

Analysis of Synonymous Codon Usage Bias in the Chloroplast Genome of *Rhododendron farrerae*

Xianlin Rong

Hangzhou Vocational and Technical College

Lu Huang

Hangzhou Vocational and Technical College

Jianshuang Shen

shenjianshuang18@163.com

Hangzhou Vocational and Technical College

Research Article

Keywords: *Rhododendron farrerae*, Chloroplast genome, Codon, Synonymous codon usage

Posted Date: July 12th, 2024

DOI: <https://doi.org/10.21203/rs.3.rs-4615402/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Analysis of Synonymous Codon Usage Bias in the Chloroplast Genome of *Rhododendron farrerae*

Rong Xianlin Huang Lu Shen Jianshuang*
Hangzhou Vocational & Technical College, Hangzhou, 310018
* Corresponding author, shenjianshuang18@163.com

Abstract

In this study, the codon usage patterns of the chloroplast genome of *Rhododendron farrerae* were analyzed to provide genetic information for the conservation and innovation of *Rhododendron* germplasm resources. Codon usage and preferences of 54 protein coding sequences (CDS) from *R. farrerae* were analyzed using software such as Codon, R package, and EMBOSS. The results showed that the average GC content of the 54 CDS in the chloroplast genome of *R. farrerae* was 38.10%, with GC₁ (46.11%), GC₂ (39.68%), and GC₃ (28.51%) representing the GC contents at different positions, indicating that the third base of codons in the chloroplast genome of *R. farrerae* is predominantly composed of A and T. ENC-plot, PR2-plot, and neutral plotting analysis revealed that natural selection pressure was the most significant factor influencing the synonymous codon usage preference in the chloroplast genome of *R. farrerae*, leading to the identification of 16 optimal codons. This study provides genetic information for the chloroplast genome of *R. farrerae* and serves as a reference for genetic improvement and germplasm innovation.

Keywords *Rhododendron farrerae*; Chloroplast genome; Codon; Synonymous codon usage

Chloroplasts are crucial organelles in plant cells, and their genomes maintain relative stability in terms of structure. Therefore, the analysis of chloroplast genomes has played a vital role in exploring scientific questions such as plant classification, species diversity, and systematic evolution (Thomas et al., 2017). Codon usage bias in chloroplast genomes refers to the phenomenon of different frequencies of usage for different codons encoding the same amino acid in the chloroplast genome. This bias may be influenced by various factors, including mutation pressure, natural selection, gene expression level, and GC content of the genome (Li et al., 2022).

The analysis of codon usage bias helps understand the regulatory mechanisms of gene expression, predict the expression levels of exogenous genes in chloroplasts, and provide a basis for optimizing codons in genetic engineering. For example, by analyzing the codon usage patterns in chloroplast genomes, optimal codons can be determined to design more efficient genetic engineering vectors and improve gene expression efficiency in chloroplasts. Moreover, the study of codon usage bias in chloroplast genomes can provide important information for molecular phylogenetic analysis in plants. Due to the conservation and semi-autonomy of chloroplast genomes in plants, comparing their codon usage patterns among different species may reveal their evolutionary relationships (Tang et al., 2021). Codon usage bias in chloroplast genomes is a complex phenomenon influenced by multiple factors and is of significant importance for understanding chloroplast biological functions and conducting plant genetic improvement. These studies provide a strong foundation for species identification, revealing codon usage preferences, and conducting biologically relevant research (Kang et al., 2019).

Rhododendron farrerae, a widely distributed plant in China, is commonly found in diverse ecologies such as shrublands and grasslands. *R. farrerae* is highly valued for its outstanding ornamental value and strong adaptability, making it promising in the fields of landscaping and landscape design (Ng and Corlett, 2003). Additionally, the leaves and flowers of *R. farrerae* contain various metabolites, especially flavonoids, which

provide significant value in the medicinal field (Arthur, 1955). Despite the significant potential of *R. farrerae* in horticulture and medicine, research on the codon analysis of its chloroplast genome is still relatively limited. Therefore, this study aims to analyze the characteristics of the chloroplast genome and codon preference information of *R. farrerae* to better understand the chloroplast genome of *R. farrerae* and provide basic data and references for genetic variation and genetic engineering breeding of this species.

1 Materials and Methods

1.1 Materials

Based on the previously obtained sequencing results from the chloroplast genome of *R. farrerae* (GenBank accession number: OQ980386.2), 80 protein-coding genes were annotated. In this study, genes shorter than 300bp were excluded, and a final set of 54 protein-coding genes was selected for codon analysis.

1.2 Analytical methods

1.2.1 Codon composition analysis

Codon quantities, effective number codons (ENC), and relative synonymous codon usage (RSCU) of each protein-coding gene were calculated using CodonW 1.4.2 software (Lu et al., 2024). The GC content of each protein-coding gene's CDS sequence was calculated using the EMBOSS online software (<http://www.bioinformatics.nl/emboss-explorer/>), including the GC content at positions 1, 2, and 3 of each codon (GC₁, GC₂, GC₃) and the overall GC content of the nucleotide sequence (GC_{all}) (Qin et al., 2023). Correlation analysis was performed using the R package.

1.2.2 Neutrality plot analysis

A two-dimensional scatter plot is created, with the GC₃ content as the x-coordinate and GC₁₂ (the average of GC₁ and GC₂) as the y-coordinate. Each scatter point represents a gene. The neutrality level is measured based on the slope of the regression line. When the regression line is close to 0 or 1, it indicates that selective pressure or mutational pressure plays a major role (Mao, et al., 2022).

1.2.3 ENC-plot analysis

An ENC-plot scatter plot is created, with ENC as the y-coordinate and GC₃ as the x-coordinate. The standard curve is drawn based on the distribution of the scatter points, and the frequency of ENC ratio is calculated based on the distance between the scatter points and the standard curve. The degree of deviation of ENC values from the standard curve is used to determine codon usage bias (Cai Yuanbao, et al., 2022). When the ENC values of genes are near the curve, it indicates that codon usage bias is mainly influenced by mutational pressure, whereas deviation from the curve indicates a influence of selective pressure.

1.2.4 PR2-plot analysis

The content of A, T, C, and G on the third position of codons is calculated for each gene. A PR2-plot analysis is created, with A₃/(A₃+T₃) as the y-coordinate and G₃/(G₃+C₃) as the x-coordinate (Duan et al., 2021). The central point on the plot represents unbiased codon usage, where A=T and C=G. The vector distance between the other points and the central point represents the degree and direction of bias.

1.2.5 Optimal codon analysis

The effective number of codons (ENC) of each gene is used as a reference standard for codon preference. The ENC values of the coding sequence (CDS) of 54 genes are sorted in descending order. The top and bottom 10% of genes are selected to construct high and low biased libraries, respectively. The Δ RSCU value is also calculated (Δ RSCU = RSCU high expression - RSCU low expression). Codons with Δ RSCU $\geq$ 0.08 are considered as highly expressed advantageous codons. Codons with an RSCU value greater than 1 and satisfying both Δ RSCU $\geq$ 0.08 and RSCU > 1 conditions are classified as optimal codons (Qin, et al., 2023).

87 **2. Results**

88 **2.1 Codon usage characteristics of the chloroplast genome of *R. farrerae***

89 The 54 protein-coding sequences (CDS) of *R. farrerae* used in this study include 26 photosynthesis genes (*psa*,
90 *psb*, *ndh*, *pet*, *atp*, *rbcL*), 15 ribosomal protein genes (*rpl*, *rps*), 4 RNA polymerase genes (*rpo*), and 9 other genes
91 (*matK*, *ccsA*, *ycf*, *accD*, *cemA*). These genes are 52,920 bp in total length, accounting for 35.41% of the genome.
92 Codon preference analysis showed that the average GC content of the 54 CDS is 38.10%, with GC₁ (46.11%) >
93 GC₂ (39.68%) > GC₃ (28.51%), and the ENC values ranging from 36.58 to 61.00, with an average of 47.00 (Table
94 1). The above results indicate that the GC content at the 3rd codon position of the chloroplast genome of *R.*
95 *farrerae* is significantly lower than the 1st and 2nd codon positions, suggesting a bias towards A and T at the 3rd
96 codon position. 39 out of the 54 genes have ENC values > 45, indicating a relatively weak codon usage bias in the
97 chloroplast genome of *R. farrerae*, with *ycf15* having an ENC value of 61, suggesting no codon preference.

98
99 Table 1 GC proportion and ENC of chloroplast genome of *Rhododendron farrerae*

基因 Gene	GC 含量 GC content					基因 Gene	GC 含量 GC content					
	GC _{all} /%	GC ₁ /%	GC ₂ /%	GC ₃ /%	ENC		Gene	GC _{all} /%	GC ₁ /%	GC ₂ /%	GC ₃ /%	ENC
<i>atpA</i>	39.96	54.33	40.16	25.39	44.53	<i>psbD</i>	42.56	51.98	43.22	32.49	46.30	
<i>atpB</i>	41.82	56.11	41.68	27.66	46.22	<i>ycf1</i>	36.29	37.50	32.26	39.11	56.64	
<i>atpE</i>	38.31	48.51	38.06	28.36	45.29	<i>pafl</i> (<i>ycf3</i>)	38.66	47.34	38.46	30.18	50.08	
<i>atpF</i>	37.30	44.86	34.05	32.97	46.37	<i>paflII</i> (<i>ycf4</i>)	38.92	48.11	40.00	28.65	50.18	
<i>atpI</i>	38.98	49.19	37.10	30.65	46.55	<i>ycf15</i>	38.94	41.58	40.59	34.65	61.00	
<i>ccsA</i>	33.02	34.89	37.07	27.10	47.81	<i>ycf68</i>	51.52	47.93	52.89	53.72	46.48	
<i>cemA</i>	33.62	39.57	28.26	33.04	46.43	<i>rpl2</i>	42.16	47.76	48.51	30.22	50.02	
<i>petA</i>	39.98	52.02	36.45	31.46	48.98	<i>rpl14</i>	38.85	53.38	36.84	26.32	48.35	
<i>petB</i>	39.66	48.61	42.13	28.24	44.44	<i>rpl16</i>	39.22	45.59	52.21	19.85	36.58	
<i>petD</i>	37.47	50.93	39.13	22.36	44.70	<i>rpl20</i>	38.60	37.72	48.25	29.82	43.93	
<i>matK</i>	32.81	38.46	31.76	28.21	47.56	<i>rpl22</i>	34.41	39.35	38.71	25.16	44.09	
<i>ndhA</i>	34.07	40.93	39.56	21.70	41.69	<i>rps2</i>	37.29	41.53	40.68	29.66	50.42	
<i>ndhB</i>	37.12	41.58	38.13	31.64	46.31	<i>rps3</i>	35.01	44.96	34.03	26.05	50.92	
<i>ndhC</i>	35.54	47.11	31.40	28.10	57.13	<i>rps4</i>	36.84	46.41	38.28	25.84	47.40	
<i>ndhD</i>	35.24	40.04	37.48	28.21	45.44	<i>rps7</i>	39.53	51.28	42.31	25.00	45.51	
<i>ndhE</i>	33.33	40.20	34.31	25.49	39.90	<i>rps8</i>	33.83	39.85	39.85	21.80	40.65	
<i>ndhF</i>	31.82	35.70	35.43	24.33	42.88	<i>rps11</i>	40.53	48.20	52.52	20.86	51.18	
<i>ndhG</i>	32.77	41.57	33.71	23.03	47.49	<i>rps12</i>	41.67	50.81	46.77	27.42	44.22	
<i>ndhH</i>	37.73	51.02	36.55	25.63	49.33	<i>rps14</i>	42.57	46.53	47.52	33.66	43.49	
<i>ndhI</i>	35.58	42.70	36.52	27.53	48.12	<i>rps18</i>	33.90	34.19	41.88	25.64	38.99	
<i>ndhJ</i>	38.57	49.69	36.48	29.56	46.51	<i>rps19</i>	33.95	41.67	37.04	23.15	51.45	
<i>ndhK</i>	37.05	43.35	42.92	24.89	49.75	<i>rbcL</i>	43.70	58.40	43.49	29.20	47.10	
<i>psaA</i>	42.57	51.80	43.14	32.76	50.28	<i>rpoA</i>	34.67	43.27	35.64	25.09	44.64	
<i>psaB</i>	40.77	48.44	42.99	30.88	48.31	<i>rpoB</i>	38.01	48.60	36.80	28.65	48.32	
<i>psbA</i>	41.15	49.44	43.79	30.23	40.12	<i>rpoC1</i>	37.56	48.83	36.59	27.26	47.95	
<i>psbB</i>	44.01	54.62	45.97	31.43	48.57	<i>rpoC2</i>	36.64	45.22	36.62	28.08	49.11	

<i>psbC</i>	44.08	54.76	46.54	30.95	45.01		<i>accD</i>	37.38	51.75	29.95	30.45	47.53
-------------	-------	-------	-------	-------	-------	--	-------------	-------	-------	-------	-------	-------

2.2 Neutrality plot analysis of chloroplast codons of *R. farrerae*

The neutrality plot analysis (Figure 1) showed that the GC₁₂ values ranged from 33.92% to 50.95%, while the GC₃ values ranged from 19.85% to 53.72%. The correlation coefficient and slope between GC₁₂ and GC₃ were 0.027 and 0.147, respectively, with the slope close to 0. This indicates that the three codon positions in the chloroplast genome of *R. farrerae* have significant differences in composition, and natural selection pressure has a large influence on the codon usage preference.

Table 2 Analysis of the correlation between codon GC Content and ENC values in the chloroplast genome of *Rhododendron farrerae*

Parameter	GC _{all}	GC ₁	GC ₂	GC ₃	ENC	condon
GC _{all}	1					
GC ₁	0.718**	1				
GC ₂	0.711**	0.29*	1			
GC ₃	0.586*	0.113	0.129	1		
ENC	0.057	0.076	-0.245	0.303*	1	
condon	0.028	0.175	-0.155	0.031	0.054	1

Note: **, a highly significant correlation at P<0.01 level; *, a highl significant correlation at P<0.05.

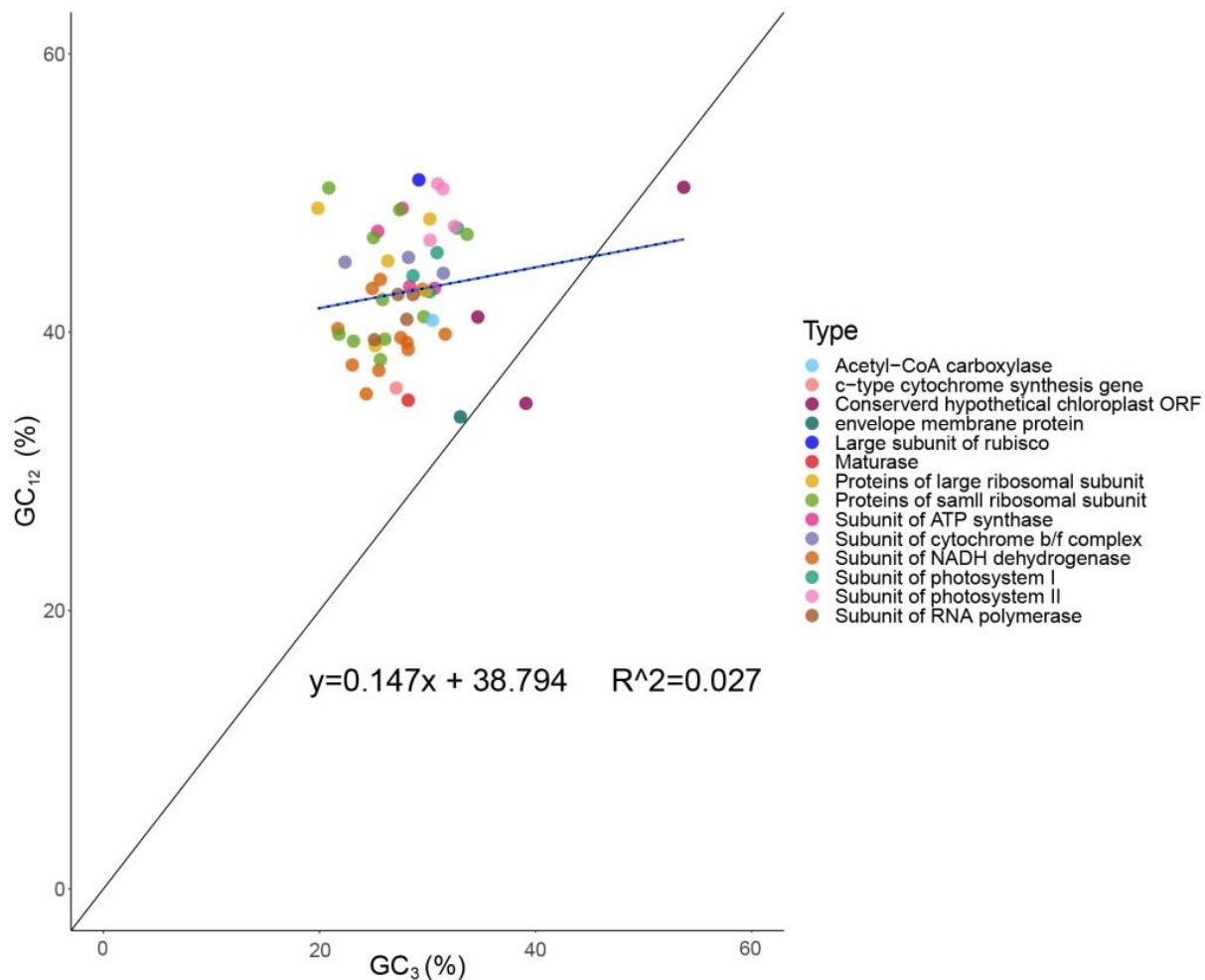


Figure 1 Neutrality plot analysis of the chloroplast genome of *Rhododendron farrerae*

113 GC₁₂ refers to the average GC content of the first and second positions of a codon; GC₃ refers to the GC content
114 of the third position of a codon.

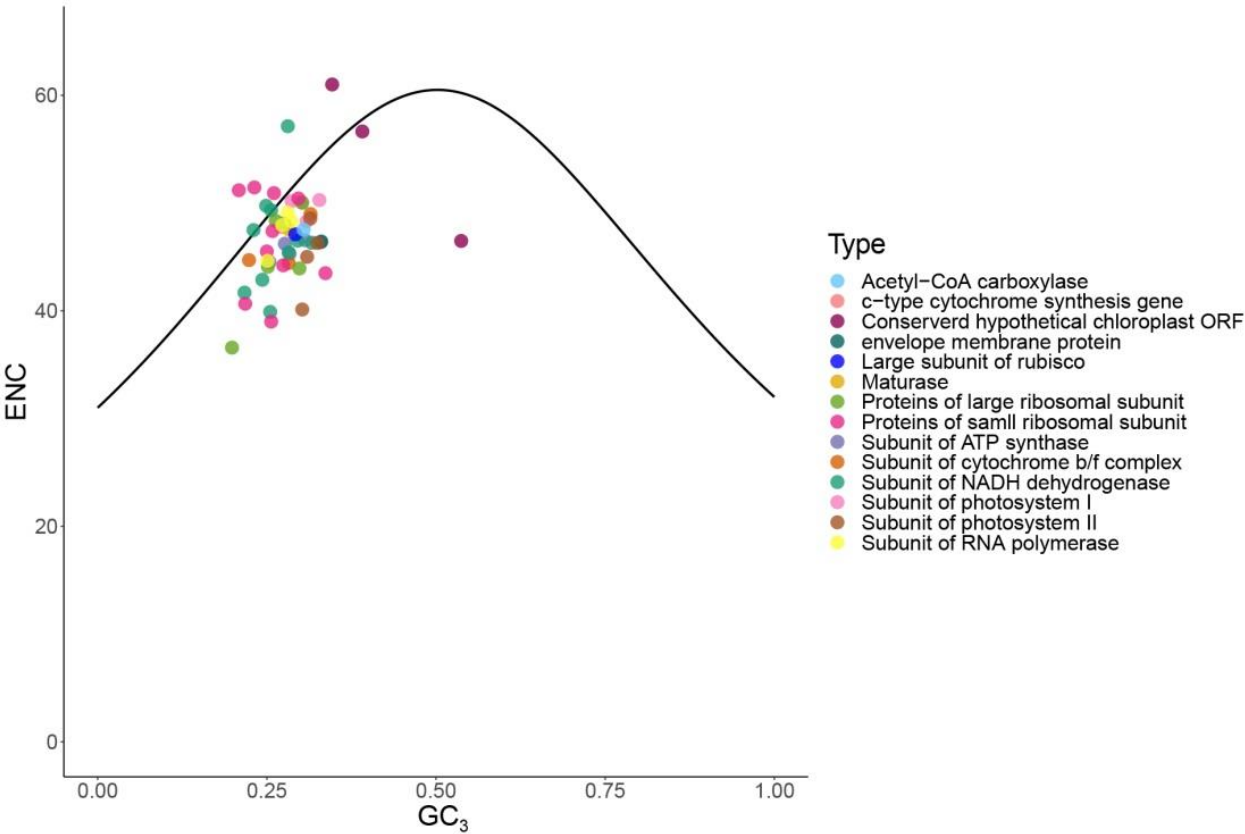
115 **2.3 ENC-plot analysis of chloroplast codons of *R. farrerae***

116 The ENC-plot analysis (Figure 2) showed that most genes cluster around the standard ENC curve, and further
117 quantitative analysis (Table 3) revealed that 25.93% of the genes have ENC values within -0.05 to 0.05 of the
118 expected ENC, while the remaining 74.07% of the genes have ENC values significantly different from the
119 expected, suggesting that the codon usage of the chloroplast genome of *R. farrerae* is mainly influenced by natural
120 selection pressure, and to a certain extent by mutation pressure.

121 Table 3 Frequency distribution of ENC ratio

组限	组数	频率%
-0.15~-0.05	24	44.44
-0.05~0.05	14	25.93
0.05~0.15	4	7.41
0.15+	12	22.22

122



123

124 Figure 2 ENC-plot analysis of the chloroplast genome of *Rhododendron farrerae*

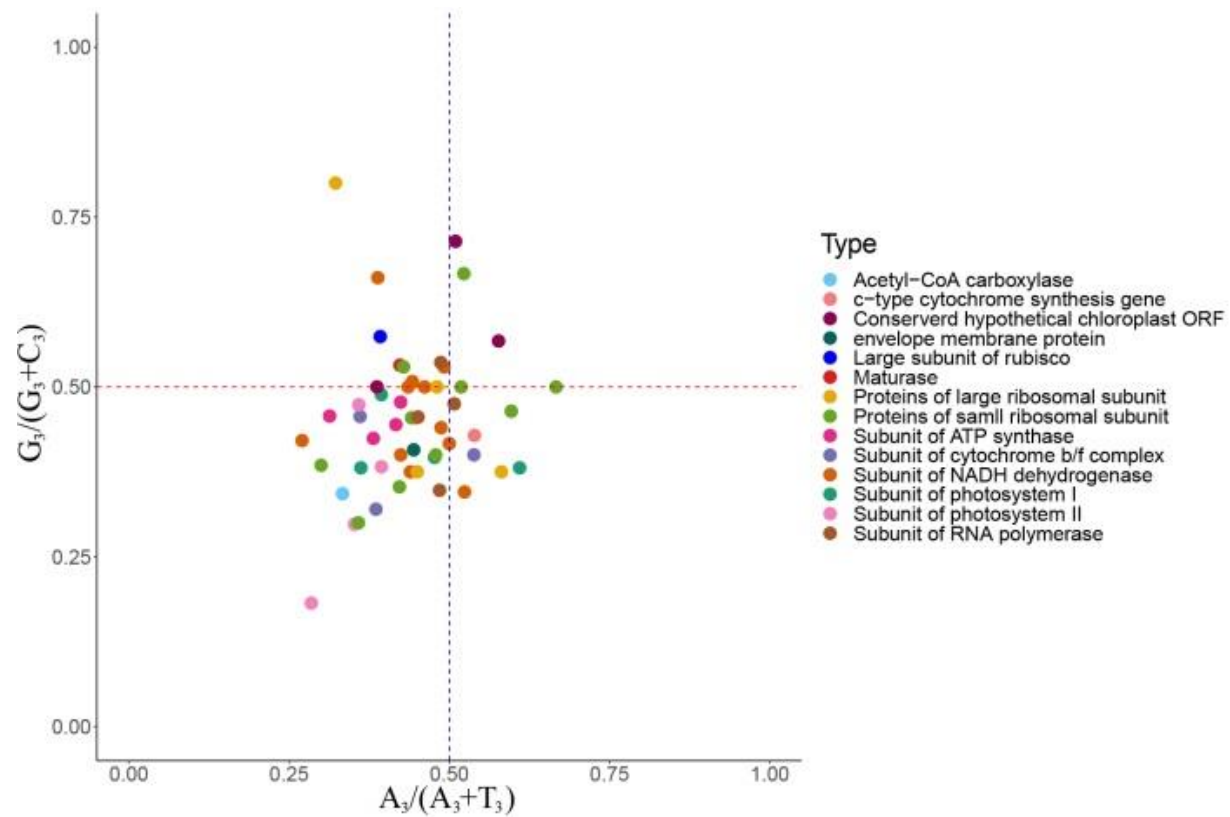
125 The curve is ENC standard curve.

126

127 **2.4 PR2-plot analysis of chloroplast codons of *R. farrerae***

128 The PR2-plot analysis (Figure 3) showed that the distribution of the 3rd codon position bases is uneven, with 28
129 genes having A₃/(A₃+T₃) and G₃/(G₃+C₃) values less than 0.5. This indicates that the usage frequency of the
130 pyrimidine bases C and T at the 3rd codon position is higher than the purine bases A and G, suggesting that the
131 codon usage preference of the chloroplast genome of *R. farrerae* is influenced not only by mutation pressure, but

132 also by other factors.



133
134 Figure 3 PR2-plot analysis of the chloroplast genome of *Rhododendron farrerae*
135

136 **2.5 Optimal codon analysis of chloroplast codons of *R. farrerae***

137 Based on the criteria of $RSCU > 1$ and $\Delta RSCU \geq 0.08$, 16 optimal codons were identified, including UUA, CGA, CGU, CAA, GAA, GGA, GGU, AUU, AAA, GUU, GUA, UAU, ACU, UCU, CCU and GCU, all of
138 which end with A or U, suggesting a preference for A or U-ending codons.
139

140
141 Table 3 Optimal codons in the chloroplast genome of *Rhododendron farrerae*

Amino acid	Codon	Low expressed gene		High expressed gene		$\Delta RSCU$	Amino acid	Codon	Low expressed gene		High expressed gene		$\Delta RSCU$
		RSCU	Number	RSCU	Number				RSCU	Number	RSCU	Number	
Ala	GCA [#]	1.67	15	0.92	15	-0.74	Leu	CUA [#]	1.08	11	0.92	13	-0.17
	GCC ⁺	0.33	3	0.49	8	0.16		CUC	0.39	4	0.00	0	-0.39
	GCG	0.11	1	0.18	3	0.07		CUG	0.79	8	0.21	3	-0.58
	GCU [#]	1.89	17	2.40	39	0.51		CUU [#]	1.28	13	1.27	18	-0.01
Arg	AGA [#]	1.61	19	1.92	24	0.31	Phe	UUA [#]	1.28	13	1.48	21	0.20
	AGG	0.68	8	0.32	4	-0.36		UUG [#]	1.18	12	1.13	16	-0.05
	CGA [#]	0.93	11	2.00	25	1.07		UUC ⁺	0.58	14	0.84	19	0.26
	CGC ⁺	0.00	0	0.16	2	0.16	Pro	UUU [#]	1.42	34	1.16	26	-0.26
	CGG	0.51	6	0.08	1	-0.43		CCA [#]	1.41	13	0.83	6	-0.58
	CGU [#]	1.27	15	1.52	19	0.25		CCC	0.76	7	0.55	4	-0.21
Asn	AAC ⁺	0.41	6	0.68	16	0.27		CCG ⁺	0.32	3	0.41	3	0.09
	AAU [#]	1.59	23	1.32	31	-0.27		CCU [#]	1.51	14	2.21	16	0.69
Asp	GAC	0.42	5	0.40	3	-0.02	Ser	AGC	0.53	4	0.47	5	-0.07

	GAU [#]	1.58	19	1.60	12	0.02		AGU [#]	0.93	7	1.31	14	0.38
	UGC	0.86	3	0.00	0	-0.86		UCA [#]	1.20	9	0.66	7	-0.54
Cys	UGU [#]	1.14	4	1.00	6	-0.14		UCC	0.67	5	0.66	7	-0.01
	CAA ^{*#}	1.71	12	1.90	20	0.19		UCG	0.80	6	0.47	5	-0.33
Gln	CAG	0.29	2	0.10	1	-0.19		UCU ^{*#}	1.87	14	2.44	26	0.57
	GAA ^{*#}	1.50	27	1.68	32	0.18		UAA [#]	1.20	3	1.20	3	0.00
Glu	GAG	0.50	9	0.32	6	-0.18	*Ter	UAG	0.80	2	0.80	2	0.00
	GGA ^{*#}	1.03	8	1.19	20	0.16		UGA	0.00	0	0.00	0	0.00
	GGC	0.65	5	0.42	7	-0.23		ACA	1.00	11	0.74	8	-0.26
Gly	GGG	0.90	7	0.30	5	-0.61		ACC ^{*#}	0.64	7	1.21	13	0.57
	GGU ^{*#}	1.42	11	2.09	35	0.67	Thr	ACG	0.73	8	0.28	3	-0.45
	CAC	0.67	5	0.71	5	0.05		ACU ^{*#}	1.64	18	1.77	19	0.13
His	CAU [#]	1.33	10	1.29	9	-0.05		UAC	0.67	9	0.25	3	-0.42
	AUA [#]	1.02	18	0.79	23	-0.23	Tyr	UAU ^{*#}	1.33	18	1.75	21	0.42
	AUC [*]	0.51	9	0.66	19	0.15		GUA ^{*#}	1.49	16	1.76	18	0.27
Ile	AUU ^{*#}	1.47	26	1.55	45	0.08		GUC	0.37	4	0.10	1	-0.28
	AAA ^{*#}	1.49	52	1.57	33	0.09	Val	GUG	0.74	8	0.29	3	-0.45
Lys	AAG	0.51	18	0.43	9	-0.09		GUU ^{*#}	1.40	15	1.85	19	0.46
Met	AUG [#]	1.90	19	1.00	29	-0.90	Trp	UGG	1.00	13	1.00	14	0.00

* indicates $\Delta\text{RSCU} \geq 0.08$; # indicates $\text{RSCU} > 1$; underlined indicates the optimal codon.

3 Discussion

Pant chloroplast codon usage bias is often used in species evolution analysis. Previous studies have found that codon usage bias is mainly influenced by mutational pressure and selective pressure. Additionally, factors such as gene length and gene expression levels also play a role (Liu et al., 2010; Wang et al., 2021; Parvathy et al., 2022). These factors are commonly used to elucidate the variation in codon usage patterns among species and within genomes. The GC content of chloroplast genomes has been found to shape codon usage bias. Mutational pressure and selective pressure are important factors influencing codon usage bias. The usage of the third codon position (GC_3) is often considered a key indicator of codon usage bias (Duan et al., 2021). Under conditions where codon usage is primarily regulated by selective pressure, GC_3 values typically exhibit a narrow range of variation, and there is no significant correlation between GC_{12} and GC_3 . Analysis of codon usage bias in *Rhododendron* chloroplast genomes revealed that the frequency of GC_3 usage ranges from 19.85% to 39.11%, with only the *ycf68* gene showing a GC_3 usage frequency of 53.72% (Table 1). The correlation between GC_1 and GC_2 was statistically significant, whereas that with GC_3 was not (Table 2), and the correlation coefficient between GC_{12} and GC_3 was 0.027. This study demonstrates significant variation in the composition of the third base of codons in *Rhododendron* chloroplast genomes, with differences from the first and second base compositions, indicating a strong influence of natural selection pressure on codon usage bias in *Rhododendron* chloroplasts. This finding is consistent with most research results (Luo et al., 2024), indicating relative conservatism in chloroplast codon usage bias.

ENC-plot, neutrality plot, and PR2-plot analyses are commonly used to analyze the usage of the third base of codons (Wang et al., 2021). Existing studies have shown that when mutational pressure dominates synonymous codon usage bias, the distribution of G and C versus A and T bases in codons tends to be uniform. In contrast, when selective pressure predominates, this distribution is not uniform. Neutrality plot, ENC-plot, and PR2-plot analyses in this study found that selective pressure is the dominant factor influencing synonymous codon usage

167 bias in *Rhododendron* chloroplast genomes, with additional influence from mutational pressure, consistent with
168 previous studies (Nie et al., 2014). The strength of synonymous codon usage bias can significantly affect gene
169 expression levels, with higher bias often correlating with higher gene expression (Shamsun et al., 2023).
170 Optimizing coding sequences based on codon bias when constructing chloroplast gene expression vectors can
171 enhance the efficiency of gene insertion in chloroplasts. Additionally, researchers can use known codon usage
172 patterns to predict the expression characteristics and functions of newly discovered genes (Khandia et al., 2024).
173 The *Rhododendron* chloroplast genome is highly conservative, and 16 optimal codons have been selected, all
174 ending in A or U, consistent with preferences observed in dicot plants such as *Paulownia* (Luo et al., 2024) and
175 alfalfa (Sun et al., 2015). These optimal codons can provide information for studying the origin and evolution of
176 species and for research on plant breeding and phylogenetic relationships.

177

178 **4 Conclusion**

179 The *Rhododendron* chloroplast genome exhibits strong codon usage bias, primarily influenced by selective
180 pressure and also affected by factors such as mutational pressure. Sixteen optimal codons have been identified,
181 which can be utilized in future studies related to *Rhododendron* breeding and evolution.

182

183 **Author Contributions:**

184 Conceptualization, Jianshuang Shen and Rong Xianlin; Methodology, data, Software, Jianshuang Shen and Lu
185 Huang; Writing—original draft, Jianshuang Shen; Writing—review & editing, Jianshuang Shen and Rong Xianlin.
186 All authors have read and agreed to the published version of the manuscript.

187

188 **Funding:** This work was funded by the National Natural Science Foundation of China [32201608], and the
189 Research Foundation for Advanced Talents of Hangzhou Vocational & Technical College [RCXY202303].

190

191 **References**

- 192 Arthur HR (1955) A new optically active flavanone from the leaves of *Rhododendron farrerae*, Tate. J Chem Soc
193 6467: 3740-3742. <https://doi.org/10.1039/jr9550003740>
- 194 Braukmann TWA, Broe MB, Stefanović S, Freudenstein JV (2017) On the brink: the highly reduced plastomes of
195 nonphotosynthetic Ericaceae. New Phytol 216(1): 254-266. <https://doi.org/10.1111/nph.14681>
- 196 Cai YB, Yang XY (2022) Analysis of codon use preference and its influencing factors
197 in chloroplast genome of *Macadamia integrifolia*. Chinese Journal of Plant Science 40(2): 229-239.
198 <https://doi.org/10.11913/PSJ.2095-0837.2022.20229>
- 199 Duan H, Zhang Q, Wang C, Li F, Tian F, Lu Y, Hu Y, Yang H, Cui G (2021) Analysis of codon usage patterns of
200 the chloroplast genome in *Delphinium grandiflorum* L. reveals a preference for AT-ending codons as a result
201 of major selection constraints. PeerJ 9:e10787. <https://doi.org/10.7717/peerj.10787>
- 202 Kang H-I, Lee HO, Lee IH, Kim IS, Lee S-W, Yang TJ, Shim D (2019) Complete chloroplast genome of *Pinus*
203 *densiflora* Siebold & Zucc. and comparative analysis with five pine trees. Forests 10(7): 600.
204 <https://doi.org/10.3390/f10070600>
- 205 Khandia R, Gurjar P, Kamal MA, Greig NH (2024) Relative synonymous codon usage and codon pair analysis of
206 depression associated genes. Sci Rep 14(1): 3502. <https://doi.org/10.1038/s41598-024-51909-8>
- 207 Li L, Peng J, Wang D, Duan A (2022) Chloroplast genome phylogeny and codon preference of *Docynia*
208 *longiunguis*. Sheng Wu Gong Cheng Xue Bao 38(1):328-342. Chinese. <https://doi.org/10.13345/j.cjb.210298>
- 209 Liu H, He R, Zhang H, Huang Y, Tian M, Zhang J (2010) Analysis of synonymous codon usage in *Zea mays*. Mol
210 Biol Rep 37(2): 677-84. <https://doi.org/10.1007/s11033-009-9521-7>

211 Luo Y, Wang R, Zhao R, Lu X, Yin G, Deng Z (2024) Analysis of synonymous codon usage bias in the
 212 chloroplast genome of *Davidia involucrata*. J Beijing Forestry Univ 46 (03): 8-16.
 213 <http://j.bjfu.edu.cn/cn/article/pdf/preview/10.12171/j.1000-1522.20220199.pdf>

214 Mao L Y, Huang Q W, Long L Y, et al. Codon bias analysis of chloroplast genome of seven *Nymphaea* species. J
 215 Northwest Forestry Univ 2022, 37(2): 98–107. <https://doi.org/10.3969/j.issn.1001-7461.2022.02.13>

216 Ng SC, Corlett R (2003) The ecology of six *Rhododendron* species (Ericaceae) with contrasting local abundance
 217 and distribution patterns in Hong Kong, China. Plant Ecol 164(2): 225-233.
 218 <https://doi.org/10.1023/A:1021292227583>

219 Nie X, Lv S, Zhang Y, Du X, Wang L, Biradar SS, Tan X, Wan F, Weining S (2014) Complete chloroplast genome
 220 sequence of a major invasive species, crofton weed (*Ageratina adenophora*). PLoS One 7(5):e36869.
 221 <https://doi.org/10.1371/journal.pone.0036869>

222 Parvathy ST, Udayasuriyan V, Bhadana V (2022) Codon usage bias. Molecular Biology Reports 49(1): 539-565.
 223 <https://link.springer.com/article/10.1007/s11033-021-06749-4>

224 Qin W, Xu T, Yan J, Ju X (2023) Analysis of synonymous codon usage bias in the chloroplast genome of *Tulipa*
 225 *buhsean*. Jiangsu Agri Sci 51 (22): 41-47. <https://doi.org/10.15889/j.issn.1002-1302.2023.22.006>

226 Shamsun N, Suruchi G, Waqas A, Ravail S (2023) Environmental implications of codon usage bias in *Crocus*
 227 *sativus* and its impact on host pathogen interactions. Rhizosphere 29(12): 100819.
 228 <https://doi.org/10.1016/j.rhisph.2023.100819>

229 Sun J, Yang G, Sogn Z, Su K, Zhao Y (2015). Analysis of codon usage in the chloroplast genome of *Medicago*
 230 *truncatula*. Acta Prataculturae Sinica 24(12): 171-179.
 231 <http://cyxb.magtech.com.cn/CN/10.11686/cyxb2015016>

232 Tang D, Wei F, Cai Z, Wei Y, Khan A, Miao J, Wei K (2021) Analysis of codon usage bias and evolution in the
 233 chloroplast genome of *Mesona chinensis* Benth. Dev Genes Evol 231(1-2): 1-9.
 234 <https://doi.org/10.1007/s00427-020-00670-9>

235 Wang P, Mao Y, Su Y, Wang J (2021) Comparative analysis of transcriptomic data shows the effects of multiple
 236 evolutionary selection processes on codon usage in *Marsipenaues japonicus* and *Marsipenaues*
 237 *pulchricaudatus*. BMC Genomics 22(1): 781. <https://doi.org/10.1186/s12864-021-08106-y>