Zhang, F., Yi, X., Zhang, J., and Wan, X. (2020). Analysis of codon usage patterns in citrus based on coding sequence data. *BMC Genomics* 21, 234. doi: 10.1186/s12864-020-6641-x

Song, H., Liu, J., Song, Q., Zhang, Q., Tian, P., and Nan, Z. (2017). Comprehensive analysis of codon usage bias in seven *Epichloë* species and their peramine-coding genes. *Front. Microbiol.* 8, 1419. doi: 10.3389/fmicb.2017.01419

Southworth, J., Armitage, P., Fallon, B., Dawson, H., Bryk, J., and Carr, M. (2018). Patterns of ancestral animal codon usage bias revealed through Holozoan protists. *Mol. Biol. Evol.* 35, 2499–2511. doi: 10.1093/molbev/msy157

Sueoka, N. (1988). Directional mutation pressure and neutral molecular evolution. *Proc. Natl. Acad. Sci. United States America* 85, 2653–2657. doi: 10.1073/pnas.85.8.2653

Sueoka, N. (1995). Intrastrand parity rules of DNA base composition and usage biases of synonymous codons. *J. Mol. Evol.* 40, 318–325. doi: 10.1007/BF00163236

Sueoka, N. (1999). Translation-coupled violation of Parity Rule 2 in human genes is not the cause of heterogeneity of the DNA G+C content of third codon position. *Gene* 238, 53–58. doi: 10.1016/s0378-1119(99)00320-0

Takahashi, S., Bauwe, H., and Badger, M. (2007). Impairment of the photorespiratory pathway accelerates photoinhibition of photosystem II by suppression of repair but not

acceleration of damage processes in *Arabidopsis. Plant Physiol.* 144, 487–494. doi: 10.1104/pp.107.097253

Tao, P., Dai, L., Luo, M., Tang, F., Tien, P., and Pan, Z. (2009). Analysis of synonymous codon usage in classical swine fever virus. *Virus Genes* 38, 104–112. doi: 10.1007/s11262-008-0296-z

Tillich, M., Lehwark, P., Pellizzer, T., Ulbricht-Jones, E. S., Fischer, A., Bock, R., et al. (2017). GeSeq - versatile and accurate annotation of organelle genomes. *Nucleic Acids Res.* W6-W11. doi: 10.1093/nar/gkx391

Timme, R. E., Kuehl, J. V., Boore, J. L., and Jansen, R. K. (2007). A comparative analysis of the *Lactuca* and *Helianthus* (Asteraceae) plastid genomes: identification of divergent regions and categorization of shared repeats. *Am. J. Bot.* 94, 302–312. doi: 10.3732/ajb.94.3.302

Tyagi, S., Kabade, P. G., Gnanapragasam, N., Singh, U. M., Gurjar, A. K. S., Rai, A., et al. (2023). Codon usage provide insights into the adaptation of rice genes under stress condition. *Int. J. Mol. Sci.* 24, 1098. doi: 10.3390/ijms24021098

Wang, Y., Geng, L., Yuan, M., Wei, J., Jin, C., Li, M., et al. (2017). Deletion of a target gene in Indica rice via CRISPR/Cas9. *Plant Cell Rep.* 36, 1333–1343. doi: 10.1007/s00299-017-2158-4

Wang, J., Lin, Y., and Xi, M. (2021). Analysis of codon usage patterns of six sequenced *Brachypodium distachyon* lines reveals a declining CG skew of the CDSs from the 5'-ends to the 3'-ends. *Genes* 12, 1467. doi: 10.3390/genes12101467

Wang, M., Liu, H., Ge, L., Xing, G., Wang, M., Weining, S., et al. (2016). Identification and analysis of RNA editing sites in the chloroplast transcripts of *Aegilops tauschii* L. *Genes* 8, 13. doi: 10.3390/genes8010013

Wang, Z., Xu, B., Li, B., Zhou, Q., Wang, G., Jiang, X., et al. (2020). Comparative analysis of codon usage patterns in chloroplast genomes of six Euphorbiaceae species. *Peerj* 8, e8251. doi: 10.7717/peerj.8251

Wang, W., Zhang, W., Wu, Y., Maliga, P., and Messing, J. (2015). RNA editing in chloroplasts of *Spirodela polyrhiza*, an aquatic monocotelydonous species. *PloS One* 10, e140285. doi: 10.1371/journal.pone.0140285

Wei, L., He, J., Jia, X., Qi, Q., Liang, Z., Zheng, H., et al. (2014). Analysis of codon usage bias of mitochondrial genome in *Bombyx mori* and its relation to evolution. *BMC Evolutionary Biol.* 14, 262. doi: 10.1186/s12862-014-0262-4

Whitney, S. M., and Sharwood, R. E. (2008). Construction of a tobacco master line to improve Rubisco engineering in chloroplasts. *J. Exp. Bot.* 59, 1909–1921. doi: 10.1093/jxb/erm311

Wright, F. (1990). The 'effective number of codons' used in a gene. *Gene* 87, 23–29. doi: 10.1016/0378-1119(90)90491-9

Wyman, S. K., Jansen, R. K., and Boore, J. L. (2004). Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* 20, 3252–3255. doi: 10.1093/bioinformatics/bth352

Xia, X. (2005). Mutation and selection on the anticodon of tRNA genes in vertebrate mitochondrial genomes. *Gene* 345, 13–20. doi: 10.1016/j.gene.2004.11.019

Xin, C., and Yang, Q. (2020). The first complete chloroplast genome sequence of *Vicia ramuliflora* (Fabaceae). *Mitochondrial DNA Part B Resour.* 5, 410–411. doi: 10.1080/23802359.2019.1705196

Yang, Q., Fu, G., Wu, Z., Li, L., Zhao, J., and Li, Q. (2021). Chloroplast genome evolution in four montane Zingiberaceae taxa in China. *Front. Plant Sci.* 12, 774482. doi: 10.3389/fpls.2021.774482

Yang, Q., Xin, C., Li, J., Peng, F., Zhang, R., Wang, J., et al. (2020). Complete chloroplast genome of *Triaenophora shennongjiaensis* (Orobanchaceae), an endangered medical herb from China. *Annales Botanici Fennici* 57, 23–31. doi: 10.5735/085.057.0104

Yang, Y., Zhu, J., Feng, L., Zhou, T., Bai, G., Yang, J., et al. (2018). Plastid genome comparative and phylogenetic analyses of the key genera in Fagaceae: highlighting the effect of codon composition bias in phylogenetic inference. *Front. Plant Sci.* 9, 82. doi: 10.3389/fpls.2018.00082

Zhang, Z., Dai, W., Wang, Y., Lu, C., and Fan, H. (2013). Analysis of synonymous codon usage patterns in torque teno sus virus 1 (TTSuV1). *Arch. Virol.* 158, 145–154. doi: 10.1007/s00705-012-1480-y

Zhang, L., Guo, Y., Luo, L., Wang, Y., Dong, Z., Sun, S., et al. (2011). Analysis of nuclear gene codon bias on soybean genome and transcriptome. *Acta Agronomica Sin.* 37, 965–974. doi: 10.1016/S1875-2780(11)60028-X

Zhang, W., Zhou, J., Li, Z., Wang, L., Gu, X., and Zhong, Y. (2007). Comparative analysis of codon usage patterns among mitochondrion, chloroplast and nuclear genes in *Triticum aestivum* L. *J. Integr. Plant Biol.* 49, 246–254. doi: 10.1111/j.1744-7909.2007.00404.x

Zhao, J., Yu, X., Kress, W. J., Wang, Y., Xia, Y., and Li, Q. (2022). Historical biogeography of the gingers and its implications for shifts in tropical rain forest habitats. *J. Biogeography* 49, 1339–1351. doi: 10.1111/jbi.14386

Zhao, Y., Zheng, H., Xu, A., Yan, D., Jiang, Z., Qi, Q., et al. (2016). Analysis of codon usage bias of envelope glycoprotein genes in nuclear polyhedrosis virus (NPV) and its relation to evolution. *BMC Genomics* 17, 677. doi: 10.1186/s12864-016-3021-7