Physiology and Molecular Biology of Plants (February 2024) 30(2):153–166
https://doi.org/10.1007/s12298-024-01421-0

RESEARCH ARTICLE

Comparative analysis of codon usage patterns in the chloroplast genomes of nine forage legumes

Mingkun Xiao[1] · Xiang Hu[2] · Yaqi Li[1] · Qian Liu[1] · Shaobin Shen[1] · Tailing Jiang[1] · Linhui Zhang[1] · Yingchun Zhou[1] · Yuexian Li[1] · Xin Luo[1] · Lina Bai[1] · Wei Yan[1]

Received: 25 September 2023 / Revised: 19 February 2024 / Accepted: 23 February 2024 / Published online: 9 March 2024

Abstract

Leguminosae is one of the three largest families of angiosperms after *Compositae* and *Orchidaceae*. It is widely distributed and grows in a variety of environments, including plains, mountains, deserts, forests, grasslands, and even waters where almost all legumes can be found. It is one of the most important sources of starch, protein and oil in the food of mankind and also an important source of high-quality forage material for animals, which has important economic significance. In our study, the codon usage patterns and variation sources of the chloroplast genome of nine important forage legumes were systematically analyzed. Meanwhile, we also constructed a phylogenetic tree based on the whole chloroplast genomes and protein coding sequences of these nine forage legumes. Our results showed that the chloroplast genomes of nine forage legumes end with A/T bases, and seven identical high-frequency (HF) codons were detected among the nine forage legumes. ENC-GC3s mapping, PR2 analysis, and neutral analysis showed that the codon bias of nine forage legumes was influenced by many factors, among which natural selection was the main influencing factor. The codon usage frequency showed that the *Nicotiana tabacum* and *Saccharomyces cerevisiae* can be considered as receptors for the exogenous expression of chloroplast genes of these nine forage legumes. The phylogenetic relationships of the chloroplast genomes and protein coding genes were highly similar, and the nine forage legumes were divided into three major clades. Among the clades *Melilotus officinalis* was more closely related to *Medicago sativa*, and *Galega officinalis* was more closely related to *Galega orientalis*. This study provides a scientific basis for the molecular markers research, species identification and phylogenetic studies of forage legumes.

Keywords Forage legume · Codon usage bias · Chloroplast genome · Phylogenetic relationships

Introduction

Forage legume is an important part of global grassland agriculture. Compared with grass forage, forage legume has some unique advantages, such as exhibiting symbiosis with *rhizobium*, which is beneficial to reduce nitrogen fertilizer input in low-nitrogen land (Phelan et al. 2015). With rising energy and fertilizer prices, stringent environmental regulations, and high feed costs, the importance of forage legume in ruminant production agricultural systems will be enhanced (Lüscher et al. 2014; Van Grinsven et al. 2013) At the same time, adding forage legume to ruminant daily feed can improve the digestibility and metabolic energy of feed to meet the protein requirements of animals (Rochon et al. 2004). Currently, the number of legumes used as forage worldwide is currently unknown; the Food and Agriculture Organization of the United Nations, online animal feed information base (http://www.feedipedia.org/content/feeds?category=13594) lists only 153 forage legumes. However, only a very small fraction of forage legume is widely considered to be of global commercial importance (Gustavo and Timothy 2017; Phelan et al. 2015; Peyraud et al. 2009). In this study, we compared the codon preferences of the chloroplast genomes of seven perennial and two annual forage legumes in world grassland agriculture.

Mingkun Xiao, Xiang Hu and Yaqi Li contributed equally to this work.

✉ Wei Yan
rjsyw@yaas.org.cn

[1] Tropical and Subtropical Cash Crops Research Institute, Yunnan Academy of Agricultural Sciences, Baoshan, Yunnan, China

[2] Tropical Eco-agricultural Research Institute, Yunnan Academy of Agricultural Sciences, Yuanmou, Yunnan, China

Chloroplasts are energy converters unique to higher plants and some algae (Zhang et al. 2012). Plants use solar energy, water, minerals, and carbon dioxide through photosynthesis to synthesize organic compounds to sustain life (Dobrogojski et al. 2020). In addition, chloroplasts are essential carriers for the synthesis of amino acids, nucleotides, plant hormones, fatty acids, and vitamins. These metabolites play an important role in plant stress response and information transmission (Daniell et al. 2016). Compared with those of nuclear and mitochondrial genomes, the structure and function of chloroplast genome are characterized by small size, high conservation, simple structure, and single-parent inheritance, which have great advantages in genetic transformation (Wang et al. 2023), thus chloroplast attracting the attention of many scholars in recent years.

Codons are the link between amino acids, proteins, and genetic material in living organisms and play an irreplaceable role in transmitting genetic information. Living organisms have 64 codons, 3 of which are termination codons, and the other 61 are translated into 20 amino acids. Except methionine and tryptophan that have unique codons, the rest of the amino acids usually have 2–6 corresponding codons, and codons coding for the same amino acid are called synonymous codons. Genes from different species or within the same species exhibit different codon usage preference patterns, and the phenomenon of uneven usage of synonymous codons that is prevalent in organisms, is known as codon bias (Wang et al. 2020; Plotkin and Kudla 2011). Codon bias to some extent can reflect the origin of species or genes, the evolutionary process, and the mode of mutation. In animals, it occurs mainly from mutational pressure in microorganisms and from natural selection; but for plants, it is influenced by natural selection and mutational pressure (Bulmer 1991; Baeza et al. 2015; Camiolo et al. 2015; Boel et al. 2016). High-throughput sequencing technologies have allowed the enrichment of chloroplast genome databases and the analysis of genome codon preference in a variety of organisms. Example include the comparative analysis of codon bias in the chloroplast genomes of *Theaceae* species (Wang et al. 2022), comparative analysis of codon usage patterns in the chloroplast genomes of ten *Epimedium* species (Wang et al. 2023), analysis of codon usage bias of chloroplast genomes in *Gynostemma* species (Zhang et al. 2021), and analysis of the codon usage bias of chloroplast genes in *Oryza* species (Chakraborty et al. 2020). Many reports are available on the chloroplast genomes of *Leguminosae*, but only a few have focused on their codon bias. Studying the codon usage patterns in plants is necessary to construct stable transgenic systems.

In this study, the codon usage preference characteristics of chloroplast genes of nine forage legumes were analyzed from the perspective of coding sequences and explored the reasons for their formation. We have also analyzed several important parameters of codon usage patterns, including GC content at three positions (GC1, GC2, and GC3), effective number of codons (ENC), relative synonymous codon usage (RSCU), and relative frequency of synonymous codon usage (RFSC). In addition, the codon usage frequencies of these nine forage legumes were compared with four model organisms i.e., *Arabidopsis thaliana*, *Nicotiana tabacum*, *Escherichia coli*, and *Saccharomyces cerevisiae*. This work provides a basis for further improving the exogenous gene expression efficiency of the nine forage legumes. The findings reveal the role of usage patterns and sources of variation in influencing the codon usage bias of the chloroplast genomes of different forage legumes during the evolutionary process, and provide a reference for the verification of gene function, genetic evolution and, transgenic engineering of chloroplast genomes.

Materials and methods

Genomes and coding sequences data acquisition and filtering of nine forage legumes

The chloroplast genomes and coding sequences (CDS) of nine forage legumes were downloaded from the National Center for Biotechnology Information database (NCBI, https://www.ncbi.nlm.nih.gov/, accessed on 15 May 2023). For ensure the accuracy of the experimental results, the CDS of the nine plants should meet the following conditions: (1) the number of bases in the CDS of each plant should be an integer multiple of 3; (2) the length of each CDS must be $\geq$ 300 bp (Wang et al. 2020); (3) the CDS should be of high quality with recognized bases, in other words, they contain only the bases of A, T, G, and C; (4) the sequence starts with the start codon ATG and ends with the stop codon TAA, TAG or TGA; 5) the CDC series does not contain intermediate termination codons (He et al. 2016; Li et al. 2016;). The GC content of three positions (GC1, GC2, and GC3) was calculated using the CUSP program in EMBOSS explorer (https://www.bioinformatics.nl/emboss-explorer/). Detailed information on the complete chloroplast genome and gene annotation of *Trifolium repens*, *Melilotus officinalis*, *Galega orientalis*, *Clitoria ternatea*, *Astragalus laxmannii*, *Galega officinalis*, *Pisum sativum*, *Stylosanthes guianensis* and *Medicago sativa* were shown in Table 1.

Analysis of RSCU, RFSC, and HF

RSCU is a statistical index that measures the ratio of the relative usage frequency of each synonymous codon (Wang et al. 2022). RSCU greater than 1, indicates that the codon is used more frequently than other synonymous codon frequencies and is positively biased for a specific amino acid.

Table 1 Genetic characterization of chloroplast genomes of nine forage legumes

Species	Accession No	CDSs number (before processing)	CDSs number (after processing)	L-aa	GC1%	GC2%	GC3%	Average GC at three locations%
Trifolium repens	NC_024036.1	75	50	20,048	45.67	37.47	27.99	37.04
Melilotus officinalis	NC_070051.1	76	44	15,855	46.25	37.85	27.00	37.04
Galega orientalis	NC_069214.1	73	47	16,817	46.26	37.87	26.54	36.89
Clitoria ternatea	NC_047365.1	81	50	20,331	44.11	36.55	25.98	35.55
Astragalus laxmannii	NC_052923.1	75	48	18,218	45.47	37.51	26.50	36.49
Galega officinalis	NC_051885.1	77	48	19,808	45.23	37.03	27.21	36.49
Pisum sativum	NC_014057.1	74	52	20,439	45.57	37.01	28.00	36.86
Stylosanthes guianensis	NC_058691.1	82	52	20,891	45.34	37.76	29.17	37.42
Medicago sativa	NC_042841.1	76	51	20,649	45.44	37.14	26.93	36.50

L-aa means the total number of amino acids; GC1, GC2, and GC3 indicate the GC content at the first, second, and third codon positions

RSCU less than 1, indicates that the codon is used less frequently than other synonymous codon frequencies and is negatively biased for a specific amino acid. RSCU of 1, indicates that the codon is not biased for a specific amino acid and is selected equally or randomly in the RNA transcript. Synonymous codons with RSCU values greater than 1.6 and less than 0.6 are considered overrepresented and underrepresented, respectively (Butt et al. 2014). RSCU is calculated using the formula of Sharp and Li (1986) as follows:

$$\mathrm{RSCU} = \frac{\mathrm{X_{ij}}}{\sum_{\mathrm{j}}^{\mathrm{n_i}} \mathrm{X_{ij}}} \mathrm{n_i}$$

where X_{ij} denotes the number of *j*th codons encoding the *i*-th amino acid, and n_i denotes the number of synonymous codons encoding the *i*-th amino acid, with the value ranging from 1 to 6.

RFSC is the ratio of the number of one codon to the number of synonymous codons in the actual observation, reflecting the usage frequency of each synonymous codon. It is calculated using the formula of Sharp and Li (1986) as follows:

$$\mathrm{RFSC} = \frac{\mathrm{X_{ij}}}{\sum_{\mathrm{j}}^{\mathrm{n_i}} \mathrm{X_{ij}}}$$

where X_{ij} denotes the number of *j*-th codons encoding the *i*-th amino acid. RFSC more than 60% or more than 0.5 times the average frequency of the synonymous codon indicates that the synonymous codon is a HF codon (Zhou 2007).

Analysis of the optimal codons

The high and low expression datasets of genes were established according to the ENC value of each gene. RSCU and ΔRSCU were then calculated using Codon W 1.4.2 software, and codons that satisfy the conditions of RSCU > 1 and ΔRSCU > 0.08 were defined as optimal codons (Romero et al. 2000).

Comparative analysis of codon usage frequency

For in-depth analysis of the codon usage patterns of nine forage legumes, the codon usage frequency data for the genomes of four model organisms, *Arabidopsis thaliana* (http://www.kazusa.or.jp/codon/cgi-bin/showcodon.cgi?species=3702), *Nicotiana tabacum* (http://www.kazusa.or.jp/codon/cgi-bin/showcodon.cgi?species=4097), *Escherichia coli* (http://www.kazusa.or.jp/codon/cgi-bin/showcodon.cgi?species=199310), and *Saccharomyces cerevisiae* (http://www.kazusa.or.jp/codon/cgi-bin/showcodon.cgi?species=4932), were downloaded from the Codon Usage Database (http://www.kazusa.or.jp/codon). The codon usage frequencies of the genomes of the nine forage legumes were compared with those of the four model organisms to investigate differences in codon usage patterns. In addition, the ratio of codon usage frequency of nine forages legumes to four model organisms was calculated. A ratio of ≥ 2 or ≤ 0.5, indicates that the two organisms have a large difference in codon usage bias and should not be selected as gene heterologous expression receptors (Pan et al 2013). A ratio between 0.5 and 2, indicates that they have a high degree of similarity in codon usage patterns and could be considered as gene heterologous expression receptors (Zhou et al. 2007a, b).

Analysis of ENC-plot

ENc is often used to analyze the degree of synonymous codon usage preference in the genes or genomes of specific species, its value ranges 20–61, in which a small value indicates a strong codon usage preference (Wu et al. 2018). GC3s refers to the ratio of the G + C content of the third

position of the codon of the CDS sequence to the total number of bases of genes other than Met and Trp. The ENC-plot was plotted with GC3s as the horizontal coordinate and ENC as the vertical coordinate to reveal the factors influencing the codon usage pattern of a gene or genome. When mutational pressure plays a major role in codon usage patterns, the ENC values lie on or near the expected curve. When natural selection and other factors play a major role in codon usage patterns, the ENC values lie well below the expected curve (Wright 1990).

$$ENC = 2 + S + \frac{29}{S^2+(1 - S)^2}$$

Parity rule 2 bias plot (PR2-plot) analysis

PR2-plot was used to analyze the usage and relationship of the four bases at position 3 of the codon encoding amino acids. Scatter plots were drawn with A3/(A3 + T3) as the vertical coordinate and G3/(G3 + C3) as the horizontal coordinate. The center of the plot refers to the position where A=T or G=C. In general, the ratio of A/T and C/G in codons of a gene or genome is balanced under a single mutational pressure (Xiang et al. 2015). The ratios vary considerably when the codon usage preferences have influenced by natural selection and other factors (Sueoka 1999a, b; Sueoka 1995).

Analysis of neutrality plot

Neutrality analysis was used to study the effects of mutational pressure and natural selection on codon preference (He et al. 2016). Here, y-axis GC12 represents the average of the GC content of the first and second codon positions, and x-axis GC3 represents the GC content of the third codon position. If GC12 and GC3 are unaffected by mutation pressure, then the points should be distributed along a diagonal (slope of 1). Conversely, if GC12 is completely nonneutral to GC3, then the points should be on the parallel lines of the horizontal coordinates (slope of 0). There trends suggest that codon preferences are affected by natural selection. When the regression coefficients converge or are equal to 1 and the correlation is significant, then mutational pressure plays a major role in the gene (Sueoka 1999a, b).

Correspondence analysis (COA)

COA is a statistical method used to analyze the relationship between variables and samples (Romero et al. 2003). Each CDS was represented as a 59-dimensional vector (excluding codons encoding the Met and Trp proteins and the three-stop codons) (Sharp and Li 1986), and each dimension corresponded to the RSCU of each synonymous codon (Wei et al. 2014). R programming language was used to draw the relationship chart of axis 1, axis 2, and the codon utilization index (the GC content of GC3s, the codon adaptation index (CAI), ENC, and L_aa) to further investigate the components affecting codon utilization preference.

Phylogenetic tree construction of nine forage legumes

To clarify the interspecific affinities and phylogenetic position of nine forage legumes, the nine forage legumes were used as ingroup, while *Rosa cymose*, *Broussonetia kaempferi*, *Fragaria viridis*, and *Elatostema dissectum* were selected as outgroup to jointly construct a phylogenetic tree. These chloroplast genomes and CDSs were subjected to multiple sequence comparisons and trimmed by MEGA 7.0 software (Shahzadi et al. 2020; Chi et al. 2023). Phylogenetic trees were constructed by neighbor-joining (NJ), and set 1000 times of Bootstrap test (BS) for the confidence of each clades, and all other parameters were set as default values. Evolutionary trees were beautified with Figtree software (http://tree.bio.ed.ac.uk/software/figtree/).

Results

Characteristics of codon base composition

The CDSs of the nine forage legumes processed by the Perl script contained 50, 44, 47, 50, 48, 48, 52, 52, and 51 (Table 1). Codon usage patterns are closely related to GC content (Shackelton et al. 2006), so we calculated the GC content of the three loci of codons GC1, GC2, and GC3 for each of the nine forage legumes species. We found that the contents of GC1, GC2, and GC3 and the average content of GC at the three loci of the nine forage legumes grasses were less than 40.00%. The order in each species was GC1 > GC2 > GC3, indicating that all nine forage legumes chloroplast genomes are inclined to use codons ending in A/T bases. In addition, the average GC content was the same for the three positions of *T. repens* and *M. officinalis* (37.04), and the average GC content was the same for the three positions of *A. laxmannii* and *G. officinalis* (36.49), but the contents of the other five species of forage legumes varied slightly (35.55–37.42; Table 1). The chloroplast genome codon preference for the use of A/T endings has been reported in 10 species of *Epimedium* species (Wang et al. 2023), 40 *Theaceae* species (Wang et al. 2022), and 18 *Oryza* species (Chakraborty et al. 2020).

RSCU and RFSC

To further understand the synonymous codon usage pattern and the preferences for G/C terminal codons and A/T terminal codons, we performed an RSCU analysis of codons in nine forage legumes. The results showed that the number of optimal codons (RSCU > 1) in the chloroplast genomes of the nine forage legumes was 30, except for *C. ternatea* that had 31. Among these, 29 of their codons ended with A/T (13 A and 16 T) except *C. ternatea* with 30 (14 A and 16 T). These values accounted for 96.68% of the total number of preferred codons, which was more than the number of codons ended with G/C. In summary, the nine forage legumes chloroplast genes showed a relatively great preference for A/T ended codons.

The range of variation in the RSCU values of the nine forage chloroplast genomes was close (Table S1). The TTA of codon encoding Leu had the highest RSCU (2.21), and the AGC of codon encoding Ser had the lowest RSCU (0.27). The following 15 commonly used HF codons were found in the chloroplast genomes of the nine forage legumes: GCT, GAT, GAA, GGA, CAT, TTA, ATG, AAT, CAA, AGA, TCT, GTT, TGG, TAT, and TAA (Fig. 1). In addition, codon overexpression (RSCU > 1.6) was found in all nine forage legumes, such as 9 in *P. sativum*, *M. officinalis*, and *S. guianensis*; 10 in *G. officinalis*; 11 in *C. ternatea* and *A. laxmannii*; 12 in *M. sativa*; and 13 in *C. ternatea* and *G. orientalis*. It is noteworthy that seven codons (GCT, TTA, AGA, TCT, ACT, TAT, and TAA) were overexpressed in all nine forages.

Determination of optimal codons

The high and low expression datasets of the chloroplast genes of nine forage legumes were established according to the ENC value of each CDS, and the optimal codons were screened by RSCU > 1 and ΔRSCU > 0.08. The results are shown in Table 2, with 149 optimal codons found for the nine forage legumes. Among them, 50 codons ended in A, 95 ended in T, and 4 ended in G. GCT, ACT, CCT, GGT, and TCT were the common optimal codons for the nine forage legumes.

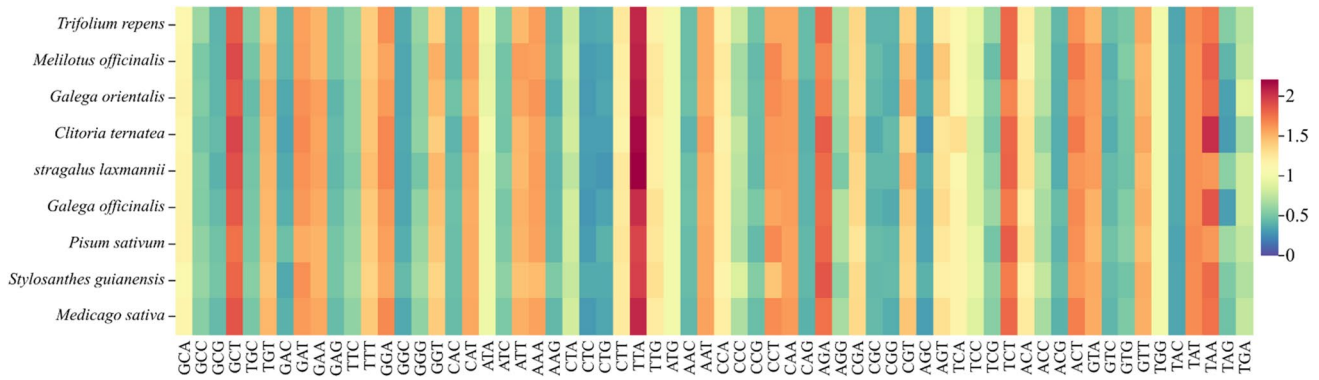

Fig. 1 Relative synonymous codon usage of chloroplast genes in nine forage legumes (blue to red indicates codons with low to high RSCU values)

Table 2 Optimal codons of chloroplast genomes of nine forage legumes

Species	Number	Optimal codon
Trifolium repens	20	TAA, GCT, TCT, CGT, CCT, AGA, TGT, ACT, GTT, GTA, TAT, CTT, CAA, GGT, GAA, TTT, GAT, ATT, TTA, GGA
Melilotus officinalis	14	CCT, CGT, GCT, GGT, ACT, TTA, TCT, TAA, ATT, TGT, CAA, GTA, GAT, TAT
Galega orientalis	14	GCT, ACT, CGT, TTG, CCT, GTT, GGT, CTT, GAT, TTA, ATT, AGT, CAA, GTA
Clitoria ternatea	14	ACT, GCT, TTA, TAA, GGT, TCT, CGA, TGT, CCT, GAA, GTT, AGA, TTT, TTG
Astragalus laxmannii	15	GCT, GTT, CGA, CTT, TAA, ACT, TCA, CAA, TTA, AGA, CCT, GGT, TCT, CCA, CAT
Galega officinalis	14	GCT, CGT, GTT, ACT, TTG, CCT, AGT, GGT, TTA, ATT, TTT, TGT, TCT, TAT
Pisum sativum	26	TTA, GGA, TAA, TCT, GAT, GCT, GTT, CAA, CCT, CGT, ACA, ATT, AGA, CTT, ACT, TGT, TAT, GAA, TTT, CCA, AAA, TTG, GTA, AGT, AAT, GGT
Stylosanthes guianensis	17	TTA, GCT, CGT, GAA, ATT, CAA, GGT, CGA, TAT, ACT, TCT、CCT,, TGT, TTT、AAA,, AGA, CCA
Medicago sativa	15	TAA, GCT, CTT, CGT, GTT, CCT, TCT, ACA, AGA, GGT, CAA, GAA, GCA, TAT, ACT

Codon usage frequency

Considering the differences in the codon usage preference of chloroplast genes of the nine forage legumes, we have compared the genomic codon usage frequencies of the nine forage grasses with those of the four model species in this study (Table 3). The results showed differences in the number of codons between the nine forage legumes and the four model species, in the descending order, *E. coli* (25–29), *A. thaliana* (13–19), *N. tabacum* (9–14), and *S. cerevisiae* (9–11). It is noteworthy that *G. orientalis*, and *A. laxmannii* had the same number of differential codons as the four model species. *G. officinalis*, *S. guianensis*, and *M. sativa* had the same number of differential codons with *A. thaliana*, *N. tabacum*, and *S. cerevisiae*. The *M. officinalis* and *G. officinalis* had the same number of differential codons with *E. coli* and *S. cerevisiae*. Only in terms of codon usage frequency, the relatively high codon usage differences between the nine forage legumes and *E. coli* can theoretically be excluded as heterologous expression receptors, the *N. tabacum* and *S. cerevisiae* can be considered receptors for the exogenous expression of chloroplast genes in the nine forage legumes.

Source analysis of variation in codon usage

ENC-plot

The effect of GC3s on codon preference was explored by drawing an ENC-GC3s plot to analyze the codon usage variation of chloroplast genes in the nine forage legumes (Fig. 2). As shown in the figure, the distribution of ENC-GC3s plots was relatively similar in the nine forage legumes. A small number of genes were located above or near the expected curve, suggesting that the mutational pressure had some effect on codon usage patterns. Most of the genes were located in the lower part of the expected curve, suggesting that natural selection plays an important role in the formation of codon usage bias.

PR2-plot

The usage between A/T and G/C at codon position 3 was analyzed to understand the effects of mutational and selection pressures on codon usage in the chloroplast genomes of the nine forage legumes (Fig. 3). If the frequencies of nucleotides A and T are equal to the frequencies of C and G in codon position 3, then the codon usage bias is only affected by mutational pressure (Kawabe and Miyashita 2003). On the contrary, the effect of natural selection would show inequality in the use of A and T bases and G and C bases. Observation of Fig. 3 shows that most genes in *T. repens* and *A. laxmannii* were distributed in the lower left region, indicating that T is used more frequently than A at codon position 3, and C is used more frequently than G. Most of the genes in the remaining seven forage legumes were distributed in the lower right region, indicating that T is used more frequently than A at codon position 3, and G is used more frequently than C. Therefore, the frequency

Table 3 Comparison of codon usage frequency between nine forage legumes and four model species

Species	Segment	*Arabidopsis thaliana*	*Nicotiana tabacum*	*Escherichia coli*	*Saccharomyces cerevisiae*
Trifolium repens	≥ 2 or ≤ 0.5	16	11	25	9
	0.5–2	48	53	39	55
Melilotus officinalis	≥ 2 or ≤ 0.5	12	9	26	10
	0.5–2	44	55	38	54
Galega orientalis	≥ 2 or ≤ 0.5	16	14	27	11
	0.5–2	48	50	37	53
Clitoria ternatea	≥ 2 or ≤ 0.5	19	14	29	12
	0.5–2	45	50	35	52
Astragalus laxmannii	≥ 2 or ≤ 0.5	16	14	27	11
	0.5–2	48	50	37	53
Galega officinalis	≥ 2 or ≤ 0.5	16	12	26	10
	0.5–2	48	52	38	54
Pisum sativum	≥ 2 or ≤ 0.5	13	10	26	8
	0.5–2	51	54	38	56
Stylosanthes guianensis	≥ 2 or ≤ 0.5	16	12	25	10
	0.5–2	48	52	38	54
Medicago sativa	≥ 2 or ≤ 0.5	16	12	27	10
	0.5–2	48	52	37	54

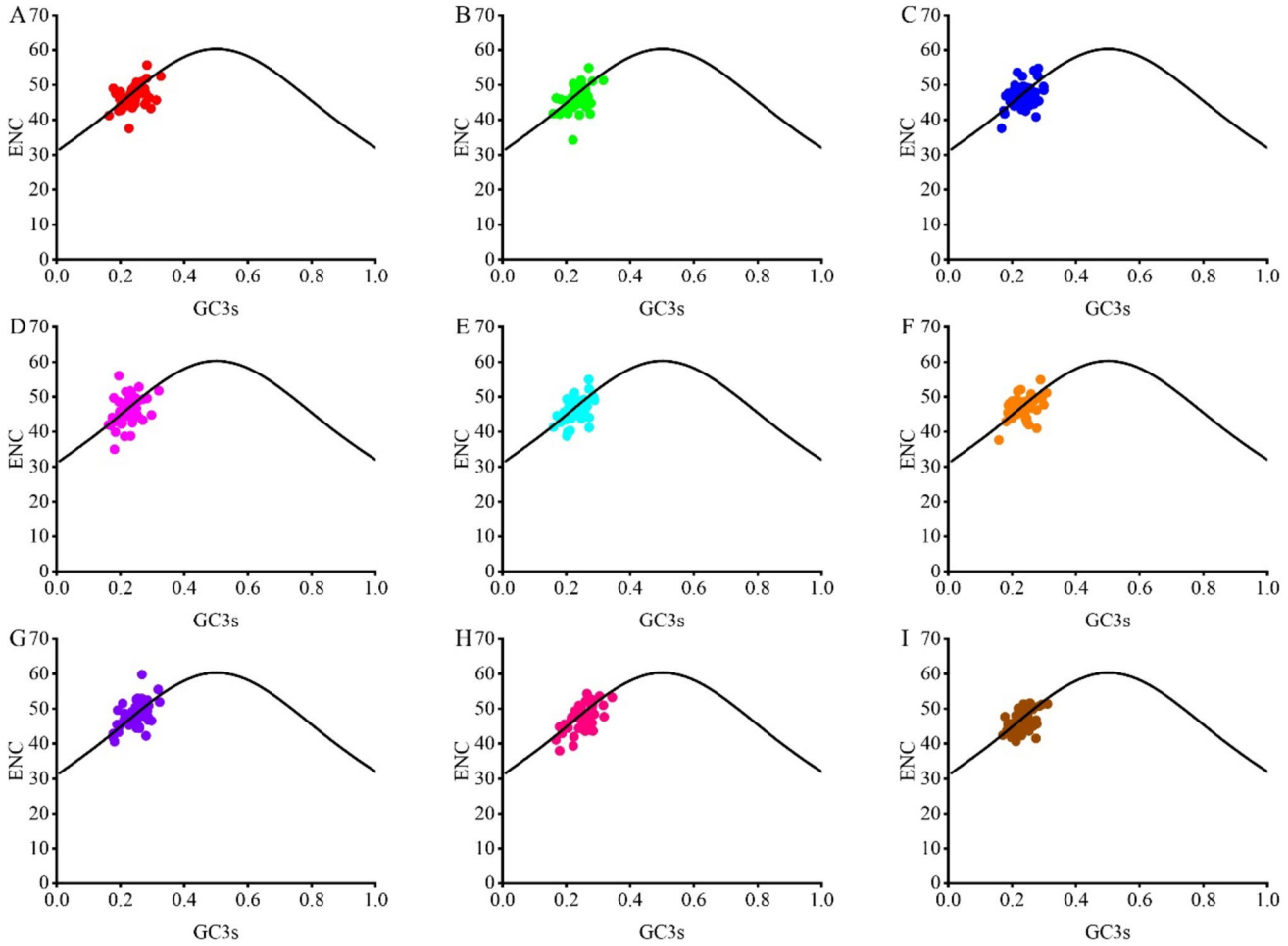

Fig. 2 ENC-plot analysis in the chloroplast genomes of nine forage legumes (**A** *T. repens*, **B** *M. officinalis*, **C** *G. orientalis*, **D** *C. ternatea*, **E** *A. laxmannii*, **F** *G. officinalis*, **G** *P. sativum*, **H** *S. guianensis*, **I** *M. sativa*)

of A/T and G/C use in the chloroplast genomes of the nine forage legumes were unbalanced, and natural selection plays an important role in the codon usage bias in the chloroplast genes of the nine forage legumes.

Neutrality plot

The chloroplast genes were subjected to neutral analysis (regression analysis between GC12 and GC3) to further investigate the extent of the influence of natural selection and mutational pressure on the codon usage bias of forage legumes. In Fig. 4, the slopes of the regression lines for the nine forage genomes were 0.2348, 0.2644, 0.4121, 0.5505, 0.4357, 0.3431, 0.2938, 0.4642, and 0.1390, and the correlation between GC12 and GC3 was weak ($r_1 = 0.159$, $r_2 = 0.164$, $r_3 = 0.267$, $r_4 = 0.327$, $r_5 = 0.248$, $r_6 = 0.222$, $r_7 = 0.189$, $r_8 = 0.325$, $r_9 = 0.088$). This finding indicated that mutational pressure accounted for only 23.48%–55.05% of the codon usage patterns of the nine forage genomes, and factors such as natural selection accounted for 44.95–76.52%. Here, natural selection plays a very important or even dominant role in codon usage patterns in the nine forage legumes, which is consistent with the results of ENC-GC3s and PR2-plot analyses.

COA

COA is a multivariate statistical method used for the relationship between variables in a sample. In this study, COA using RSCU revealed the main factors affecting the formation of codon usage patterns in the chloroplast genomes of forage legume plants. In the COA of nine forage legumes (Fig. 5), the coordinate origin indicates the average RSCU of all genes relative to axes 1 and 2. The first four axes accounted for 42.22%, 42.82%, 41.18%, 44.04%, 42.04%, 42.70%, 40.57%, 41.43%, and 42.44% of the overall variation. Axis 1 accounted for 19.20%, 16.45%, 19.46%, 19.77%, 18.67%, 20.11%, 18.12%, 19.07%, and 18.95% of the total variation in the nine forage legumes and thus was the major source of variation, accounting for approximately 20% of the

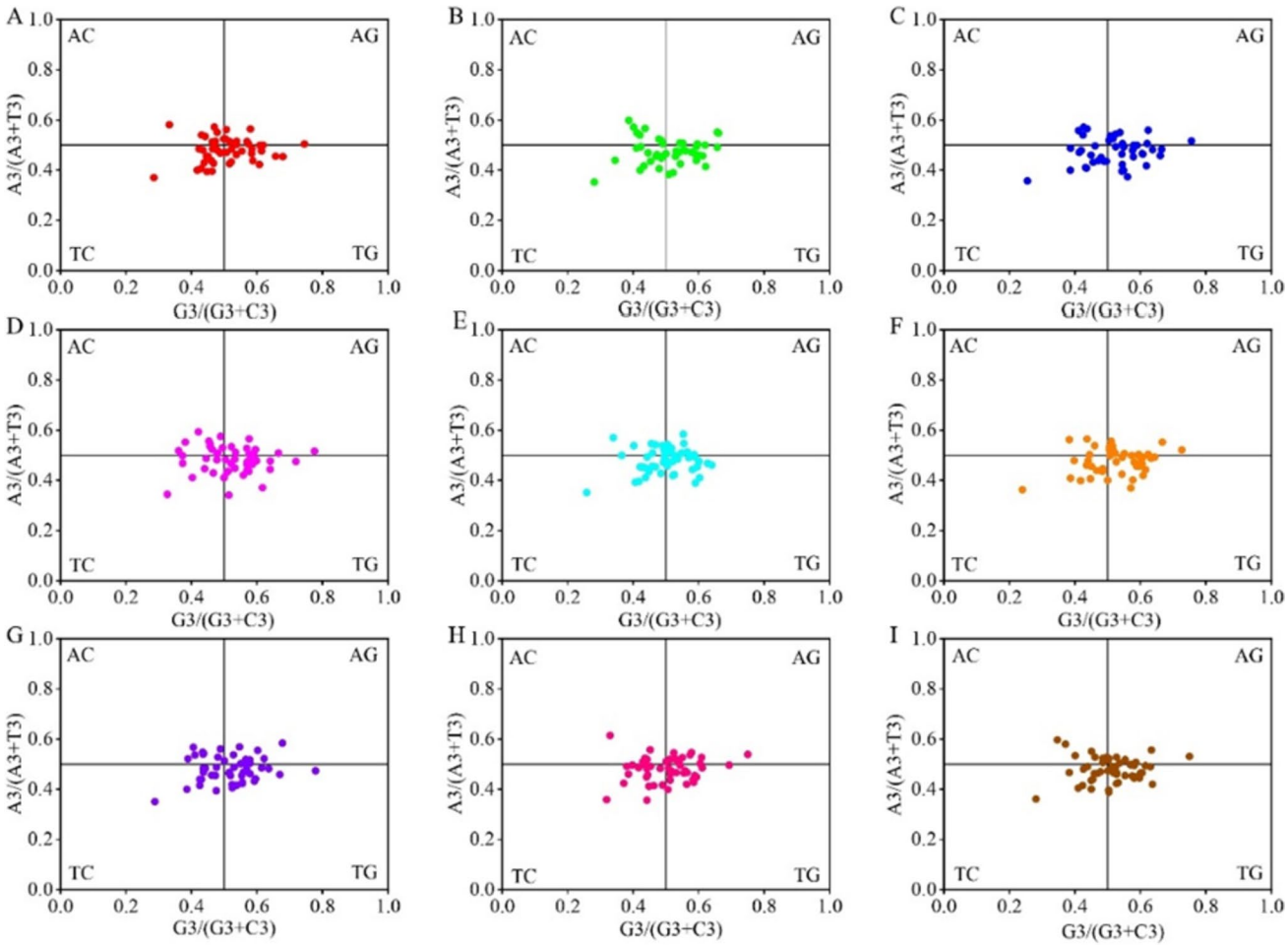

Fig. 3 PR2-polt analysis of the chloroplast genomes of nine forage legumes (**A** *T. repens*, **B** *M. officinalis*, **C** *G. orientalis*, **D** *C. ternatea*, **E** *A. laxmannii*, **F** *G. officinalis*, **G** *P. sativum*, **H** *S. guianensis*, **I** *M. sativa*)

total variation. This result suggested that codon usage may not be influenced by a single factor. Each gene in the chloroplasts of nine legume species was color-coded in different colors on a plane with axis 1 as the horizontal coordinate and axis 2 as the vertical coordinate to investigate the effect of GC content on CUB (Fig. 5). Among the nine legume species, only one gene in *S. guianensis* with a GC content of between 45 and 60% was color-coded in red, and all the other genes had a GC content of less than 45%. This high degree of consistency indicated that the sources of codon variation in the chloroplast genomes of the nine forage legumes are extremely similar.

Correlation analysis was performed on axis 1, axis 2 and codon index to further explore the factors affecting codon use preference (Fig. 6). The results showed that the CAI of *G. orientalis*, *P. sativum*, and *M. sativa* had a highly significant negative correlation with axis 1, the CAI of *M. officinalis* showed a significant negative correlation with axis 1; the CAI of *T. repens*, *C. ternatea*, *A. laxmannii*, *G. officinalis* and *S. guianensis* showed a highly significant positive correlation with axis 1; the L_aa of *M. officinalis*, *C. ternatea*, and *P. sativum* showed a significant positive correlation with axis 2, the L_aa of *M. sativa* showed a significant negative correlation with axis 2, *T. repens*, *G. orientalis*, *A. laxmannii*, *G. officinalis* and *S. guianensis* showed a non-significant correlation with axis 2. These findings indicated that gene expression and gene length have an effect on the codon bias of the nine forage legumes, with a greater effect on gene expression.

Phylogenetic relationships

In phylogenetic trees, species are clustered together according to different genera or proximity of kinship, and the nodes of the constructed phylogenetic tree, all have very high support. The phylogenetic relationships based on the chloroplast genome (Fig. 7a) and protein-coding genes (Fig. 7b) were highly similar. The nine forage legumes constituted a monophyletic clade, which significantly differed from the outgroups. Most clades were strongly supported by high

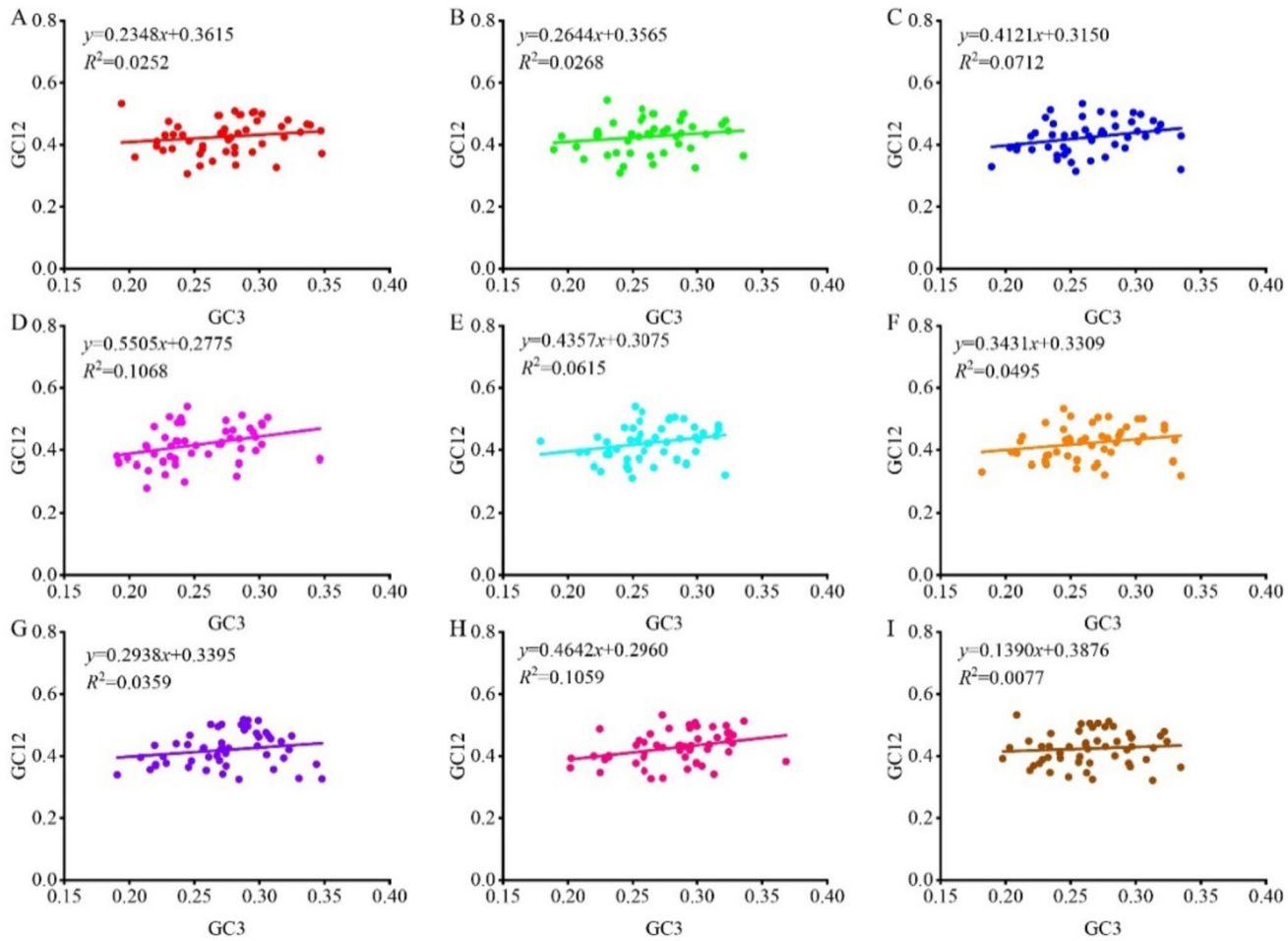

Fig. 4 Neutrality plot analysis in the chloroplast genomes of nine forage legumes (**A** *T. repens*, **B** *M. officinalis*, **C** *G. orientalis*, **D** *C. ternatea*, **E** *A. laxmannii*, **F** *G. officinalis*, **G** *P. sativum*, **H** *S. guianensis*, **I** *M. sativa*)

BS and posterior probabilities (PP). In both phylogenetic trees, the nine forage legumes were categorized into three major clades. The first clade consisted of *M. officinalis*, *M. sativa*, *T. repens*, *A. laxmannii*, *P. sativum*, *G. officinalis*, and *G.orientalis*. *C. ternatea* and *S.guianensis* formed the second and third clade, respectively. A sister relationship was observed between *M. officinalis* and *M. sativa*, and between *G. officinalis and G. orientalis*, and this relationship was strongly supported (BS = 100 and PP = 100). *S. guianensis* was distantly phylogenetically related to the remaining eight forage legumes.

Discussion and conclusion

Codon usage bias is an important feature of genome evolution, and is significant for the study of plant genetic engineering and genetic evolution (Liu et al. 2020). The chloroplast genomes of dicotyledonous plants prefer codons ending in A or U, whereas monocotyledonous plants prefer codons ending in G or C (Murray et al. 1989). The frequency of synonymous codon use varies among organisms, and GC composition, tRNA abundance, gene expression, and gene length are the key factors influencing codon use bias (Niu et al. 2021). Several studies focused on the codon usage preference and its causes in various plants, such as *Camellia japonica* (Wang et al. 2022), *Gynostemma pentaphyllum* (Zhang et al. 2021), and *Manihot esculenta* (Geng et al. 2022).

The codon usage pattern is closely related to the GC content. In this study, we calculated the GC content of the three loci of the forage chloroplast genome codons GC1, GC2, and GC3 to explore the codon usage preference pattern. The nine forage legumes codons had the highest percentage of GC1 and the lowest percentage of GC3, and the average GC content was below 40%. This finding indicated that the chloroplast genes of the nine forage legumes preferred codons ending in A/T, which was consistent with previous studies. The chloroplast genomes of other plants, such as *Panicum incomtum* (Li et al. 2021), *Oryza australiensis* (Chakraborty

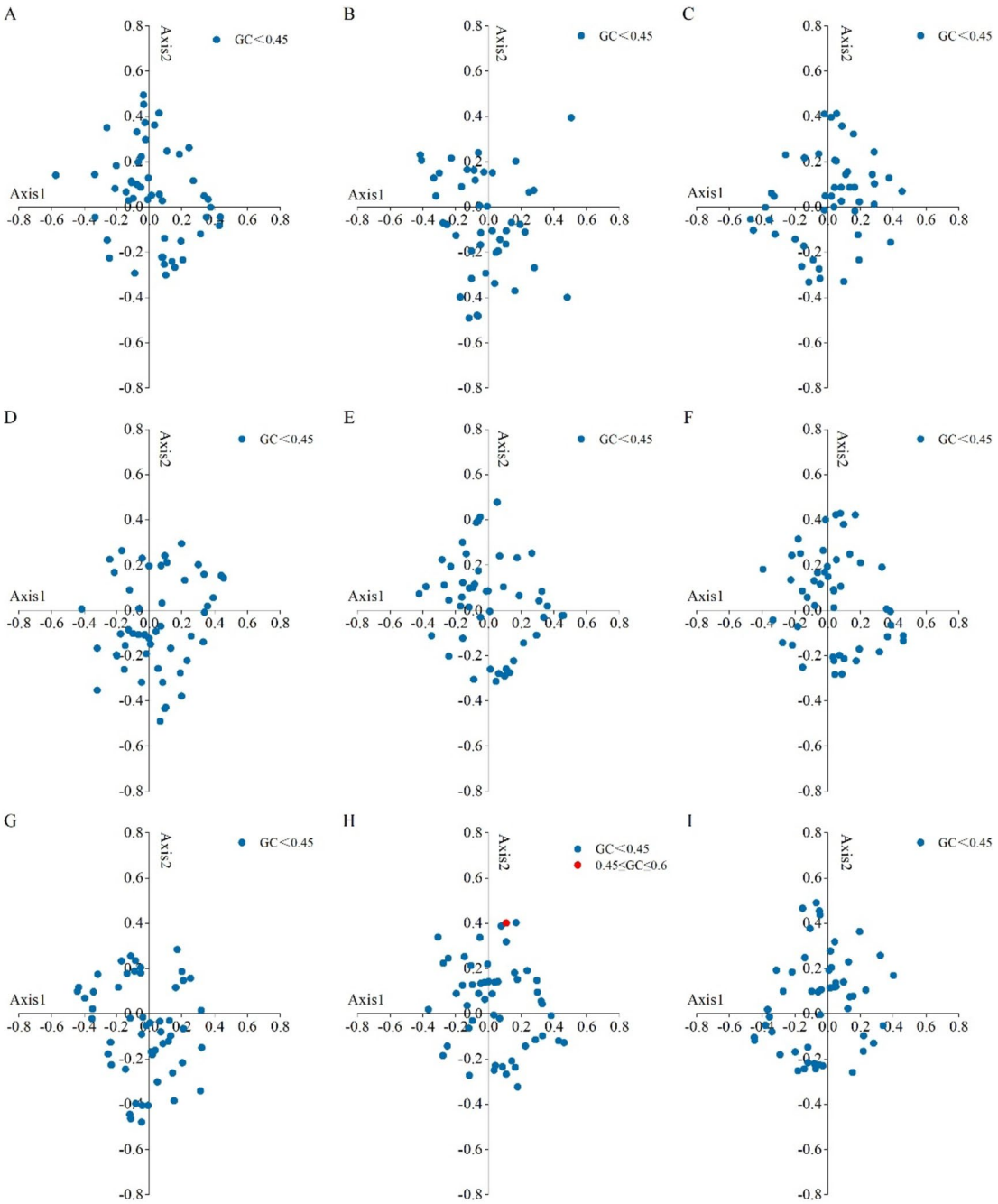

Fig. 5 Correspondence analysis of the chloroplast genomes of nine forage legumes. (**A** *T. repens*, **B** *M. officinalis*, **C** *G. orientalis*, **D** *C. ternatea*, **E** *A. laxmannii*, **F** *G. officinalis*, **G** *P. sativum*, **H** *S. guianensis*, **I** *M. sativa*)

et al. 2020), and *Euphorbia esula* (Wang et al. 2020) also tend to use codons ending in A or T. RSCU analysis revealed that most of the optimal codons (RSCU > 1) of the nine forage legumes ended in A/T, and the less common codons (RSCU < 1) ended in G/C. This finding is consistent with the results of the base composition analysis. Analysis of GC content indicated that among the nine forage legumes, *T. repens* was more similar to *M. officinalis*, *A. laxmannii*, and *G. officinalis* in terms codon preference. Hence, these species were presumed to be closely related. Comparative analysis of RSCU values revealed that the nine forage grasses had 29 common preferred codons, of which 96.68% ended in A/T.

Codon usage preference is mainly influenced by natural selection and mutational pressure (Rao et al. 2011). However, the main factors of codon usage preference differ

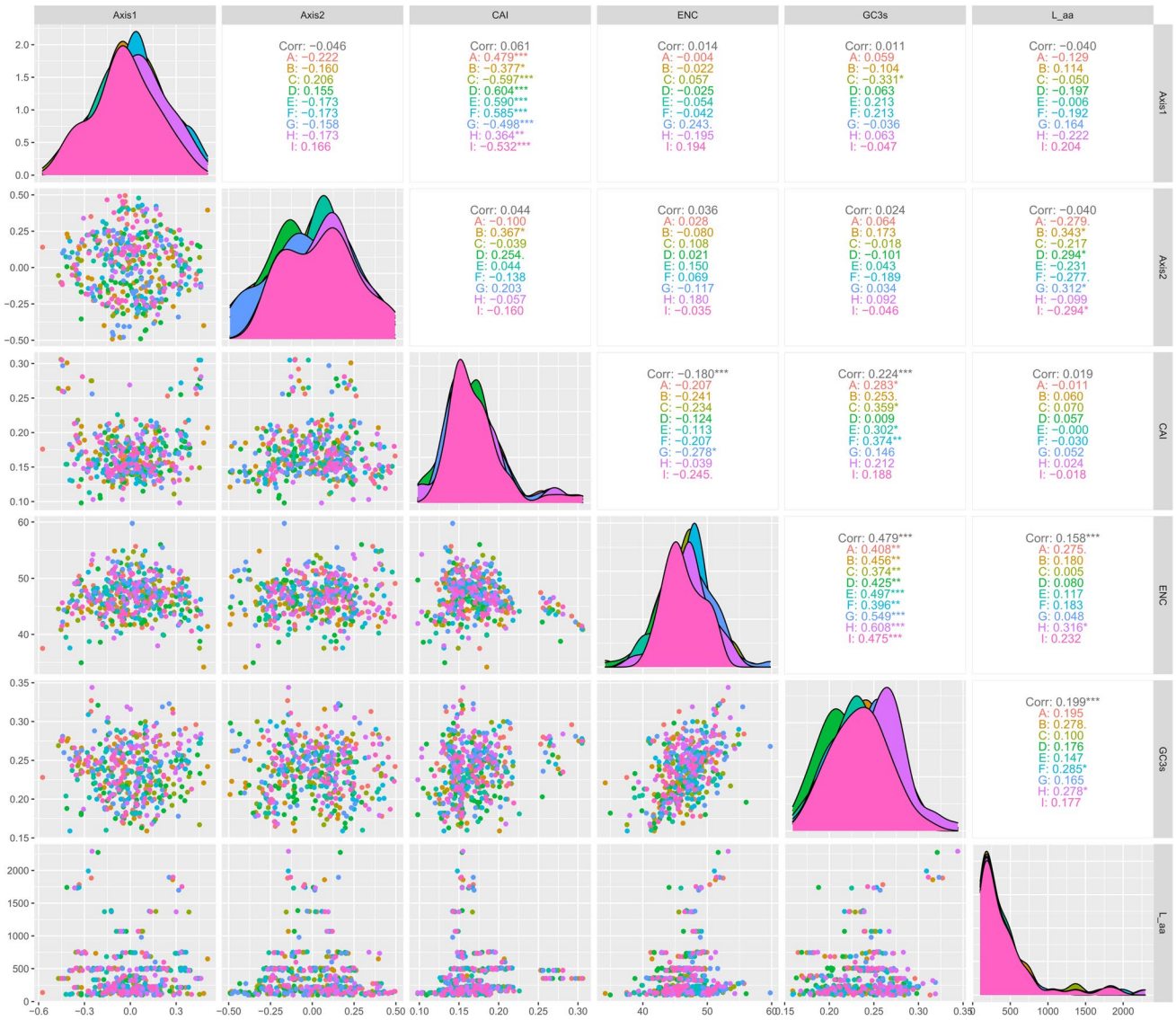

Fig. 6 Correlation analysis of axis 1, axis 2 and codon usage index of the chloroplast genomes of nine forage legumes. (**A** *T. repens*, **B** *M. officinalis*, **C** *G. orientalis*, **D** *C. ternatea*, **E** *A. laxmannii*, **F** *G. officinalis*, **G** *P. sativum*, **H** *S. guianensis*, **I** *M. sativa*)

among species. COA of the three codon sites in a neutral plot revealed that when mutational pressure plays a dominant role, the trends of base composition changes in the three codon sites should be similar. On the contrary, when natural selection plays a dominant role, there is no correlation between the three codon sites (Sharp et al. 2010). In our study, no significant correlation was found between GC12 and GC3 in the chloroplast genomes of the nine forage legumes. Mutation pressure accounted for 23.48%–55.05% of the codon usage pattern, and factors such as natural selection accounted for 44.95%–76.52%, suggesting that natural selection may have influenced codon selection in forage grasses. The slopes of the regression lines indicated that natural selection played a dominant role in the formation of codon usage patterns in constructing the codon bias of the forage chloroplast genomes. In conclusion, natural selection played a dominant role in determining codon usage bias in the chloroplast genomes of the nine forage legumes, followed by mutational pressure. This finding is similar to previous studies on the chloroplast genes of *P. sativum* (Bhattacharyya et al. 2019), *Medicago truncatula* (Song et al. 2018), and *Elaeis guineensis* (Aditama et al. 2020).

The PR2-plot showed that the frequency of use between A/T and G/C in codon position 3 of nine forage legumes was not uniform in the number of genes distributed in the four quadrants. Most genes were located below the midline in the vertical direction, and the number of genes on the right side of the midline was higher than that on the left

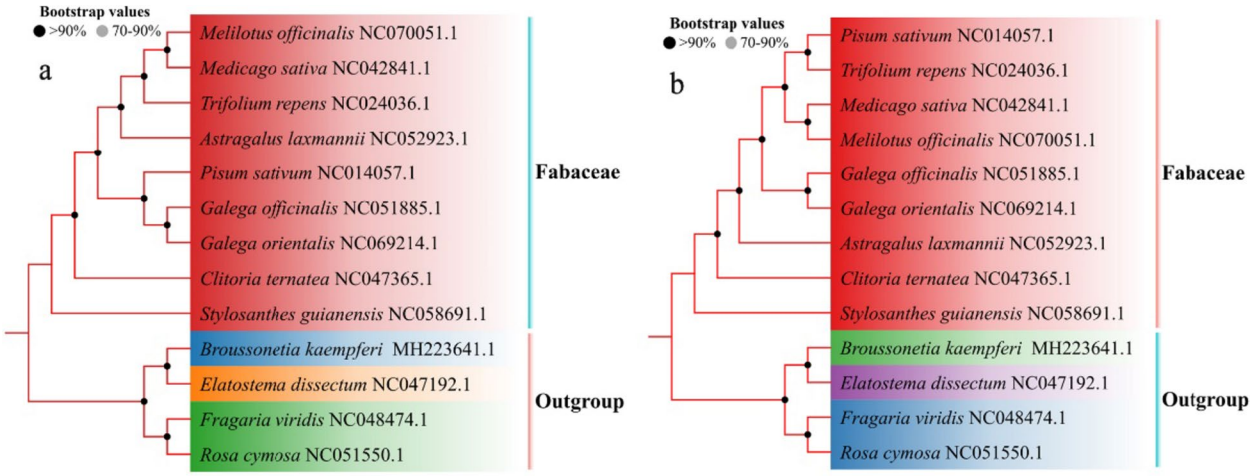

Fig. 7 Phylogenetic tree based on the complete chloroplast genome (**a**) and protein coding sequences (**b**)

side of the midline in the horizontal direction. Therefore, the frequency of G and T in the third position of the codon is higher than that of C and A. Natural selection is the main factor of codon usage bias in the chloroplast genomes of nine forage legumes (Zhang et al. 2007).

Neutral plot analysis, PR2-plot analysis, and ENC plot analysis revealed that the codon preferences of the chloroplast genomes of the nine forage legumes were jointly affected by natural selection and mutational pressure, with natural selection playing a dominant role. The result that natural selection is the main driver of chloroplast genome was also observed in other plants such as *Miscanthus* (Sheng et al. 2021), *P. sativum* (Bhattacharyya et al. 2019), and *M. truncatula* (Song et al. 2018). On the basis of these results, different genomes may be affected by different pressures. A total of 149 optimal codons were found in the nine forage legumes evaluated in this paper, with GCT, ACT, CCT, GGT, and TCT as the common ones. These results are important for improving the expression of chloroplast genes in host cells. The selection of heterologous expression hosts directly affects the expression of imported genes in host organisms. To realize the successful expression of exogenous genes and improve their expression, we should select the ones with few differences in codon usage preference as the receptor. By comparing the differences in the codon usage frequency of the nine forage genes and the four model organisms, we have found that *S. cerevisiae* is the best heterologous expression host.

The study of plant phylogeny from the perspective of chloroplast whole genome has become one of the hotspots in current plant classification research. In this study, the phylogenetic relationships based on the chloroplast genome and protein-coding genes were highly similar. The nine forage legumes constituted a monophyletic clade, which significantly differed from the outgroups, and most clades were strongly supported by high BS and PP.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12298-024-01421-0.

Acknowledgements The authors would like to thank Associate Professor C.F. Duan, Assistant Research Fellow J.M. Song and Z.S. Shen for their help and support during the text writing process.

Author contributions M.K. Xiao, X.H. and Y.Q. Li. Writing-original draft, Methodology, Drawing. Q.L. Data curation, Supervision, Funding acquisition. S.B. Shen. and T.L. Jiang. Conceptualization, Data curation and analysis. L.H. Zhang. and Y.C. Zhou. Data curation, Validation, Resources. Y.X. Li, X.L. and L.N. Bai. Data curation, Supervision. W.Y. Writing-review and editing, Funding acquisition.

Funding This research was funded by Xingdian Talent Support Plan project-Integrated demonstration of key technologies for breeding and industrialization of excellent forage varieties in Yunnan, the National key research and development program (2022YFF1302404), the 10th batch of young and middle-aged academic and technical leaders in Baoshan City-Yan Wei (Bao Zhengfu [2021]35), the Baoshan City Science and Technology plan project-Baoshan fine forage variety screening and high-yield cultivation technology (2022zc07) and Baoshan City Science and Technology plan project-Baoshan fine forage variety screening and high-yield cultivation technology (2022zc04).

Data availability Data will be made available on request.

Declarations

Conflict of interest The authors declare no potential conflicts of interest in this paper.

References

Aditama R, Tanjung ZA, Sudania WM, Nugroho YA, Utomo C, Liwang T (2020) Analysis of codon usage bias reveals optimal codons in *Elaeis guineensis*. Biodiversitas 21(11):5311–5337. https://doi.org/10.13057/biodiv/d211138

Baeza M, Alcaíno J, Barahona S, Sepúlveda D, Cifuentes V (2015) Codon usage and codon context bias in *Xanthophyllomyces dendrorhous*. BMC Genomics 16(1):293. https://doi.org/10.1186/s12864-015-1493-5

Bhattacharyya D, Uddin A, Das SK, Chakraborty S (2019) Mutation pressure and natural selection on codon usage in chloroplast genes of two species in Pisum L. (Fabaceae: Faboideae). Mitochondrial DNA Part B 30(4):664–673. https://doi.org/10.1080/24701394.2019.1616701

Boel G, Letso RR, Neely H, Price WN, Wong KH, Su M et al (2016) Codon influence on protein expression in *E. coli* correlates with mRNA levels. Nature 529(7586):358. https://doi.org/10.1038/nature16509

Bulmer MG (1991) The selection-mutation-drift theory of synonymous codon usage. Genetics 129(3):897–907. https://doi.org/10.1016/1050-3862(91)90016-K

Butt AM, Nasrullah I, Tong Y (2014) Genome-wide analysis of codon usage and influencing factors in chikungunya viruses. PLoS ONE 9(3):e90905. https://doi.org/10.1371/journal.pone.0090905

Camiolo S, Melito S, Porceddu A (2015) New insights into the interplay between codon bias determinants in plants. DNA Res 22(6):461–470. https://doi.org/10.1093/dnares/dsv027

Chakraborty S, Yengkhom S, Uddin A (2020) Analysis of codon usage bias of chloroplast genes in Oryza species: codon usage of chloroplast genes in Oryza species. Planta 252(4):67. https://doi.org/10.1007/s00425-020-03470-7

Chi XF, Chen RS, Zhang FQ, Chen SL (2023) Comparative plastomes of species from Phrymaceae and Mazaceae: insights into adaptive evolution, codon usage bias and phylogenetic relationships. Genome 66(11):281–294. https://doi.org/10.1139/GEN-2023-0014

Daniell H, Lin CS, Yu M, Chang WJ (2016) Chloroplast genomes: diversity, evolution, and applications in genetic engineering. Genome Biol 17:134. https://doi.org/10.1186/s13059-016-1004-2

Dobrogojski J, Adamiec M, Luciński R (2020) The chloroplast genome: a review. Acta Physiol Plant 42:98. https://doi.org/10.1007/s11738-020-03089-x

Geng XS, Huang N, Zhu Y, Liu Q, Hui LS (2022) Codon usage bias analysis of the chloroplast genome of cassava. S Afr J Bot 151(PA):970–975. https://doi.org/10.1016/j.sajb.2022.11.022

Gustavo GS, Timothy DC (2017) Flooding tolerance of forage legumes. J Exp Bot 68(8):1851–1872. https://doi.org/10.1093/jxb/erw239

He B, Dong H, Jiang C, Cao FL, Tao ST, Xu LA (2016) Analysis of codon usage patterns in *Ginkgo biloba* reveals codon usage tendency from A/U-ending to G/C-ending. Sci Rep 3(6):35927. https://doi.org/10.1038/srep35927

Kawabe A, Miyashita NT (2003) Patterns of codon usage bias in three dicot and four monocot plant species. Genes Genet Syst 78(5):343–352. https://doi.org/10.1266/ggs.78.343

Li N, Sun MH, Jiang ZS, Shu HR, Zhang SZ (2016) Genome-wide analysis of the synonymous codon usage patterns in apple. J Integr Agric 15:983–991. https://doi.org/10.1016/s2095-3119(16)61333-3

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

Liu HB, Lu YZ, Lan BL, Xu JF (2020) Codon usage by chloroplast gene is bias in *Hemiptelea davidii*. J Genet 99:8. https://doi.org/10.1007/s12041-019-1167-1

Lüscher A, Mueller HI, Soussana JF, Rees RM, Peyraud JL (2014) Potential of legume-based grassland livestock systems in Europe: a review. Grass Forage Sci 69:206–228. https://doi.org/10.1111/gfs.12124

Murray EE, Lotzer J, Eberle M (1989) Codon usage in plant genes. Nucleic Acids Res 17(2):477–498. https://doi.org/10.1093/nar/17.2.477

Niu Y, Luo YY, Wang CL, Liao WB (2021) Deciphering codon usage patterns in genome of *Cucumis sativus* in comparison with nine species of Cucurbitaceae. Agronomy 11(11):2289. https://doi.org/10.3390/agronomy11112289

Pan LL, Wang Y, Hu JH, Ding ZT, Li C (2013) Analysis of codon use features of stearoylacyl carrier protein desaturase gene in *Camellia sinensis*. J Theor Biol 334:80–86. https://doi.org/10.1016/j.jtbi.2013.06.006

Peyraud JL, Gall AL, Lüscher A (2009) Potential food production from forage legume-based systems in Europe: an overview. Irish J Agric Food Re 48(2):115–135. https://doi.org/10.3109/09637480802165650

Phelan P, Moloney AP, McGeough EJ, Humphreys J, Bertilsson J, O'Riordan EG et al (2015) Forage legumes for grazing and conserving in ruminant production systems. Crit Rev Plant Sci. https://doi.org/10.1080/07352689.2014.898455

Plotkin JB, Kudla G (2011) Synonymous but not the same: the causes and consequences of codon bia. Nat Rev Genet 12:32–42. https://doi.org/10.1038/nrg2899

Rao YH, Wu GZ, Wang ZF, Chai XW, Nie QH, Zhang XQ (2011) Mutation bias is the driving force of codon usage in the *Gallus gallus* genome. DNA Res 18(6):499–512. https://doi.org/10.1093/dnares/dsr035

Rochon JJ, Doyle CJ, Greef JM, Hopkins A, Molle G, Sitzia M et al (2004) Grazing legumes in Europe: a review of their status, management, benefits, research needs and future prospects. Grass Forage Sci 59:197–214. https://doi.org/10.1111/j.1365-2494.2004.00423.x

Romero H, Zavala A, Musto H (2000) Codon usage in *Chlamydia trachomatis* is the result of strand-specific mutational biases and a complex pattern of selective forces. Nucleic Acids Res 28(10):2084–2090. https://doi.org/10.1093/nar/28.10.2084

Romero H, Zavala A, Musto H, Bernardi G (2003) The influence of translational selection on codon usage in fishes from the family Cyprinidae. Gene 317:141–147. https://doi.org/10.1016/s0378-1119(03)00701-7

Shackelton LA, Parrish CR, Holmes EC (2006) Evolutionary basis of codon usage and nucleotide composition bias in vertebrate DNA viruses. J Mol Evol 62:551–563. https://doi.org/10.1007/s00239-005-0221-1

Shahzadi I, Abdullah MF, Ali Z, Ahmed I, Mirza B (2020) Chloroplast genome sequences of *Artemisia maritima* and *Artemisia absinthium*: comparative analyses, mutational hotspots in genus Artemisia and phylogeny in family Asteraceae. Genomics 112(2):1454–1463. https://doi.org/10.1016/j.ygeno.2019.08.016

Sharp PM, Li WH (1986) An evolutionary perspective on synonymous codon usage in unicellular organisms. J Mol Evol 24:28–38. https://doi.org/10.1007/BF02099948

Sharp PM, Emery LR, Zeng K (2010) Forces that influence the evolution of codon bias. Philos Trans R Soc B 365:1203–1212. https://doi.org/10.1098/rstb.2009.0305

Sheng JJ, She X, Liu XY, Wang J, Hu ZL (2021) Comparative analysis of codon usage patterns in chloroplast genomes of five Miscanthus species and related species. Peer J 9:e12173. https://doi.org/10.7717/peerj.12173

Song H, Liu J, Chen T, Nan ZB (2018) Synonymous codon usage pattern in model legume *Medicago truncatula*. J Integr Agric 17(9):2074–2081. https://doi.org/10.1016/S2095-3119(18)61961-6

Sueoka N (1995) Intrastrand parity rules of DNA base composition and usage biases of synonymous codons. J Mol Evol 40(3):318–325. https://doi.org/10.1007/BF00163236

Sueoka N (1999a) Two aspects of DNA base composition: G+C content and translation-coupled deviation from intra-strand rule of A=T and G=C. J Mol Evol 49(1):49–62. https://doi.org/10.1007/PL00006534

Sueoka N (1999b) 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(1):53–58. https://doi.org/10.1016/s0378-1119(99)00320-0

Van Grinsven HJM, Holland M, Jacobsen BH, Klimont Z, Sutton MA, Willems WJ (2013) Costs and benefits of nitrogen for Europe and implications for mitigation. Environ Sci Technol 47(8):3571–3579. https://doi.org/10.1021/es303804g

Wang ZJ, Xu BB, Li B, Zhou QQ, Wang GY, Jiang XZ et al (2020) Comparative analysis of codon usage patterns in chloroplast genomes of six Euphorbiaceae species. Peer J 8:e8251. https://doi.org/10.7717/peerj.8251

Wang ZJ, Cai QW, Wang Y, Li MH, Wang CC, Wang ZX et al (2022) Comparative analysis of codon bias in the chloroplast genomes of Theaceae species. Front Genet 13:824610. https://doi.org/10.3389/fgene.2022.824610

Wang YZ, Jiang DC, Guo K, Zhao L, Meng FF, Xiao JL et al (2023) Comparative analysis of codon usage patterns in chloroplast genomes of ten Epimedium species. BMC Genomic Data 24(1):3. https://doi.org/10.1186/s12863-023-01104-x

Wei L, He J, Jia X, Qi Q, Liang ZS, Zheng H et al (2014) Analysis of codon usage bias of mitochondrial genome in *Bombyx moriand* its relation to evolution. BMC Evol Biol 14(1):262. https://doi.org/10.1186/s12862-014-0262-4

Wright F (1990) The 'Effective Number of Codons' used in a gene. Gene 87(1):23–29. https://doi.org/10.1016/0378-1119(90)90491-9

Wu YQ, Li ZY, Zhao DQ, Tao J (2018) Comparative analysis of flower-meristem-identity geneAPETALA2 (AP2) codon in different plant species. J Integr Agric 17:867–877. https://doi.org/10.1016/S2095-3119(17)61732-5

Xiang H, Zhang RZ, Butler RR, Liu T, Zhang L, Pombert JF et al (2015) Comparative analysis of codon usage bias patterns in Microsporidian genomes. PLoS ONE 10(6):e0129223. https://doi.org/10.1371/journal.pone.0129223

Zhang WJ, Zhou J, Li ZF, Wang L, Gu X, 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. https://doi.org/10.1111/j.1744-7909.2007.00404.x

Zhang YR, Ni XJ, Jia XO, Zhao CZ, Biradar SS, Wang L et al (2012) Analysis of codon usage patterns of the chloroplast genomes in the Poaceae family. Aust J Bot 60:461–470. https://doi.org/10.1071/BT12073

Zhang PP, Xu WB, Lu X, Wang L (2021) Analysis of codon usage bias of chloroplast genomes in Gynostemma species. Physiol Mol Biol Plants 27(12):2727–2737. https://doi.org/10.1007/s12298-021-01105-z

Zhou M, Tong CF, Shi JS (2007a) A preliminary analysis of synonymous codon usage in poplar species. J Plant Physiol Mol Biol 33(4):285–293. https://doi.org/10.3321/j.issn:1671-3877.2007.04.003

Zhou M, Tong CF, Shi JS (2007b) Analysis of codon usage between different poplar species. J Genet Genomics 34(6):555–561. https://doi.org/10.1016/S1673-8527(07)60061-7