

## Analysis of codon usage pattern in *Lonicera* × *heckrottii* 'Gold Flame' based on chloroplast genome

Jiaqiang ZHANG\*, Huichun LIU, Wenting XU, Kaiyuan ZHU

Zhejiang Institute of Landscape Plants and Flowers, Hangzhou, 311251, China; [zhangqiang414@126.com](mailto:zhangqiang414@126.com) (\*corresponding author);  
[lhuichun@163.com](mailto:lhuichun@163.com); [xuwenting\\_1128@126.com](mailto:xuwenting_1128@126.com); [kzhu1999@163.com](mailto:kzhu1999@163.com)

### Abstract

Codon usage bias (CUB) was a unique feature of the genome, and revealing chloroplast genome codon usage bias can provide useful information for the evolution of plant species. *Lonicera* × *heckrottii* 'Gold Flame' is one of the members of the *Lonicera* genus with important ornamental value. However, the codon usage bias of chloroplast genome of this genotype has not been investigated. In this study, base compositions and various codon usage indices of 51 coding sequences from *Lonicera* × *heckrottii* 'Gold Flame' chloroplast genome was calculated, by Codon W, DnaSP, CUSP of EMBOSS and SPSS software. A total of 51 CDS of the *Lonicera* × *heckrottii* 'Gold Flame' chloroplast genome was selected to analyse codon usage bias. The results showed that the average GC content of 51 CDS sequences was 39.27%, and the average value of ENC was 48.75%. The chloroplast gene codon usage bias was weak, and preferred A/T ending. The general GC content order was GC1 (47.72%) > GC2 (39.89%) > GC3 (30.19%). Correlation analysis results showed that there was a significant positive correlation between GC content and GC1, GC2 and GC3 content. Combined with neutral plot analysis, ENC-plot and PR2-plot analysis, it is found that chloroplast gene codon usage preference was affected by mutations pressure and natural selection. In addition, the eight optimal codons in chloroplast genome were finally identified, codon bias ending with A/T. The study on the codon usage bias of *Lonicera* × *heckrottii* 'Gold Flame' provides a demonstration for exploring its genetic structure and molecular evolution mechanism, and provides a reference for molecular breeding.

**Keywords:** genetic code; influence factor; optimal codons; synonymous codon usage

### Introduction

The genetic code is an important link between DNA (nucleic acid) and protein. It is a unique attribute of genes and genomes. The unequal use of synonymous codons is known as codon usage bias (CUB). CUB widely existed in different organisms, such as diatoms (Krasovec and Filatov, 2019), bacteria (Dilucca Pavlopoulou and Georgakilas, 2020), animals (Galtier *et al.*, 2020), human (Dhindsa *et al.*, 2020), plants (Zhang *et al.*, 2012; Nie *et al.*, 2014; Wang *et al.*, 2020b; Chakraborty *et al.*, 2017), etc. Quantitative analysis of codon usage bias can provide an important reference for species classification and evolutionary mechanisms.

Chloroplasts are subcellular organelles essential for the photosynthesis and metabolism of green plants. Compared to the nuclear genome, the chloroplast genome, which possesses many characteristics, including simple, small size and highly conserved. It is widely used in researches such as identification, phylogenetic and

Received: 14 Oct 2021. Received in revised form: 16 Feb 2022. Accepted: 21 Feb 2022. Published online: 24 Feb 2022.

From Volume 49, Issue 1, 2021, Notulae Botanicae Horti Agrobotanici Cluj-Napoca journal uses article numbers in place of the traditional method of continuous pagination through the volume. The journal will continue to appear quarterly, as before, with four annual numbers.

adaptive analysis (Dobrogojski, Adamiec and Luciński, 2020). In recent years, with the development of high-throughput sequencing technology for chloroplast genomes, a large number of chloroplast genomes have been sequenced, and the codon usage bias of chloroplast genomes has become possible. Therefore, the chloroplast genome is a special tool for plant systems evolution research (Iriarte *et al.*, 2021). Many plant chloroplast genomes have been sequenced and analysed for codon usage characteristics, including *Helianthus annuus* (Chen *et al.*, 2021), *Oryza* (Chakraborty *et al.*, 2020), *Delphinium grandiflorum* L. (Duan *et al.*, 2021), *Paeonia suffruticosa* (Guo *et al.*, 2020), *Euphorbiaceae* (Wang *et al.*, 2020b), *Hemiptalea davidii* (Liu *et al.*, 2020), *Asteraceae* (Nie *et al.*, 2014), *Solanum* (Zhang *et al.*, 2018), and *Lonicera macranthoides* (Hu *et al.*, 2018).

Codon usage bias was the adaptation mechanism of species to environmental selection pressure. It was affected by many different factors. The possible evolutionary forces based on codon usage patterns have been studied in the genomes of many organisms (Zhang *et al.*, 2007; Iriarte *et al.*, 2021). There are differences in codon usage bias among species. It may be affected by mutation pressure, natural selection and genetic drift in the population. Previous research found that the codon usage bias of some plant's chloroplast genome was mainly driven by natural selection, as exemplified in *Delphinium grandiflorum* L. (Duan *et al.*, 2021), *Hemiptalea davidii* (Liu *et al.*, 2020), *Euphorbiaceae* (Wang *et al.*, 2020b), *Helianthus annuus* (Chen *et al.*, 2021) and *Mesona chinensis* Benth (Tang *et al.*, 2020). *Asteraceae* plastomes chloroplast genome codon usage bias was affected by gene length (Nie *et al.*, 2014), and influenced by translation level. the chloroplast genes of *Triticum aestivum* L. (Zhang *et al.*, 2007). Additionally, in plant species, the codon usage bias in the *Porphyrumbilicalis* and *Nitrariaceae* chloroplast genome were affected by natural selection and mutation pressure (Li *et al.*, 2019; Chi *et al.*, 2020). In addition, codon usage bias was affected by many factors, including GC content, gene length, gene expression level and so on (Wang *et al.*, 2018). Codon usage bias reflects a selection-mutation balance, which is affected by natural selection, mutation pressure and genetic drift in a population (Bulmer, 1991; Eyre-Walker, 1991; Duan *et al.*, 2021). Analysing the dominant role of codon usage bias can provide a suitable strategy for identifying the main driving force. Therefore, selecting which one is the dominant factor is still a focus of research.

Caprifoliaceae contains > 800 species. These are grown as an ornamental plants around the world. Among them, *Lonicera japonica* Thunb., is the most famous, mainly as anti-inflammatory herbs (Wang *et al.*, 2019; Pu *et al.*, 2020). *Lonicera × heckrottii* 'Gold Flame' belongs to *Lonicera* and is a semi-evergreen vine. *Lonicera × heckrottii* 'Gold Flame' (also known as Goldflame honeysuckle) is a cross between *Lonicera sempervirens* and *Lonicera americana*. It is characterized by the habit of winding and climbing, and it blooms continuously during the growing season in most areas (Bruner *et al.*, 2001). Oval blue-green leaves and red stems make it very ornamental. It is known as the most handsome climbing honeysuckle. It is a popular garden plant. As a vine, it can trim into various shapes. Therefore, it is an excellent hedge plant. It is also original material for producing jelly (Bruner *et al.*, 2002). There have been a few studies on cultivation and cutting propagation (Bruner *et al.*, 2001, 2002). We had already sequenced the complete chloroplast genome sequence of *Lonicera × heckrottii* 'Gold Flame' (GenBank accession: MZ522723) before this study. The total length of *Lonicera × heckrottii* 'Gold Flame' is 155,437 bp and includes 125 genes, with a typical quadripartite structure. However, the codon usage bias of chloroplast genome in *Lonicera × heckrottii* 'Gold Flame' has not been reported.

## Materials and Methods

### *Sequence data*

The complete chloroplast genome sequences of *Lonicera × heckrottii* 'Gold Flame' (accession numbers: MZ522723) were downloaded from GenBank (<https://www.ncbi.nlm.nih.gov/>) in FASTA format. A total of 125 genes were obtained. After filtering genes with sequence length <300 bp and the repeated sequences, and deleted the genes of non-coding regions. Finally, a total of 51 qualified *CDS* (coding DNA sequences) were obtained for subsequent codon usage bias analysis (shown in Table S1).

### *Analysis of codon composition*

A number of the codon usage indices were calculated using Codon W 1.4.2, DnaSP 6.0 software and the online CUSP program of EMBOSS (<http://emboss.toulouse.inra.fr/cgi-bin/emboss/cusp>), including Effective number of codons (*ENC*), Relative synonymous codon usage (*RSCU*), Codon bias index (*CBI*), The codon adaptation index (*CAI*), Frequency of optimal codon (*Fop*), The frequency of codon usage (*FCU*), at the 3rd codon position of the four-codon bases (*A3*, *T3*, *C3* and *G3*) and *GC3* (the proportion of GC nucleotides at the 3rd position of synonymous codons). The GC content at the 1st, 2nd, 3rd of codons (*GC1*, *GC2*, *GC3*) and the overall GC content of genes (*GC*) and so on.

### *Neutrality plot*

GC12 represents the average value of GC contents GC1 and GC2, drawing scatter plot with GC3 as the abscissa and GC12 as the ordinate, each point is considered to be an independent gene, in neutrality plots. If GC12 and GC3 were significantly correlated, and the codon usage patterns was mainly affected by mutation pressure. Conversely, if there is no correlation between GC12 and GC3, indicating the codon usage patterns as mainly affected by natural selection (Sueoka, 1988; Duan *et al.*, 2021).

### *ENC-plot*

ENC-plot mapping analysis was an effective codon number plotting analysis, which can reflect the relationship between codon usage bias and base composition. Drawing a two-dimensional scatter chart with the GC3 values as the abscissa and ENC values as the ordinate. And add a standard curve in the scatter chart:  $ENC = 2 + GC3 + 29/[GC3^2 + (1 - GC3)^2]$ . The corresponding point close to the standard curve means that the codon usage bias might be only driven by mutations pressure, on the contrary, besides mutations, there are other factors, especially natural selection (Wright, 1990; Gupta, Bhattacharyya and Ghosh, 2004).

### *PR2-plot*

PR2-plot analysis was called parity preference analysis, it explores whether the A, T, C, and G base mutations in the third base of each gene codon was balanced.  $A3/(A3+T3)$  and  $G3/(G3+C3)$  were drawing a scatter diagram as the ordinate and as the abscissa in PR2-plot analysis, respectively. The vector from the center represents the range and direction of PR2 bias. If  $A=T$  and  $C=G$ , there was no bias by mutation and selection affected (Sueoka, 1999).

### *Optimal codons*

Taking ENC as the preferred standard, and rank ENC. A 10% of the total CDS with extremely high/low ENC value were regarded as two datasets (high/low expression). Calculate the RSCU value of each codon in the two datasets, and record it as  $\Delta RSCU$ . The codons with  $RSCU > 1.00$  were defined as high-frequency codons. The codons with  $\Delta RSCU > 0.08$  and  $RSCU > 1.00$  were defined as optimal codon (Liu *et al.*, 2020; Wang *et al.*, 2020a).

### Correspondence analysis

Correspondence analysis (COA) is a type of the multivariate statistical methods, that allows used to explore the codon usage bias among genes, and also reflect the distribution of genes in a multi-dimensional space (Romero *et al.*, 2003; Zhang *et al.*, 2012), it can as vectors consisting of columns and rows. COA was performed on the RSCU values. The data is plotted in 59-dimensional space on different axes, and the RSCU value of one sense codon corresponds to each dimension (Nair *et al.*, 2012; Nie *et al.*, 2014; Wang *et al.*, 2020b).

### Statistical analysis

SPSS 19.0 and Microsoft Excel software were used for statistical analyses. Correlation analysis based on Pearson's rank. The charts were depicted in Microsoft EXCEL 2010, including Neutrality plot, ENC-plot and PR2-plot. The box plot of GC contents was depicted using the online program of imageGP (<http://www.ehbio.com/ImageGP/#opennewwindow>).

## Results

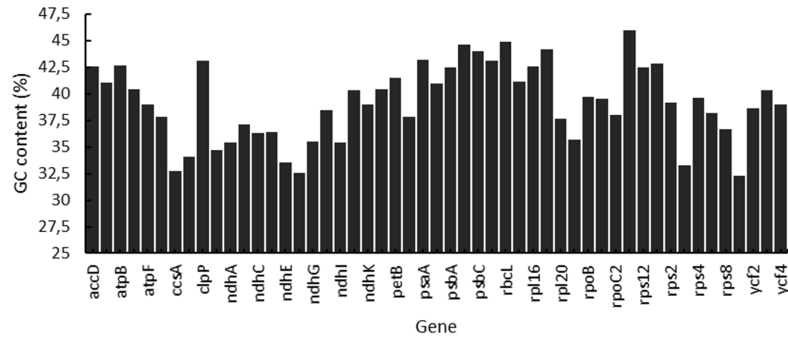
### Base composition in *Lonicera × heckrottii* 'Gold Flame' chloroplast genome

The assembled chloroplast genome of *Lonicera × heckrottii* 'Gold Flame' was 155,437 bp in length. After removing genes less than 300bp in length and repetitive sequences, a total of 51 CDS sequences of *Lonicera × heckrottii* 'Gold Flame' selected were analysed (Table 1). The average GC content of chloroplast genome was 39.27% (Figure 1A; Table S2). The distribution of the four deoxynucleotides was unequal, thymine (T), adenine (A), cytosine (C) and guanine (G) nucleotide of the 51 CDS sequences were recorded 30.82%, 29.91%, 21.32% and 17.94%, respectively. We have summarized the CDS numbers of different GC Content level, the GC content of 51 CDS sequences were contained between 32.00-47.00%, and the GC content was divided into three groups, the 39-40% groups contained the most (seven genes), followed by the 42-43% groups with 6 genes (Figure 1B). We also calculated the GC contents of the first, second, and third sites of codons (GC1, GC2 and GC3) (Figure 1C), the GC1 content of most genes was higher than the GC2 and GC3 content. The order of GC content was GC1>GC2>GC3. The average content of GC2 and the range between the upper and lower quartiles was the highest, followed by GC1, and GC3 was the lowest.

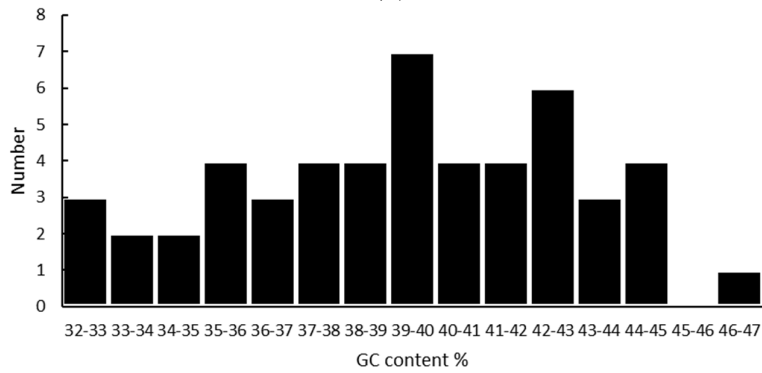
**Table 1.** GC content and ENC value of 51 genes from the chloroplast genome of *Lonicera × heckrottii* 'Gold Flame'

Gene	GC	GC1	GC2	GC3	ENC	Gene	GC	GC1	GC2	GC3	ENC
accD	42.70	55.85	38.06	34.19	47.11	psbA	42.56	50.56	43.50	33.62	42.73
atpA	41.14	55.71	39.96	27.76	50.18	psbB	44.73	55.40	46.17	32.61	51.11
atpB	42.75	55.31	41.88	31.06	48.59	psbC	44.09	52.74	46.20	33.33	46.40
atpE	40.55	50.75	41.79	29.10	48.35	psbD	43.22	52.54	43.22	33.90	45.39
atpF	39.10	49.19	32.97	35.14	45.41	rbcL	44.96	58.19	43.28	33.40	51.07
atpI	37.90	48.79	38.31	26.61	48.19	rpl14	41.19	55.28	38.21	30.08	45.48
ccsA	32.82	33.85	38.51	26.09	50.03	rpl16	42.65	50.74	52.21	25.00	35.55
cemA	34.20	41.74	27.39	33.48	57.50	rpl2	44.24	51.10	48.90	32.72	55.36
clpP	43.16	54.84	37.79	36.87	56.61	rpl20	37.73	38.76	44.96	29.46	46.92
matK	34.79	42.18	32.87	29.31	49.05	rpl22	35.82	43.48	37.27	26.71	52.19
ndhA	35.53	44.78	37.91	23.90	45.16	rpoB	39.78	50.80	39.06	29.48	48.16
ndhB	37.18	41.88	38.36	31.31	47.37	rpoC1	39.58	50.88	38.50	29.35	50.59
ndhC	36.36	45.45	33.06	30.58	57.74	rpoC2	38.09	46.36	37.20	30.71	50.96
ndhD	36.46	42.51	37.72	29.14	46.56	rps11	46.04	52.52	58.27	27.34	52.90
ndhE	33.66	41.18	33.33	26.47	47.34	rps12	42.59	52.38	47.62	27.78	48.71
ndhF	32.67	37.14	36.34	24.54	44.44	rps14	42.90	42.57	48.51	37.62	40.72
ndhG	35.59	45.76	33.90	27.12	48.49	rps2	39.24	43.46	43.04	31.22	49.80
ndhH	38.58	50.76	35.79	29.19	51.02	rps3	33.33	44.93	30.84	24.23	46.99

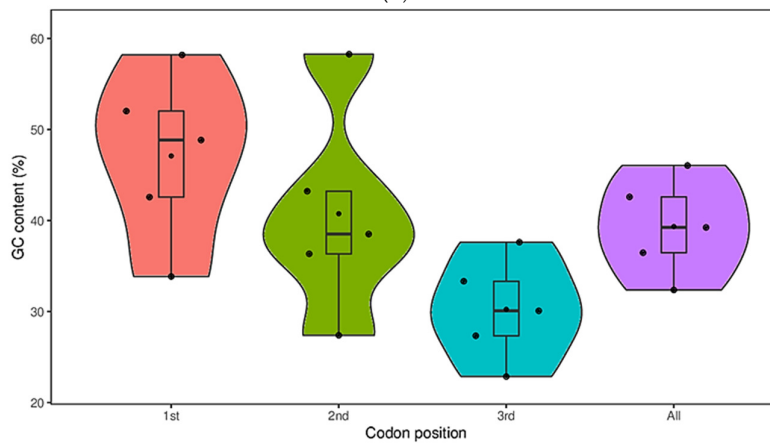
ndhI	35.54	41.57	35.54	29.52	47.03	rps4	39.73	52.02	36.36	30.81	53.11
ndhJ	40.46	50.94	37.74	32.70	51.09	rps7	38.25	48.08	42.95	23.72	44.25
ndhK	39.09	44.69	43.81	28.76	53.23	rps8	36.74	41.61	41.61	27.01	40.21
petA	40.50	52.02	35.51	33.96	57.28	ycf1	32.37	37.59	31.68	27.83	48.50
petB	41.56	48.90	42.73	33.04	47.82	ycf2	38.71	44.68	34.69	36.76	52.12
petD	37.90	50.29	40.57	22.86	38.00	ycf3	40.43	47.93	39.05	34.32	50.48
psaA	43.28	52.73	43.41	33.69	51.91	ycf4	39.10	41.62	42.70	32.97	53.71
psaB	41.09	48.84	42.99	31.43	47.30	Average	39.27	47.72	39.89	30.19	48.75



(A)



(B)



(C)

**Figure 1.** Base composition of *Lonicera × heckrottii* ‘Gold Flame’ chloroplast genome (A) GC content of different protein coding genes; (B) Distribution of different GC content levels; (C) Box plot of GC contents variation in different codon positions (1st, 2nd, 3rd and all (overall chloroplast genome)).

The ENC value of 51 CDS sequences ranged from 35.55 to 57.74%, and the average ENC was 48.75%. Which shows that the codon usage bias of *Lonicera × heckrottii* ‘Gold Flame’ chloroplast genes was weak.

Correlation analysis results show that GC was extreme positive correlation with GC1, GC2 and GC3, GC3 were extreme positive correlation with ENC ( $r=0.3575$ ,  $p<0.01$ ), and GC1 were significantly correlation with GC2, GC3 ( $r=0.2759$ ,  $p<0.05$ ) (Table 2). This indicates that GC content has an impact on codon usage bias.

**Table 2.** Correlation coefficients between codon indices of chloroplast genes in *Lonicera × heckrottii* ‘Gold Flame’

Indices	GC	GC1	GC2	GC3	ENC
GC	1				
GC1	0.8175**	1			
GC2	0.7305**	0.3481*	1		
GC3	0.5219**	0.3018*	0.0387	1	
ENC	0.0466	0.0953	-0.2477	0.3752**	1

Notes: \*Significant difference at  $p < 0.05$ . \*\*Significant difference at  $p < 0.01$ .

*The codon usage pattern of Lonicera × heckrottii ‘Gold Flame’ chloroplast genome*

The amino acids number of 51 genes ranged from 63 to 822, with an average of 345. The RSCU values of different codons were varied greatly. We identified a total of 29 codons with an RSCU value  $> 1$  (Table 3). The codon TAA encoding Leu shows the highest RSCU value of 1.83, and the RSCU value of the termination codon TAC is only 0.32. Among 29 codons, the number of codons ending in T, A and G is 16, 12 and 1, respectively. This indicates that in the codon usage bias of *Lonicera × heckrottii* ‘Gold Flame’ chloroplast genome tended to end with A/T.

**Table 3.** RSCU analysis of protein coding region in *Lonicera × heckrottii* ‘Gold Flame’

Amino acid	Codon	Number	RSCU	Amino acid	Codon	Number	RSCU
Ala (A)	GCT	524	1.67	Asn (N)	AAC	218	0.46
	GCC	205	0.74	Pro (P)	CCT	372	1.36
	GCA	333	1.14		CCC	184	0.76
	GCG	135	0.45		CCA	251	1.15
Cys (C)	TGT	159	1.45		CCG	145	0.58
	TGC	63	0.43	Gln (Q)	CAA	578	1.56
Asp (D)	GAT	744	1.53		CAG	188	0.44
	GAC	209	0.47	Arg (R)	CGT	274	1.65
Glu (E)	GAA	798	1.47		CGC	96	0.49
	GAG	348	0.50		CGA	288	1.36
Phe (F)	TTT	782	1.39		CGG	88	0.40
	TTC	411	0.61		AGA	370	1.46
Gly (G)	GGT	494	1.30		AGG	146	0.65
	GGC	176	0.47	Ser (S)	TCT	480	1.69
	GGA	579	1.51		TCC	250	0.99
	GGG	283	0.72		TCA	294	1.16
His (H)	CAT	385	1.49		TCG	133	0.45
	CAC	118	0.47		AGT	327	1.36
Ile (I)	ATT	822	1.46		AGC	95	0.35

	ATC	344	0.57	Thr (T)	ACT	410	1.54
	ATA	560	0.97		ACC	201	0.79
Lys (K)	AAA	806	1.46		ACA	324	1.16
	AAG	295	0.50		ACG	122	0.52
Leu (L)	TTA	675	1.83	Val (V)	GTT	441	1.49
	TTG	470	1.28		GTC	141	0.44
	CTT	481	1.35		GTA	430	1.51
	CTC	161	0.42		GTG	171	0.56
	CTA	309	0.80	Tyr (Y)	TAT	653	1.56
	CTG	130	0.32		TAC	161	0.36
Asn (N)	AAT	730	1.50				

### Optimal codons

In this study (Table 4), 28 high-expression codons were identified in the *Lonicera × heckrottii* ‘Gold Flame’ chloroplast genome (RSCU>1 in high expression groups). Among them, 15 codons end with T, 10 codons end with A, 2 codons end with G and 1 codon end with C. 8 optimal codons were selected (RSCU>1 and  $\Delta$ RSCU>0.08), including TCT, AGT, CCA, CAA, GAT and GGA codons preferred an ending with A/T, only ACG and CGC ending with G/C. It further confirmed that the codons ending with G/C were lacking in the chloroplast genome of *Lonicera × heckrottii* ‘Gold Flame’.

**Table 4.** The codons statistics with high and low expression genes of the *Lonicera × heckrottii* ‘Gold Flame’

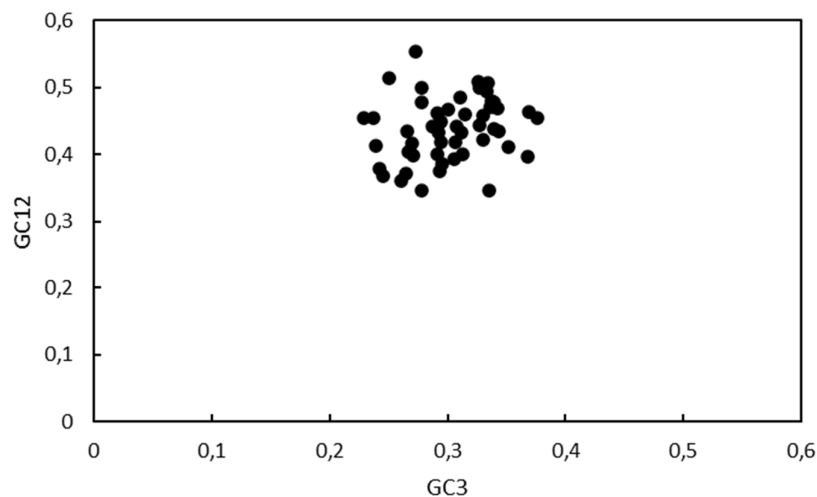
Amino acid	Codon	RSCU		$\Delta$ RSCU	Amino acid	Codon	RSCU		$\Delta$ RSCU
		High expressed gene	Low expressed gene				High expressed gene	Low expressed gene	
Ala (A)	GCT	1.08	1.56	-0.47	Asn (N)	AAC	0.55	0.56	-0.01
	GCC	0.86	0.27	0.60	Pro (P)	CCT	1.30	1.69	-0.38
	GCA	1.58	1.67	-0.08		CCC	0.91	0.81	0.10
	GCG	0.47	0.51	-0.04		<b>CCA</b>	<b>1.09</b>	<b>0.79</b>	<b>0.30</b>
Cys (C)	TGT	1.10	2.00	-0.90		CCG	0.70	0.71	-0.01
	TGC	0.50	0.00	0.50	Gln (Q)	<b>CAA</b>	<b>1.57</b>	<b>1.43</b>	<b>0.14</b>
Asp (D)	<b>GAT</b>	<b>1.44</b>	<b>0.85</b>	<b>0.59</b>		CAG	0.43	0.57	-0.14
	GAC	0.57	1.15	-0.59	Arg (R)	CGT	1.73	2.10	-0.37
Glu (E)	GAA	1.47	1.68	-0.21		<b>CGC</b>	<b>1.02</b>	<b>0.30</b>	<b>0.72</b>
	GAG	0.53	0.32	0.21		CGA	0.99	1.25	-0.27
Phe (F)	TTT	1.44	1.37	0.07		CGG	0.30	0.22	0.09
	TTC	0.56	0.63	-0.07		AGA	0.99	1.61	-0.62
Gly (G)	GGT	1.30	1.74	-0.44		AGG	0.96	0.52	0.44
	GGC	0.60	0.55	0.06	Ser (S)	<b>TCT</b>	<b>1.87</b>	<b>1.56</b>	<b>0.31</b>
	<b>GGA</b>	<b>1.50</b>	<b>1.22</b>	<b>0.29</b>		TCC	0.81	1.43	-0.63
	GGG	0.59	0.50	0.10		TCA	0.91	0.98	-0.07
His (H)	CAT	1.17	1.46	-0.29		TCG	0.43	0.79	-0.36
	CAC	0.43	0.54	-0.11		<b>AGT</b>	<b>1.53</b>	<b>0.93</b>	<b>0.60</b>
Ile (I)	ATT	1.38	1.43	-0.06		AGC	0.46	0.31	0.15
	ATC	0.59	0.51	0.08	Thr (T)	ACT	0.92	1.04	-0.12

	ATA	1.03	1.05	-0.02		ACC	0.72	1.01	-0.29
Lys (K)	AAA	1.32	1.34	-0.02		ACA	1.06	1.68	-0.61
	AAG	0.68	0.26	0.42		<b>ACG</b>	<b>1.30</b>	<b>0.28</b>	<b>1.02</b>
	TTA	1.05	1.37	-0.32	Val (V)	GTT	1.70	2.27	-0.57
Leu (L)	TTG	1.43	1.37	0.06		GTC	0.43	0.11	0.32
	CTT	1.77	2.11	-0.33		GTA	1.36	1.46	-0.09
	CTC	0.63	0.00	0.63		GTG	0.52	0.17	0.35
	CTA	0.62	1.02	-0.40	Tyr (Y)	TAT	1.44	1.83	-0.39
	CTG	0.49	0.12	0.37		TAC	0.56	0.17	0.39
	Asn (N)	AAT	1.05	1.44	-0.39				

Notes: The optimal codons are in bold.

#### Neutrality plot analysis

The relationship between GC3 and GC12 was analysed by the neutrality plot (Figure 2). The range of GC3 was 0.35-0.55, and GC12 was 0.23-0.38. The average values of GC12 and GC3 were 0.44 and 0.30, respectively. The correlation analysis of GC3 and GC12 shows that the correlation coefficient was 0.2067 ( $\alpha=0.05$ ,  $r=0.2759$ ), there were not significant. It shows that the chloroplast genome of *Lonicera × heckrottii* ‘Gold Flame’ was mainly affected by natural selection.

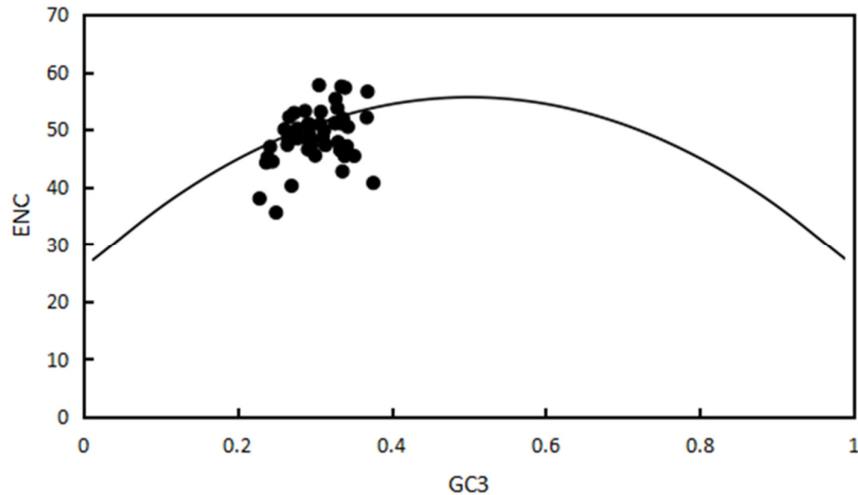


**Figure 2.** Neutrality plot analysis for chloroplast genes in *Lonicera × heckrottii* ‘Gold Flame’

#### ENC-plot analysis

The ENC plot was used to analyse the codon usage patterns of the 51 CDS in *Lonicera × heckrottii* ‘Gold Flame’ (Figure 3). The values of ENC ranged from 35.55 to 55.74, and the average value of ENC was 48.75. The contents of GC3 ranged from 0.23 to 0.38, and the average content of GC3 was 0.30. The results of correlation analysis showed that the correlation between ENC and GC3 showed an extreme positive correlation ( $r=0.3575$ ,  $P<0.01$ ). Most of the genes were located near the standard curve, 27 genes were located below the standard curve. This indicates that mutation pressure might play an important factor in determining codon usage patterns.

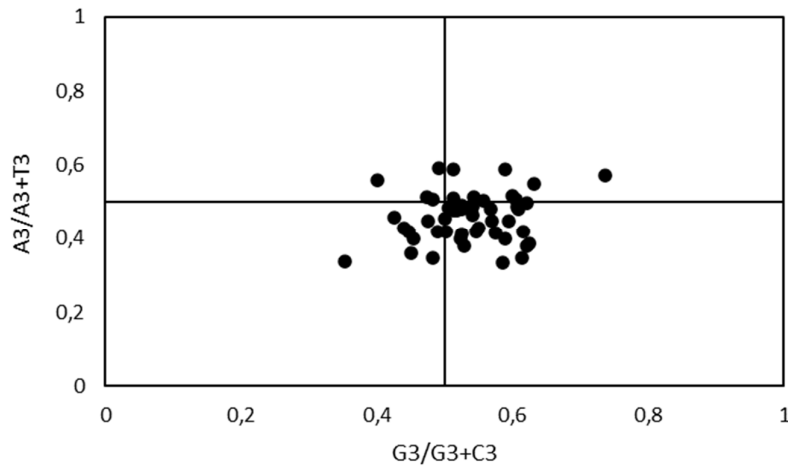




**Figure 3.** ENC-plot analysis for chloroplast genes in *Lonicera × heckrottii* ‘Gold Flame’

*PR2-plot analysis*

In the present study, Using PR2-plot analysis, points of PR2-plot fell among 0.33 to 0.59 on  $A3/(A3+T3)$ , and 0.35 to 0.74  $G3/(G3+C3)$  (Figure 4). The four regions in the PR2-plane were not evenly distributed. Most points fall on the lower right of the four quadrants centered on 0.5. Therefore, indicating that in terms of the use frequency of base,  $T>A$  and  $G>C$ . This further confirms that the codon usage pattern of *Lonicera × heckrottii* ‘Gold Flame’ chloroplast genome was not only affected by the mutation pressure, but also by other factors, such as selection.



**Figure 4.** Analysis of PR2 bias plot for chloroplast genes in *Lonicera × heckrottii* ‘Gold Flame’

*Correspondence analysis (COA)*

In this study, our correspondence analysis reveals the main trends in codon usage of chloroplast genes in *Lonicera × heckrottii* ‘Gold Flame’ based on the variation of RSCU values. Axis 1 accounted for 11.60% and Axis 2 accounted for 10.30% of the overall variation, whereas the next two axes accounted for 8.40% and 7.70%, respectively. It is confirming that Axis 1 and Axis 2 represented the main factor in explaining the data. The correlation analysis suggested that Axis 1 and Axis 2 have no significant correlation with GC, GC1, GC2, GC3, ENC ( $r=0.2759$ ,  $P<0.05$ ) (Table 5). Axis3 had significant correlation with GC, GC1. Axis4 had negative significant correlation with ENC, and Axis4 had negative significant correlation with GC3 ( $r=0.3575$ ,

$P < 0.01$ ). This result indicated that there is no obvious single trend in codon usage bias, and the reason may be several factors. In other words, this also implies that the formation process of *Lonicera × heckrottii* ‘Gold Flame’ chloroplast codon usage pattern was more complicated.

**Table 5.** The correlation coefficients between the axis and indices of chloroplast genes in *Lonicera × heckrottii* ‘Gold Flame’

Axis	GC	GC1	GC2	GC3	ENC
Axis1	0.0063	-0.0274	0.1356	-0.1505	-0.1844
Axis2	0.0387	0.2239	-0.1755	0.0408	0.1136
Axis3	0.2869*	0.2985*	0.1040	0.2154	0.0595
Axis4	-0.0701	0.0246	0.2672	-0.6601**	-0.3448*

## Discussion

### *Codon usage patterns in chloroplast genomes of Lonicera × heckrottii* ‘Gold Flame’

Codon usage bias was formed in the long-term evolutionary process of organisms (Krasovec and Filatov, 2019). Among various species and genes has different codon usage bias (Karumathil *et al.*, 2018). Many quantitative indicators reflected the pattern of codon usage, including GC, GC3 and ENC. ENC reflects the degree of codon deviation from random selection. So ENC is often used as an important indicator of codon usage bias. The ENC value ranges from 20 to 61. An ENC value of 35 was usually used as the criterion to distinguish between strong and weak codon usage bias (Wright, 1990). The smaller the ENC value means that the codon usage was more biased. In the present study, the average ENC value of 51 CDS sequences was 48.75%, which shows that *Lonicera × heckrottii* ‘Gold Flame’ chloroplast genome has a weak codon usage bias. Which was consistent with *Hemiptalea davidii* (Liu *et al.*, 2020), *Helianthus annuus* (Chen *et al.*, 2021), *Oryza* (Chakraborty *et al.*, 2020) and *Delphinium grandiflorum* L. (Duan *et al.*, 2021).

The RSCU value reflects the codon usage pattern of different genes. If the RSCU value was equal to 1.0, this codon was chosen randomly. The codon has negative codon usage bias while the RSCU value was less than 1.0. The codons with RSCU value greater than 1.0 have positive codon usage bias (Karumathil *et al.*, 2018). In the present study, we have identified 28 high-frequency codons in *Lonicera × heckrottii* ‘Gold Flame’ chloroplast genome with RSCU value greater than 1.0. Among these codons, 25 were A/T-ending codons (T-ending: fifteen; A-ending: ten), two with G-ending and only one with G-ending. In addition, 8 optimal codons were identified in our investigation. The majority of optimal codons were A/T ending. Which were similar to other plant species, such as *Hemiptalea davidii* (Liu *et al.*, 2020), *Helianthus annuus* (Chen *et al.*, 2021), the Poaceae family (Zhang *et al.*, 2012), the Asteraceae family (Nie *et al.*, 2014), *Populus alba* (Zhou *et al.*, 2008) and *Delphinium grandiflorum* L. (Duan *et al.*, 2021). However, high frequency codons are not always conservative in various species. For instance, TCC (Ser) was not in dicots, but was a high-frequency codon commonly used in monocots (Liu *et al.*, 2020). Comparably, in the nuclear genome, the opposite pattern of codon ending bases might reflect differences between dicot and monocot species. The codons in dicot were biased towards A/T bases, and were biased towards G/C bases in monocot (Zhang *et al.*, 2012; Mazumdar *et al.*, 2017). The codon characteristics showed both species and genomes specificity (Kawabe & Miyashita 2003; Chakraborty *et al.*, 2020; Liu *et al.*, 2020). In the present study, the chloroplast genomes of dicot, monocot and woody plants were not systematically analyzed, which need to be further verified in the following work.

### *Effect factors on codon usage in chloroplast genomes of Lonicera × heckrottii* ‘Gold Flame’

Codon usage bias is an important evolutionary feature in genome, it has been widely documented in various organisms (Wang *et al.*, 2020b). Previous studies have shown that various biological factors were involved in the synonymous of codon usage patterns, such as ribosomal frameshifting, gene length, tRNA

abundance, methylation, gene expression level, GC composition and mutation bias (Zhang *et al.*, 2012; Mazumdar *et al.*, 2017; Bergman and Tuller, 2020). However, the pressure of mutation and the natural selection were the main factors that affect the codon bias of many organisms in *Solanum* (Zhang *et al.*, 2018). In the chloroplast genome of *Camellia sinensis* var. (Yengkhom, Uddin and Chakraborty, 2019), wetland plants (Deng *et al.*, 2020), *Hippophae* (Wang *et al.*, 2020b), *Pisum* L. (Bhattacharyya *et al.*, 2019) and Nitrariaceae (Chi *et al.*, 2020), natural selection plays an important role that affected the codon usage patterns, but the codon usage patterns of *Coffea arabica* (Nair *et al.*, 2012) and *Populus alba* (Zhou *et al.*, 2008) chloroplast genome were dominated by mutational pressure. Neutrality plot analysis, ENC-plot and PR2-plot analysis, indicated that the codon usage patterns of *Lonicera × heckrottii* ‘Gold Flame’ chloroplast genome was formed under the effect of mutational pressure and natural selection. This was similarly observed in *Mesona chinensis* Benth (Tang *et al.*, 2020), Poaceae family (Zhang *et al.*, 2012), Biebersteiniaceae and Nitrariaceae (Chi *et al.*, 2020). The results of the COA analysis further suggest that there was no obvious single trend in the codon usage bias of *Lonicera × heckrottii* ‘Gold Flame’, probably due to several factors. Therefore, it can be seen that the plant codon usage patterns was affected by many factors, and merits further study.

#### *The optimal codons of the chloroplast genomes in Lonicera × heckrottii ‘Gold Flame’*

The optimal codon can provide an important representative of codon usage patterns. The 8 optimal codons in *Lonicera × heckrottii* ‘Gold Flame’ chloroplast genome was finally identified, and most preferred A/T ending. The number, type and distribution of optimal codons together with chromosomes were also considered to be related to gene expression. Codon usage patterns affect the expression of genes. The higher the gene expression level, the stronger the codon preference (Hershberg and Petrov, 2008; Tang *et al.*, 2020). Codon preference affects the expression level of exogenous genes by regulating the accuracy and efficiency of gene translation (Zhang *et al.*, 2012). Codon usage patterns and optimal codons were identified in *Lonicera × heckrottii* ‘Gold Flame’ chloroplast genome. This may be important for exploring species evolution and increasing the expression of exogenous genes in host cells.

## **Conclusions**

This study systematically analysed the codon usage patterns in the chloroplast genome of *Lonicera × heckrottii* ‘Gold Flame’, and explored the influencing factors of codon usage bias. The codon usage bias of *Lonicera × heckrottii* ‘Gold Flame’ chloroplast genome was weak, and it tends to use A/T ending. The codon usage pattern is mainly affected by mutation pressure and natural selection. This study reveals for the first time the codon usage patterns and influencing factors in the chloroplast genome of *Lonicera × heckrottii* ‘Gold Flame’, and provides theoretical support for further research on the optimization of codons and the phylogenetic development mechanism.

## **Authors’ Contributions**

Conceptualization: JQZ; Data curation and Writing-Original draft: JQZ; Investigation: HCL and JQZ; Methodology: HCL and WTX; Writing-review and editing: JQZ and KYZ. All authors read and approved the final manuscript.

## **Ethical approval** (for researches involving animals or humans)

Not applicable.

## Acknowledgements

This work was supported by the Demonstration and Promotion Project of New Scientific and Technological Achievements from Zhejiang Academy of Agricultural Sciences in 2021 (No. tg2021009), the Program of Cooperation between Zhejiang Academy of Agricultural Sciences and Pujiang County (No. PJ202104) and the Program of Young Talent from Zhejiang Academy of Agricultural Sciences in 2020 (No. 2020R25R08E03).

## Conflict of Interests

The authors declare that there are no conflicts of interest related to this article.

## References

- Bergman S, Tuller T (2020). Widespread non-modular overlapping codes in the coding regions. *Physical Biology* 17(3):031002. <http://doi.org/10.1088/1478-3975/ab7083>
- Bhattacharyya D, Uddin A, Das S, Chakraborty S (2019). Mutation pressure and natural selection on codon usage in chloroplast genes of two species in *Pisum* L. (Fabaceae: Faboideae). *Mitochondrial DNA* 30(4):664-673. <http://doi.org/10.1080/24701394.2019.1616701>
- Bruner LL, Keever GJ, Kessler JR, Gilliam CH (2002). Atrimmec suppresses shoot length and promotes branching of *Lonicera × heckrottii* 'goldflame' (goldflame honeysuckle). *Journal of Environmental Horticulture* 20(2):73-76. <http://doi.org/10.24266/0738-2898-20.2.73>
- Bruner LL, Keever GJ, Kessler JR, Gilliam CH (2001). Shoot suppression of *Lonicera × heckrottii* 'Goldflame' (Goldflame honeysuckle) using growth retardants. *Journal of Environmental Horticulture* 19(4):203-206. <http://doi.org/10.24266/0738-2898-19.4.203>
- Bulmer M (1991). The selection-mutation-drift theory of synonymous codon usage. *Genetics* 129:897-907. <http://doi.org/10.1093/genetics/129.3.897>
- Chakraborty S, Nag D, Mazumder TH, Uddin A (2017). Codon usage pattern and prediction of gene expression level in *Bungarus* species. *Gene* 604:48-60. <http://doi.org/10.1016/j.gene.2016.11.023>
- Chakraborty S, Yengkhom S, Uddin DA (2020). Analysis of codon usage bias of chloroplast genes in *Oryza* species. *Planta* 252(4):67. <http://doi.org/10.1007/s00425-020-03470-7>
- Chen SY, Zhang HX, Wang X, Zhang YH, Ruan GH, Ma J (2021). Analysis of codon usage bias in the chloroplast genome of *Helianthus annuus* J-01. *IOP Conference Series: Earth and Environmental Science* 792(1):012009. <http://doi.org/10.1088/1755-1315/792/1/012009>
- Chi XF, Zhang FQ, Dong Q, Chen SL (2020). Insights into comparative genomics, codon usage bias, and phylogenetic relationship of species from Biebersteiniaceae and Nitrariaceae based on complete chloroplast genomes. *Plants* 9(11):1605. <http://doi.org/10.3390/plants9111605>
- Deng N, Liu CX, Tian YX, Song Q, Niu YD, Ma FF (2020). Complete chloroplast genome sequences and codon usage pattern among three wetland plants. *Agronomy Journal* 13(2):840-851. <http://doi.org/10.1002/agj2.20499>
- Dhindsa R, Copeland B, Mustoe A, Goldstein D (2020). Natural selection shapes codon usage in the human genome. *American Journal of Human Genetics* 107(1):83-95. <http://doi.org/10.1016/j.ajhg.2020.05.011>
- Dilucca M, Pavlopoulou A, Georgakilas AG, Giansanti A (2020). Codon usage bias in radioresistant bacteria. *Gene* 742: 144554. <http://doi.org/10.1016/j.gene.2020.144554>
- Dobrogojski J, Adamiec M, Luciński R (2020). The chloroplast genome: a review. *Acta Physiologiae Plantarum* 42(6). <http://doi.org/10.1007/s11738-020-03089-x>
- 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 the chloroplast genome in *Delphinium grandiflorum* L. reveals a preference for AT-ending codons as a result of major selection constraints. *Peer J* 9:e10787. <http://doi.org/10.7717/peerj.10787>
- Eyre-Walker AC (1991). An analysis of codon usage in mammals, selection or mutation bias? *Journal of Molecular Evolution* 33:442-449. <http://doi.org/10.1007/BF02103136>

- Galtier N, Roux C, Rousselle M, Romiguier J, Figuet E, Glémin S, Bierne N, Duret L (2018). Codon usage bias in animals: disentangling the effects of natural selection, effective population size, and GC-biased gene conversion. *Molecular Biology and Evolution* 35(5):1092-1103. <http://doi.org/10.1093/molbev/msy015>
- Guo LL, Guo S, Xu J, He LX, Carlson JE, Hou XG (2020). Corrigendum to phylogenetic analysis based on chloroplast genome uncover evolutionary relationship of all the nine species and six cultivars of tree peony. *Industrial Crops and Products* 154: 112710. <http://doi.org/10.1016/j.indcrop.2020.112567>
- Gupta SK, Bhattacharyya TK, Ghosh TC (2004). Synonymous codon usage in *Lactococcus lactis*, mutational bias versus translational selection. *Journal of Biomolecular Structure and Dynamics* 21:527-536. <http://doi.org/10.1080/07391102.2004.10506946>
- Hershberg R, Petrov DA (2008). Selection on codon bias. *Annual Review of Genetics* 42(1):287-299. <http://doi.org/10.1146/annurev.genet.42.110807.091442>
- Hu H, Liu J, An J, Wang M, Wang Q (2018). Characterization of the complete chloroplast genome of *Lonicera macranthoides*. *Mitochondrial DNA Part B* 3(2):1000-1001. <http://doi.org/10.1080/23802359.2018.1507643>
- Iriarte A, Lamolle G, Musto H (2021). Codon usage bias: an endless tale. *Journal of Molecular Evolution* <http://doi.org/10.1007/s00239-021-10027-z>
- Karumathil S, Raveendran NT, Ganesh D, Kumar NS, Nair RR, Dirisala VR (2018). Evolution of SCU bias in West African and Central African strains of monkeypox virus. *Evolutionary Bioinformatics* 14:1-22. <http://doi.org/10.1177/1176943318761368>
- Kawabe A, Miyashita NT (2003). Patterns of codon usage bias in three dicot and four monocot plant species. *Genes & Genetic Systems* 78(5):343-352. <http://doi.org/10.1266/ggs.78.343>
- Krasovec M, Filatov DA (2019). Evolution of codon usage bias in *Diatoms*. *Genes* 10(11):894. <http://doi.org/10.3390/genes10110894>
- Li GL, Pan ZL, Gao SC, He YY, Xia QY, Yan J, Yao HP (2019). Analysis of SCU of chloroplast genome in *Porphyra umbilicalis*. *Genes Genomics* 41:1173-1181. <http://doi.org/10.1007/s13258-019-00847-1>
- Liu HB, Lu YZ, Lan BL, Xu JC (2020). Codon usage by chloroplast gene is bias in *Hemiptera davidii*. *Journal of Genetics* 99:8. <http://doi.org/10.1007/s12041-019-1167-1>
- Mazumdar P, Othman RYB, Mebus K, Ramakrishnan N, Harikrishna JA (2017). Codon usage and codon pair patterns in non-grass monocot genomes. *Annals of Botany* 120(6):893-909. <http://doi.org/10.1093/aob/mcx112>
- Nair RR, Nandhini MB, Monalisha E, Murugan K, Sethuraman T, Nagarajan S, ... Ganesh D (2012). Synonymous codon usage in chloroplast genome of *Coffea arabica*. *Bioinformation* 8(22):1096-104. <http://doi.org/10.6026/97320630081096>
- Nie XJ, Deng PC, Feng KW, Liu PX, Du XH, Frank MY, Song WN (2014). Comparative analysis of codon usage patterns in chloroplast genomes of the Asteraceae family. *Plant Molecular Biology Reporter* 32:828-840. <http://doi.org/10.1007/s11105-013-0691-z>
- Pu X, Li Z, Tian Y, Gao R, Hao L, Hu Y, ... Song JY (2020). The honeysuckle genome provides insight into the molecular mechanism of carotenoid metabolism underlying dynamic flower coloration. *New Phytologist* 227(3):930-943. <http://doi.org/10.1111/nph.16552>
- Romero H, Zavala A, Musto H, Bernadi G (2003). The influence of translational selection on codon usage in fishes from the family Cyprinidae. *Gene* 317:141-147. [http://doi.org/10.1016/S0378-1119\(03\)00701-7](http://doi.org/10.1016/S0378-1119(03)00701-7)
- Sueoka N (1988). Directional mutation pressure and neutral molecular evolution. *Proceedings of the National Academy of Sciences of the United States of America* 85:2653-2657. <http://doi.org/10.1073/pnas.85.8.2653>
- Sueoka N (1999). Translation-coupled violation of Parity Rule 2 in human genes is not the cause of heterogeneity of the DNA G+C content of third codon position. *Gene* 238:53-58. [http://doi.org/10.1016/S0378-1119\(99\)00320-0](http://doi.org/10.1016/S0378-1119(99)00320-0)
- Tang DF, Wei F, Cai ZQ, Wei YY, Khan A, Miao JH, Wei KH (2020). Analysis of codon usage bias and evolution in the chloroplast genome of *Mesona chinensis* Benth. *Development Genes and Evolution* (4):1-9. <http://doi.org/10.1007/s00427-020-00670-9>
- Wang LY, Wang J, He CY, Zhang JG, Zeng YF (2020a). Characterization and comparison of chloroplast genomes from two sympatric Hippophae species (Elaeagnaceae). *Journal of Forestry Research* 32(7):307-318. <http://doi.org/10.1007/s11676-019-01079-5>
- Wang LY, Xing HX, Yuan YC, Wang XL, Saeed M, Tao JC, ... Sun XZ (2018). Genome-wide analysis of codon usage bias in four sequenced cotton species. *PLoS One* 13(3):e0194372. <http://doi.org/10.1371/journal.pone.0194372>

- Wang TT, Yang BX, Guan QJ, Chen X, Zhong ZZ, Huang W, ... Tian JK (2019). Transcriptional regulation of *Lonicera japonica* Thunb. during flower development as revealed by comprehensive analysis of transcription factors. *BMC Plant Biology* 19(1):198. <http://doi.org/10.1186/s12870-019-1803-1>
- Wang Z, Xu B, Li B, Zhou Q, Wang G, Jiang X, Wang C, Xu Z (2020b). Comparative analysis of codon usage patterns in chloroplast genomes of six Euphorbiaceae species. *Peer J* 8:e8251. <http://doi.org/10.7717/peerj.8251>
- Wright F (1990). The 'effective number of codons' used in a gene. *Gene* 87:23-29. [http://doi.org/10.1016/0378-1119\(90\)90491-9](http://doi.org/10.1016/0378-1119(90)90491-9)
- Yengkhom S, ArifUddin Y, Chakraborty S (2019). Deciphering codon usage patterns and evolutionary forces in chloroplast genes of *Camellia sinensis* var. *assamica* and *Camellia sinensis* var. *sinensis* in comparison to *Camellia pubicosta*. *Journal of Integrative Agriculture* 18(12):2771-2785. [http://doi.org/10.1016/S2095-3119\(19\)62716-4](http://doi.org/10.1016/S2095-3119(19)62716-4)
- Zhang R, Zhang L, Wang W, Zhang Z, Du H, Qu Z, Li XQ, Xiang H (2018). Differences in codon usage bias between photosynthesis-related genes and genetic system-related genes of chloroplast genomes in cultivated and wild solanum species. *International Journal of Molecular Sciences* 19(10):3142. <http://doi.org/10.3390/ijms19103142>
- Zhang WJ, Zhou J, Li ZF, Wang L, Gu X, Zhong Y (2007). Comparative analysis of codon usage patterns among mitochondrion, chloroplast and nuclear genes in *Triticum aestivum* L. *Journal of Integrative Plant Biology* 49:246-254. <http://doi.org/10.1111/j.1672-9072.2007.00404.x>
- Zhang Y, Nie X, Jia X, Zhao C, Biradar SS, Wang L, Du X, Weining S (2012). Analysis of codon usage patterns of the chloroplast genomes in the Poaceae family. *Australian Journal of Botany* 60:461-470. <http://doi.org/10.1071/BT12073>
- Zhou M, Long W, Li X (2008). Analysis of synonymous codon usage in chloroplast genome of *Populus alba*. *Journal of Forestry Research* 19:293-297. <http://doi.org/10.1007/s11676-008-0052-1>



The journal offers free, immediate, and unrestricted access to peer-reviewed research and scholarly work. Users are allowed to read, download, copy, distribute, print, search, or link to the full texts of the articles, or use them for any other lawful purpose, without asking prior permission from the publisher or the author.

**License** - Articles published in *Notulae Botanicae Horti Agrobotanici Cluj-Napoca* are Open-Access, distributed under the terms and conditions of the Creative Commons Attribution (CC BY 4.0) License.

© Articles by the authors; UASVM, Cluj-Napoca, Romania. The journal allows the author(s) to hold the copyright/to retain publishing rights without restriction.