Check for updates

Analysis of codon usage patterns in complete plastomes of four medicinal *Polygonatum* species (Asparagaceae)

Naixing Shi, Yiwen Yuan, Renjie Huang and Guosong Wen*

College of Agronomy and Biotechnology, Yunnan Agricultural University, Kunming, China

Polygonati Rhizoma and Polygonati odorati Rhizoma, known as "Huangjing" and "Yuzhu" in China, are medicinal *Polygonatum* species resources with top-grade medical and edible properties. The chloroplast (cp) genome has been used to study species diversity, evolution, and breeding of species for applications in genetic engineering. Codon usage bias (CUB), a common and complex natural phenomenon, is essential for studies of codon optimization of exogenous genes, genetic engineering, and molecular evolution. However, the CUB of medicinal *Polygonatum* species chloroplast genomes has not been systematically studied. In our study, a detailed analysis of CUB was performed in the medicinal *Polygonatum* species chloroplast genomes. We investigated the codon bias of 204 plastid protein-coding genes (PCGs) in 4 medicinal *Polygonatum* species using CodonW and CUSP online software. Through the analysis of the codon bias index, we found that the medicinal *Polygonatum* species chloroplast genomes had weak codon usage bias. In addition, our results also showed a high preference for AT bases in medicinal *Polygonatum* species chloroplast genomes, and the preference to use AT-ending codons was observed in these species chloroplast genomes. The neutrality plot, ENC plot, PR2-Bias plot, and correspondence analysis showed that compared with mutation pressure, natural selection was the most important factor of CUB. Based on the comparative analysis of high-frequency codons and high expression codons, we also determined the 10-11 optimal codons of investigative medicinal *Polygonatum* species. Furthermore, the result of RSCU-based cluster analysis showed that the genetic relationship between different medicinal *Polygonatum* species could be well reflected. This study provided an essential understanding of CUB and evolution in the medicinal *Polygonatum* species chloroplast genomes.

1 Introduction

The *Polygonatum* Mill. comprises approximately 80 species and is species-rich genus in Polygonateae, which occur in temperate Northern Hemisphere regions (Wang J et al., 2022; Wang et al., 2023; Floden and Schilling, 2018). Many species of this genus are significant medicinal plants in China, and have been widely used as Traditional Chinese Medicines (TCM) to treat fatigue, weakness, diabetes, cough, and loss of appetite for thousands of years (Zhao et al., 2018; Chen et al., 2021). The Chinese Pharmacopoeia embodied four species

(*Polygonatum odoratum*, *Polygonatum sibiricum*, *Polygonatum cyrtonema*, *Polygonatum kingianum*) as medicinal *Polygonatum* species (Chinese Pharmacopoeia Commission, 2020). In addition, as medicinal and edible plants, most of this genus greatly benefits human health, which is often used as one of the main raw materials of health food (Cui et al., 2018; Bai et al., 2021; Xiao et al., 2023). Because of the remarkable medicinal and edible value of medicinal *Polygonatum* species, hundreds of drugs and functional foods, such as medicine, food, and drink whose raw materials are based on *Polygonati Rhizoma* and *Polygonati odorati Rhizoma*, have been developed and applied in China (https://db.yaozh.com/zhongyaocai) (Guo et al., 2022).

As important organelles in green plants, chloroplasts play a crucial role in photosynthesis. According to their functions, there are four categories of chloroplast genes: self-replication genes, photosynthesis genes, biosynthesis-related genes, and unknown function genes (Cao et al., 2022; Sugiura, 1995). The chloroplast genome has the characteristics of relatively small size, uniparental inheritance, multicopy numbers, and highly conserved genomic structure (Daniell et al., 2016). These features are of great advantage in genetic transformation (James and Daniell, 2005; Wang et al., 2009). Therefore, it has attracted extensive attention from relevant researchers in recent years (Kumar and Ling, 2021; Singhal et al., 2023; Basso et al., 2020). To date, chloroplast genome engineering has been used in many biotechnology applications (Daniell et al., 2016). With the rapid development of high-throughput sequencing technologies, the number of chloroplast genomes has been increased using next-generation sequencing (NGS) technology in the NCBI (https://www.ncbi.nlm.nih.gov/) and NGDC (https://ngdc.cncb.ac.cn/?lang = zh) databases. The chloroplast genomes of four medicinal *Polygonatum* species (*P. odoratum*, *P. sibiricum*, *P. cyrtonema*, and *P. kingianum*) have also attracted significant attention and have been reported by many researchers (Wang et al., 2023; Xia et al., 2022; Floden and Schilling, 2018).

The frequency of synonymous codons probability differs in all genes and organisms due to the comprehensive effects of mutation, selection, and genetic drift (Parvathy et al., 2021; Salim and Cavalcanti, 2008). Organisms and genes tend to encode amino acids using one or more specific synonymous codons called codon usage bias (CUB). Studying the CUB of organisms has important biological implications. Because closely related species have similar patterns of codon bias, therefore, CUB analysis can reveal species evolutionary, horizontal gene transfer and so on (Athey et al., 2017; Tuller, 2011). Besides, studies have found that CUB is vital for a multitude of cellular processes, for example, influencing the stability and efficiency of transcription, as well as the structure, expression, function of proteins (Chakraborty et al., 2020). Hence analyzing CUB can be utilized to optimize expression vectors in genetic engineering, thereby enhancing the expression levels of target genes (Kwon et al., 2016). As a universal natural phenomenon, it exists widely in different organisms (Wang et al., 2011). At present, numerous studies on CUB of medical plant have been reported, e.g., *Dendrobium* species (Wang X. S et al., 2022), *Lilium* species (Dai et al., 2024), *Aconitum* species (Yang et al., 2023). However, CUB in medicinal *Polygonatum* species has not been exhaustively investigated. The quality of Polygonati Rhizoma and Polygonati odorati Rhizoma products is critical to the pharmaceutical and food

industries. By studying the CUB of medicinal *Polygonatum* species cp genomes, stable genetic transformation can be more efficiently constructed and increase the expression level of the target gene in genetic transformation research. Moreover, it will also provide an important research basis for revealing the genetic relationships and evolutionary analysis of these medicinal *Polygonatum* species.

2 Materials and methods

2.1 Sequence data

The voucher specimen (SNX3) was collected from the Yuxin TCM plantation base (N25.76104 and E103.75594) and deposited at Yunnan Agricultural University, Kunming, China (Figure 1). Leaf tissues from SNX3 were dried with silica gel. The total DNA was extracted from 100 mg of dried leaves using the modified CTAB method (Porebski et al., 1997). Then, a paired-end (PE) library of 150 bp was constructed using the total DNA and performed on the Illumina Hiseq 2500 sequencing platform. The PE reads were trimmed for adapter and low-quality reads (Phred score <30) using NGS QC Toolkit v.2.3.3 software (Patel and Jain, 2012). The *P. kingianum* cp genome was assembled using GetOrganelle v.1.6.4 (Jin et al., 2020). After assembly, the circular cp genomes were annotated using online tools CpGAVAS2 (Shi et al., 2019) and GeSeq (Tillich et al., 2017) based on the reference cp genome (GenBank: MZ286313). The annotated cp genome sequence was submitted to the GenBank database of the NCBI (Accession: PP315902). The plastome map (Figure 2) was drawn by OrganellarGenomeDRAW (OGDRAW) online (Lohse et al., 2007). In addition, we also downloaded the other medicinal *Polygonatum* species chloroplast genomes from the NGDC database (https://ngdc.cncb.ac.cn/?lang = zh), including two species (*P. cyrtonema*, GenBank number: ON872700; *P. sibiricum*, GenBank number: OQ532972) previously reported by us and one species (*P. odoratum*, GenBank number: MZ150858) reported by other scholars. Information on medicinal *Polygonatum* species is shown in Table 1.

2.2 Sequence processing

In order to reduce sampling errors and redundancy of the CUB analysis, all protein-coding sequences (CDSs) from chloroplast genomes were selected for subsequent analysis based on the following principles: 1) excluding duplicate genes, 2) CDSs with length greater than 300 bp were retained, 3) retaining CDSs without intermediate stop codons and erroneous bases. 4) CDSs starting with a start codon (ATG) and ending with a stop codon (TAG, TGA, TAA) were retained (Wang et al., 2020; Li et al., 2016).

2.3 Analysis of codon usage index

All the codon usage indicators were calculated through CodonW 1.4.2 software (Peden, 1999) and the CUSP program of EMBOSS (Rice et al., 2000), including (1) effective number of codon (ENC), (2) relative synonymous codon usage (RSCU), (3) codon number

FIGURE 1
Polygonatum kingianum Coll. et Hemsl. **(A)** individual, **(B)** whorled leaf, **(C)** flower and umbel, **(D)** anatomy of flower, **(E)** leaves.

(L_aa), (4) overall GC content of genes and the GC content at the first, second and third base position of codons (GC1, GC2, GC3), (5) the G or C content of the third base of a synonymous codon (GC3s), (6) The frequency of base A, T, G, and C at the third position of codons ($A3_S$, $T3_S$, $G3_S$, $C3_S$).

2.4 Neutrality plot analysis

A neutrality plot was conducted using GC12 as the ordinate and GC3 as the abscissa to evaluate the degree of influences from mutation pressure and natural selection. GC12 represents the average GC content of GC1 and GC2. If there is a significant correlation between GC12 and GC3, it indicates that codon usage is mainly affected by mutation pressure. In contrast, if there is no correlation between GC12 and GC3, it means that natural selection is the main driving force in influencing codon usage bias (Sueoka, 1988).

2.5 ENC-plot analysis

ENC-plot analysis is a routine analysis to explore whether the codon preference of a specific gene is caused by mutation or selection pressure. The ENC-plot is used to draw a two-dimensional scatter plot with GC3 as the abscissa and ENC value as the ordinate. The expected curve was calculated based on the following formula: $ENC_{exp} = 2+GC3_S+29/[GC3_S^2+(1-GC3_S)^2]$. If the corresponding points are distant from the expected curve, this reveals that the gene will be influenced by natural selection. Mutational effects may be the

dominant factor if the points fall near the standard curve (Wright, 1990).

We also analyzed the ENC frequency ratio distribution to determine further the main factors affecting the CUB. The value of the ENC ratio was calculated according to the formula: $ENC_{ratio} = (ENC_{exp}-ENC_{obs})/ENC_{exp}$.

2.6 Parity Rule 2 plot analysis

Parity rule 2 (PR2) plot analysis is one of the methods to examine the effects of mutation pressure and natural selection on codon bias. PR2 plot, as a graphical analysis, is widely used to reveal the relationship between the variation of the four bases at the third codon position and the codon usage bias (Sueoka, 1999). In our study, the graphic were established with $A3_S/(A3_S + T3_S)$ and $G3_S/(G3_S + C3_S)$ as the x- and y-axis, respectively. In the PR2 plot, the center point means that the codon usage of genes is only affected by mutational pressure (Galtier and Lobry, 1997). The vector emitted from the center point represents the degree and direction of the codon usage bias (Li et al., 2023).

2.7 Statistical analysis

Correspondence analysis (COA) is a commonly multivariate statistical approach to analyze codon usage patterns (James and McCulloch, 1990). We performed a correlation analysis based on the parameters (GC1, GC2, GC3, GCall, ENC, and L_aa) obtained by the CodonW 1.4.4. SPSS 20.0 software was used to COA based on Spearman's rank correlation method.
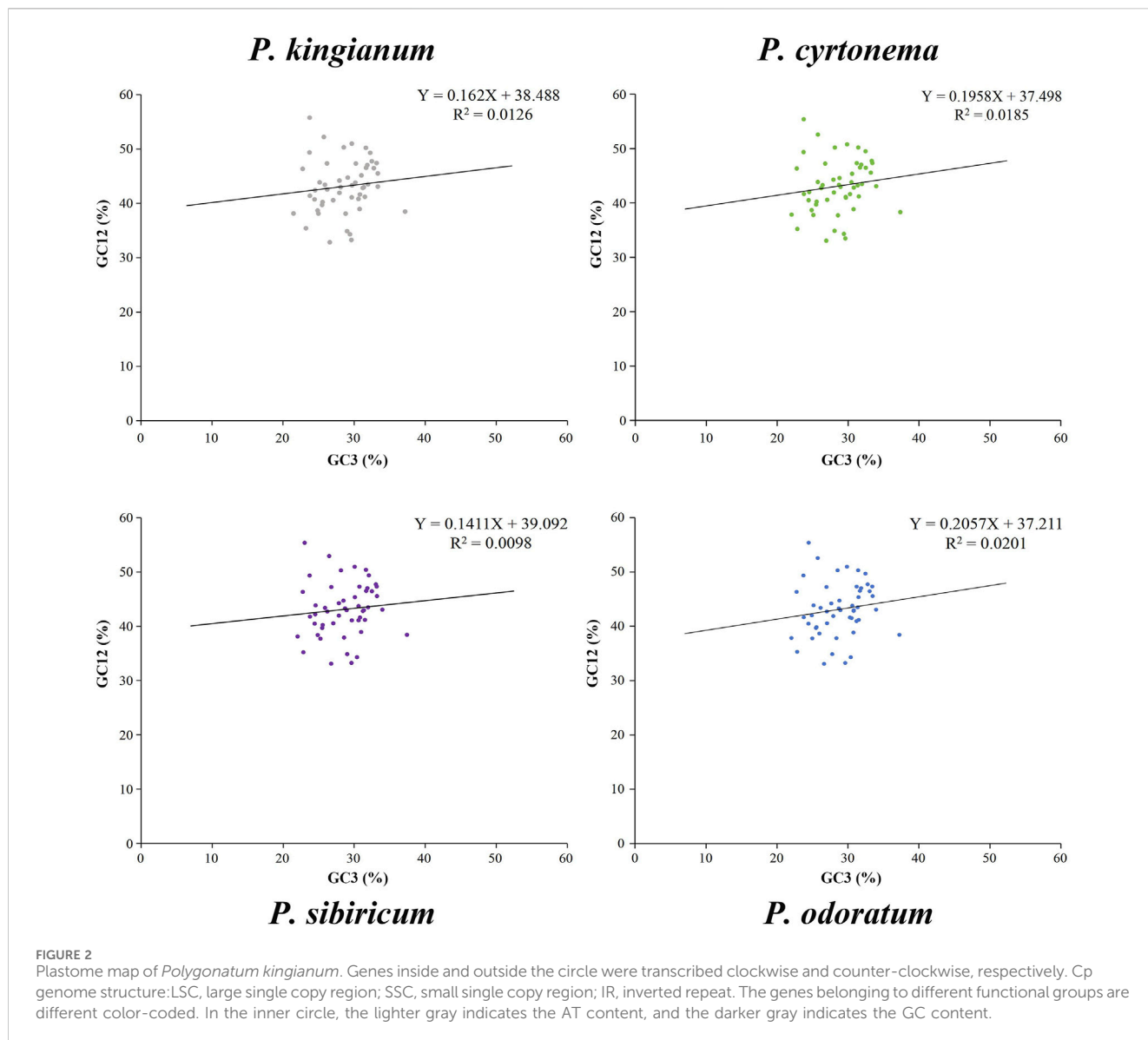
FIGURE 2
Plastome map of *Polygonatum kingianum*. Genes inside and outside the circle were transcribed clockwise and counter-clockwise, respectively. Cp genome structure:LSC, large single copy region; SSC, small single copy region; IR, inverted repeat. The genes belonging to different functional groups are different color-coded. In the inner circle, the lighter gray indicates the AT content, and the darker gray indicates the GC content.

TABLE 1 Detailed information of four chloroplast genomes of medicinal Polygonatum species.

Species	GenBank number	Total sequence	CDSs number (before fltering)	CDSs number (after fltering)	Total sequence (after fltering)	L_aa (after fltering)
P. kingianum	PP315902	1,55,714	85	51	62,736	20,912
P. cyrtonema	ON872700	1,55,521	85	51	62,703	20,901
P. sibiricum	OQ532972	1,55,513	86	51	62,703	20,901
P. odoratum	MZ150858	1,54,577	85	51	62,652	20,884

2.8 Determination of optimal codons

The codon with RSCU>1 is considered as the high-frequency codon (Yu et al., 2012). Furthermore, all tested genes were arranged from large to small according to the ENC values. The 10% of the genes were screened from the top and bottom to establish high-and low-expression gene groups, respectively (Gao et al., 2024). The RSCU values of the two datasets were then calculated. Codons with $\Delta RSCU \geq 0.08$ ($\Delta RSCU = RSCU_{high\text{-}expression}\text{-}RSCU_{low\text{-}expression}$) were defined as high-expression codons (Liu et al., 2020). Finally,

TABLE 2 List of screened genes in the chloroplast genomes of medicinal Polygonatum species.

Category of Genes	Group of gene	Name of gene
Self-replication	Ribosomal protein (small subunit)	*rps2, rps3, rps4, rps7c, rps8, rps11, rps14, rps18*
	Ribosomal protein (large subunit)	*rpl14, rpl16, rpl20, rpl22*
	RNA polymerase	*rpoA, rpoB, rpoC1, rpoC2*
Genes for photosynthesis	Subunits of photosystem I	*psaA, psaB*
	Subunits of photosystem II	*psbA, psbB, psbC, psbD*
	Subunits of cytochrome	*petA, petB, petD*
	Subunits of ATP synthase	*atpA, atpB, atpE, atpF[a], atpI*
	Large subunit of Rubisco	*rbcL*
	Subunits of NADH dehydrogenase	*ndhA, ndhB[a,c], ndhC, ndhD, ndhE, ndhF ndhG, ndhH, ndhI, ndhJ, ndhK*
Other genes	Maturase	*matK*
	Envelope membrane protein	*cemA*
	Subunit of acetyl-CoA	*accD*
	Synthesis gene	*ccsA*
	ATP-dependent protease	*clpP*
	Component of TIC complex	*ycf1[d]*
Genes of unknown function	Conserved open reading frames	*ycf2[c], ycf3[b], ycf4*

[a]Gene containing one intron.
[b]Gene containing two introns.
[c]Duplicated genes.
[d]Partial duplicated gene.

the codon meeting both high frequency and high expression was identified as the optimal codon (Liu and Xue, 2004) for the chloroplast genome of medicinal *Polygonatum* species.

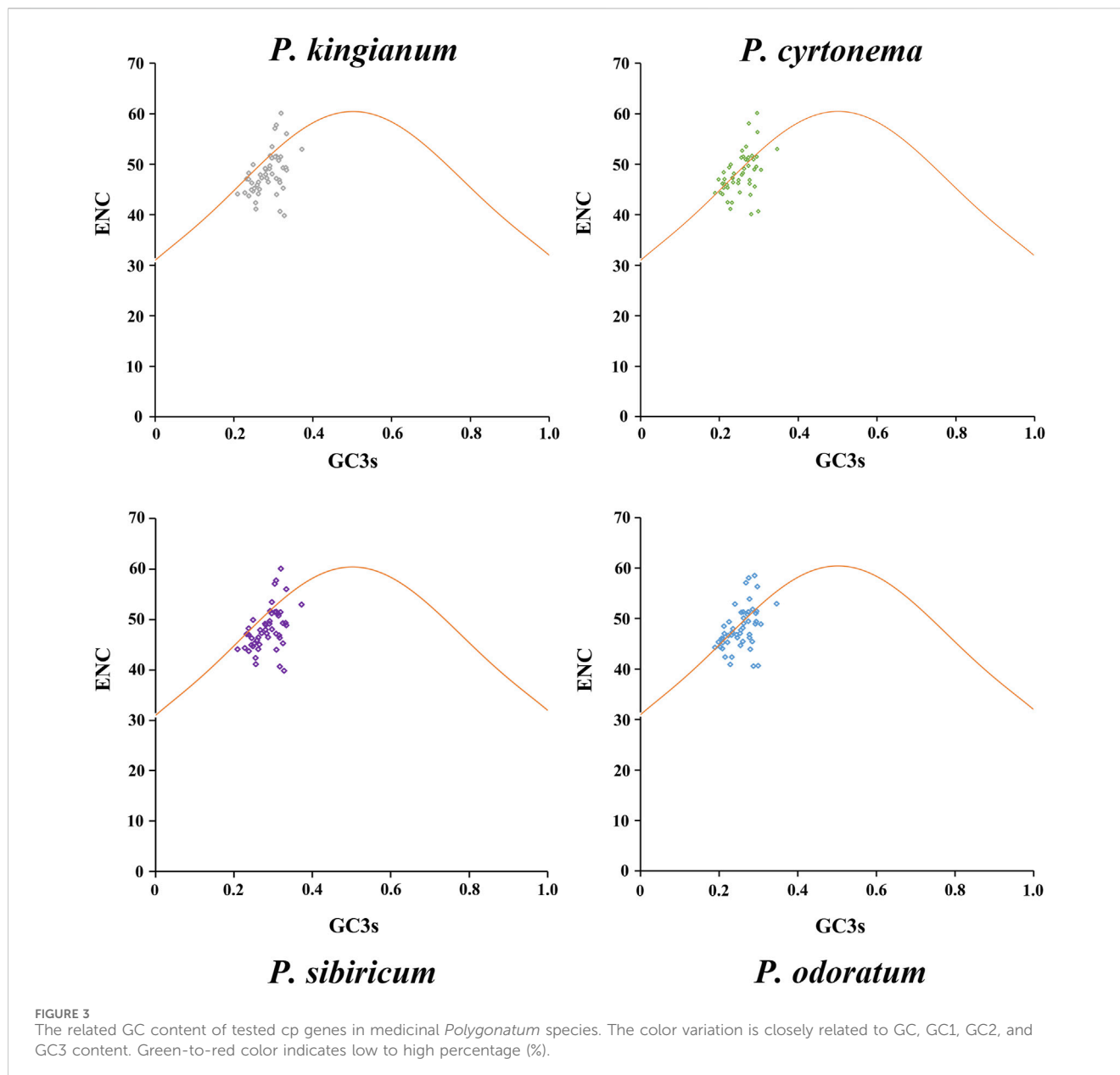## 2.9 RSCU-based cluster analysis

We explored the relationship of the four medicinal *Polygonatum* species using a hierarchical clustering method. The OriginPro 2019b (OriginLab Corporation, 2019) was employed to construct the RSCU value matrix, and calculate the average distance of these species.

3 Results

3.1 Nucleotide composition analysis

The 204 CDSs were screened from four medicinal *Polygonatum* species for subsequent analysis (Table 2). These tested CDSs mainly included photosynthesis genes, ribosome genes, self-replication genes, hypothetical chloroplast reading frames, and other genes. All codon usage indexes (ENC, RSCU, L_aa, GCall, GC1, GC2, GC3, GC3s, $A3_S$, $T3_S$, $G3_S$, $C3_S$) were calculated as shown in Supplementary Tables S3, S4. The third codon positions (GC1, GC2, and GC3) for each CDS in the four medicinal *Polygonatum* species exhibited a

similar trend of GC content, with GC1 > GC2 > GC3 (Figure 3). It is considered that when ENC≤35, the gene has significant codon bias and *vice versa* (Comeron and Aguade, 1998). The ENC values of the four medicinal *Polygonatum* species had a weak preference for codons (40.1–60.11 for *P. sibiricum*, 40.58–58.53 for *P.odoratum*, 40.1–60.11 for *P.cyrtonema*, and 39.83–60.11 for *P. kingianum*, respectively). The average ENC values for self-replication, photosynthesis, unknown function, and other genes ranged from 47.33 to 47.42, 47.42 to 47.66, 49.38 to 49.72, and 53.86 to 54.74. Among the above four categories of genes, *psbA* gene has the strongest codon usage preference (ENC values: 39.83–40.58), while *ycf3* gene has the weakest codon bias, (ENC values: 58.53.83–60.11), as shown in Supplementary Table S3. RSCU is calculated as the ratio of the expected frequency of amino acid synonymous codon usage to its observed frequency (Shields et al., 1988). We calculated the RSCU values of each amino acid in the four medicinal *Polygonatum* species (Supplementary Table S5; Figure 4). There were 30 codons with an RSCU greater than 1.00, which were defined as high-frequency codons. Among these high-frequency codons, 29 of these ended with A/T, including UUU, UUA, CUU, AUU, GUU, GUA, UCU, UCA, AGU, CCU, CCA, ACU, ACA, GCU, GCA, UAU, UAA, CAU, CAA, AAU, AAA, GAU, GAA, UGU, CGU, CGA, AGA, GGU, GGA, while only one ended with G (UUG). The codon UUA (Leu, RSCU = 1.92–1.93) had the highest RSCU value in these high-frequency codons.

FIGURE 3
The related GC content of tested cp genes in medicinal *Polygonatum* species. The color variation is closely related to GC, GC1, GC2, and GC3 content. Green-to-red color indicates low to high percentage (%).
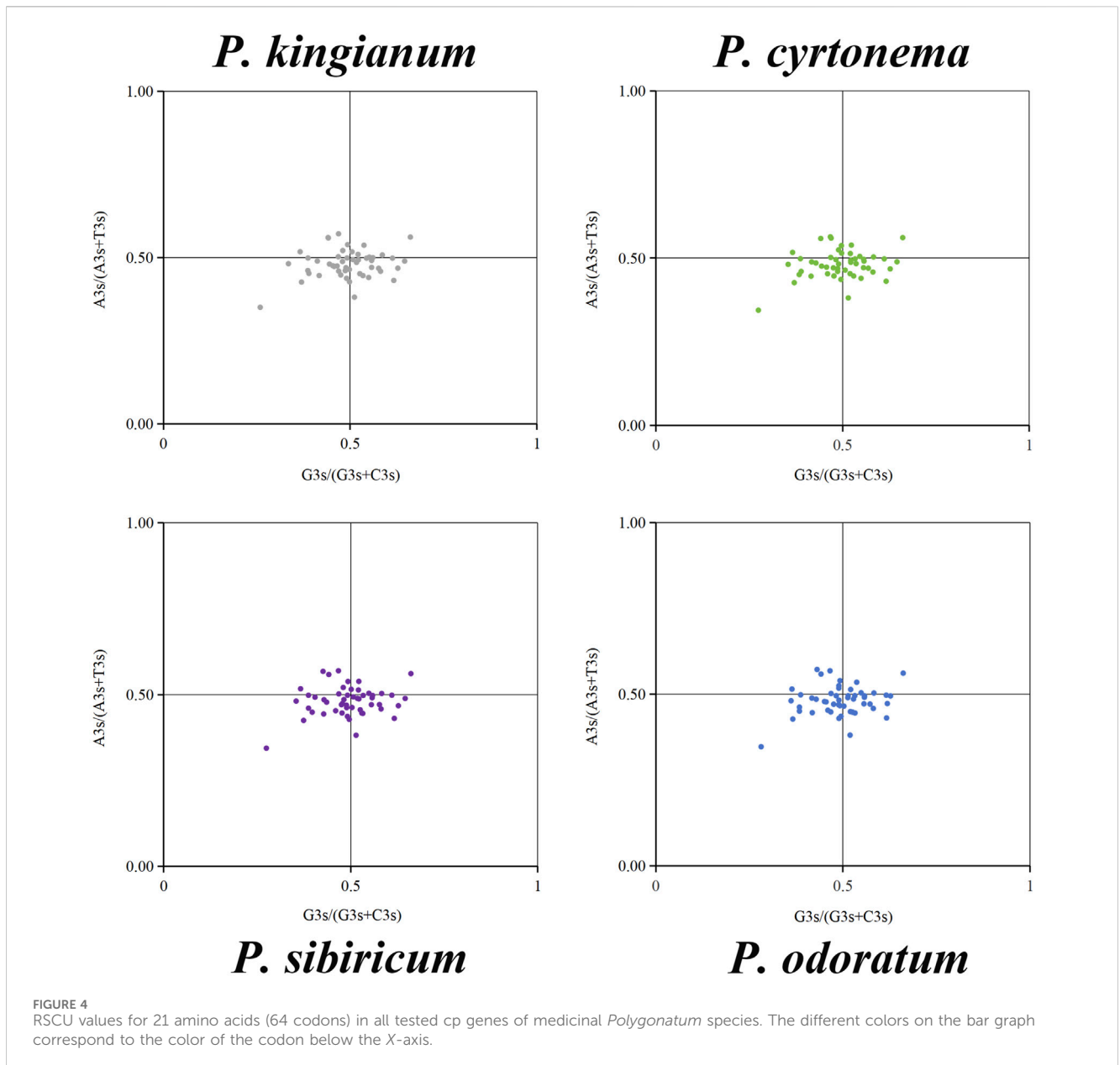
3.2 Neutrality plot analysis

As shown in Supplementary Table S4, the value ranges of GC3 were 22.03%–37.36% (*P. cyrtonema*), 21.47%–37.12% (*P. kingianum*), 22.03%–37.24% (*P. odoratum*), 22.03%–37.42% (*P. sibiricum*), and the GC12 content of *P. cyrtonema*, *P. kingianum*, *P. odoratum* and *P. sibiricum* varied from 33.06% to 55.40%, 32.88%–55.76%, 33.09%–55.40%, 33.12%–55.40%, respectively. We constructed a neutrality plots (GC3 vs. GC12) for the cp genes in four medicinal *Polygonatum* species (Figure 5). The correlation coefficient (R) between GC12 and GC3 of these medicinal *Polygonatum* species were 0.0185 (*P. cyrtonema*), 0.0126 (*P. kingianum*), 0.0201 (*P. odoratum*), 0.0098 (*P. sibiricum*), respectively. The regression coefficients ranged from 0.1411 to 0.2057, indicating that the contribution of mutation pressure effect accounted for 14.11%–20.57%. The results of

neutrality plot analysis suggest that the contribution of natural selection was more significant than that of mutation selection.

3.3 ENC-plot analysis

The results of ENC-plot analysis of the four medicinal *Polygonatum* species were shown in Figure 6. All tested species had similar patterns in the ENC-plots. We found that most points were under the standard curve, and only a few were near or above the standard curve. It indicates that the CUB of most genes was mainly affected by selection pressure, while a few genes were affected by mutation pressure.

The ENC ratio frequency was calculated to observe the variation range between the actual and expected values of ENC. Distribution of all sample's ENC ratios was shown in Figure 7 and showed almost

FIGURE 4
RSCU values for 21 amino acids (64 codons) in all tested cp genes of medicinal *Polygonatum* species. The different colors on the bar graph correspond to the color of the codon below the *X*-axis.

the same distribution. Majority of cp genes (30–31 genes) had ENC ratios between −0.05 and 0.05, meaning their CUB was primarily influenced by mutation pressure. In contrast, natural selection influenced the CUB of the remaining cp genes.

3.4 Parity rule 2 (PR2) plot analysis

The PR2 plot analysis mainly analyzes the bias between the third base A, T, G, and C in chloroplast genes. The PR2-plots of the medicinal *Polygonatum* species were shown in Figure 8. As can be seen from the PR2-plot, the points of these medicinal *Polygonatum* species showed a similar distribution. Besides, the distribution of dots in the four quadrants was unbalanced, and most of the dots were distributed in the third and fourth quadrants. These results suggested a few differences in the same genes among different

medicinal *Polygonum* species. And the third base position of the synonymous codon preferred to end in T/C. In summary, the formation of codon usage patterns was influenced by natural selection and mutation pressure.

3.5 Correspondence analysis (COA)

The correlation analysis of four medicinal *Polygonatum* species was shown in Figure 9. Correlation analysis results show that GCall was extremely significantly correlated with the GC1, GC2, and GC3 in all samples ($p < 0.01$). A extremely significant positive correlation was also observed between GC3 and ENC of all tested medicinal *Polygonatum* species ($p < 0.01$). In the three tested medicinal *Polygonatum* species (*P. odoratum, P. sibiricum, P. cyrtonema*), a extremely significant positive correlation was observed between GC1 and GC2 ($p < 0.01$). In *P.*
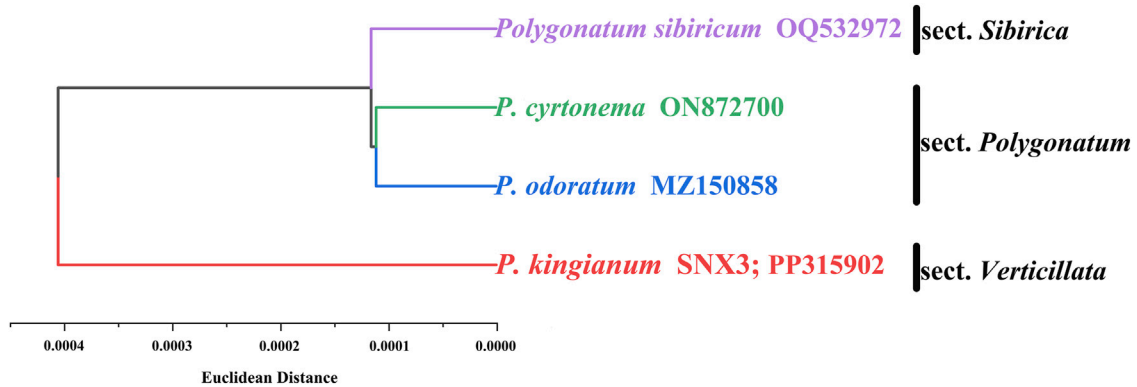
FIGURE 5
Neutrality plot analysis (GC12 vs. GC3) of cp genes in the medicinal *Polygonatum* species. R represents the Pearson's correlation coefficient.
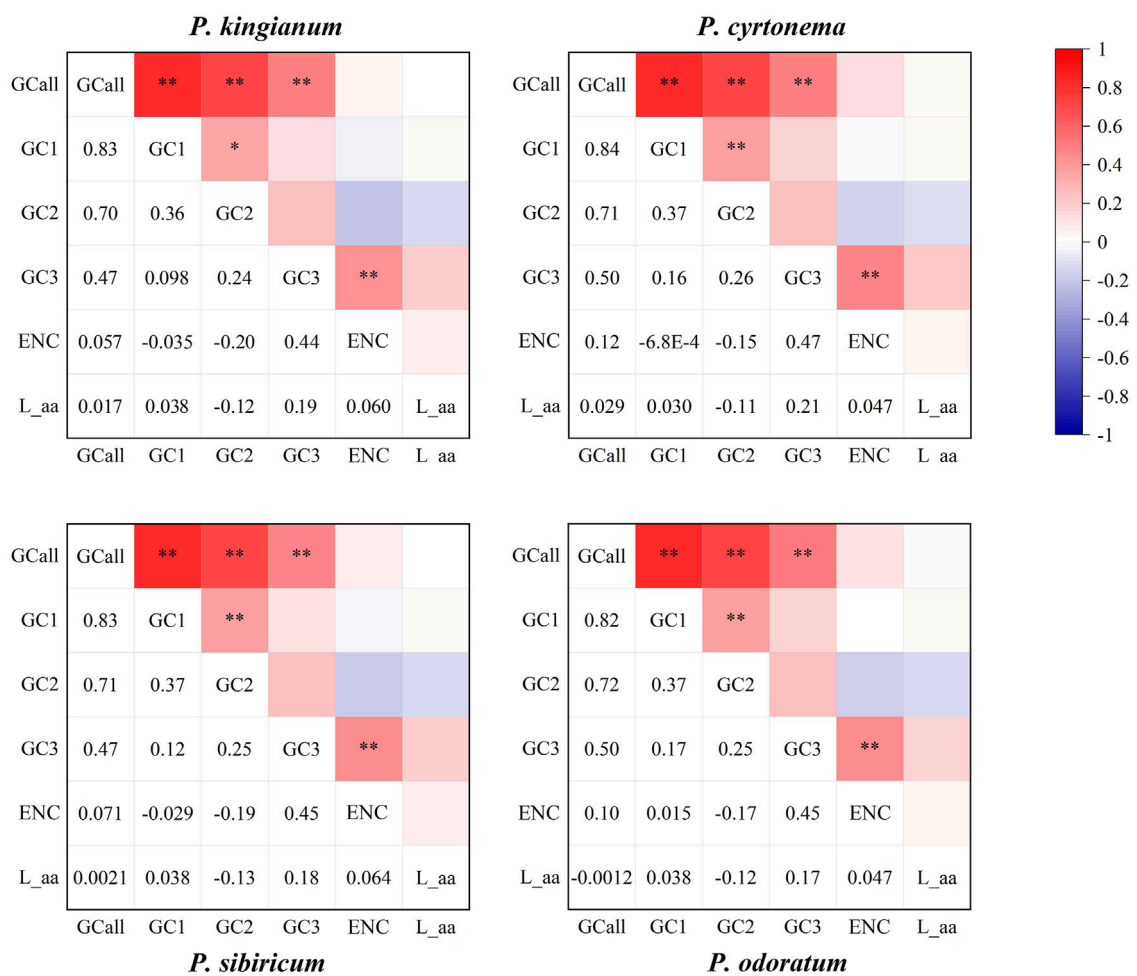

FIGURE 6
ENC-GC3s plot analysis of 51 tested cp genes in medicinal *Polygonatum* species. The yellow expected curve shows codon usage bias is only affected by mutation pressure.

kingianum, GC1 had a significant positive correlation with GC2 ($p < 0.05$). However, there was no correlation between GC3 and GC1 or GC2 in all medicinal *Polygonum* species. Furthermore, codon number (L_aa) was not significantly associated with all tested parameters. In all *Polygonatum* species tested, there was no significant correlation between ENC and GC1 or GC2.

FIGURE 7
Cp genes number distribution of ENC frequency ratio.
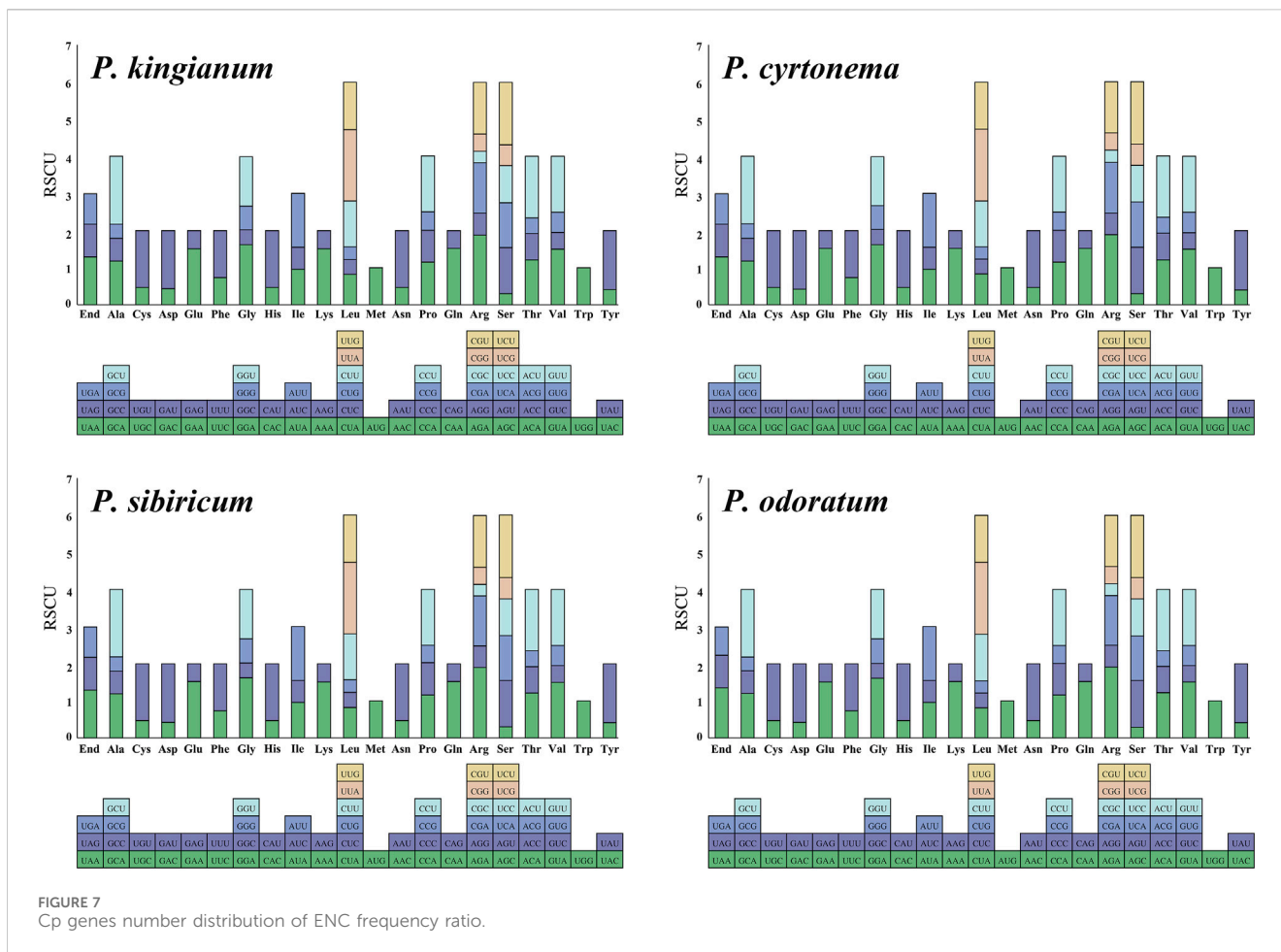
3.6 Optimal codons of medicinal *Polygonatum* species chloroplast genes

We calculated the RSCU difference (△RSCU) per codon for high-expression and low-expression groups. According to the calculation results of △RSCU (Supplementary Table S6), 30 high-expression codons were screened in the each medicinal *Polygonatum* species. We detected 10 to 11 optimal codons based on high-frequency codons (RSCU>1) and high-expression codons (△RSCU>0.08). The results showed that there were differences in the number and types of the optimum codons of the medicinal *Polygonatum* species. The optimal codons of *P. cyrtonema* were AAU, ACA, AGA, CAU, CCA, CUU, GAU, GGA, UAU, UCA. The optimal codons of other medicinal *Polygonatum* species (*P. sibiricum*, *P. odoratum*, and *P. kingianum*) were AAU, ACA, CAA, CAU, CCA, CUU, GAU, GGA, UAU, UCA, UUU. Most of the optimal codons of medicinal *Polygonatum* species ended in A/U.
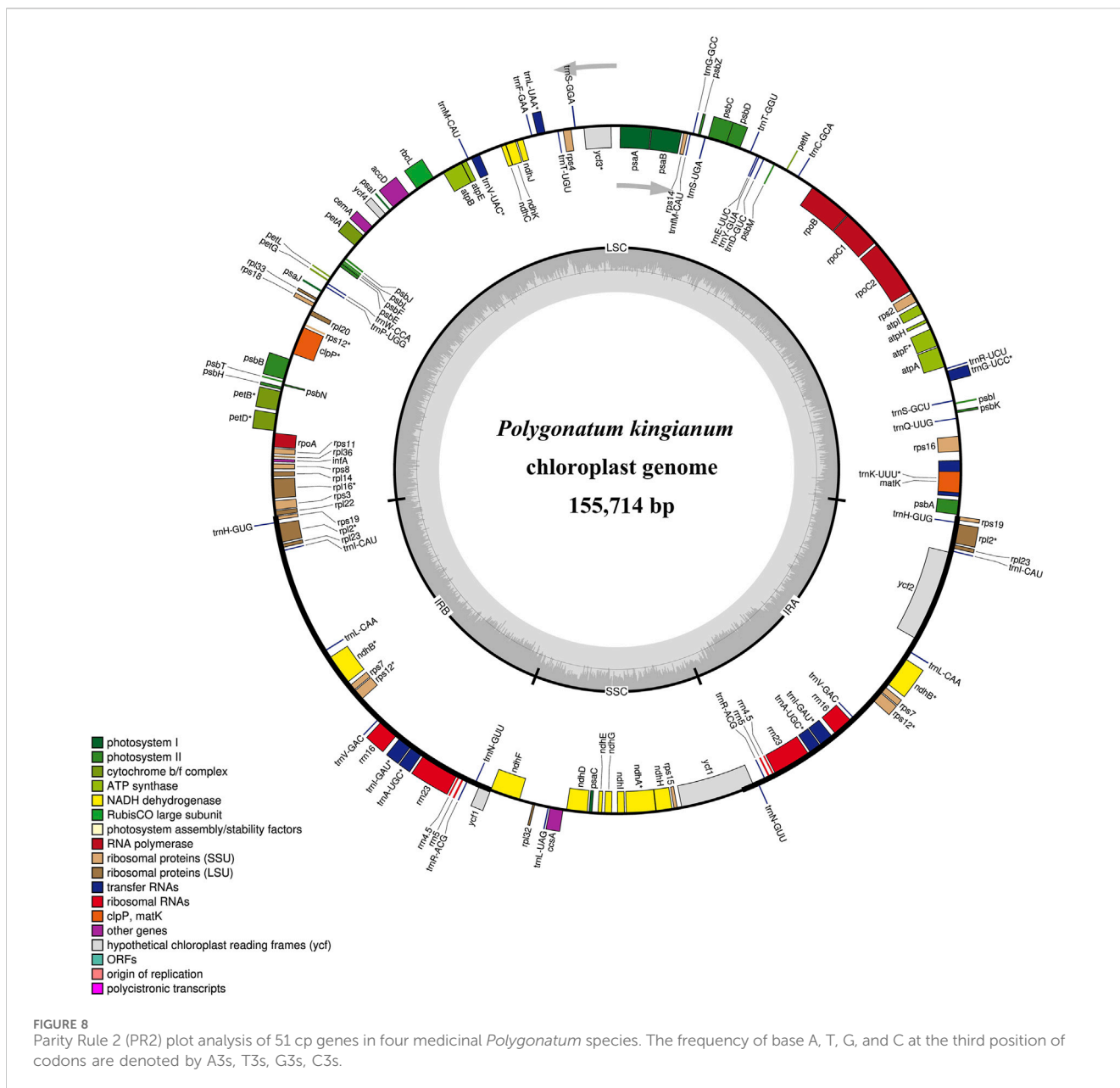
3.7 RSCU-based cluster analysis

The codon-based hierarchical cluster tree is shown in Figure 10. *P. kingianum* (sect. *Verticillate*) was located at the outermost position of the tree, followed by the *P. sibiricum* (sect.

Verticillate). Finally, the next node included the *P. cyrtonema* and *P. odoratum* from the sect. *Polygonatum*. Among the four medicinal species of the genus *Polygonatum*, the relationship between *P. cyrtonema* and *P. odoratum* was closer than *P. sibiricum*; the relationship with *P. kingianum* was weaker for *P. sibiricum*, *P. cyrtonema* and *P. odoratum*. The species relationship inferred based on RSCU is consistent with other studies built based on the sequences (Xia et al., 2022; Wang J et al., 2022).

4 Discussion

Genetic codon is the link between nucleic acid and protein, and plays an important role in the transmission of genetic information in organisms (Tang et al., 2021). Many studies have confirmed that codon bias varies in different species and genes which is considered to be an evolutionary behavior caused by adapting to their environment (Chaney et al., 2016; Zhang et al., 2019; Chakraborty et al., 2020; Shi et al., 2022). Previous studies have found that many specific factors affect codon preference, including gene length, gene expression level, tRNA abundance, mutation, selection, and so on, among which mutation and selection pressure are the main influencing factors (Wan et al., 2004; Meyer, 2021). Research on the codon bias characteristics and variations is of great significance for the understanding of

FIGURE 8
Parity Rule 2 (PR2) plot analysis of 51 cp genes in four medicinal *Polygonatum* species. The frequency of base A, T, G, and C at the third position of codons are denoted by A3s, T3s, G3s, C3s.

molecular evolution and the exogenous expression of species (Wang Z et al., 2022; Chen et al., 2023).

In the present study, we systematically investigated the codon usage patterns of CDS in the medicinal *Polygonatum* chloroplast genomes, as well as the shaping factors influencing the codon usage bias patterns. Codon usage index analysis showed that the base composition shows different characteristics at the three codon positions. We found that the GCall and GC3 content of the tested chloroplast genomes was less than the AT content, which means these medicinal *Polygonatum* species exhibited a preference for the use of A or T nucleotides. This finding was consistent with a previous study that higher plants have a tendency to use codons that end in A or T (Campbell and Gowri, 1990). RSCU is an essential parameter in studying the codon usage bias of species (Sharp and Li, 1986). According to RSCU analysis, most high-frequency and highly-expressed codons also end with A/U, which further demonstrates the A/U preference trend of the third codon of

medicinal *Polygonatum* species. ENC values analysis of chloroplast genomes revealed weak codon bias in these species. According to the ENC value, the codon preference of genes responsible for photosynthesis and self-replication is stronger than that of others. The reason for the difference of CUB among four kinds of genes is probably related to the adaptive adjustment of species to the environment. We also detected the correlation between GC content (GC1, GC2, GC3, GCall), ENC, and codon number (L_aa) from four medicinal *Polygonatum* spices, indicating the base composition influenced codon usage bias. Combined with analysis of the neutrality plot, ENC-plot, and PR2-plot, it revealed that both mutation pressure and natural selection had affected CUB in the chloroplast genomes of the four medicinal *Polygonatum* species, of which natural selection made great contributions in framing the CUB. Correlation analysis showed that GC3 has no correlation with GC1 or GC2, implying that the base composition in the third codon position
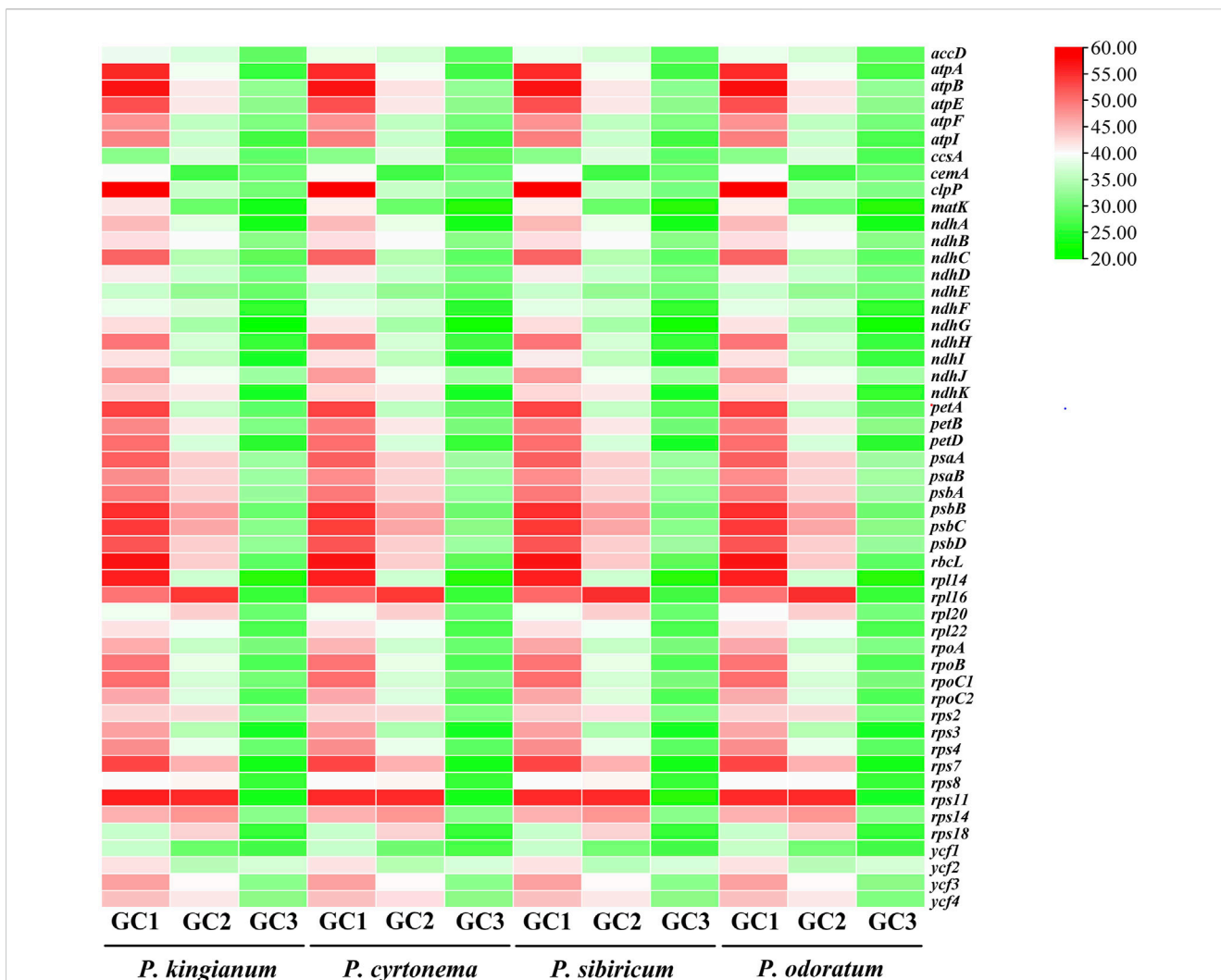
FIGURE 9
Spearman's correlation analysis heatmap of different codon usage indicators of four medicinal *Polygonatum* species. The color of the color block changes from blue to red, indicating that the correlation index is rising. An asterisk indicates a significant association (*p* < 0.05); two asterisks indicate a extremely significant correlation (*p* < 0.01).

significantly differed from those of the former two. Moreover, no correlation was observed between L_aa and ENC, indicating that gene sequence length has no effect on CUB. According to all the resulting data we obtained, the codon usage bias of chloroplast genomes was similar among the four *Polygonatum* species, indicating that natural selection and mutation pressure have similar effects between closely related species. This may be related to the conservation of cp genome evolution.

Codon optimization, which adjusts synonymous codons of foreign genes according to the codon usage pattern of the cp genome, to improve the efficiency and accuracy of related gene expression products (Tang et al., 2021). The 10-11 optimal codons were identified in medicinal *Polygonatum* spices chloroplast genomes. However, the optimal codon of *P. cyrtonema* was somewhat different from that of the remaining three species. The reason for the variations is due to the inconsistency between the defined high and low expression gene groups. The relationship between different medicinal *Polygonatum* species was constructed based on the RSCU values and found that the three species *P. sibiricum*,

P. cyrtonema, and *P. odoratum* were closely related to each other, but distantly related to *P. kingianum*. We believe that the reason for this relationship may be due to the high geographical overlap among *P. sibiricum* (Northwestern, Northeastern, North, and South China), *P. cyrtonema* (Southwest, Central, East, North, and South China), and *P. odoratum* (Northwestern, Central, East, and North China), which is higher than that of *P. kingianum* (Southwest and South China) (Chen and Tamura, 2000). Therefore, it can provide a reference for inferring their genetic relationship according to the difference in RSCU values among different medicinal *Polygonatum* species. All the above results can provide a theoretical basis for subsequent exogenous gene improvement and genetic evolution in the chloroplast genome of medicinal *Polygonatum* species.

Codon bias is only one of many factors that affect gene expression. Furthermore, CUB itself is influenced by a variety of factors. Hence, although CUB has important applications in enhancing heterologous gene expression, it has not been well studied in most medicinal plants. To date, the RSCU-based cluster analysis has been applied to explore the relationship of
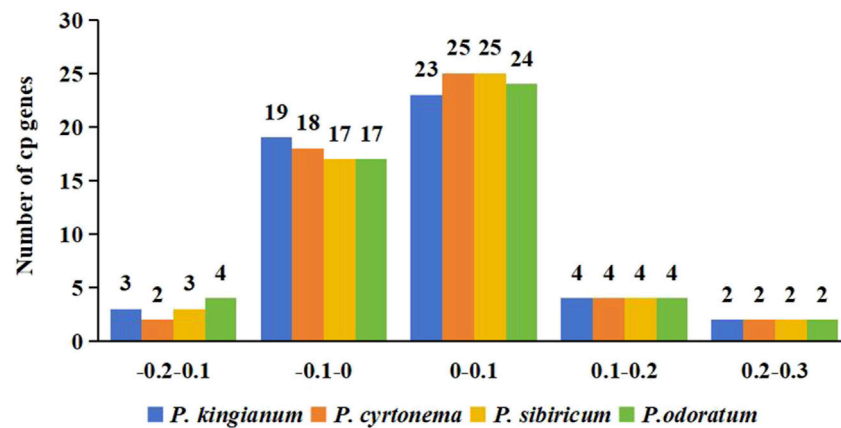
FIGURE 10
Hierarchical clustering analysis based on the RSCU of codons in chloroplast CDSs.

plants. However, some studies have shown that RSCU clustering results were inconsistent with CDS phylogenetic tree results (Niu et al., 2021; Zhang et al., 2024). Therefore, more *Polygonum* species are needed to examine its reliability in future studies.

5 Conclusion

This was the first report to systematically compare codon usage characteristics and patterns in the chloroplast genomes of medicinal *Polygonatum*. In the present study, 204 PCGs from chloroplast genomes of four *Polygonatum* species were screened and calculated to analyze the CUB. The results of the present study exhibited similar weak codon usage preferences prevalent in chloroplast genomes of these *Polygonatum* species. We discussed the formation of codon usage patterns of medicinal *Polygonatum* species and confirmed that natural selection was the determining factor affecting codon preference. Besides, the mutation factor could not be ignored. Based on the analysis results of CUB, we identified 10-11 optimal codons in these chloroplast genomes of four *Polygonatum* species, and most of them ending with A or U. Our finding indicated that the phylogenetic results based on RSCU cluster analysis not only can be used for understanding the evolutionary relationships of different medicinal *Polygonatum* species but also can be an important supplement to the phylogenetic results based on sequences. In summary, this study provides an important theoretical reference about the codon usage tendency and variation, codon optimization of exogenous genes, as well as evolutionary analysis in these important medicinal *Polygonatum* species.

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 in the article/supplementary material.

Author contributions

NS: Writing–original draft. YY: Writing–original draft. RH: Writing–original draft. GW: Writing–review and 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/fgene.2024.1401013/full#supplementary-material

References

Athey, J., Alexaki, A., Osipova, E., Rostovtsev, A., Santana-Quintero, L. V., Katneni, U., et al. (2017). A new and updated resource for codon usage tables. *BMC Bioinform* 18, 391. doi:10.1186/s12859-017-1793-7

Bai, J. B., Ge, J. C., Zhang, W. J., Liu, W., Luo, J. P., Xu, F. Q., et al. (2021). Physicochemical, morpho-structural, and biological characterization of polysaccharides from three *Polygonatum* spp. *RSC Adv.* 11, 37952–37965. doi:10.1039/d1ra07214e

Basso, M. F., Arraes, F. B. M., Grossi-de-Sa, M., Moreira, V. J. V., Alves-Ferreira, M., and Grossi-de-Sa, M. F. (2020). Insights into genetic and molecular elements for transgenic crop development. *Front. Plant Sci.* 11, 509. doi:10.3389/fpls.2020.00509

Campbell, W. H., and Gowri, G. (1990). Codon usage in higher plants, green Algae, and Cyanobacteria. *Plant Physiol.* 92, 1–11. doi:10.1104/pp.92.1.1

Cao, D. L., Zhang, X. J., Xie, S. Q., Fan, S. J., and Qu, X. J. (2022). Application of chloroplast genome in the identification of Traditional Chinese Medicine *Viola philippica*. *BMC Genom* 23, 540. doi:10.1186/s12864-022-08727-x

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

Chaney, L., Mangelson, R., Ramaraj, T., Jellen, E. N., and Maughan, P. J. (2016). The complete chloroplast genome sequences for four Amaranthus Species (Amaranthaceae). *Appl. Plant Sci.* 4, 1600063. doi:10.3732/apps.1600063

Chen, D. H., Han, Z. G., and Si, J. P. (2021). Huangjing (Polygonati rhizoma) is an emerging crop with great potential to fight chronic and hidden hunger. *Sci. China Life Sci.* 64, 1564–1566. doi:10.1007/s11427-021-1958-2

Chen, J., Ma, W., Hu, X., and Zhou, K. (2023). Synonymous codon usage bias in the chloroplast genomes of 13 Oil-Tea *Camellia* samples from south China. *Forests* 14, 794. doi:10.3390/f14040794

Chen, S. C., and Tamura, M. N. (2000). "Missouri botanical garden press," in *Polygonatum*. Editors Z. Y. Wu and P. H. Raven (Beijing: Science Press), 223–232.

Chinese Pharmacopoeia Commission (2020). *Pharmacopoeia of the people's Republic of China*. Beijing: China medicine science and technology press.

Comeron, J. M., and Aguade, M. (1998). An evaluation of measures of synonymous codon usage bias. *J. Mol. Evol.* 47, 268–274. doi:10.1007/pl00006384

Cui, X., Wang, S., Cao, H., Guo, H., Li, Y., Xu, F., et al. (2018). Review: the bioactivities and pharmacological applications of *Polygonatum sibiricum* polysaccharides. *Molecules* 23, 1170. doi:10.3390/molecules23051170

Dai, G., Shang, M., Wang, J., Zheng, J., Liao, B., Liu, Y., et al. (2024). Codon bias and phylogenesis analysis of chloroplast genome in medicinal plants of *Lilium*. *Chin. Tradit. Herb. Drugs* 55 (11), 3835–3844. doi:10.7501/j.issn.0253-2670.2024.11.025

Daniell, H., Lin, C. S., Yu, M., and Chang, W. J. (2016). Chloroplast genomes: diversity, evolution, and applications in genetic engineering. *Genome Biol.* 17, 134. doi:10.1186/s13059-016-1004-2

Floden, A., and Schilling, E. E. (2018). Using phylogenomics to reconstruct phylogenetic relationships within tribe Polygonateae (Asparagaceae), with a special focus on *Polygonatum*. *Mol. Phylogenet. Evol.* 129, 202–213. doi:10.1016/j.ympev.2018.08.017

Galtier, N., and Lobry, J. (1997). Relationships between genomic G+C content, RNA secondary structures, and optimal growth temperature in prokaryotes. *J. Mol. Evol.* 44, 632–636. doi:10.1007/pl00006186

Gao, W., Chen, X., He, J., Sha, A., Luo, Y., Xiao, W., et al. (2024). Intraspecific and interspecific variations in the synonymous codon usage in mitochondrial genomes of 8 *pleurotus* strains. *BMC Genomics* 25, 456. doi:10.1186/s12864-024-10374-3

Guo, X. R., Shi, N. X., Xie, P. X., Zhang, G. F., Liu, H. Y., and Ji, Y. H. (2022). Plastome sequencing for accurate and effective authentication of *Polygonatum kingianum* (Asparagaceae). *Ind. Crops Prod.* 184, 115056. doi:10.1016/j.indcrop.2022.115056

James, F. C., and McCulloch, C. E. (1990). Multivariate analysis in ecology and systematics: panacea or Pandora's box? *Annu. Rev. Ecol. Syst.* 21, 129–166. doi:10.1146/annurev.es.21.110190.001021

James, G. J., and Daniell, H. (2005). Chloroplast genetic engineering: recent advances and future perspectives. *CRC Crit. Rev. Plant Sci.* 24, 83–107. doi:10.1080/07352680590935387

Jin, J. J., Yu, W. B., Yang, J. B., Song, Y., dePamphilis, C. W., Yi, T. S., et al. (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

Kumar, A. U., and Ling, A. P. K. (2021). Gene introduction approaches in chloroplast transformation and its applications. *J. Genet. Eng. Biotechnol.* 19, 148. doi:10.1186/s43141-021-00255-7

Kwon, K. C., Chan, H. T., León, I. R., Williams-Carrier, R., Barkan, A., and Daniell, H. (2016). Codon optimization to enhance expression yields insights into chloroplast translation. *Plant Physiol.* 172, 62–77. doi:10.1104/pp.16.00981

Li, N., Sun, M. H., Jiang, Z. S., Shu, H. R., and Zhang, S. Z. (2016). Genome-wide analysis of the synonymous codon usage patterns in apple. *J. Integr. Agric.* 15, 983–991. doi:10.1016/s2095-3119(16)61333-3

Li, Q., Luo, Y., Sha, A., Xiao, W., Xiong, Z., Chen, X., et al. (2023). Analysis of synonymous codon usage patterns in mitochondrial genomes of nine *Amanita* species. *Front. Microbiol.* 14, 1134228. doi:10.3389/fmicb.2023.1134228

Liu, H. B., Lu, Y. Z., Lan, B. L., and Xu, J. C. (2020). Codon usage by chloroplast gene is bias in *Hemiptelea davidii*. *J. Genet.* 99, 8. doi:10.1007/s12041-019-1167-1

Liu, Q. P., and Xue, Q. Z. (2004). Codon usage in the chloroplast genome of rice (*Oryza sativa* L. ssp. japonica). *Acta Agron. Sin.* 30, 1220–1224.

Lohse, M., Drechsel, O., and Bock, R. (2007). OrganellarGenomeDRAW (OGDRAW): a tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Curr. Genet.* 52, 267–274. doi:10.1007/s00294-007-0161-y

Meyer, M. M. (2021). Revisiting the relationships between genomic G+C content, RNA secondary structures, and optimal growth temperature. *J. Mol. Evol.* 89, 165–171. doi:10.1007/s00239-020-09974-w

Niu, Y., Luo, Y., Wang, C., and Liao, W. (2021). Deciphering codon usage patterns in genome of cucumis sativus in comparison with nine species of cucurbitaceae. *Agronomy* 11 (11), 2289. doi:10.3390/agronomy11112289

OriginLab Corporation (2019). OriginPro 2019b. OriginLab. Available at: https://www.originlab.com.

Parvathy, S. T., Udayasuriyan, V., and Bhadana, V. (2021). Codon usage bias. *Mol. Biol. Rep.* 49, 539–565. doi:10.1007/s11033-021-06749-4

Patel, R. K., and Jain, M. (2012). NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. *PLoS One* 7, e30619. doi:10.1371/journal.pone.0030619

Peden, J. F. (1999). *Analysis of codon usage*. UK: University of Nottingham. PhD thesis.

Porebski, S., Bailey, L. G., and Baum, B. R. (1997). Modification of a CTAB DNA extraction protocol for plants containing high polysaccharide and polyphenol components. *Plant Mol. Biol. Rep.* 15, 8–15. doi:10.1007/BF02772108

Rice, P., Longden, I., and Bleasby, A. (2000). EMBOSS: the European molecular biology open software suite. *Trends Genet.* 16 (6), 276–277. doi:10.1016/S0168-9525(00)02024-2

Salim, H. M., and Cavalcanti, A. R. (2008). Factors influencing codon usage bias in genomes. *J. Braz. Chem. Soc.* 19, 257–262. doi:10.1590/S0103-50532008000200008

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

Shi, L. C., Chen, H. M., Jiang, M., Wang, L. Q., Wu, X., Huang, L. F., et al. (2019). CPGAVAS2, an integrated plastome sequence annotator and analyzer. *Nucleic Acids Res.* 47, W65-W73–W73. doi:10.1093/nar/gkz345

Shi, S. L., Liu, Y. Q., Xia, R. X., and Qin, L. (2022). Comprehensive analysis of codon usage in *Quercus* chloroplast genome and focus on *psbA* gene. *Genes.* 13, 2156. doi:10.3390/genes13112156

Shields, D. C., Sharp, P. M., Higgins, D. G., and Wright, F. (1988). "Silent" sites in Drosophila genes are not neutral: evidence of selection among synonymous codons. *Mol. Biol. Evol.* 5, 704–716. doi:10.1093/oxfordjournals.molbev.a040525

Singhal, R., Pal, R., and Dutta, S. (2023). Chloroplast engineering: fundamental insights and its application in amelioration of environmental stress. *Appl. Biochem. Biotechnol.* 195, 2463–2482. doi:10.1007/s12010-022-03930-8

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

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

Sugiura, M. (1995). The chloroplast genome. *Essays Biochem.* 30, 49–57.

Tang, D., Wei, F., Cai, Z., Wei, Y., Khan, A., Miao, J., et al. (2021). Analysis of codon usage bias and evolution in the chloroplast genome of *Mesona chinensis* Benth. *Dev. Genes. Evol.* 231, 1–9. doi:10.1007/s00427-020-00670-9

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.* 45, W6-W11–W11. doi:10.1093/nar/gkx391

Tuller, T. (2011). Codon bias, tRNA pools and horizontal gene transfer. *Mob. Genet. Elem.* 1 (1), 75–77. doi:10.4161/mge.1.1.15400

Wan, X., Xu, D., Kleinhofs, A., and Zhou, J. (2004). Quantitative relationship between synonymous codon usage bias and GC composition across unicellular genomes. *BMC Ecol. Evol.* 4, 19. doi:10.1186/1471-2148-4-19

Wang, B., Yuan, J., Liu, J., Jin, L., and Chen, J. Q. (2011). Codon usage bias and determining forces in green plant mitochondrial genomes. *J. Integr. Plant Biol.* 53 (4), 324–334. doi:10.1111/j.1744-7909.2011.01033.x

Wang, H. H., Yin, W. B., and Hu, Z. M. (2009). Advances in chloroplast engineering. *J. Genet. Genomics* 36, 387–398. doi:10.1016/S1673-8527(08)60128-9

Wang, J., Qian, J., Jiang, Y., Chen, X. C., Zheng, B. J., Chen, S. L., et al. (2022). Comparative analysis of chloroplast genome and new insights into phylogenetic relationships of *Polygonatum* and tribe Polygonateae. *Front. Plant Sci.* 13, 882189. doi:10.3389/fpls.2022.882189

Wang, S. Y., Zhou, N., Shi, N. X., Zhang, G. F., Liu, H. Y., Guo, X. R., et al. (2023). Testing and using complete plastomes for authentication of medicinal Polygonatum species (Asparagaceae). *Ind. Crops Prod.* 197, 116557. doi:10.1016/j.indcrop.2023.116557

Wang, X. S., Yang, F., Luo, Q., Liu, Y. L., and Zhu, B. (2022). Analysis of codon usage bias and phylogenetic in the chloroplast genome of *Dendrobium*. *Mol. Plant Breed.* 20 (23), 7759–7769. doi:10.13271/j.mpb.020.007759

Wang, Z., Cai, Q., Wang, Y., Li, M., Wang, C., Wang, Z., et al. (2022). Comparative analysis of codon bias in the chloroplast genomes of Theaceae species. *Front. Genet.* 13, 824610. doi:10.3389/fgene.2022.824610

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

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

Xia, M., Liu, Y., Liu, J., Chen, D., Shi, Y., Chen, Z., et al. (2022). Out of the himalaya-hengduan mountains: phylogenomics, biogeography and diversification of *Polygonatum* Mill. (Asparagaceae) in the northern Hemisphere. *Mol. Phylogenet. Evol.* 169, 107431. doi:10.1016/j.ympev.2022.107431

Xiao, L. L., Run, H. M., Fan, Z., Zhi, J. N., Kiran, T., Wang, S. Y., et al. (2023). Evolutionary research trend of *Polygonatum* species: a comprehensive account of their transformation from traditional medicines to functional foods. *Crit. Rev. Food Sci. Nutr.* 63, 3803–3820. doi:10.1080/10408398.2021.1993783

Yang, M., Liu, J., Yang, W., Li, Z., Hai, Y., Duan, B., et al. (2023). Analysis of codon usage patterns in 48 *Aconitum* species. *BMC Genom* 24, 703. doi:10.1186/s12864-023-09650-5

Yu, T., Li, J., Yang, Y., Qi, L., Chen, B., Zhao, F., et al. (2012). Codon usage patterns and adaptive evolution of marine unicellular cyanobacteria Synechococcus and Prochlorococcus. *Mol. Phylogenet. Evol.* 62, 206–213. doi:10.1016/j.ympev.2011.09.013

Zhang, K., Wang, Y., Zhang, Y., and Shan, X. (2024). Codon usage characterization and phylogenetic analysis of the mitochondrial genome in *Hemerocallis citrina*. *BMC Genom Data* 25, 6. doi:10.1186/s12863-024-01191-4

Zhang, W., Zhao, Y., Yang, G., Peng, J., Chen, S., and Xu, Z. (2019). Determination of the evolutionary pressure on *Camellia oleifera* on Hainan island using the complete chloroplast genome sequence. *Peer J.* 7, e7210. doi:10.7717/peerj.7210

Zhao, P., Zhao, C., Li, X., Gao, Q., Huang, L., Xiao, P., et al. (2018). The genus *Polygonatum*: a review on ethnopharmacology, phytochemistry and pharmacology. *J. Ethnopharmacol.* 214, 274–291. doi:10.1016/j.jep.2017.12.006