
Analysis of Chloroplast Genome Characteristics and Codon Usage Bias of *Styphnolobium japonicum f. oligophyllum*

[Zhi-Qiang Mu](#)¹, [Xiao Zhang](#)¹, Jing-Jing Yan¹, Bao-Ping Chen¹, [Hong-Wei Wang](#)² *

Posted Date: 30 April 2026

doi: 10.20944/preprints202604.2102.v1

Keywords: *Styphnolobium japonicum f. oligophyllum*; chloroplast genome; codon usage bias; optimal codons



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC, OpenAlex.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Analysis of Chloroplast Genome Characteristics and Codon Usage Bias of *Styphnolobium japonicum* f. *oligophyllum*

Zhi-Qiang Mu ¹, Xiao Zhang ¹, Jing-Jing Yan ¹, Bao-Ping Chen ¹ and Hong-Wei Wang ^{2,*}

¹ Kashi University, Kashi 844000, China

² Henan Agricultural University, Zhengzhou 450000, China

* Correspondence: whwcas@163.com

Abstract

To investigate the codon usage bias (CUB) and its influencing factors in the chloroplast genome of *Styphnolobium japonicum* f. *oligophyllum*, we sequenced, assembled and annotated the genome using Illumina high-throughput sequencing, and systematically analyzed 52 protein-coding sequences. The chloroplast genome is 158,837 bp with a typical quadripartite structure, containing 129 functional genes. It presents a mean GC₃ content of 28.26% and a mean ENC value of 45.40, indicating weak CUB and low gene expression. Among 31 preferred codons (RSCU > 1), 29 (93.5%) end with A/U. Neutrality plot, ENC-plot and PR2-plot analyses reveal that natural selection is the primary regulator of CUB. A total of 19 optimal codons were identified. These results provide fundamental data for the genetic engineering of *S. japonicum* f. *oligophyllum*.

Keywords: *Styphnolobium japonicum* f. *oligophyllum*; chloroplast genome; codon usage bias; optimal codons

1. Introduction

Chloroplasts are essential organelles within plant cells, serving as the primary sites for photosynthesis. They possess their own genome, which encodes dozens of proteins that play critical roles in various physiological and biochemical processes of chloroplasts [1,2]. The chloroplast genome exhibits several advantageous features, including structural stability, high gene expression efficiency, a relatively small molecular size, multiple copies per cell, and amenability to efficient genetic transformation. These characteristics have facilitated its widespread application in plant identification, molecular evolution studies, genetic engineering, and phylogenetic analyses [3,4].

Among these applications, chloroplast genetic engineering—leveraging the unique properties of the plastid genome, such as maternal inheritance, absence of gene silencing, and the capacity for multi-gene co-expression—has emerged as a powerful platform for crop improvement and biopharmaceutical production [5–7]. In agriculture, the introduction of genes involved in photosynthetic efficiency optimization, osmotic protection, and redox regulation has significantly enhanced CO₂ assimilation capacity and tolerance to abiotic stresses such as drought, high temperature, and salinity [4]. For instance, the accumulation of glycine betaine in tobacco effectively protects the photosynthetic apparatus from drought-induced damage [8]. Moreover, chloroplast engineering has successfully conferred resistance to insects, pathogens, and herbicides in various crops [4]. To date, plastid transformation systems have been established in tobacco, soybean, carrot, cotton, potato, and lettuce [9–11]. Notably, a “marker-free” one-step transformation strategy developed in potato has paved the way for commercial applications [10].

In the field of biopharmaceuticals, chloroplasts serve as efficient biofactories for the high-level expression of vaccine antigens, therapeutic proteins, and industrial enzymes [9,12,13]. For example, an antifungal enzyme cocktail (chitinase, glucanase, and mannanase) expressed in lettuce

chloroplasts effectively degrades the cell wall of *Candida albicans* and inhibits fungal growth in clinical samples from oral cancer patients [11]. Additionally, a biological containment system utilizing the rare UGA codon in microalgal chloroplasts enables the safe expression of target genes that are toxic to *Escherichia coli* [14].

Despite remaining challenges, such as low transformation efficiency in some species and time-consuming homoplasmy selection [3,4], the continued development of standardized frameworks and novel transformation vectors positions chloroplast engineering to play a pivotal role in climate-resilient agriculture and sustainable biomanufacturing [3,10]. Therefore, in-depth investigation into the codon usage bias and optimal codon composition of plant chloroplast genomes is of great significance, as it will enhance the translational efficiency of chloroplast transgenes and facilitate the stable, high-level expression of heterologous genes of interest.

Codon usage bias refers to the non-random usage of different synonymous codons in species or genes [14]. Its formation is comprehensively affected by multiple factors such as genomic nucleotide composition (GC content), gene length, expression level, and environmental stress [10,13], and is the result of the dynamic balance among mutational pressure, natural selection, and genetic drift [8]. Studies have shown that CUB widely exists among different species, tissues, organs, and genes [11,15,16], and is more significant in highly expressed genes, with its intensity increasing with the elevation of gene expression level [17].

S. japonicum f. *oligophyllum* (Fabaceae), a form of *S. japonicum* in the genus *Styphnolobium*, is also known as butterfly pagoda tree. Its leaflets are clustered palmately with a unique leaf shape, showing high ornamental value [18]. Meanwhile, it has the characteristics of drought tolerance, barren tolerance, and strong stress resistance, and is widely used in landscaping, building material processing, and other fields. At present, the structural characteristics of the chloroplast genome in the genus *Styphnolobium* have been reported [19,20], but research on the codon usage bias of *S. japonicum* f. *oligophyllum* remains blank. In this study, by analyzing the structural characteristics and codon usage patterns of its chloroplast genome, we aim to provide a reference for genome evolution and genetic improvement of *Styphnolobium* plants.

2. Materials and Methods

2.1. Sampling, DNA Extraction Sequencing and Annotation

In this study, fresh leaves of *S. japonicum* f. *oligophyllum* were collected from the Chinese Pagoda Tree Garden in Shenqiu, Henan Province, China. Genomic DNA was extracted and subjected to Illumina paired-end (PE) sequencing (2×150 bp). Using the chloroplast genome of *S. japonicum* (GenBank accession no. MG784459) as a reference, genome assembly was performed with SPAdes v3.15.3 [21]. Annotation was conducted using the online tool CPGAVAS2, followed by manual correction with Geneious v9.0 [22]. The complete chloroplast genome sequence obtained in this study was deposited in the National Center for Biotechnology Information (NCBI) under the accession number ON571618.

2.2. Calculation of Codon Nucleotide Composition and Preference Parameters

Short coding DNA sequences (CDSs) tend to cause large deviations in codon usage analysis; thus, CDSs shorter than 300 bp were excluded prior to calculation. Protein-coding sequences from the annotated chloroplast genome of *S. japonicum* f. *oligophyllum* were manually extracted using Geneious software. Redundant sequences and sequences shorter than 300 bp were removed, and 52 CDSs were ultimately retained for codon usage bias analysis. The 52 qualified CDSs were combined into a single FASTA file. Codon composition, codon adaptation index (CAI), codon bias index (CBI), and frequency of optimal codons (Fop) were determined using CodonW 1.4.2. The CUSP program in the EMBOSS online tool was used to analyze codon preferences, including GC content at the first (GC₁), second (GC₂), and third (GC₃) codon positions, as well as the mean GC content across all three positions (GC_{all}). Statistical analysis was performed with SPSS 22.0.

2.3. Analysis of Relative Synonymous Codon Usage

The relative usage of synonymous codons was analyzed as the ratio of the observed usage frequency to the theoretical usage frequency of codons in the chloroplast genome of *S. japonicum* f. *oligophyllum*. The relative synonymous codon usage (RSCU) values of the 52 CDSs were calculated. An RSCU value of 1 indicates no codon usage bias; an RSCU value > 1 indicates higher usage frequency of the codon; otherwise, the codon usage frequency is lower [23].

2.4. Neutrality Plot Analysis

Neutrality plot can be used to analyze factors affecting codon usage bias [24]. A scatter plot was generated with GC_3 as the abscissa and GC_{12} ($GC_{12} = (GC_1 + GC_2)/2$) as the ordinate; each point on the plot represents one gene. If the scatter points are uniformly distributed along the diagonal line, the regression coefficient of the standard curve is close to 1, suggesting no significant difference between GC_{12} and GC_3 and no obvious divergence in base-pair composition at the three codon positions. Under this circumstance, the gene is less affected by selection pressure and more prone to mutational bias. Conversely, when the scatter points deviate from the diagonal, the standard regression coefficient approaches zero, indicating a remarkable difference between GC_{12} and GC_3 , and the codon usage of this gene is mainly shaped by selection pressure.

2.5. ENC-Plot Analysis

The effective number of codons (ENC) is also an important indicator for analyzing codon usage bias [24,25], and its value can be used to identify highly and lowly expressed genes. ENC-plot analysis was performed with ENC as the ordinate and GC content at the third codon position (GC_3) as the abscissa. The plot contains a standard curve and scatter points, where each circle represents one protein-coding gene.

2.6. PR2-Plot Bias Analysis

The contents of A, T, G, and C at the third position of each codon in the chloroplast genome of *S. japonicum* f. *oligophyllum* were first determined. Then, a PR2-plot scatter diagram was constructed with the proportion of G at the third position relative to G+C ($G_3/(C_3+G_3)$) as the abscissa and the proportion of A at the third position relative to A+T ($A_3/(T_3+A_3)$) as the ordinate. The central reference point (0.5, 0.5) represents complete base equilibrium ($A=T, G=C$), which is the neutral state without bias. The direction and degree of bias of other genes are indicated by the vector distance between this point and the central reference point [26].

2.7. Determination of Optimal Codons

Sequences were sorted by the effective number of codons (ENC) values, and the top and bottom 10% of sequences at both ends were selected to construct the high-expression library and low-expression library, respectively. The $\Delta RSCU$ values of the gene were then calculated. A codon was defined as an optimal codon if it met two criteria: $\Delta RSCU \geq 0.08$ (high-expression codon) and $RSCU > 1$ (high-frequency codon)[27].

3. Results

3.1. Characteristics of the Chloroplast Genome of *S. japonicum* f. *oligophyllum*

The sequencing results showed that the complete chloroplast genome of *S. japonicum* f. *oligophyllum* was 158,739 bp in length, exhibiting a typical quadripartite structure of angiosperm chloroplast genomes (Figure 1), with distinct characteristics in each structural partition. The total GC content of this chloroplast genome was 36.1%, and significant differences in GC content were detected among different structural regions: the large single-copy (LSC) region had a GC content of 33.5%, the

small single-copy (SSC) region had the lowest GC content at only 29.61%, and the inverted repeat (IR) region had the highest GC content, reaching 43.15%. This distribution pattern was highly consistent with the GC content characteristics of most angiosperm chloroplast genomes.

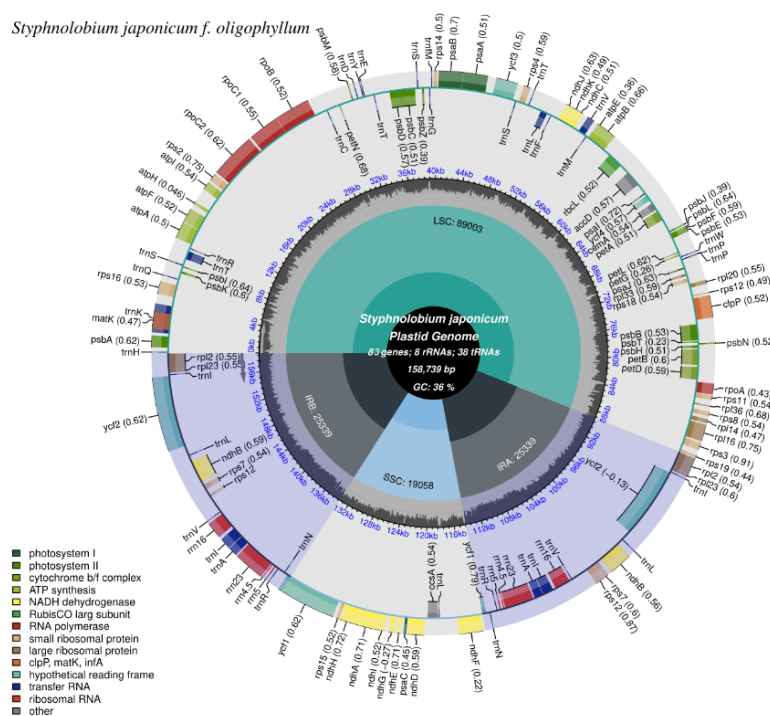


Figure 1. Circular map of the complete chloroplast genome of *S. japonicum* f. *oligophyllum*.

Gene composition analysis indicated that the chloroplast genome of *S. japonicum* f. *oligophyllum* had an identical gene composition to that of *S. japonicum* (accession number: MG784459). Both genomes contained 129 unique functional genes, which fell into three major categories: 83 protein-coding genes (PCGs), 38 tRNA genes, and 8 rRNA genes. According to the basic biological functions of the genes, the functional genes in the chloroplast genome of *S. japonicum* f. *oligophyllum* could be divided into four categories, as detailed in (Table 1).

Table 1. Gene List of the Chloroplast Genome of *S. japonicum* f. *oligophyllum*.

Category	Gene group	Gene group
Photosynthesis	Subunits of photosystem I	<i>psaA, psaB, psaC, psaI, psaJ</i>
	Subunits of photosystem II	<i>psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ</i>
	Subunits of NADH dehydrogenase	<i>ndhA*, ndhB*(2), ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK</i>
	Subunits of cytochrome b/f complex	<i>petA, petB*, petD*, petG, petL, petN</i>
	Subunits of ATP synthase	<i>atpA, atpB, atpE, atpF*, atpH, atpI</i>
	Large subunit of rubisco	<i>rbcl</i>
	Proteins of large ribosomal subunit	<i>rpl14, rpl16*, rpl2*(2), rpl20, rpl23(2), rpl33, rpl36</i>
	Proteins of small ribosomal subunit	<i>rps11, rps12*(2), rps14, rps15, rps16*, rps18, rps19, rps2, rps3, rps4, rps7(2), rps8</i>
	Subunits of RNA polymerase	<i>rpoA, rpoB, rpoC1*, rpoC2</i>
	Ribosomal RNAs	<i>rrn16S(2), rrn23S(2), rrn4.5S(2), rrn5S(2)</i>
Self-replication	Transfer RNAs	<i>trnA-UGC*(2), trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, trnG-GCC, trnH-GUG, trnI-CAU(2), trnI-GAU*(2), trnK-UUU*, trnL-CAA(2), trnL-UAA*, trnL-UAG, trnM-CAU, trnN-GUU(2), trnP-GGG, trnP-UGG, trnQ-UUG, trnR-ACG(2), trnR-UCU, trnS-GCU, trnS-GGA, trnS-UGA, trnT-CGU*, trnT-GGU, trnT-UGU, trnV-GAC(2), trnV-UAC*, trnW-CCA, trnY-GUA, trnY-CAU</i>

	Maturase	<i>matK</i>
	Protease	<i>clpP**</i>
Other genes	Envelope membrane protein	<i>cemA</i>
	Acetyl-CoA carboxylase	<i>accD</i>
Genes of unknown function	c-type cytochrome synthesis gene	<i>ccsA</i>
	Conserved hypothetical chloroplast ORF	<i>ycf1(2),ycf2(2),ycf3**,ycf4</i>

Notes: Gene*: Gene with one introns; Gene**: Gene with two introns; Gene(2): Number of copies of multi-copy genes.

Further analysis of gene copy number characteristics revealed that the genome contained 18 duplicated genes, each with two copies. These genes belonged to three categories: seven protein-coding genes, namely *ycf2*, *ycf1*, *rps12*, *rps7*, *rpl23*, *rpl2*, and *ndhB*; four rRNA genes, namely *rrn23S*, *rrn16S*, *rrn5S*, and *rrn4.5S*; seven tRNA genes, namely *trnA-UGC*, *trnI-CAU*, *trnI-GAU*, *trnL-CAA*, *trnN-GUU*, *trnR-ACG*, and *trnV-GAC*.

3.2. Codon Composition and Preference Parameter Analysis

Analysis was conducted on 52 CDS sequences from the chloroplast genome of *S. japonicum* f. *oligophyllum*. The results are shown in (Table S1). The average contents of GC₁, GC₂, GC₃ and GC_{all} were 46.56%, 39.37%, 28.26% and 38.06%, respectively. Among them, the GC₂ content was relatively close to the overall level of GC_{all}, while the difference between GC₁ and GC₃ contents was significant. The GC contents at all three codon positions were lower than 50%, and showed an overall decreasing pattern of GC₁ > GC₂ > GC₃, with obvious differences in nucleotide composition among different positions.

Further analysis of the variation range of GC content at each codon position revealed distinct differences in the extreme GC content values among different genes. GC₁ content ranged from 31.79% to 58.38%, with the highest value in *clpP* and the lowest in *ccsA*; GC₂ content ranged from 33.72% to 53.24%, with the highest value in *rps11* and the lowest in *ycf1*; GC₃ content ranged from 21.69% to 36.81%, with the highest value in *ycf2* and the lowest in *ndhF*; GC_{all} content ranged from 29.14% to 45.10%, with the highest value in *rps12* and the lowest in *ycf1*. Combined with the GC content distribution characteristics at the above positions, it can be confirmed that the chloroplast genome of *S. japonicum* f. *oligophyllum* prefers codons ending in A or U, and the third codon position shows a stronger bias toward A or U.

The ENC values of the codons ranged from 36.63 to 52.65 (Table S1), with an average of 45.40, indicating a weak codon usage bias. The CAI values ranged from 0.1 to 0.29 (average 0.17); the CBI values ranged from -0.22 to 0.17 (average -0.10), reflecting that the usage frequency of optimal codons was lower than the average usage frequency of all codons.

The Fop values ranged from 0.27 to 0.52 (average 0.35), indicating that the proportion of optimal codons among synonymous codons was small. The above results demonstrate a weak codon usage bias in the chloroplast genome of *S. japonicum* f. *oligophyllum*.

GC₁, GC₂ and GC₃ were all highly significantly correlated with GC_{all} (P<0.001) (Figure 2). GC₁ exhibited a highly significant positive correlation with GC₂, whereas GC₃ showed only weak correlation with GC₁ and GC₂. This indicated that the base composition and usage at the first and second codon positions were similar, but distinct from those at the third position. ENC was significantly positively correlated with GC₁, GC₃ and GC_{all} (P<0.05), but showed little correlation with GC₂, implying that the GC content at the second codon position had little effect on codon usage bias. GC₁ showed extremely significant positive correlations with CAI, CBI and Fop (P<0.01); GC₂ was significantly positively correlated with Fop; GC₃ was significantly positively correlated with CBI and Fop; GC_{all} was highly significantly positively correlated with CBI and Fop, and significantly positively correlated with CAI. These results indicated that base composition had a strong influence on codon usage bias in the chloroplast genome of *S. japonicum* f. *oligophyllum*.

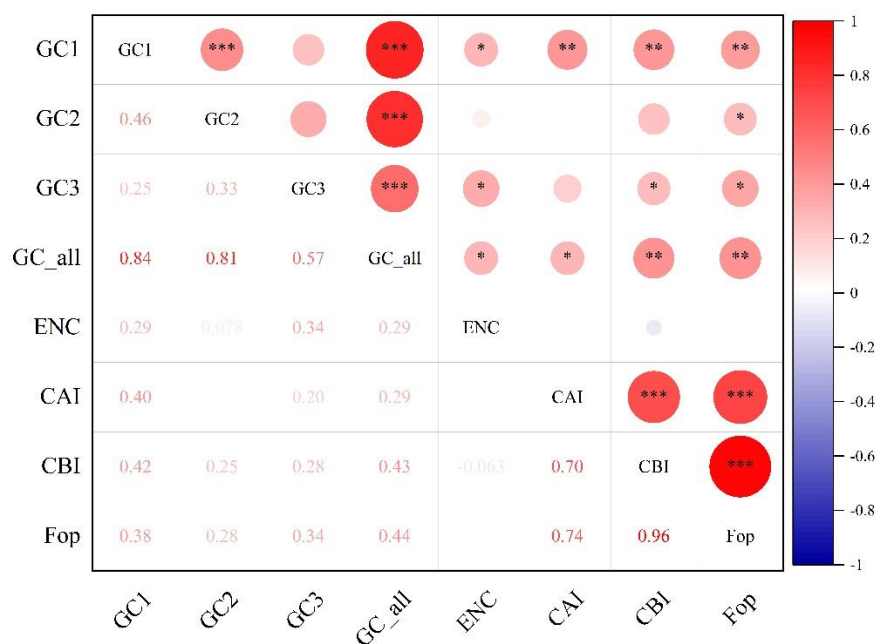


Figure 2. Correlation Analysis of Parameters in the Genome. **Note:** *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

3.3. Analysis of Relative Synonymous Codon Usage

According to RSCU data analysis (Figure 3), 31 codons in *S. japonicum* f. *oligophyllum* had RSCU > 1. Of these, 13 (41.9%) ended with A, 16 (51.6%) with U, 1 (3.25%) with G, and 1 (3.25%) with C. The proportion of codons ending in A+U reached 93.5%, indicating a strong preference for A/U-ending codons in its chloroplast genome.

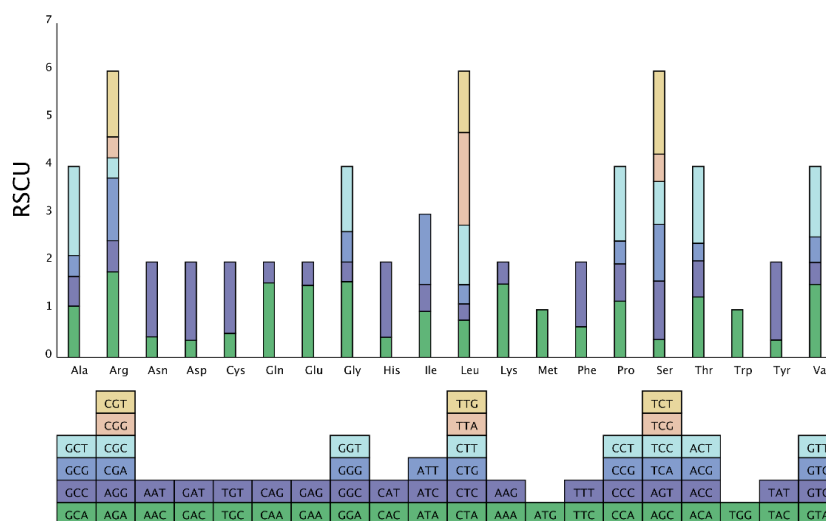


Figure 3. RSCU of codon in chloroplast genome.

3.4. Neutrality Plot Analysis

GC_{12} (the average of GC_1 and GC_2) ranged from 31% to 54%, while GC_3 varied narrowly from 22% to 37%, showing the typical AT bias of chloroplast genomes (Figure 4A). The first and second codon positions are under stronger functional constraints for amino acid coding, whereas the third synonymous site exhibited a more distinct AT preference. The correlation coefficient (R^2) between GC_{12} and GC_3 was 0.095, and the regression slope ($y = 0.524$) was significantly less than 1, indicating

that codon usage bias in the chloroplast genome was mainly driven by natural selection rather than mutation pressure.

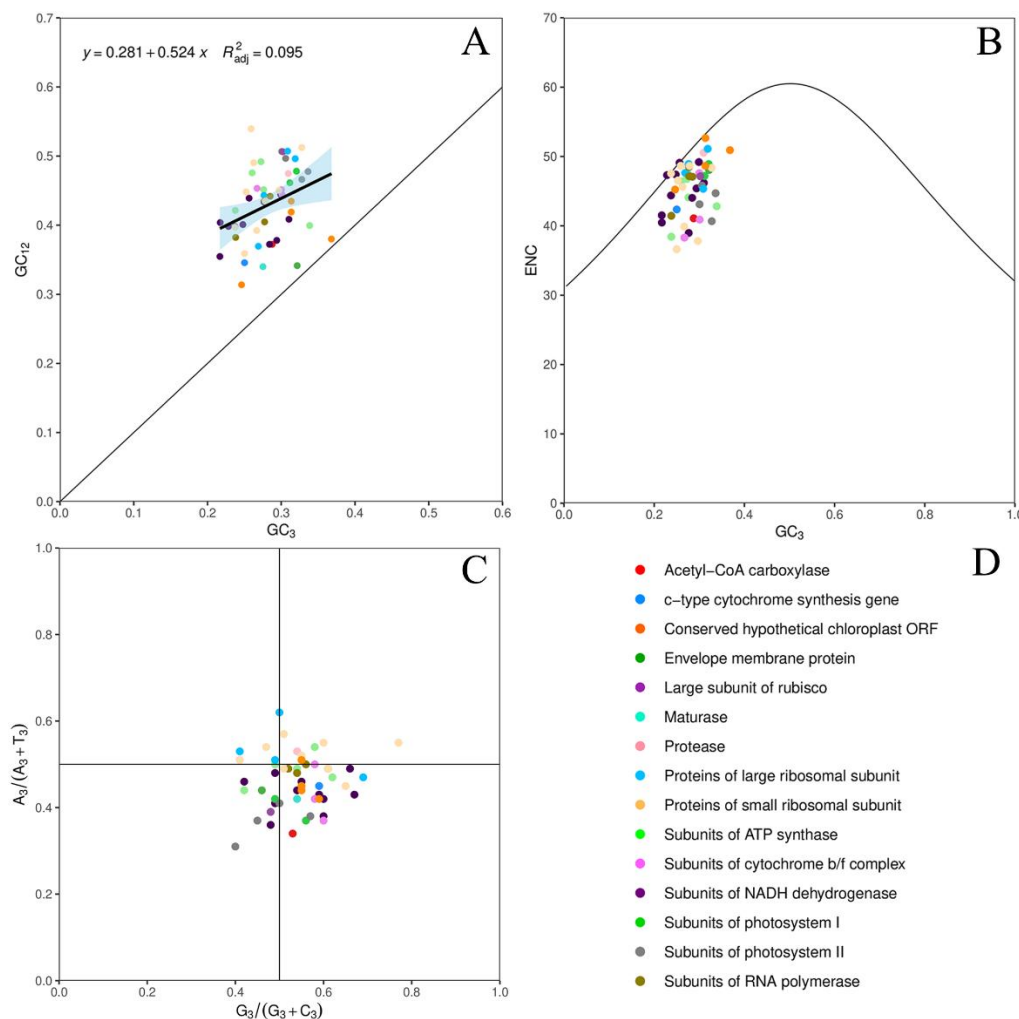


Figure 4. Plots of the causes of codon preference in the chloroplast genome. (A) Neutrality plot analysis; (B) ENC plot analysis; (C) PR2-plot analysis; (D) Functional classification legend for chloroplast genome proteins/genes.

3.5. ENC-Plot Analysis

The ENC values of all chloroplast genes of *S. japonicum* f. *oligophyllum* were significantly lower than the theoretical maximum of 61, indicating an obvious codon usage bias in its chloroplast genome (Figure 4B). All gene points were concentrated in the low GC range of 22%–37% for GC₃, and lay significantly below the standard curve with only a few genes near the curve. This is a typical feature that the codon usage bias of the chloroplast genome is dominated by natural selection, excluding the possibility that it is driven solely by mutation pressure.

Among the 52 protein-coding genes in the chloroplast genome of *S. japonicum* f. *oligophyllum*, 77% had ENC ratios ranging from -0.05 to 0.15 (40 genes) (Table 2). Only 23% showed a deviation > 0.15 , and just one gene (2%) fell in the high-deviation range of 0.25 – 0.35 . This indicates that the observed ENC values of most genes differed slightly from the theoretical values predicted by GC₃, suggesting a weak effect of mutation pressure. Combined with the ENC-plot analysis, natural selection was confirmed as the main factor shaping codon usage bias.

Table 2. Frequency distribution of ENC ratios.

Class range	Class mid value	number	Frequency
-0.05~0.05	0	13	0.25
0.05~0.15	0.1	27	0.52
0.15~0.25	0.2	11	0.21
0.25~0.35	0.3	1	0.02
Total		52	0.25

3.6. PR2-Plot Bias Analysis

PR2-plot analysis showed that the third codon base usage bias of all protein-coding genes was moderate (Figure 4C). All gene points clustered around the central reference point (0.5, 0.5) without extreme aggregation in a single quadrant. $G_3/(G_3+C_3)$ was mainly distributed at 0.4–0.6 (slightly > 0.5), indicating a slightly higher G than C content; $A_3/(A_3+T_3)$ ranged from 0.3 to 0.5 (slightly < 0.5), showing a slightly higher T than A content, which is consistent with the typical AT bias of chloroplast genomes. These results suggest that the base usage bias at the third codon position is shaped by both natural selection and mutation pressure. Natural selection restricts extreme bias and maintains functional stability of codon usage, which is consistent with the results of neutrality plot and ENC-plot analyses.

3.7. Determination of Optimal Codons

There are a total of 22 high-expression codons with $\Delta RSCU \geq 0.08$ (Table S2). Among these codons, 19 have an RSCU value greater than 1 and are identified as optimal codons, namely *GCU*, *CGU*, *AAU*, *GAU*, *UGU*, *CAA*, *GAA*, *GGU*, *CAU*, *AUU*, *UUA*, *AAA*, *UUU*, *CCC*, *AGU*, *UCU*, *ACU*, *GUA*, and *GUU*. Among these optimal codons, 13 ended with U, 5 ended with A, and 1 ended with C, indicating a strong preference for codons ending in A/U.

4. Discussion

The chloroplast genome of *S. japonicum* f. *oligophyllum* presents a typical circular quadripartite structure. Its core characteristics, including genome length, structural partitioning, GC content and gene composition, are highly consistent with previous studies on intraspecific taxa of *S. japonica* [19,28,29], further verifying the relatively conserved evolution of its chloroplast genome. The GC content varies significantly across different structural regions, following the pattern IR regions > LSC region > SSC region. This pattern has been confirmed in diverse angiosperms, such as *Quercus lamellosa* [30] and *Cinnamomum mollifolium* [31]. IR regions have a high GC content, which may be related to the enrichment of abundant RNA-coding genes and higher sequence conservation in the IR regions. Nevertheless, this GC distribution pattern is not universal, as some plant species exhibit specificity. A study on *Schefflera octophylla* showed that its GC content decreased in the order of SSC region (41.13%), IR regions (38.73%) and LSC region (36.11%) [10]. This may be attributed to the active evolution of chloroplast genomes in Araliaceae, where frequent shifts of IR boundaries lead to changes in gene content and further alter the nucleotide composition of each region.

Numerous studies have confirmed that codon optimization can effectively improve the expression efficiency of plant target genes to a certain extent [27,31,32]. Analysis of the chloroplast genome of *S. japonicum* f. *oligophyllum* in this study showed that the GC content differed among different codon positions, with an average order of $GC_1 > GC_2 > GC_3$, meaning the GC content at the first two positions of codons was higher than that at the third position. This pattern is consistent with the results of codon preference studies in *Macadamia integrifolia* [23], *Actinostemma tenerum* [33], *Sphaerophysa salsula* [34], *Koelreuteria bipinnata* [35] and other plant species. In this study, 31 synonymous codons with $RSCU > 1$ were identified in the chloroplast genome of *S. japonicum* f. *oligophyllum*, and 93.5% of the high-frequency codons ended with A+U. A total of 19 optimal codons were detected, among which 94.74% ended with A/U, demonstrating the core characteristics of strong

AT bias, predominant U-ending, and formation mainly driven by natural selection. This further confirms that the codon structure of the chloroplast genome is relatively conserved in higher plants during evolution, with high similarity in codon usage patterns. These findings are consistent with those from studies on *Koelreuteria bipinnata* [35], *Picea brachytyla* var. *complanata* [36]. Meanwhile, some differences exist in the type and number of optimal codons among different plants, suggesting that distinct plant lineages have experienced different selective pressures during evolution.

In this study, neutrality plot, ENC-plot and PR2-plot were comprehensively applied to analyze the formation mechanism of codon usage bias in its chloroplast genome. Neutrality plot analysis showed that the correlation coefficient between GC_{12} and GC_3 was extremely low ($R^2 = 0.095$), and the regression slope (0.524) was significantly less than 1. This indicated that the first and second codon positions were under strong functional constraints, the third codon position showed more significant AT bias, and codon usage bias was mainly dominated by natural selection. In the ENC-plot analysis, the ENC values of all genes were lower than the theoretical maximum, and gene points were concentrated in the low GC_3 range (22%–37%) and distributed below the standard curve. The ENC ratio distribution showed that 77% of the genes had ratios concentrated in the range of -0.05 to 0.15 with small deviations, further confirming that codon usage bias was dominated by natural selection, while mutation pressure played only a weak role. PR2-plot analysis revealed that gene points were evenly distributed around the central reference point (0.5, 0.5). $G_3/(G_3+C_3)$ was slightly higher than 0.5, and $A_3/(A_3+T_3)$ was slightly lower than 0.5, showing a trend that G was slightly more abundant than C and T was slightly more abundant than A, which was consistent with the typical AT bias characteristic. These results suggested that base usage at the third codon position was shaped by both natural selection and mutation pressure, and natural selection effectively restricted extreme bias. The above patterns are consistent with the findings of studies on plants such as *Actinostemma tenerum* [33] *Kadsura* [37], *Cladrastis yunchunii* [38], and Juglandaceae (walnut)[39]. In contrast, the codon usage bias of species including *Eriobotrya fragrans* [16] and *Acer amplum* subsp. *catalpifolium* [40] is jointly regulated by natural selection and mutation pressure, indicating that the dominant driving factors differ significantly among species.

5. Conclusions

This study systematically analyzed the structural characteristics and codon usage bias pattern of the chloroplast genome of *S. japonicum* f. *oligophyllum*. It was clarified that the chloroplast genome of this species exhibits high evolutionary conservation in sequence composition, structural arrangement and gene constitution, and natural selection was the dominant driving force underlying codon usage bias. These findings provide a theoretical basis for dissecting the molecular evolutionary mechanism, optimizing target gene codons and conducting genetic breeding research in *Styphnolobium* species. Future studies can carry out comparative analyses of codon usage bias among closely related *Styphnolobium* species, further explore the molecular differentiation characteristics during their evolutionary process, and improve the research system of molecular evolution for the genus *Styphnolobium*.

Supplementary Materials: The following supporting information can be downloaded at the website of this paper posted on Preprints.org.

Author Contributions: Conceptualization, Z.-Q.M. and H.-W.W.; formal analysis, Z.-Q.M.; investigation, Z.-Q.M., J.-J.Y., X.Z. and B.-P.C.; data curation, Z.-Q.M.; writing—original draft preparation, Z.-Q.M.; writing—review and editing, H.-W.W.; visualization, Z.-Q.M.; supervision, H.-W.W.; funding acquisition, Z.-Q.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Scientific Research Project of Kashi University (grant number 20242937). The APC was funded by the Scientific Research Project of Kashi University (grant number 20242937).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The genome sequence of this study was deposited in GenBank of NCBI (<https://www.ncbi.nlm.nih.gov/> (accessed on 15 August 2025)) under accession no. ON571618.

Acknowledgments: The authors are grateful to the colleagues who provided technical support for this study.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Arimura, S.; Nakazato, I. Genome Editing of Plant Mitochondrial and Chloroplast Genomes. *Plant And Cell Physiology* **2024**, *65*, 477–483, doi:10.1093/pcp/pcad162.
2. Xu, C.; Dong, W.; Li, W.; Lu, Y.; Xie, X.; Jin, X.; Shi, J.; He, K.; Suo, Z. Comparative Analysis of Six *Lagerstroemia* Complete Chloroplast Genomes. *Front. Plant Sci.* **2017**, *8*, doi:10.3389/fpls.2017.00015.
3. Daniell, H.; Lin, C.-S.; Yu, M.; Chang, W.-J. Chloroplast Genomes: Diversity, Evolution, and Applications in Genetic Engineering. *Genome Biol* **2016**, *17*, 134, doi:10.1186/s13059-016-1004-2.
4. Raubeson, L.A.; Jansen, R.K. Chloroplast genomes of plants. In *Plant Diversity and Evolution. Genotypic and Phenotypic Variation in Higher Plants*; Henry, R.J., Ed.; CABI Publishing: Wallingford, UK, 2005; Volume 42, pp. 45–68.
5. Daniell, H.; Kumar, S.; Dufourmantel, N. Breakthrough in Chloroplast Genetic Engineering of Agronomically Important Crops. *Trends Biotechnol* **2005**, *23*, 238–245, doi:10.1016/j.tibtech.2005.03.008.
6. Bock, R. Genetic Engineering of the Chloroplast: Novel Tools and New Applications. *Current Opinion in Biotechnology* **2014**, *26*, 7–13, doi:10.1016/j.copbio.2013.06.004.
7. Bulle, M.; Rahman, Md.M.; Kota, S.; Islam, Md.R.; Keya, S.S.; Abbagani, S.; Kirti, P.B. Advancing Chloroplast Bioengineering: Innovations, Regulatory Challenges, and Translational Pathways for Sustainable Agriculture. *International Journal of Biological Macromolecules* **2026**, *350*, 150873, doi:10.1016/j.ijbiomac.2026.150873.
8. Ingvarsson, P.K. Molecular Evolution of Synonymous Codon Usage in *Populus*. *BMC Evol Biol* **2008**, *8*, 307, doi:10.1186/1471-2148-8-307.
9. Wang, J.; Qian, J.; Jiang, Y.; Chen, X.; Zheng, B.; Chen, S.; Yang, F.; Xu, Z.; Duan, B. Comparative Analysis of Chloroplast Genome and New Insights Into Phylogenetic Relationships of *Polygonatum* and Tribe Polygonateae. *Front. Plant Sci.* **2022**, *13*, doi:10.3389/fpls.2022.882189.
10. Li, C.-L.; Zhang, P.; Cai, C.-L.; Qin, S.-F. Genomic Characteristics and Codon Preference Analysis of *Schefflera Octophylla* (Lour.) Harms Chloroplasts. *Journal of Agricultural Science and Technology* **2024**, *26*, 63–76.
11. Geng, X.; Huang, N.; Zhu, Y.; Qin, L.; Hui, L. Codon Usage Bias Analysis of the Chloroplast Genome of Cassava. *South African Journal of Botany* **2022**, *151*, 970–975, doi:10.1016/j.sajb.2022.11.022.
12. Yao, S.; Zhang, Q.; Su, M. Chloroplast Genome Capture History and Genetic Diversity of *Camelia Sinensis* Var. *Sinensis* 'Liupao'. *Guihaia* **2025**, *45*, 527–541.
13. Mazumdar, P.; Binti Othman, R.; Mebus, K.; Ramakrishnan, N.; Ann Harikrishna, J. Codon Usage and Codon Pair Patterns in Non-Grass Monocot Genomes. *Ann Bot* **2017**, *120*, 893–909, doi:10.1093/aob/mcx112.
14. Hershberg, R.; Petrov, D.A. Selection on Codon Bias. *Annu Rev Genet* **2008**, *42*, 287–299, doi:10.1146/annurev.genet.42.110807.091442.
15. Li, R.; Wang, B.; Xiao, S.; Chen, L.; Yin, F.; Li, J.; Jiang, C.; Zhang, D.; Zhong, Q.; Zhang, Y.; et al. Characterization of the Complete Chloroplast Genome and Comparative Analysis of the Phylogeny and Codon Usage Bias of Three Yunnan Wild Rice Species. *Front. Plant Sci.* **2025**, *16*, doi:10.3389/fpls.2025.1555104.
16. Qu, Y.-Y.; Xin, J.; Feng, F.-F. Codon Usage Bias in Chloroplast Genome of *Eriobotrya Fragrans* Champ. Ex Benth. *Journal of Northwest Forestry University* **2021**, *36*, 138–144.
17. Liang, Y.; Kong, Y.-G.; Wang, Y.-H.; Yan, Y.-P. Characteristics and Codon Usage Bias in Chloroplast Genome of *Robinia Neo-Mexicana* Var. *Luxurians*. *Journal of Central South University of Forestry & Technology* **2026**, *46*, 199–210.

18. Zhao, Y. Study on Taxonomy and Reproduction Characteristics of *Sophora Japonica* L. Abstract, Shandong Normal University, 2007.
19. Mu, Z.-Q. Genetic Diversity of Ancient *Styphnolobium Japonicum* Trees in Henan, Henan Agricultural University, 2023.
20. Mu, Z.; Zhang, Y.; Zhang, B.; Cheng, Y.; Shang, F.; Wang, H. Intraspecific Chloroplast Genome Variation and Domestication Origins of Major Cultivars of *Styphnolobium Japonicum*. *Genes* **2023**, *14*, doi:10.3390/genes14061156.
21. Bankevich, A.; Nurk, S.; Antipov, D.; Gurevich, A.A.; Dvorkin, M.; Kulikov, A.S.; Lesin, V.M.; Nikolenko, S.I.; Pham, S.; Pribelski, A.D.; et al. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J Comput Biol* **2012**, *19*, 455–477, doi:10.1089/cmb.2012.0021.
22. Kearse, M.; Moir, R.; Wilson, A.; Stones-Havas, S.; Cheung, M.; Sturrock, S.; Buxton, S.; Cooper, A.; Markowitz, S.; Duran, C.; et al. Geneious Basic: An Integrated and Extendable Desktop Software Platform for the Organization and Analysis of Sequence Data. *Bioinformatics* **2012**, *28*, 1647–1649, doi:10.1093/bioinformatics/bts199.
23. Cai, Y.-B.; Yang, X.-Y. Codon Usage Bias and Its Influencing Factors in the Chloroplast Genome of *Macadamia Integrifolia* Maiden & Betche. *Plant Science Journal* **2022**, *40*, 229–239, doi:10.11913/PSJ.2095–0837.2022.20229.
24. Li, R.-X.; Wang, B.; Xiao, S.-Q. Assembly, Codon Usage Bias, and Phylogenetic Analysis of Chloroplast Genome of *Oryza Meyeriana*. *Chinese Journal of Rice Science* **2025**, *39*, 801–812, doi:10.16819/j.1001-7216.2025.240612.
25. Chen, W.-Q.; Liu, Z.-H.; Wang, M.-J. Analysis of Codon Usage Bias in Chloroplast Genome of *Halenia Elliptica*. *Molecular Plant Breeding* **2024**.
26. Ma, Y.-D.; Yang, Y.-R.; Guo, C.-C. Analysis of Codon Bias in Chloroplast Genome of *Cercis Gigantea*. *Journal of Agricultural Science and Technology* **2026**, doi:10.13304/j.nykjdb.2024.0787.
27. Dai, J.-F.; Dang, L.; Zhao, X. Characteristics and Codon Usage Bias in *Larix Principis-Rupprechtii* Chloroplast Genome. *Journal of Northwest Forestry University* **2026**, *41*, 81–92, doi:10.3969/j.issn.1001-7461.2026.01.08.
28. Lu, Y.; Li, W.-Q.; Xie, X.-M. The Complete Chloroplast Genome Sequence of *Sophora Japonica* Var. *Violacea*: Gene Organization and Genomic Resources. *Conservation Genetics Resources* **2018**, *10*, 1–4.
29. Shi, Y.; Liu, B. Complete Chloroplast Genome Sequence of *Sophora Japonica* 'Jinhuij2' (Papilionaceae), an Important Traditional Chinese Herb. *Mitochondrial DNA B Resour* **5**, 319–320, doi:10.1080/23802359.2019.1703588.
30. Li, P.-Y.; Hong, T.; Chen, X.-L.; He, J.-Y. Characteristics and Phylogenetic Analysis of the Complete Chloroplast Genome of *Quercus Lamellosa*. *South China Agriculture* **2023**, *17*, 1–7, doi:10.19415/j.cnki.1673-890x.2023.23.001.
31. Chen, C.-H.; Yu, X.-L.; Fu, C. Genomic Sequence Characteristics and Phylogeny of Chloroplast in *Cinnamomum Mollifolium*. *Molecular Plant Breeding* **2023**, *21*, 8066–8074, doi:10.13271/j.mpb.021.008066.
32. Bhattacharyya, D.; Uddin, A.; Das, S.; Chakraborty, S. Mutation Pressure and Natural Selection on Codon Usage in Chloroplast Genes of Two Species in *Pisum* L. (Fabaceae: Faboideae). *Mitochondrial DNA A DNA Mapp Seq Anal* **2019**, *30*, 664–673, doi:10.1080/24701394.2019.1616701.
33. Mu, J.-J.; Zhang, J.-S. Codon Usage Bias Analysis in the Chloroplast Genome of *Actinostemma Tenerum* (Cucurbitaceae). *Curr Issues Mol Biol* **2025**, *47*, 833, doi:10.3390/cimb47100833.
34. Liang, X.-L.; Guo, S. Codon Usage Bias in the Chloroplast Genome of *Sphaerophysa Salsula*. *Journal of Northwest Forestry University* **2022**, *37*, 121–126.
35. Xiao, M.-K.; Nie, K.-H.; Shen, Z.-B. Analysis of Codon Usage Bias in the Chloroplast Genome of *Koelreuteria Bipinnata*. *Journal of Southwest Forestry University(Natural Sciences)* **2023**, *43*, 56–63, doi:10.11929/j.swfu.202111031.
36. Ding, S.-J.; Wei, J.-S.; Lu, Y.-L. Characteristics and Codon Usage Bias of *Picea Brachytyla* Var. *Complanata* chloroplast Genome. *Journal of Central South University of Forestry & Technology* **43**, 156–163+190, doi:10.14067/j.cnki.1673-923x.2023.04.016.
37. Liu, T.; Yin, D.-P.; Jin, J.-F. Analysis of Codon Usage Bias in the Chloroplast Genome of *Kadsura*. *Journal of Northwest Forestry University* **2023**, *38*, 102–109, doi:10.3969/j.issn.1001-7461.2023.05.14.

38. Li, J.-F.; Yi, X.-L.; Li, X.-Y. Analysis on Codon Usage Bias of Chloroplast Genome in *Cladrastis Yun-Chunii* X. W. Li & G. S. Fan. *Molecular Plant Breeding* **2023**, *21*, 2583–2590, doi:10.13271/j.mpb.021.002583.
39. Zeng, Y.; Shen, L.; Chen, S.; Qu, S.; Hou, N. Codon Usage Profiling of Chloroplast Genome in Juglandaceae. *Forests* **2023**, *14*, doi:10.3390/f14020378.
40. Long, T.; Dong, W.-P.; Chao, M. Codon Usage Bias Analysis in the *Acer Amplum* Subsp. *Catalpifolium* Genome. *Journal of Northwest Forestry University* **2023**, *38*, 61-66+80, doi:10.3969/j. issn.1001-7461.2023.06.08.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.