

Article

# Analysis of Codon Usage Bias in Chloroplast Genomes of *Dryas octopetala* var. *asiatica* (Rosaceae)

Lizhen Ling, Shudong Zhang \* and Tao Yang

Key Laboratory for Specialty Agricultural Germplasm Resources Development and Utilization of Guizhou Province, Liupanshui Normal University, Liupanshui 553004, China; primula\_ling@foxmail.com (L.L.); 19885348955@163.com (T.Y.)

\* Correspondence: sdchang@foxmail.com

**Abstract:** *Dryas octopetala* var. *asiatica*, a dwarf shrub belonging to the Rosaceae family and native to Asia, exhibits notable plasticity in photosynthesis in response to temperature variations. However, the codon usage patterns and factors influencing them in the chloroplast genome of this species have not yet been documented. This study sequenced and assembled the complete genome of *D. octopetala* var. *asiatica*. The annotated genes in the chloroplast genome were analyzed for codon composition through multivariate statistical methods including a neutrality plot, a parity rule 2 (PR2) bias plot, and an effective number of codons (ENC) plot using CodonW 1.4.2 software. The results indicated that the mean GC content of 53 CDSs was 38.08%, with the average GC content at the third codon base position being 27.80%, suggesting a preference for A/U(T) at the third codon position in chloroplast genes. Additionally, the chloroplast genes exhibited a weak overall codon usage bias (CUB) based on ENC values and other indicators. Correlation analysis showed a significant negative correlation between ENC value and GC2, an extremely positive correlation with GC3, but no correlation with GC1 content. These findings highlight the importance of the codon composition at the third position in influencing codon usage bias. Furthermore, our analysis indicated that the CUB of the chloroplast genome of *D. octopetala* var. *asiatica* was primarily influenced by natural selection and other factors. Finally, this study identified UCA, CCU, GCU, AAU, GAU, and GGU as the optimal codons. These results offer a foundational understanding for genetic modification and evolutionary dynamics of the chloroplast genome of *D. octopetala* var. *asiatica*.

**Keywords:** *D. octopetala* var. *asiatica*; chloroplast genome; codon usage bias; optimal codon



**Citation:** Ling, L.; Zhang, S.; Yang, T. Analysis of Codon Usage Bias in Chloroplast Genomes of *Dryas octopetala* var. *asiatica* (Rosaceae). *Genes* **2024**, *15*, 899. <https://doi.org/10.3390/genes15070899>

Academic Editors: Guiqi Bi, Luodong Huang and Min Cao

Received: 22 June 2024  
Revised: 4 July 2024  
Accepted: 7 July 2024  
Published: 9 July 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The genetic material of DNA typically transmits genetic information to proteins through RNA in the form of triplets. In the genetic code, there are 64 standard codons, including UAA, UAG, and UGA translation termination signal codons, and the other 61 codons encode 20 common amino acids. Among these amino acids, methionine and tryptophan are encoded by one codon, AUG and UGG, respectively, and the other 18 amino acids are encoded by at least two codons. The degeneracy of the genetic code allows the same amino acid to be encoded by multiple different codons, termed as synonymous codons [1]. However, these synonymous codons are not utilized equally, leading to codon usage bias (CUB). Significant variations in the usage of synonymous codons have been observed among plant genes and different organisms [2]. For example, a gradient of GC content and codon usage exists in correlation with the direction of transcription in Graminae genes. Monocot genes show up to 25% higher GC content at their 5' ends compared to their 3' ends, a pattern not observed in dicot genes [3]. Codon bias plays a crucial role in various cellular processes, such as transcription, mRNA stability, translation efficiency and accuracy, and protein expression and co-translational folding [4–6]. The codon preference is often affected by many factors such as environmental conditions, genetic drift, base mutations, genome size, tRNA abundance, and gene expression levels [7,8].

Chloroplasts (cps) are crucial organelles within plant cells, serving essential functions in photosynthesis, metabolic regulation, and signal transduction. These organelles possess distinct structural and compositional features, as well as a remarkable ability to adjust and thrive in response to diverse environmental factors. Chloroplasts have their own genomes and can carry out some metabolic reactions independently. In land plants, the chloroplast genome size typically ranges from 120 to 180 kb, varying in specific sizes among species. The chloroplast genome encompasses critical genes for photosynthesis and plays an important role in plant evolution [9]. The chloroplast genome in land plants originates from endophytic cyanobacteria and typically features a circular structure, consisting of two inverted repeats (IRs), a large single copy region (LSC), and a small single copy region (SSC) [9,10]. Studies have shown that the chloroplast genome of land plants contains approximately 110–130 different genes, with over 80 coding genes primarily involved in photosynthesis or metabolic processes [9]. Therefore, a comprehensive knowledge of plant chloroplast genomes can aid researchers in understanding the expression of functional proteins, thereby contributing to advancements in genetics and molecular biology. The advancement of high-throughput sequencing technology and reduced sequencing costs have enabled the sequencing of chloroplast genomes in numerous plant species. Codon preferences in multiple plant chloroplast genomes, such as *Gynostemma pentaphylla* [11], *Trichoderma pentaphylla* [12], and *Lespedeza* [13], have been analyzed and documented.

*Dryas octopetala*, a dwarf shrub in the Rosaceae family, thrives in the alpine and Arctic regions of the Northern Hemisphere. The living environment of *D. octopetala* is extremely harsh, with a damp climate, low temperature, and thin soil layers with little nutrients. Therefore, *D. octopetala* has a short growth cycle and slow growth rate. In the long-term evolutionary process, *D. octopetala* has developed special abilities to adapt to the harsh polar environment. Studies have reported that it can acquire essential nutrients like nitrogen and phosphorus through a symbiotic relationship with ectomycorrhizae (ECM) for their own growth [14–17]. At present, *D. octopetala* has become a key player in studying the role of fungi in community ecology [14]. In addition, *D. octopetala* is a living fossil that records climate change and is very important in studying it [18,19]. There are two varieties of *D. octopetala*: the original *D. octopetala* L. var. *octopetala* and the unique *D. octopetala* var. *asiatica* sporadically found in alpine tundra regions in the northeast of Asia. Recent studies have shown that the photosynthesis of *D. octopetala* var. *asiatica* is very sensitive to climate warming, demonstrating relatively strong plasticity to temperature increases [20]. To better understand *D. octopetala* var. *asiatica*, this research specifically investigated the codon usage bias of the genes within the chloroplast genome, clarifying the main factors influencing this bias in this species. This study is significant for understanding the molecular evolution of chloroplast genes and optimizing the incorporation of foreign genes in *D. octopetala* var. *asiatica*.

## 2. Materials and Methods

### 2.1. Sequence of Chloroplast Genome of *D. octopetala* var. *asiatica*

The leaves of *D. octopetala* var. *asiatica* were collected from Changbai Mountain in Jilin province, China, using standard botanical collection methods, and the voucher specimen was deposited in the Herbarium of Kunming Institute of Botany, Chinese Academy of Sciences. Genomic DNA was extracted from silica gel-dried leaves using a modified CTAB method [21], and the constructed library was used to be sequenced on the Illumina NovaSeq PE150 platform (Illumina, Inc., San Diego, CA, USA) following the manufacturer's protocols. The assembly and annotation of the chloroplast genome of this species were conducted as previously described [22]. The annotated chloroplast genome information of *D. octopetala* var. *asiatica* was submitted to the NCBI GenBank database with the accession number OQ420424.

## 2.2. Analysis of Codon Usage Characteristics

The coding protein sequences (CDSs) were extracted from the *D. octopetala* var. *asiatica* chloroplast genome according to the annotation information. The CDSs of genes were removed for analysis based on the following criteria: (1) genes contained the stop codons in the CDS (i.e., pseudogenes); (2) genes were duplicated in chloroplast genome (i.e., duplicated genes in the IR regions); (3) the gene sequences were less than 300 bp in length (gene length < 300 bp). The remaining CDSs were selected from the chloroplast genome and analyzed for subsequent analysis.

CodonW 1.4.2 software was used to calculate the occurrence frequency of GC1, GC2, and GC3 content and the average GC (GCall) base content. Codon usage bias was often assessed using relative synonymous codon usage (RSCU) and effective codon number (ENC) with the CodonW 1.4.2 software [23,24]. ENC provides an indication of the codon bias strength, with values ranging between 20 and 61, where ENC values  $\leq 35$  suggest significant codon bias within genes. Conversely, a higher ENC value indicates weaker codon usage bias [25]. The RSCU value was computed to assess the bias in codon usage among genes by comparing the observed frequency of a synonymous codon with its expected frequency. It is derived using the following formula:

$$\text{RSCU} = \frac{X_{ij}}{\sum_j^{n_i} X_{ij}} n_i$$

$X_{ij}$  represents the frequency of codon  $i$  encoding the  $j$ -th amino acid, while  $n_i$  refers to the number of codons encoding the  $i$ -th amino acid. In the absence of codon usage bias, the RSCU value for the codon would be 1. An RSCU value greater than 1 for a specific codon indicates that this codon is being used at a relatively higher frequency [26,27].

## 2.3. Analysis of the Source of Codon Usage Bias

A neutral plot was used to scrutinize the impact of mutation pressure and natural selection on codon usage. A regression line was drawn with GC3 as the x-coordinate and the average of GC1 and GC2 (GC12) as the y-coordinate. When the slope of the regression curve is close to 0, this suggests that natural selection plays a key role in shaping codon usage. When the slope is near 1, this implies that the base composition at positions 1, 2, and 3 of GC codons is similar and codon bias is mainly influenced by mutation [28].

Parity rule 2 (PR2) bias plot analysis is utilized to assess the impact of natural selection and mutation pressure on the third codon position. The composition of the four bases A, T, C, and G at the third codon position of each gene was calculated to obtain GC bias  $[G3/(G3 + C3)]$  and AU bias  $[A3/(A3 + T3)]$ . A plot was generated by plotting AU bias against GC bias, illustrating the correlation between the purine (A and G) and pyrimidine (C and T) content in genes. The central point in the plot represents the equilibrium state ( $A = T, C = G$ ), where both coordinates are 0.5. The distance and direction of a gene site from this central point indicate the degree of deviation from PR2 [29]. When more genes cluster near the standard curve, it suggests that the codon usage preference is likely to be entirely caused by mutation, whereas a deviation from the curve indicates a stronger role of natural selection in the codon usage.

ENC reflects the degree of codon dispersal and is widely used as an indicator to measure codon preference. The ENC plot was generated with ENC values on the y-axis and GC3 on the x-axis. This is a two-dimensional scatter plot that includes a reference line representing the expected values. By analyzing the distance between the actual ENC values and the expected values for various functional genes, we can observe the relationship between codon preference and different functional genes [30]. When the scatter points are close to the reference line, it indicates that codon preference is mainly influenced by mutation, while a greater distance suggests a stronger influence of natural selection. The standard curve used in this study was obtained from the equation  $\text{ENC}_{\text{exp}} = 2 + \text{GC3s} + 29/[\text{GC3s}^2 + (1 - \text{GC3s})^2]$ , which represents the expected ENC value [29].

### 2.4. Identification of Optimal Codons

The 53 CDSs that met the criteria were ranked based on their ENC values. The top and bottom 5% of the ranking were selected to establish the high-expression gene (HEG) library and low-expression gene (LEG) library, respectively. Subsequently, the RSCU values of the high- and low-expression libraries were calculated, and the difference between the high- and low-expression libraries was determined, denoted as  $\Delta$ RSCU. In the gene library, codons with  $\Delta$ RSCU > 0.08 were identified as high-expression advantageous codons [31]. The optimal codons are determined by combining codons with  $\Delta$ RSCU > 0.08 and RSCU > 1, combining high-expression advantageous codons with high-frequency codons [31].

## 3. Results

### 3.1. Annotation and Analysis of Protein-Coding Genes

The complete chloroplast genome of *D. octopetala* var. *asiatica* was assembled based on high-throughput sequencing data. A total of 84 protein-coding genes (PCGs) were annotated in the cp genome of this species. After filtering out genes with sequence lengths shorter than 300 bp, duplicated genes, and pseudogenes, 53 PCGs were selected for codon analysis. These genes were mainly categorized into three functional classifications, including photosynthetic, self-replication, and other functional groups (Table 1). Of them, eight genes (*petB*, *petD*, *atpF*, *ndhA*, *ndhB*, *rpoC*, *rpl2* and *rpl16*) contained one intron, whereas the remaining genes were all coding sequences (Table 1).

**Table 1.** The functional categories of chloroplast genes in *Dryas octopetala* var. *asiatica*.

Classification	Gene Group	Gene Name
Genes related to photosynthesis	Photosystem I	<i>psaA</i> , <i>psaB</i> , <i>psaC</i>
	Photosystem II	<i>psbA</i> , <i>psbB</i> , <i>psbC</i> , <i>psbD</i> ,
	Cytochrome b/f complex	<i>petA</i> , <i>petB</i> <sup>b</sup> , <i>petD</i> <sup>b</sup>
	ATP synthase	<i>atpA</i> , <i>atpB</i> , <i>atpE</i> , <i>atpF</i> <sup>b</sup> , <i>atpI</i>
	NADH-dehydrogenase	<i>ndhA</i> <sup>b</sup> , <i>ndhB</i> <sup>ab</sup> , <i>ndhC</i> , <i>ndhD</i> , <i>ndhE</i> , <i>ndhF</i> , <i>ndhG</i> , <i>ndhH</i> , <i>ndhI</i> , <i>ndhJ</i> , <i>ndhK</i>
	Subunit of rubisco	<i>rbcL</i>
Self-replication genes	RNA polymerase	<i>rpoA</i> , <i>rpoB</i> , <i>rpoC1</i> <sup>b</sup> , <i>rpoC2</i>
	Small subunit of ribosome	<i>rps2</i> , <i>rps3</i> , <i>rps4</i> , <i>rps7</i> <sup>a</sup> , <i>rps8</i> , <i>rps11</i> , <i>rps12</i> <sup>ac</sup> , <i>rps14</i> , <i>rps18</i> ,
	Large subunit of ribosome	<i>rpl2</i> <sup>ab</sup> , <i>rpl14</i> , <i>rpl16</i> <sup>b</sup> , <i>rpl20</i> , <i>rpl22</i> ,
Other genes	C-type cytochrome synthesis gene	<i>ccsA</i>
	Subunits of Acetyl-CoA-carboxylase	<i>accD</i>
	Envelope membrane protein	<i>cemA</i>
	Protease	<i>ClpP</i> <sup>c</sup>
Unknown genes	Conserved open reading frames	<i>ycf1</i> , <i>ycf2</i> <sup>a</sup> , <i>ycf3</i> <sup>c</sup> , <i>ycf4</i>

Note: <sup>a</sup> indicates the double-copy gene, <sup>b</sup> indicates genes with one intron, and <sup>c</sup> indicates genes with two introns.

### 3.2. Analysis of Codon Usage Bias

In this study, codon bias was primarily evaluated by analyzing GC content, the effective number of codons (ENC), and relative synonymous codon usage (RSCU). In the *D. octopetala* var. *asiatica* chloroplast genome, the GC contents of 53 genes (GCall) ranged from 29.4% to 44.1%, with an average of 38.08% (Figure 1). Further analysis of the three positions of codons (GC1, GC2, and GC3) in these genes revealed GC contents ranging from 34.47% to 57.77% for GC1, 27.39% to 56.12% for GC2, and 21.43% to 36.78% for GC3 (Table 2). On average, the GC contents of these three positions were 46.80%, 39.43%, and 28.03%, respectively (Figure 1). In addition, the frequencies of four nucleotides on the third position of codons were as follows: U(T)3s (35.19~59.49%) > A3s (29.76~56.25%) >

G3s (7.66~26.15%) > C3s (9.00~25.00%). The expression levels of U(T) and A were higher compared to G and C (Table S1). These results show an uneven distribution of the four bases across the three codon positions, with an overall preference for A/U(T) bases. Notably, a lower GC content was observed at the third nucleotide position in codons, suggesting a bias towards A or U(T) at the third codon position.

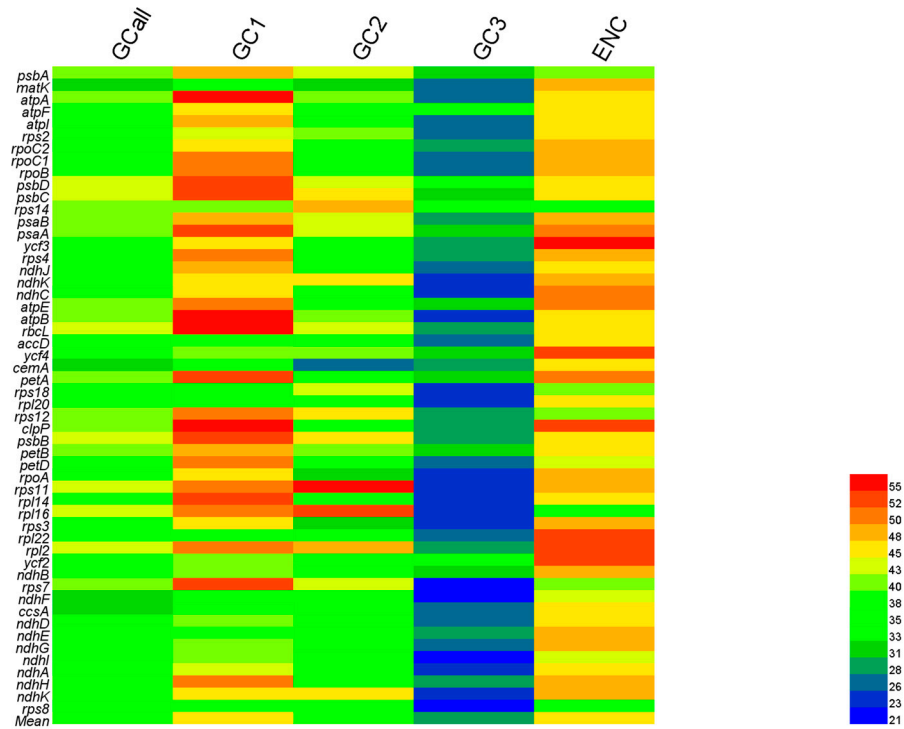


Figure 1. Heat map of GC contents and ENC of each *D. octopetala* var. *asiatica* chloroplast gene.

Table 2. RSCU analysis of each amino acid in *D. octopetala* var. *asiatica* chloroplast genes.

Amino Acid	Codon	Number	RSCU	Amino Acid	Codon	Number	RSCU
Phe	UUU	796	1.34	Tyr	UAU	657	1.61
	UUC	392	0.66		UAC	161	0.39
	UUA	748	2.02	His	CAU	398	1.53
UUG	460	1.24	CAC		122	0.47	
Leu	CUU	465	1.26	Gln	CAA	585	1.53
	CUC	134	0.36		CAG	180	0.47
	CUA	277	0.75	Asn	AAU	796	1.53
	CUG	139	0.38		AAC	244	0.47
Arg	AGA	394	1.87	Lys	AAA	852	1.54
	AGG	123	0.58		AAG	257	0.46
Met	AUG	490	1.00	Asp	GAU	706	1.61
	GUU	437	1.47		GAC	173	0.39
Val	GUC	125	0.42	Glu	GAA	870	1.51
	GUA	468	1.57		GAG	284	0.49
	GUG	160	0.54	Cys	UGU	178	1.52
	UCU	458	1.74		UGC	56	0.48
Ser	UCC	234	0.89	Trp	UGG	378	1.00
	UCA	311	1.18		CGU	296	1.41
	UCG	144	0.55	Arg	CGC	88	0.42
	CCU	355	1.61		CGA	280	1.33
Pro	CCC	159	0.72	Ser	CGG	83	0.39
	CCA	252	1.14		AGU	329	1.25
	CCG	118	0.53	AGC	100	0.38	

Table 2. Cont.

Amino Acid	Codon	Number	RSCU	Amino Acid	Codon	Number	RSCU
Thr	ACU	442	1.63	Gly	GGU	493	1.33
	ACC	188	0.69		GGC	160	0.43
	ACA	344	1.27		GGA	586	1.58
	ACG	109	0.40		GGG	244	0.66
	GCU	536	1.85		AUU	929	1.49
Ala	GCC	173	0.60	Ile	AUC	338	0.54
	GCA	328	1.13		AUA	603	0.97
	GCG	125	0.43				

The ENC values of 53 genes ranged from 33.14 to 57.39, with an average of 47.20 (Figure 1). A higher ENC value indicates a weaker codon bias, and an ENC value  $\leq 35$  is indicative of genes with significant codon bias [29]. Therefore, *D. octopetala* var. *asiatica* exhibited a weak codon preference. Moreover, analysis of RSCU values showed CUU as the most frequent codon (RSCU = 2.02) and CUC as the least frequent (RSCU = 0.36) (Table 2). There were 31 high-frequency codons with RSCU values greater than 1, predominantly ending in A/U, while codons with RSCU values less than 1 were mainly ending in G/C (Figure 1). Overall, these results indicated that the cp genome of *D. octopetala* var. *asiatica* was rich in A/U bases and exhibited a preference for codons ending with A/U in this species.

Our study delved deeper into the correlation between GCall and GC1, GC2, and GC3. The results indicated a strong positive correlation between GCall and the GC content at all three positions (GC1, GC2, and GC3) (Table 3). Specifically, a very significant correlation was observed between GC1 and GC2, while a significant correlation was found between GC1 and GC3 (Table 3). On the other hand, the correlation between GC2 and GC3 was not deemed significant (Table 3). These findings suggested that the base compositions at the first codon position had a higher similarity to those of the second position than the third position, with a notable difference in base composition between the second and third positions. Additionally, it was noted that ENC did not show a correlation with GCall and GC1 ( $p < 0.05$ ). However, ENC displayed a significant negative correlation with GC2 content and a very significant positive correlation with GC3 content (Table 3). Moreover, the size of codons did not exhibit any correlation with GC1, GC2, GC3, GCall, or ENC. These results highlighted that the GC content of codons did not vary based on the size of codons. Furthermore, the GC compositions at the first position of the codon were linked to those at the second and third positions, while the GC compositions at the third position did not show any relation to the GC content at the second position.

Table 3. Correlation analysis of codon parameters in *D. octopetala* var. *asiatica* chloroplast genes.

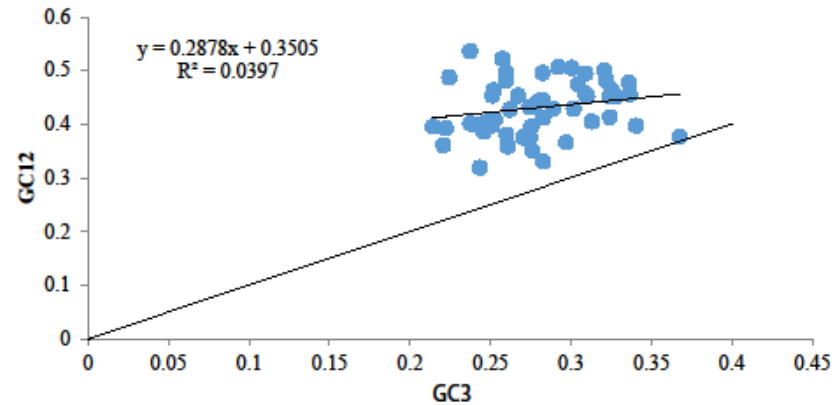
	GC1	GC2	GC3	GCall	ENC
GC2	0.43 **				
GC3	0.24 *	0.09			
GCall	0.84 **	0.77 **	0.49 **		
ENC	0.15	−0.30 *	0.37 **	0.04	
N	−0.14	−0.27 *	0.21 *	−0.15	0.18

Note: The asterisk (\*) signifies a statistically significant correlation at the 0.05 level ( $p < 0.05$ ), while the double asterisk (\*\*) denotes an extremely significant correlation at the 0.01 level ( $p < 0.01$ ).

### 3.3. Analysis of the Source of Codon Usage Bias

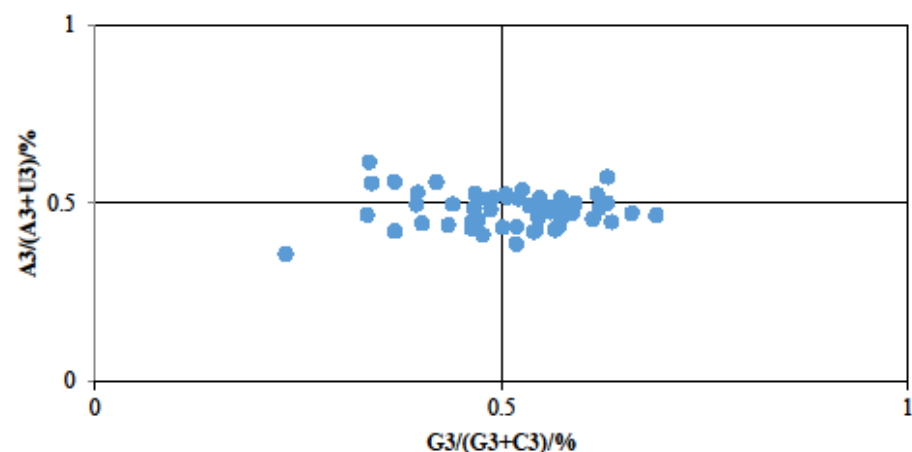
A neutral plot was used to examine the relationship between GC12 and GC3 to assess the impact of mutation and natural selection on codon preference. The results indicated that GC12 had a distribution range of 0.3–0.6, while GC3 ranged from 0.2 to 0.4 (Figure 2). Pearson correlation analysis demonstrated a weak correlation between GC12 and GC3, with a regression coefficient (R2) of 0.039 across all coding genes (Figure 2).

Additionally, most genes were situated above the diagonal line, indicating a deviation of the correlation coefficient value from 1. These results imply that natural selection has a more substantial influence than mutation on codon preference in the *D. octopetala* var. *asiatica* chloroplast genome.



**Figure 2.** Neutral plot analysis of chloroplast codons of *D. octopetala* var. *asiatica*. Note: The slope curve shows that GC12 is equal to GC3. Different coding genes are marked by blue colors.

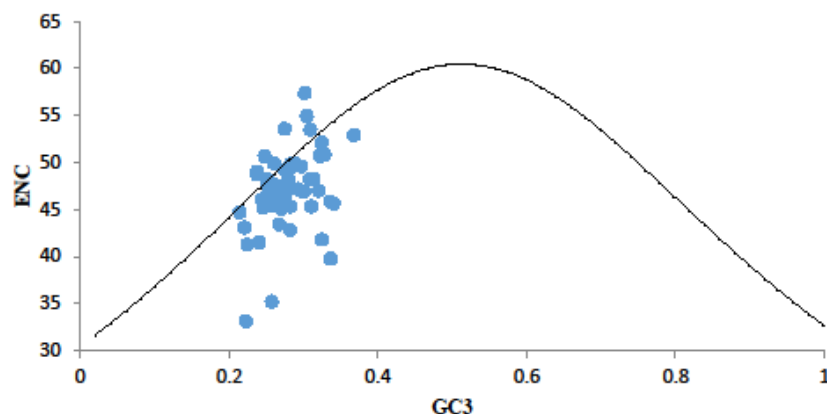
The impact of mutation and selection pressure on gene codon usage bias was further examined through PR2 plots. The analysis revealed that 53 genes exhibited uneven distribution across the four regions, with most genes being distant from the central region of 0.5, except for a few near the center (Figure 3). This observation suggests that natural selection might play a significant role in influencing the utilization of the third base of these gene codons. Moreover, the  $A3/(A3 + T3)$  ratio for most codon bases was below 0.5, while the  $G3/(G3 + C3)$  ratio for more than half of the gene codon bases exceeded 0.5 (Figure 3). These findings revealed that the frequency of codon base usage, particularly the third codon base, lean towards T(U) over A and G over C. Overall, the codon usage pattern of *D. octopetala* var. *asiatica* chloroplast genes was also affected by other factors such as mutations apart from natural selection.



**Figure 3.** PR2 plot bias analysis of chloroplast codons of *D. octopetala* var. *asiatica*. Genes were plotted based on their GC bias [ $G3/(G3 + C3)$ ] and AU bias [ $A3/(A3 + U3)$ ] in the third codon position.

To evaluate codon usage preference in the chloroplast genes of *D. octopetala* var. *asiatica*, we calculated ENC values and examined their frequency distribution. Our results showed that ENC values ranged from  $-0.09$  to  $0.29$ , with 23 genes (43%) having an ENC value below 0.05 (Figure 4). By analyzing the ENC plot, we observed that most genes fell below the standard curve, while a few genes were scattered near it, indicating diverse codon preferences across genes. The discrepancies between expected and actual ENC value

distributions suggested that natural selection played a significant role in shaping codon preference in the chloroplast genes of *D. octopetala* var. *asiatica*.



**Figure 4.** ENC plot analysis of chloroplast codons of *D. octopetala* var. *asiatica*. Note: The curve shows the relationship between ENC values and GC3S under random codon usage assumption. Different coding genes are marked by blue colors.

### 3.4. Identification of Optimal Codons

In this study, optimal codons were identified by analyzing  $\Delta$ RSCU values in high- and low-expression genes. A total of 23 codons with values exceeding 0.08 were found in high-expression genes. By combining with 31 high-frequency codons (Table 2), 6 of 23 codons were determined as optimal codons and included UCA, CCU, GCU, AAU, GAU, and GGU (Table 4). Notably, five out of the six optimal codons ended with U and one optimal codon ended with A, which was consistent with the RSCU result.

**Table 4.** Optimal codons in the chloroplast genome of *D. octopetala* var. *asiatica*.

No.	Codon	Amino Acid
1	UCA ***	Ser
2	CCU ***	Pro
3	GCU *	Ala
4	AAU	Asn
5	GAU ***	Asp
6	GGU	Gly

Note: \* indicates  $\Delta$ RSCU  $\geq$  0.1; \*\*\* indicates  $\Delta$ RSCU  $\geq$  0.5.

## 4. Discussion

Synonymous codon usage preference is a common phenomenon in the genomes of various organisms. The frequency of codon usage can differ between species and even genes within the same species [32]. Although synonymous mutations at the third base of codons do not alter the amino acid, they play a significant role in determining amino acid types. Consequently, GC3 is often utilized as a crucial indicator of codon bias [33]. Our results indicated that the chloroplast genome of *D. octopetala* var. *asiatica* had an average GC content of 38.08%, suggesting a preference for codons ending in A/U(T). This pattern is also observed in numerous plant chloroplast genomes, such as *Camellia* [34] and species from the Euphorbiaceae family [35]. Moreover, the order of GC content at the three codon positions typically follows GC1 > GC2 > GC3, consistent with patterns found in most angiosperms [29].

The neutral theory of molecular evolution suggests that base mutations and natural selection have neutral or nearly neutral effects on changes in the third base of codons. When codon usage is shaped by natural selection, GC3 values typically fall within a narrow range, with no significant correlation between GC12 and GC3 [36]. The results of the neutral analysis in this study suggest that codon bias in the chloroplast genome of *D. octopetala*



var. *asiatica* is primarily driven by natural selection, possibly as an adaptation to the harsh alpine environments where this species thrives. This base preference could potentially help researchers identify genetic markers or traits that confer resilience against stressors related to changing climate conditions, such as increased temperatures, by unraveling the genetic code and its functions in this species. In addition, PR2 and ENC plot analyses indicated that codon bias in the chloroplast genome of this species may also be influenced by other factors, including mutations apart from natural selection. In the case of *Delphinium grandiflorum* L., it is proposed that codons ending with A/T are predominantly shaped by natural selection [37]. Therefore, the codon bias in plants can be affected by various factors, which may differ among species.

The identification of optimal codons in the chloroplast genome of *D. octopetala* var. *asiatica* is influenced by strong positive selection or mutation. Studies have shown that codon usage patterns play a role in gene expression, with higher expression levels correlating with stronger codon preferences. This preference impacts gene regulation by influencing the accuracy and efficiency of translation [8,38]. Plant chloroplast gene expression vectors are typically designed using optimal codons to enhance gene expression. Furthermore, analyzing codon usage patterns can provide insights into the expression of genes with unknown functions [39]. Therefore, the investigation of optimal codons and codon usage patterns in this study is crucial for understanding species evolution and maximizing the expression of foreign genes in *D. octopetala* var. *asiatica*.

## 5. Conclusions

In this study, the codons in the chloroplast genome of *D. octopetala* var. *asiatica* were comprehensively analyzed for the first time. The results showed that the gene base compositions did not mainly consist of G and C, but instead had an average of 38.08%. Specifically, the GC content at the third codon position was only 27.08%, suggesting a preference for ending with A/U(T) in chloroplast genes of this species. Further analysis indicated that the primary factor influencing the codon usage preference was natural selection, with mutation also playing a role during evolution. Finally, this study identified UCA, CCU, GCU, AAU, GAU, and GGU as the optimal codons. In this study, the identified optimal codons and codon usage biases could potentially serve as targets for genetic engineering strategies aimed at enhancing the tolerance of *D. octopetala* var. *asiatica* to climate change-related stressors (i.e., increased temperatures), thereby contributing to conservation efforts and supporting its survival in the face of environmental change. Therefore, these findings provided a foundational dataset for future studies on species evolution and chloroplast genetic engineering in *D. octopetala* var. *asiatica*, particularly in relation to adapting to climate changes.

**Supplementary Materials:** The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/genes15070899/s1>: Table S1: GC contents of four nucleotides on the third position of codons in each gene.

**Author Contributions:** Conceptualization, S.Z.; methodology, L.L. and T.Y.; formal analysis, S.Z. and T.Y.; data curation, S.Z.; writing—original draft preparation, L.L.; writing—review and editing, S.Z. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the National Natural Science Foundation of China, grant number 31860052, and the Science and Technology Program of Liupanshui, grant number 52020-2022-PT-20.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The dataset generated and analyzed during the current study is available in the NCBI GenBank database (<https://www.ncbi.nlm.nih.gov/genbank/>, accessed on 26 March 2023) and the accession number is OQ420424.

**Acknowledgments:** We thank the Germplasm Bank of Wild Species, Kunming Institute of Botany, Chinese Academy of Sciences, for providing the samples.

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

## References

1. Duret, L. Evolution of synonymous codon usage in metazoans. *Curr. Opin. Genet. Dev.* **2002**, *12*, 640–649. [[CrossRef](#)] [[PubMed](#)]
2. Novoa, E.M.; Ribas de Pouplana, L. Speeding with control: Codon usage, tRNAs, and ribosomes. *Trends Genet. TIG* **2012**, *28*, 574–581. [[CrossRef](#)] [[PubMed](#)]
3. Wong, G.K.; Wang, J.; Tao, L.; Tan, J.; Zhang, J.; Passey, D.A.; Yu, J. Compositional gradients in Gramineae genes. *Genome Res.* **2002**, *12*, 851–856. [[CrossRef](#)] [[PubMed](#)]
4. Liu, Y. A code within the genetic code: Codon usage regulates co-translational protein folding. *Cell Commun. Signal. CCS* **2020**, *18*, 145. [[CrossRef](#)] [[PubMed](#)]
5. Komar, A.A. A Code Within a Code: How Codons Fine-Tune Protein Folding in the Cell. *Biochemistry. Biokhimiia* **2021**, *86*, 976–991. [[CrossRef](#)] [[PubMed](#)]
6. Quax, T.E.; Claassens, N.J.; Söll, D.; van der Oost, J. Codon Bias as a Means to Fine-Tune Gene Expression. *Mol. Cell* **2015**, *59*, 149–161. [[CrossRef](#)] [[PubMed](#)]
7. Romero, H.; Zavala, A.; Musto, H. Codon usage in *Chlamydia trachomatis* is the result of strand-specific mutational biases and a complex pattern of selective forces. *Nucleic Acids Res.* **2000**, *28*, 2084–2090. [[CrossRef](#)] [[PubMed](#)]
8. Fages-Lartaud, M.; Hundvin, K.; Hohmann-Marriott, M.F. Mechanisms governing codon usage bias and the implications for protein expression in the chloroplast of *Chlamydomonas reinhardtii*. *Plant J. Cell Mol. Biol.* **2022**, *112*, 919–945. [[CrossRef](#)]
9. 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.
10. Green, B.R. Chloroplast genomes of photosynthetic eukaryotes. *Plant J. Cell Mol. Biol.* **2011**, *66*, 34–44. [[CrossRef](#)]
11. Zhang, P.; Xu, W.; Lu, X.; Wang, L. Analysis of codon usage bias of chloroplast genomes in *Gynostemma* species. *Physiol. Mol. Biol. Plants Int. J. Funct. Plant Biol.* **2021**, *27*, 2727–2737. [[CrossRef](#)]
12. Hao, J.; Liang, Y.; Ping, J.; Li, J.; Shi, W.; Su, Y.; Wang, T. Chloroplast gene expression level is negatively correlated with evolutionary rates and selective pressure while positively with codon usage bias in *Ophioglossum vulgatum* L. *BMC Plant Biol.* **2022**, *22*, 580. [[CrossRef](#)] [[PubMed](#)]
13. Somaratne, Y.; Guan, D.L.; Wang, W.Q.; Zhao, L.; Xu, S.Q. The complete chloroplast genomes of two *lespedeza* species: Insights into codon usage bias, RNA editing sites, and phylogenetic relationships in *Desmodieae* (Fabaceae: *Papilionoideae*). *Plants* **2019**, *9*, 51. [[CrossRef](#)] [[PubMed](#)]
14. Bjorbækmo, M.F.; Carlsen, T.; Brysting, A.; Vrålstad, T.; Høiland, K.; Ugland, K.I.; Geml, J.; Schumacher, T.; Kausrud, H. High diversity of root associated fungi in both alpine and arctic *Dryas octopetala*. *BMC Plant Biol.* **2010**, *10*, 244–255. [[CrossRef](#)] [[PubMed](#)]
15. Billault-Penneteau, B.; Sandré, A.; Folgmann, J.; Parniske, M.; Pawlowski, K. *Dryas* as a model for studying the root symbioses of the Rosaceae. *Front. Plant Sci.* **2019**, *10*, 661. [[CrossRef](#)] [[PubMed](#)]
16. Li, H.L.; Wang, W.; Li, R.Q.; Zhang, J.B.; Sun, M.; Naeem, R.; Su, J.X.; Xiang, X.G.; Mortimer, P.E.; Li, D.Z. Global versus Chinese perspectives on the phylogeny of the N-fixing clade. *J. Syst. Evol.* **2016**, *54*, 392–399. [[CrossRef](#)]
17. Gardes, M.; Dahlberg, A. Mycorrhizal diversity in arctic and alpine tundra: An open question. *New Phytol.* **1996**, *133*, 147–157. [[CrossRef](#)]
18. McGraw, J.B.; Turner, J.B.; Chandler, J.L.; Vavrek, M.C. Disturbances as hot spots of ecotypic variation: A case study with *Dryas octopetala*. *Arct. Antarct. Alp. Res.* **2014**, *46*, 542–547. [[CrossRef](#)]
19. Panchen, Z.A.; Gorelick, R. Prediction of Arctic plant phenological sensitivity to climate change from historical records. *Ecol. Evol.* **2017**, *7*, 1325–1338. [[CrossRef](#)] [[PubMed](#)]
20. Zhou, Y.; Deng, J.; Tai, Z.; Jiang, L.; Han, J.; Meng, G.; Li, M.H. Leaf anatomy, morphology and photosynthesis of three tundra shrubs after 7-year experimental warming on changbai mountain. *Plants* **2019**, *8*, 271. [[CrossRef](#)]
21. Yang, J.-B.; Li, D.-Z.; Li, H.-T. Highly effective sequencing whole chloroplast genomes of angiosperms by nine novel universal primer pairs. *Mol. Ecol. Resour.* **2015**, *14*, 1024–1031. [[CrossRef](#)]
22. Zhang, S.-D.; Ling, L.-Z. Molecular structure and phylogenetic analyses of the plastomes of eight *Sorbus* Sensus Stricto species. *Biomolecules* **2022**, *12*, 1648. [[CrossRef](#)] [[PubMed](#)]
23. Wright, F. The ‘effective number of codons’ used in a gene. *Gene* **1990**, *87*, 23–29. [[CrossRef](#)] [[PubMed](#)]
24. Fuglsang, A. Accounting for background nucleotide composition when measuring codon usage bias: Brilliant idea, difficult in practice. *Mol. Biol. Evol.* **2006**, *23*, 1345–1347. [[CrossRef](#)] [[PubMed](#)]
25. Liao, W. Deciphering Codon Usage Patterns in Genome of *Cucumis sativus* in Comparison with Nine Species of Cucurbitaceae. *Agronomy* **2021**, *11*, 2289. [[CrossRef](#)]
26. Liu, H.; He, R.; Zhang, H.; Huang, Y.; Tian, M.; Zhang, J. Analysis of synonymous codon usage in *Zea mays*. *Mol. Biol. Rep.* **2010**, *37*, 677–684. [[CrossRef](#)]
27. Duret, L. tRNA gene number and codon usage in the *C. elegans* genome are co-adapted for optimal translation of highly expressed genes. *Trends Genet. TIG* **2000**, *16*, 287–289. [[CrossRef](#)] [[PubMed](#)]

28. Li, L.; Peng, J.; Wang, D.; Duan, A. Chloroplast genome phylogeny and codon preference of *Docynia longiunguis*. *Chin. J. Biotechnol.* **2022**, *38*, 328–342.
29. Parvathy, S.T.; Udayasuriyan, V.; Bhadana, V. Codon usage bias. *Mol. Biol. Rep.* **2022**, *49*, 539–565. [[CrossRef](#)] [[PubMed](#)]
30. Fuglsang, A. Impact of bias discrepancy and amino acid usage on estimates of the effective number of codons used in a gene, and a test for selection on codon usage. *Gene* **2008**, *410*, 82–88. [[CrossRef](#)] [[PubMed](#)]
31. Tian, L.; Baima, Y.; Yang, C.; Ma, C. Analysis of Codon Usage Bias in Chloroplast Genome of *Codonopsis convolvulacea* Kurz var. *vinciflora* (Kom.) L. T. Shen. *Mol. Plant Breed.* **2023**, 1–14. Available online: <http://kns.cnki.net/kcms/detail/46.1068.S.20230130.1902.012.html> (accessed on 6 July 2024).
32. Suzuki, H.; Morton, B.R. Codon adaptation of plastid genes. *PLoS ONE* **2016**, *11*, e0154306. [[CrossRef](#)]
33. Sablok, G.; Nayak, K.C.; Vazquez, F.; Tatarinova, T.V. Synonymous codon usage, GC(3), and evolutionary patterns across plastomes of three pooid model species: Emerging grass genome models for monocots. *Mol. Biotechnol.* **2011**, *49*, 116–128. [[CrossRef](#)]
34. Wang, Z.; Cai, Q.; Wang, Y.; Li, M.; Wang, C.; Wang, Z.; Jiao, C.; Xu, C.; Wang, H.; Zhang, Z. Comparative analysis of codon bias in the chloroplast genomes of Theaceae species. *Front. Genet.* **2022**, *13*, 824610. [[CrossRef](#)]
35. Wang, Z.; Xu, B.; Li, B.; Zhou, Q.; Wang, G.; Jiang, X.; Wang, C.; Xu, Z. Comparative analysis of codon usage patterns in chloroplast genomes of six Euphorbiaceae species. *PeerJ* **2020**, *8*, e8251. [[CrossRef](#)] [[PubMed](#)]
36. Kawabe, A.; Miyashita, N.T. Patterns of codon usage bias in three dicot and four monocot plant species. *Genes Genet. Syst.* **2003**, *78*, 343–352. [[CrossRef](#)] [[PubMed](#)]
37. Duan, H.; Zhang, Q.; Wang, C.; Li, F.; Tian, F.; Lu, Y.; Hu, Y.; Yang, H.; Cui, G. Analysis of codon usage patterns of the chloroplast genome in *Delphinium grandiflorum* L. reveals a preference for AT-ending codons as a result of major selection constraints. *PeerJ* **2021**, *9*, e10787. [[CrossRef](#)] [[PubMed](#)]
38. Zhou, Z.; Dang, Y.; Zhou, M.; Li, L.; Yu, C.H.; Fu, J.; Chen, S.; Liu, Y. Codon usage is an important determinant of gene expression levels largely through its effects on transcription. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, e6117–e6125. [[CrossRef](#)] [[PubMed](#)]
39. Wu, X.M.; Wu, S.F.; Ren, D.M.; Zhu, Y.P.; He, F.C. The analysis method and progress in the study of codon bias. *Hereditas* **2007**, *29*, 420–426. [[CrossRef](#)]

**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.