LRF-776
[1-9]

RESEARCH ARTICLE Legume Research- An International Journal

Analysis of Codon Preferences in *Medicago ruthenica* based on Transcriptome Data

Xin Peng[1], Yingtong Mu[1], Feifei Wu[1], Nana Fu[2], Fengling Shi[1], Yutong Zhang[1]

10.18805/LRF-776

ABSTRACT

Background: The study investigated codon usage bias in *Medicago ruthenica* transcriptome coding sequences, aiming to lay the foundation for optimizing codon composition and enhancing heterologous gene expression in *Medicago ruthenica*.

Methods: In this research, *Medicago ruthenica* was used as the research material and 11,581 high-quality transcript gene sequences were selected from transcriptome data. Codon usage patterns and preferences were analyzed using software such as CodonW, R and Excel.

Result: The study revealed that the effective number of codons (ENC) ranged from 28.8 to 61.0. The average GC content of codons in expressed genes of *Medicago ruthenica* was 0.40 and the average GC content of the third nucleotide position of synonymous codons (GC3s) was 0.33. Analysis through ENC-plot, neutrality plot and bias analysis suggested that codon usage bias in the *Medicago ruthenica* transcriptome may be the result of a combination of factors including selection and mutation. Fifteen optimal codons were selected, with ten ending in ‘A’ and five ending in ‘U’, indicating a preference for ‘A/U’ ending codons in the *Medicago ruthenica* transcriptome. The frequency of codon usage in *Medicago ruthenica* was compared to five other organisms, including *Arabidopsis thaliana*, *Glycine max*, *Nicotiana tabacum*, yeast and *Escherichia coli*, revealing significant differences with E. coli and relatively smaller differences with *Nicotiana tabacum*.

Key words: Codon usage preference, *Medicago ruthenica*, Optimal codons, Transcriptome.

INTRODUCTION

Medicago ruthenica L. is a perennial diploid (2n=16) herb of the genus *Medicago*, also known as flowering alfalfa (Zhang, 2023). With its wide ecological adaptability, rich nutritional value and high-stress tolerance (Liu *et al*., 2013), as well as its high strain variation and high intraspecific genetic diversity, some studies have shown that lentil beans can provide valuable genes for genetic improvement of *Medicago sativa* (Wang *et al*., 2008). *Medicago ruthenica* can be used as a high-quality genetic resource for improving the stress resistance of alfalfa and other pastures (Xu *et al*., 2021). Transcriptomics is currently the most direct and universal way to study genome-level changes (Jing *et al*., 2016), which broadly refers to the study of the abundance of all the RNA transcripts of an organ or tissue under a particular condition. Transcriptome high-throughput sequencing technology has become an important means of studying plant development and elaborating gene mining and expression regulation related to active components and secondary metabolic biosynthesis pathways in herbaceous plants (Kapoor *et al*., 2021) and in recent years, transcriptome sequencing has begun to be applied in the evaluation of forage grass germplasm resources, the mechanism of formation of important traits, as well as the mining of excellent genetic resources and the development of molecular markers, for example, in *Medicago sativa* (Zhang *et al*., 2015), *Setaria viridis* (Martin *et al*., 2016), *Leymus chinensis* (Sun *et al*., 2013), *Elymus nutans* (Fu, 2017) and *Stipa breviflora* (Cao, 2015).

Exploring codon preferences can provide important information for studying biological evolution, gene function and exogenous gene expression (Lu *et al*., 2023). The transmission of genetic information is an evolving process, amino acids are structural units of proteins encoded by triplet codons, most amino acids contain two to six triplet codons ranging from two to six, each amino acid corresponds to at least one codon and different codons coding for the same amino acid are known as synonymous codons. In the process of evolution and adaptation of organisms to their environment, the phenomenon of unequal use of synonymous codons is known as codon preference and the efficiency of translating proteins by synonymous codons

[1]College of Grassland, Resources and Environment, Inner Mongolia Agricultural University, Key Laboratory of Grassland Resources, Ministry of Education, Hohhot, Inner Mongolia 010011 China.
[2]Inner Mongolia M-Grass Ecology And Environment (Group) Co., Ltd, Hohhot, Inner Mongolia 010010, China.

Corresponding Author: Fengling Shi; Yutong Zhang, College of Grassland, Resources and Environment, Inner Mongolia Agricultural University, Key Laboratory of Grassland Resources, Ministry of Education, Hohhot, Inner Mongolia 010011 China.
Email: nmczysfl@126.com; zyt7567@126.com

How to cite this article: Peng, X., Mu, Y., Wu, F., Fu, N., Shi, F. and Zhang, Y. (2024). Analysis of Codon Preferences in *Medicago ruthenica* based on Transcriptome Data. Legume Research. DOI: 10.18805/LRF-776.

Submitted: 03-11-2023 **Accepted:** 12-01-2024 **Online**: 31-01-2024

varies and the codon with the highest frequency of use is known as the optimal codon (Lai *et al*., 2019; Liang *et al*., 2019). In the course of evolution, organisms prefer to select the optimal codon, thus ensuring codon effectiveness. Mutational pressure and selective pressure are now generally recognized as the main reasons for the formation of codon preferences. In this study, we analyzed the codon preference analysis of the transcriptome data of *Medicago ruthenica* to reveal its codon preference characteristics and find its optimal codons, which will help its subsequent genetic breeding of *Medicago ruthenica*.

MATERIALS AND METHODS

Materials for testing

Medicago ruthenica 'Zhilixing' was bred by Inner Mongolia Agricultural University (IMAU) by using multiple hybrid selection methods and in December 1992, it was validated and registered as a cultivar. *Medicago ruthenica* 'Zhilixing' seeds were collected from the pasture trial site of Inner Mongolia Agricultural University (2008). Full, uniform and consistent seeds were selected and tested in 2020-2021. Due to the presence of hard solidity in the seeds of *Medicago ruthenica*, all the seeds were soaked in concentrated H_2SO_4 for 5-8 min and germinated at a constant temperature of 20°C. After germination, the seeds were transferred to seedling trays and placed in the greenhouse and the 4th or 5th fully expanded leaf with consistent growth was selected when the seedlings grew to the 6-8 leaf stage and then quick-frozen in liquid nitrogen and then stored at -80°C for total RNA extraction and reverse transcription was performed to obtain CDNA and sequencing of the CDNA yielded 11,581 gene sequences. In addition, biologically relevant codon information from *Arabidopsis thaliana*, *Glycine max*, *Nicotiana tabacum*, yeast and *Escherichia coli*, using the Codon Usage Database online software was utilized for comparison with lentil beans.

Relative synonymous codon usage degree

The concept of relative synonymous codon usage (RSCU) is used to detect changes in the pattern of usage of all synonymous codons across a gene, reflecting the ratio of the number of times a synonymous codon in a gene sample is equivalent to the number of times the value of a synonymous codon is observed to be used in practice to its average number of times it is expected to be used in theory.

Relative codon fitness

The codon adaptation index (CAI) is a commonly used geometric method to measure the relative adaptation values of individual codons. The CAI method is commonly used in various aspects of biology (Sharp *et al*., 1987).

ENC-plot priority cipher number plot analysis

Effective number of codons (ENC) refers to the degree of codon deviation from random selection and is also a key indicator of the degree of preference for unequal use of synonymous codons. Usually, for highly expressed genes, the preference is larger because they contain slightly fewer kinds of rare codons, so the ENC value is relatively small; for lowly expressed genes, the codon preference is smaller, thus leading to a larger ENC value (Wright, 1990).

PR2-plot plot analysis

PR2-plot plot analysis, The main purpose of PR2 bias analysis is to efficiently avoid linear mutational imbalances between adenine A and thymine T as well as cytosine C and guanine G at base 3 of the codon (Sueoka, 2001).

Neutral mapping analysis

For neutral plotting the analysis method is to obtain GC12 as a line of vertical coordinates and to obtain GC3 as a line of horizontal coordinates before plotting. The main factors that have a direct effect on codon preference are investigated by analyzing the correlation between the composition of bases at codon positions 1, 2 and 3. When there is a significant correlation between GC12 and GC3, it can indicate that there is no major difference in the base composition at the 3 different positions and that the application of the codon will be affected by the mutation. When there is no significant difference in the genetic correlation between GC12 as well as GC3, the regression coefficient is very close to 0, which can indicate that the base composition is different at positions 1, 2 and 3 and the application of codons is more affected by selection factors (Sueoka *et al*., 1988).

Correspondence analysis

Correspondence analysis (CA) is a method used to analyze the main reasons for the generation of usage preferences of synonymous codons between genes in research and this method is commonly used in research (Grantham *et al*., 1981). By analyzing the correlations between various genes obtained by isolation, the main reasons for the generation of preferences in the use of synonymous codons between genes can be precisely identified and determined.

Optimal codon analysis

Frequency of optimal codons (FOP), FOP is defined as the codons that are used most frequently in highly expressed genes of a species. FOP is species-specific and all require a set of gene sequences and their corresponding expression information to determine the optimal codons (Wu *et al*., 2007).

Data processing

CodonW and other software tools were used for basic codon characterization, codon bias analysis neutral mapping analysis, *etc*. in R. The codon usage preference of individual genes was also analyzed.

RESULTS AND DISCUSSION

GC content analysis

Analyzing codon usage patterns is important for the study of gene expression levels, protein structures and translation

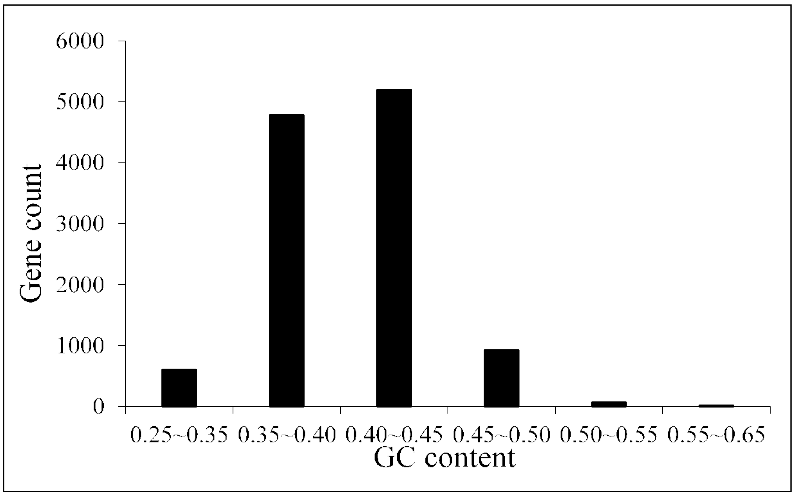

Fig 1: GC content.

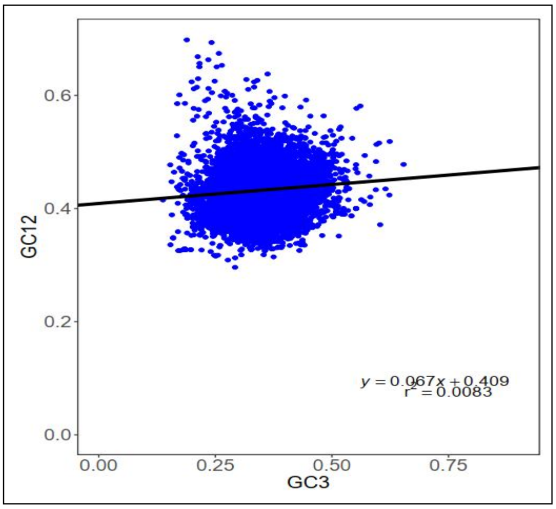

Fig 2: Neutral plot analysis.

rates in organisms (Wang *et al*., 2023). Codon usage preference is an adaptive choice formed by species during long-term natural selection and evolutionary processes, which is mainly influenced by gene mutation pressure and natural selection pressure (Sueoka *et al*., 1988). The phenomenon of synonymous codon preference usage exists in all types of plants. Studies have shown that GC content, Tr-NA abundance, protein structure and amino acid composition all have some influence on codon usage preference (Zhao *et al*., 2020). In this study, codon usage preference of 11581 complete coding sequences of *Medicago ruthenica* was analyzed in combination with second-generation sequencing technology. GC content is an important indicator of codon-base composition in organisms (Feng *et al*., 2019). The average GC content of *Medicago ruthenica* in this study was 0.40, indicating a weak codon bias in the *Medicago ruthenica* genome; and GC1 (47.93%) > GC2 (38.49%) > GC3 (34.39%). The codon usage preference analysis of the 11,581 Unigenes that had been obtained was performed using CodonW (Fig 1) and the results showed that the average total GC amount of all Unigenes in the *Medicago ruthenica* was 40.40% and the distribution of the total GC content ranged from 25.00% to 65.00%. The main distribution was between 40.00% and 45.00%. The average GC content of codon 3 nucleotides (GC3) was 32.88% and the distribution of average GC3 content ranged from 12.67% to 65.38%.

Neutral plot analysis

The neutral plotting analysis of the coding sequence of *Medicago ruthenica* is shown in Fig 2, with the values of GC3 ranging from 12.67% to 65.38% and GC12 ranging from 29.62% to 85.74%. The mean value of the GC content of the 1st and 2nd GCs of the codons of the 11581 genes, GC12 and the content of the third base, GC3, were analyzed. The results of the relationship showed that the fitted curve equation was y= 0.067× + 0.409 (r^2= 0.0083), with a regression coefficient of 0.067 and the overall trend line conformed to the GC3= GC12 diagonal, with most of the gene points falling near GC3= GC12, but some of them fell on the diagonal line, which confirmed that the codons of *Medicago ruthenica* had some This confirms that Medicago ruthenica codons have certain favoritism and that mutational pressure, in addition to natural selection pressure, also affects the use of favoritism in *Medicago ruthenica* codons.

Relative codon adaptation

Base mutation, genetics and natural selection are also important influences on codon usage preference and the ENC values of Medicago ruthenica codons were significantly negatively correlated with GC content, suggesting that base composition also affects codon preference of Medicago ruthenica genes to a certain extent. The CAI of the *Medicago ruthenica* transcriptome took values ranging from 0.097 to 0.411, indicating that the *Medicago ruthenica* gene expression level was not high. Meanwhile, the correlation analysis of CAI and several other important parameters

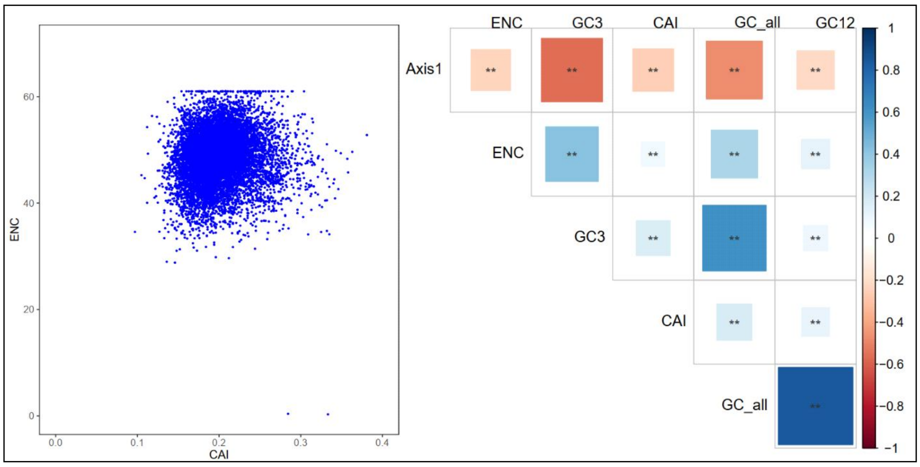

Fig 3: Relative fitness of codons.

(ENC, GC3, GC12) was carried out (Fig 3) and the results showed that there was a significant negative correlation between the CAI value and ENC and GC content. Therefore, the process of codon preference formation in *Medicago ruthenica* was influenced by the gene expression level and the higher the GC content and the higher the expression level, the higher the degree of codon preference of the gene.

ENC-plot preferred codon number mapping analysis

Neutral mapping and ENC-plot and PR2- plot analyses showed that factors other than natural selection and mutational pressure also affect the codon preference of *Medicago ruthenica* transcriptome. Therefore, it was concluded that the transcriptome codon preference of *Medicago ruthenica* was mainly dominated by the effects of natural selection and mutation and the results of this study were similar to the results of previous studies on *Mangifera indica* (Tang *et al*., 2021), *Arabis paniculata* Franch. (Luo *et al*., 2022), *etc*.; and *Amaranthus caudatus* L. (Feng *et al*., 2019) was mainly dominated by mutational effects and *Medicago sativa* (Yu *et al*., 2021) was mainly dominated by selective effects. In this way, it is again inferred that the codon preference influencing factors may be related to the species, but the specific influencing mechanism needs to be further explored. The effective codon count ENC indicates the number of effective codons used in a gene, with larger values indicating that each codon is used equally and the use preference is weaker. The results of the Codon W analysis showed that the ENC values of the *Medicago ruthenica* bean transcriptome ranged from 28.8 to 61.0 and all of them were greater than 28, so it was concluded that the *Medicago ruthenica* transcriptome has a weak codon bias. A graph was made with GC3 as the horizontal coordinate and ENC as the vertical coordinate (Fig 4) and the points in the graph show the distribution of genes. Most of the points of the representative genes are far away from the expectation curve and some of the gene points are distributed around the expectation curve, indicating that in addition to mutational pressure playing

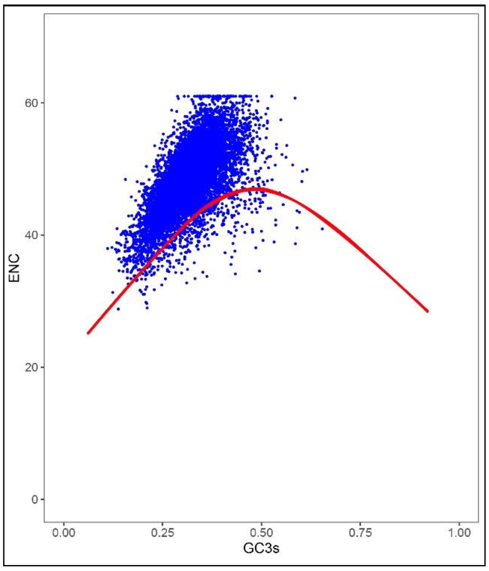

Fig 4: Association analysis between ENC and GC3.

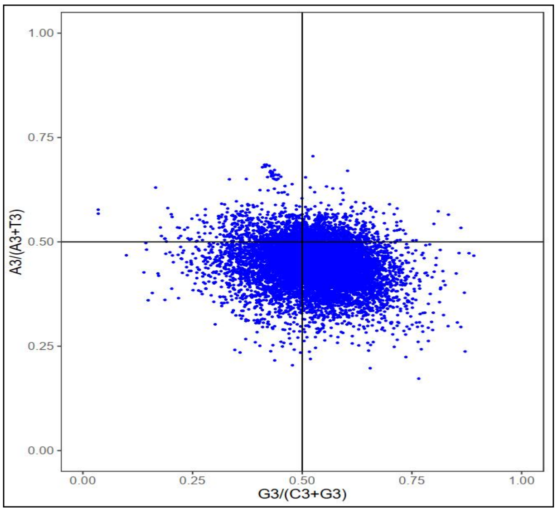

Fig 5: PR2-plot bias analysis.

an important role in the formation of codon bias in *Medicago ruthenica*, other factors such as the role of genetic selection also play an important role in the formation of codon bias in *Medicago ruthenica*.

PR2-plot bias analysis

PR2-plot bias analysis of codons of *Medicago ruthenica* (Fig 5) shows the preference of bases in codon position 3 of *Medicago ruthenica* transcriptome gene sequences, as can be seen in the figure, the frequency of base C is lower than that of G and the frequency of base T is higher than that of A. Most of the genes are located below the y-axis 0.5, with the vector downward and left-right bias indicating that the codon third position of *Medicago ruthenica* transcriptome genes has a higher content of C, G and T. The frequency of using A, T, C and G in the codon third position is not equal, which indicates that the codon bias of *Medicago ruthenica* is not only caused by mutation but

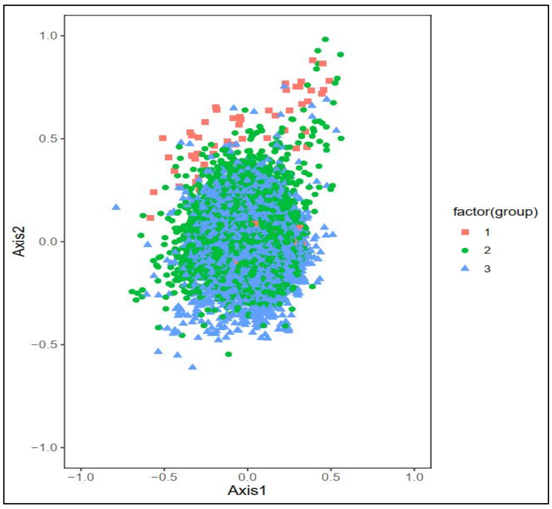

Fig 6: Correspondence analysis.

Table 1: Relative frequency of synonymous codon usage in *Medicago ruthenica.*

Amino	Codon	High expression		Low expression		Amino	Codon	High expression		Low expression	
acids		RSCU	Amount	RSCU	Amount	acids		RSCU	Amount	RSCU	Amount
Phe	UUU	1.28	8185	1.26	3869	Ser	UCU	1.59	5551	1.82	5353
	UUC	0.72	4581	0.74	2263		UCC	0.60	2090	0.72	2104
Leu	UUA*	1.36	5514	0.75	2190		UCA*	2.03	7093	1.24	3635
	UUG	1.42	5744	1.72	5050		UCG	0.20	703	0.46	1356
	CUU*	1.67	6736	1.54	4525	Pro	CCU	1.51	5542	1.73	3774
	CUC	0.65	2639	0.66	1921		CCC	0.31	1124	0.56	1231
	CUA*	0.72	2897	0.52	1538		CCA*	2.00	7354	1.27	2787
	CUG	0.18	740	0.81	2360		CCG	0.19	707	0.44	953
Ile	AUU	1.53	8540	1.56	5173	Thr	ACU	1.37	4315	1.66	4344
	AUC	0.56	3153	0.77	2538		ACC	0.58	1827	0.86	2249
	AUA*	0.91	5069	0.67	2212		ACA*	1.91	6015	1.16	3031
Met	AUG	1.00	5208	1.00	4992		ACG	0.14	450	0.32	825
Val	GUU*	1.96	6701	1.80	5922	Ala	GCU	1.73	4973	1.91	7482
	GUC	0.45	1541	0.56	1836		GCC	0.46	1328	0.57	2254
	GUA*	0.76	2603	0.59	1946		GCA*	1.68	4845	1.19	4672
	GUG	0.83	2850	1.05	3463		GCG	0.13	370	0.32	1274
Tyr	UAU	1.30	5613	1.29	2830	Cys	UGU*	1.43	3150	1.22	1410
	UAC	0.70	3053	0.71	1547		UGC	0.57	1262	0.78	896
TER	UAA	3.00	580	3.00	579	TER	UGA	0.00	0	0.00	0
	UAG	0.00	0	0.00	0	Trp	UGG	1.00	3139	1.00	1799
His	CAU*	1.44	3697	1.35	2907	Arg	CGU	0.66	668	0.92	2070
	CAC	0.56	1422	0.65	1408		CGC	0.23	234	0.46	1028
Gln	CAA*	1.71	5845	1.08	5124		CGA	0.48	487	0.62	1401
	CAG	0.29	996	0.92	4375		CGG	0.14	137	0.49	1090
Asn	AAU	1.32	8880	1.29	6440	Ser	AGU	1.09	3815	1.14	3355
	AAC	0.68	4541	0.71	3563		AGC	0.50	1737	0.63	1851
Lys	AAA*	1.31	7219	0.89	9839	Arg	AGA*	3.20	3226	1.76	3961
	AAG	0.69	3833	1.11	12209		AGG	1.28	1290	1.75	3921
Asp	GAU*	1.55	6083	1.44	10704	Gly	GGU	1.54	6882	1.54	4185
	GAC	0.45	1767	0.56	4189		GGC	0.41	1833	0.57	1542
Glu	GAA*	1.39	5254	1.10	13296		GGA*	1.68	7517	1.27	3468
	GAG	0.61	2293	0.90	10780		GGG	0.36	1613	0.62	1690

Note: * is the optimal codon RSCU: relative synonymous codon usage

Table 2: Comparison of codon preferences of *Medicago ruthenica* with those of other organisms.

Amino acid	Codon	Frequency of codon usage (1/1 000)						Analysis of codon bias				
		Ak	At	Gm	Nt	S	Ec	Ak/At	Ak/Gm	Ak/Nt	Ak/S	Ak/Ec
Ala	GCG	4.07	9.00	6.10	5.80	6.20	32.30	0.45*	0.67	0.70	0.66	0.13*
	GCA	20.72	17.50	14.60	23.10	16.20	20.70	1.18	1.42	0.90	1.28	1.00
	GCT	26.13	28.30	49.70	31.20	21.20	15.40	0.92	0.53	0.84	1.23	1.70
	GCC	8.32	10.30	11.70	12.50	12.60	25.20	0.81	0.71	0.67	0.66	0.33*
Cys	TGT	12.26	10.50	5.90	9.80	8.10	5.20	1.17	2.08*	1.25	1.51	2.36*
	TGC	6.45	7.20	11.70	7.20	4.80	6.40	0.90	0.55	0.90	1.34	1.01
Asp	GAT	39.99	36.70	32.20	36.90	37.60	32.80	1.09	1.24	1.08	1.06	1.22
	GAC	13.60	17.20	16.20	16.90	20.20	19.20	0.79	0.84	0.80	0.67	0.71
Glu	GAG	24.30	32.20	29.20	29.40	19.20	18.70	0.75	0.83	0.83	1.27	1.30
	GAA	39.23	34.30	32.20	36.00	45.60	39.30	1.14	1.22	1.09	0.86	1.00
Phe	TTT	28.74	21.80	40.90	25.10	26.10	22.20	1.32	0.70	1.15	1.10	1.29
	TTC	15.61	20.70	26.30	18.00	18.40	15.90	0.75	0.59	0.87	0.85	0.98
Gly	GGG	8.50	10.20	11.70	10.50	6.00	11.80	0.83	0.73	0.81	1.42	0.72
	GGA	22.39	24.20	23.40	23.20	10.90	8.90	0.93	0.96	0.96	2.05*	2.52*
	GGT	22.48	22.20	14.70	22.30	23.90	24.20	1.01	1.53	1.01	0.94	0.93
	GGC	7.76	9.20	11.00	11.20	9.80	28.10	0.84	0.71	0.69	0.79	0.28*
His	CAT	17.08	13.80	14.40	13.40	13.60	12.80	1.24	1.19	1.27	1.26	1.33
	CAC	7.47	8.70	12.20	8.70	7.80	9.40	0.86	0.61	0.86	0.96	0.79
Ile	ATA	16.32	12.60	16.20	14.00	17.80	5.50	1.30	1.01	1.17	0.92	2.97*
	ATT	30.51	21.50	24.00	27.80	30.10	29.70	1.42	1.27	1.10	1.01	1.03
	ATC	12.70	18.50	11.60	13.90	17.20	23.90	0.69	1.09	0.91	0.74	0.53
Lys	AAG	29.31	32.70	40.90	33.50	30.80	11.00	0.90	0.72	0.87	0.95	2.66*
	AAA	37.14	30.80	33.80	32.60	41.90	34.00	1.21	1.10	1.14	0.89	1.09
Leu	TTG	24.99	20.90	25.30	22.30	27.20	13.00	1.20	0.99	1.12	0.92	1.92
	TTA	16.93	12.70	13.40	13.40	26.20	13.80	1.33	1.26	1.26	0.65	1.23
	CTG	7.37	9.80	10.40	9.80	10.50	51.10	0.75	0.71	0.75	0.70	0.14*
	CTA	10.17	9.90	8.20	9.40	13.40	3.90	1.03	1.24	1.08	0.76	2.61*
	CTT	25.45	24.10	18.80	24.00	12.30	11.40	1.06	1.35	1.06	2.07*	2.23*
	CTC	10.17	16.10	11.30	12.30	5.40	10.50	0.63	0.90	0.83	1.88	0.97
Met	ATG	24.16	24.50	31.50	25.00	20.90	27.20	0.99	0.77	0.97	1.16	0.89
Asn	AAT	34.81	22.30	29.10	28.00	35.70	19.20	1.56	1.20	1.24	0.98	1.81
	AAC	17.69	20.90	17.90	17.90	24.80	21.70	0.85	0.99	0.99	0.71	0.82
Pro	CCG	4.36	6.90	3.60	5.00	5.30	22.40	0.63	1.21	0.87	0.82	0.19*
	CCA	18.03	16.20	19.20	19.80	18.30	8.40	1.11	0.94	0.91	0.99	2.15*
	CCT	19.06	18.70	9.00	18.70	13.50	7.20	1.02	2.12*	1.02	1.41	2.65*
	CCC	5.15	5.30	2.80	6.60	6.80	5.60	0.97	1.84	0.78	0.76	0.92
Gln	CAG	11.58	15.20	18.60	15.00	12.10	29.40	0.76	0.62	0.77	0.96	0.39*
	CAA	24.77	19.50	25.50	20.70	27.30	14.70	1.27	0.97	1.20	0.91	1.69
Arg	AGG	10.91	11.00	15.10	12.20	9.20	1.80	0.99	0.72	0.89	1.19	6.06*
	AGA	17.62	19.00	17.80	16.00	21.30	2.90	0.93	0.99	1.10	0.83	6.08*
	CGG	2.98	4.90	2.60	3.70	1.70	6.20	0.61	1.15	0.81	1.75	0.48*
	CGA	5.12	6.30	4.80	5.30	3.00	3.80	0.81	1.07	0.97	1.71	1.35
	CGT	7.03	9.00	5.90	7.50	6.40	20.20	0.78	1.19	0.94	1.10	0.35*
	CGC	3.37	3.80	5.30	3.90	2.60	20.80	0.89	0.64	0.86	1.30	0.16*
Ser	AGT	16.48	14.00	14.10	13.30	14.20	9.40	1.18	1.17	1.24	1.16	1.75
	AGC	8.19	11.30	12.20	10.00	9.80	16.00	0.72	0.67	0.82	0.84	0.51
	TCG	5.20	9.30	3.60	5.30	8.60	8.80	0.56	1.44	0.98	0.60	0.59
	TCA	22.67	18.30	15.50	17.60	18.70	8.10	1.24	1.46	1.29	1.21	2.80*

Table 2: Continue...

Table 2: Continue...

	TCT	24.49	25.20	13.40	20.00	23.50	8.70	0.97	1.83	1.22	1.04	2.81*
	TCC	9.60	11.20	7.20	10.20	14.20	8.90	0.86	1.33	0.94	0.68	1.08
Thr	ACG	3.59	7.70	4.20	4.50	8.00	15.00	0.47*	0.86	0.80	0.45*	0.24*
	ACA	19.03	15.70	32.50	17.40	17.80	8.10	1.21	0.59	1.09	1.07	2.35*
	ACT	19.39	17.50	13.90	20.30	20.30	9.10	1.11	1.39	0.96	0.96	2.13*
	ACC	9.07	10.30	6.90	9.70	12.70	22.80	0.88	1.32	0.94	0.71	0.40*
Val	GTG	14.83	17.40	22.80	16.70	10.80	26.20	0.85	0.65	0.89	1.37	0.57
	GTA	11.13	9.90	9.70	11.40	11.80	10.90	1.12	1.15	0.98	0.94	1.02
	GTT	30.35	27.20	28.50	26.80	22.10	18.10	1.12	1.06	1.13	1.37	1.68
	GTC	8.38	12.80	9.00	11.10	11.80	14.80	0.65	0.93	0.75	0.71	0.57
Trp	TGG	12.71	12.50	14.20	12.20	10.40	15.30	1.02	0.90	1.04	1.22	0.83
Tyr	TAT	19.82	14.60	20.50	17.80	18.80	16.50	1.36	0.97	1.11	1.05	1.20
	TAC	10.04	13.70	13.50	13.50	14.80	12.30	0.73	0.74	0.74	0.68	0.82
Stop	TAA	2.21	0.90	0.40	1.10	0.80	1.10	2.46*	5.53*	2.01*	2.77*	2.01*
	TAG	0.002	0.50	0.60	0.50	0.70	0.50	0.004*	0.003*	0.004*	0.003*	0.004*
	TGA	0.004	1.20	0.70	1.00	0.50	0.70	0.003*	0.01*	0.004*	0.008*	0.006*

Note: Ak: *Medicago ruthenica*; AT: *Arabidopsis thaliana*; GM: *Glycine max*; NT: *Nicotiana tabacum*; S: yeast; EC: *E. coli*; *: The ratio is less than or equal to 0.5 or greater than or equal to 2.0

also influenced by other factors such as hereditary and selective factors.

Correspondence analysis

The results of correspondence analysis showed that one axis caused the greatest effect on the codon bias of the *Medicago ruthenica* transcriptome. To further illustrate the effect of GC content on codon bias of *Medicago ruthenica*, genes with different GC content were colored differently, genes with GC content higher than 50% were marked in red, genes with GC content between 40% and 50% were marked in blue and genes with GC content lower than 40% were marked in green. As shown in Fig 6, the genes with GC content higher than 50% are more dispersed in the coordinate system, while the genes with GC content less than 50% are more concentrated.

Optimal codon analysis

RSCU refers to the codon bias for a particular codon in coding the corresponding amino acid in the synonymous relative probability among codons, which removes the effect of amino acid composition on codon usage. When the RSCU of a particular codon is >1, it indicates that the codon is used more frequently. In this study, a total of 28 codons with RSCU>1 were found in the transcriptome of *Medicago ruthenica*, which is similar to the results of Liu *et al*. (2005). On *Arabidopsis thaliana* as well as Zhou *et al*. (2008) on *Populus alba*, both of which were based on the chloroplast genome and they showed that the number of codons with RSCU>1 was 30 the reason for the difference may be due to the difference in species and based on different levels. (Tian *et al.,* 2021). found that there were 30 codons with relative synonymous codon usage RSCU>1 in the chloroplast genome of *Medicago ruthenica*, of which 29 ended in A/U, which was similar to the results of this study and in the present study, we found that there were 28 codons with RSCU>1 in *Medicago ruthenica*, of which 27 ended in A/U. The results of the RSCU analysis of the high\low expression sequence library of the *Medicago ruthenica* transcriptome are shown in Table 1. ΔRSCU≥0.08, *i.e*., there are 18 codons for high expression superiority codons (* marked), of which 13 end in A, 5 end in U, 0 end in G and 0 end in C, which suggests that the *Medicago ruthenica* genes prefer using codons ending in A or U. The codons with ΔRSCU≥0.08 and RSCU≥1 are the optimal codons in the lentil bean transcriptome. codons ending in A or U. Codons with ΔRSCU≥0.08 and RSCU≥1 are optimal codons and there are a total of 15 optimal codons in the *Medicago ruthenica* transcriptome (underlined), which are UUA, CUU, GUU, CAU, CAA, AAA, GAU, GAA, UCA, CCA and ACA, GCA, GCA. UGU, AGA and GGA; 10 of these codons end in A and 5 end in U.

Relative synonymous codon usage degree

Comparison of codon usage frequency of *Medicago ruthenica* with five organisms, *Arabidopsis thaliana*, *Glycine max*, Nicotiana tabacum, yeast and *Escherichia coli*, revealed that there were large differences with *Escherichia coli* and small differences with *Nicotiana tabacum*. Zhang *et al*. (2022) found that the frequency of Coix lacryma-jobi codon usage also differed less from that of Arabidopsis thaliana, which is consistent with the results of this paper. There are 15 optimal codons in the *Medicago ruthenica* transcriptome, of which 10 end in A and the remaining 5 end in U. There is no codon ending in G/C in the optimal codon and there is no codon ending in G/C in the optimal codon. codons ending in G/C. The above results were consistent with the results of the analysis of optimal codons in the chloroplast genomes of most species of *Medicago ruthenica* (Tian *et al*., 2021) and Medicago sativa (Yu *et al*., 2021). The codon preferences of *Arabidopsis thaliana*, *Glycine max*, *Nicotiana tabacum*, yeast and *Escherichia coli* were extracted from the Codon Usage Database and the key information of the codons of the above species were selected and compared with those of *Medicago ruthenica* codon preferences for comparison. If there is a big difference between the codon preference of *Medicago ruthenica* and that of the species, then the codon usage frequency ratio is ≤0.5 or ≥2.0 and when the ratio is in the range of 0.5~2.0, then it can be proved that the codon usage preference of the two is similar. The results, as shown in Table 2, showed that the frequency of codon usage of *Medicago ruthenica* codons had a significant deviation from that of other species. Among them, there are five species with codon usage frequency ratios ≥2 or ≤0.5 with *Arabidopsis thaliana*, five with *Glycine max*, three with *Nicotiana tabacum*, 28 with *E. coli* and six with yeast. There are different levels of differences between the codons of *Medicago ruthenica* and these several organisms, with the smallest difference with *Nicotiana tabacum* and the largest difference with *E. coli*.

CONCLUSION

Medicago ruthenica transcriptome codon preference is mainly affected by natural selection and mutation pressure, but other factors also affect *Medicago ruthenica* transcriptome codon preference. At the same time, the study identified 15 optimal codons in the *Medicago ruthenica* transcriptome and when using *Medicago ruthenica* for genetic engineering research to design exogenous genes, the selection of codons ending in A/U can to some extent improve the efficiency of exogenous gene expression and transformation and also provide a certain degree of basis for the selection and utilization of the best genes.

ACKNOWLEDGEMENT

Fengling Shi and Yutong Zhang designed the experiments, Xin Peng did the experiments and wrote the manuscript and others contributed equally to the manuscript. Project Grant: “Inner Mongolia Autonomous Region Seed Industry Science and Technology Innovation Major Demonstration Project of Unveiling and Hanging Project” (2022JBGS0040).

Conflict of interest

The article is original and has not been published previously, is not under consideration for publication elsewhere, and if accepted, it will not be published elsewhere in the same form, in English or any other language. The submission of the article has the approval of the all the authors and the authorities of the host institute where work had been carried out. All the authors have made substantive and intellectual contributions to the article and assume full responsibility for all opinions, conclusion and statements expressed in the articles.

REFERENCES

Cao, L. (2015). The Transcriptome Study of Stipa Breviflora in Desert Steppe under Simulated Warming. (Master), Inner Mongolia Agricultural University.

Feng, R.Y., Mei, C., Wang, H.J., Lei, M.L., Tian, X., *et al*. (2019). Analysis of codon usage in the chloroplast genome of grain amaranth (*Amaranthus h ypochondriacus* L). Chinese Journal of Grassland. 41(4): 8-15.

Fu, J.J. (2017). The physiological and molecular mechanisms of Tibetan wild *Elymus nutans* responses to cold stress. (Doctor), Northwest Agriculture and Forestry University.

Grantham, R., Gautier, C., Gouy, M. (1981). Codon catalog usage is a genome strategy modulated for gene expressivity. Nucleic Acids Research. 9(1): 43-74.

Jing, L., Jiang, H., Du, H.H., Meng, Y.F. (2016). Research Progress on *Dioscorea opposita* in China. Journal of Anhui Agricultural Sciences. 44(15): 114-117.

Kapoor, B., Kumar, A., Kumar, P. (2021). Transcriptome repository of North-Western Himalayan endangered medicinal herbs: A paramount approach illuminating molecular perspective of phytoactive molecules and secondary metabolism. Mol Genet Genomics. 2(96): 1177-1202.

Lai, R.L., Feng, X., Chen, J., Zhong, C.S., Chen, Y.T., *et al*. (2019). Codon Usage Bias of *Canarium album* (Lour) R. Transcriptome and its influence Factors. Journal of Nuclear Agricultural Sciences. 33(01): 31-38.

Liang, E., Qi, M.J., Ding, Y.Q., Zhang, L. (2019). Analysis of codon bias in *Panax japonicus* transcriptome. Jiangsu Agricultural Sciences. 47(02): 59-63.

Liu, J., Hu, D., Chu, H.J., Yan, J., Li, J.Q. (2013). Screening of Drought-and Salinity-responsive EST-SSR Markers in *Medicago ruthenica* Trautv. Plant Science Journal. 31(05): 493-499.

Liu, Q.P., Xue, Q.Z. (2005). Comparative studies on codon usage pattern of chloroplasts and their host nuclear genes in four plant species. Journal of Genetics. 84(1): 55-62.

Lu, Z.L., Tian, W.Y., Yang, C.D. (2023). Chloroplast genome characteristics and codon usage bias analysis of *Panax* Linn. Molecular Plant Breeding. 1-24.

Luo, Q., Fang, Y., Yu, J., Wen, Q.B., Zhu, B. (2022). Analysis of Chloroplast genome characteristics and codon usage bias of *Arabis paniculata* Franch. Molecular Plant Breeding. 20(07): 2261-2270.

Martin, A.P., Palmer, W.M., Brown, C., Abel, J.E., Lunn, *et al*. (2016). A developing Setariaviridis internode: An experimental system for the study of biomass generation in a C4 model species. Biotechnology for Biofuels. 9(1): 45-57.

Sharp, P.M., Li, W.H. (1987). The codon Adaptation Index-a measure of directional synonymous codon usage bias and its potential applications. Nucl Acids Res. 15(3): 1281-1295.

Sueoka, N. (2001). Near homogeneity of PR2-Bias fingerprints in the human genome and their implications in phylogenetic analyses. Journal of Molecular Evolution. 53(4-5): 469-476.

Sueoka, N., Suoeka, N. (1988). Directional mutation pressure and neutral molecular evolution. Proceedings of the National Academy of Sciences. 85(8): 2653-2657.

Sun, Y.P., Wang, F.W., Wang, N., Dong, Y.Y., Liu, Q., *et al*. (2013). Transcriptome exploration in *Leymus chinensis* under saline-alkaline treatment using 454 pyrosequencing. PLoS One. 8(1): 1-12.

Tang, Y.J., Zhao, Y., Huang, G.D., Fu, H.T., Song, E.L., *et al*. (2021). Analysis on Codon Usage Bias of Chloroplast Genes from Mango. Chinese Journal of Tropical Crops. 42(08): 2143-2150.

Tian, C.Y., Wu, Z.N., Li, X.S., Li, Z.Y. (2021). Codon usage bias of chloroplast genome in *Medicago ruthenica*. Acta Agrestia Sinica. 29(12): 2678-2684.

Wang, D.K., Li, H., Luo, X.Y. (2008). Crossbreeding of *Melilotoides ruthenicus* and *Medicago sativa.* Acta Agrestia Sinica. (05): 458-465.

Wang, P., Guo, M.J., Niu, J.M., Wang, X.Y., Yue, J.R., *et al*. (2023). Analysis of codon usage bias in the chloroplast genome of dianxizeqin (*Sium ventricosum*). Molecular Plant Breeding. 1-23.

Wright, F. (1990). The ‘effective number of codons’ used in a gene. Gene. 87(1): 23-29.

Wu, X.M., Wu, X.F., Ren, D.M., Zhu, Y.P., He, F.C. (2007). The analysis method and progress in the study of codon bias. Hereditas (Beijing). 29(4): 420-426.

Xu, B., Wu, R.N., Gao, C.P., Shi, F.L. (2021). Establishment of tissue culture regeneration system for *Medicago ruthenica* L. cv. ‘Zhilixing’. Legume Research. 45(2): 162-167.

Yu, F., Han, M. (2021). Analysis of codon usage bias in the chloroplastgenome of alfalfa (*Medicago sativa*). Guihaia. 41(12): 2069-2076.

Zhang, C.J., Zhao, Y.C., Zhao, D.G. (2022). Analysis of Codons Use Preference of Transcriptome in *Coix Lachryma -jobi* L. Seed. 41(04): 13-19.

Zhang, S.H., Shi, Y.H., Cheng, N.N., Du, H.Q., Fan, W.N., *et al*. (2015). De novo characterization of fall dormant and nondormant alfalfa (*Medicago sativa* L.) leaf transcriptome and identification of candidate genes related to fall dormancy. PLoS One. 10(3): 1-25.

Zhang, Y.T. (2023). Analysis on plant type characteristics and erect stem regulation mechanism of *Medicago ruthenica*. (Doctor), Inner Mongolia Agricultural University.

Zhao, X., Deng, L.H., Chen, F. (2020). Codon usage bias of chloroplast genome in *Kandelia obovata*. Journal of Forest and Environment. 40(5): 534-541.

Zhou, M., Long, W., Li, X. (2008). Analysis of synonymous codon usage in chloroplast genome of *populus alba*. Journal of Forestry Research. 19(4): 293-297.