



Comparison of the Complete Chloroplast Genomes in the *Astragalus*

Dabao Yin^{1,2}, Yingtong Mu¹, Xue Li^{1,2}, Mei Hua^{1,2}, Xiaojie Li^{1,2},
Peiqing Zhang^{1,2}, Xiaoming Zhang^{1,2}, Junjie Wang^{1,2}

10.18805/LRF-763

ABSTRACT

Background: *Astragalus* is one of the largest angiosperms and has important economic value. The chloroplast genomes (Cp) of most plants in its family have been sequenced and annotated, but there are fewer studies on the characteristics of the chloroplast genomes and codon usage bias of *Astragalus*. In this study, we sequenced and annotated the complete chloroplast genomes of three *Astragalus* species, that systematically compared the chloroplast genome and codon usage characteristics of three *Astragalus*.

Methods: In this study, we used three *Astragalus* species as materials. Firstly, we sequenced and assembled their chloroplast genomes. Subsequently, we analyzed the codon usage bias of the chloroplast genomes of the three medicinal *Astragalus* species using software such as CodonW, CUSP and SPSS.

Result: The results illustrated that the length of *Astragalus* chloroplast genome ranges from 122,815 bp (*A. dahuricus*) to 123,729 bp (*A. melilotoides* Pall.). There were 107-110 genes, including 75-76 protein-coding genes (PCGs), four ribosomal RNA genes (rRNAs) and 28-30 transfer RNA genes (tRNAs). In addition, the results illustrated that codons in three Cp genomes from *Astragalus* had the same 11 Optimal codons and ending with A/U. The codon usage frequency of five model creatures were compared and the results revealed that the codon preferences of *Astragalus* and five model creatures were pretty significant, the natural selection was the main factor in codon preference. Our research provides some value for the researches of chloroplast genetic engineering and molecular breeding in *Astragalus*.

Key words: *Astragalus*, Chloroplast genome, Codon usage bias, Codon usage frequency.

INTRODUCTION

Chloroplast is an important intracellular organelle of plants for photosynthesis (Wan *et al.*, 2009). The genome of chloroplast is circular in structure and its size varies from 120 kb to 170 kb (Olmstead *et al.*, 1994). The genome features of chloroplast are usually quadripartite configurations, including small single copy (SSC) region, large single copy (LSC) region, inverted repeats *IRa* and *IRb* (Tangphatsornruang *et al.*, 2010). In addition Chloroplast is haploid, maternally inherited and possess high conservation in gene content, with low nucleotide substitution rate which is beneficial to study evolutionary relationship in plants at taxonomic level (Liu *et al.*, 2019). The central rule is an important biological rule, the triplet codon transfers genetic information from mRNA to protein in the process of translation and plays an important role in the life activities of organisms (Wolfe *et al.*, 1987). Different codons that encode the same amino acid are called synonymous codons. There are differences in the frequency of use of synonymous codons in encoding amino acids, that is, codon usage bias (Sharm *et al.*, 1987; Archtti, 2004). For example, in monocotyledonous species such as maize, the codon often ends with G/C (Fennoy *et al.*, 1993) and in dicotyledonous plants such as tea tree, it often ends with A/U (Wang *et al.*, 2018a). Natural selection, gene mutation and genetic drift are the causes of codon usage bias and the influencing factors of codon usage bias are different in

¹Engineering Research Center for the Seed breeding of Chinese and Mongolian Medicinal Materials in Inner Mongolia, Inner Mongolia, 010011, China.

²College of Grassland, Resource and Environmental Science, Ministry of Education, Ministry of Agriculture, Inner Mongolia Agricultural University, Hohhot, China.

Corresponding Author: Junjie Wang, Engineering Research Center for the Seed breeding of Chinese and Mongolian Medicinal Materials in Inner Mongolia, Inner Mongolia, 010011, China. Email: jjw62@163.com

How to cite this article: Yin, D., Mu, Y., Li, X., Hua, M., Li, X., Zhang, P., Zhang, X. and Wang, J. (2024). Comparison of the complete chloroplast genomes in the *Astragalus*. *Legume Research*. 47(3): 420-427. doi: 10.18805/LRF-763.

Submitted: 12-08-2023 **Accepted:** 05-12-2023 **Online:** 09-01-2024

different organisms (Bulmer *et al.*, 1991). Codon usage bias is an important evolutionary feature of organisms, which can provide important information for studying biological evolution, gene function and foreign gene expression. Since the chloroplast genome is easier to sequence than the nuclear genome, the research on codon preference mainly focuses on the chloroplast genome.

Astragalus is the largest genus in *Leguminosae* and is widely distributed in the Northern hemisphere, south America and Africa. At present, more than 30 medicinal plants of

Astragalus have been found, in which diterpenoids have obvious clinical effects (Lei *et al.*, 2016). *Astragalus Mongolicus* and *Astragalus Membranaceus* are both medicinal and edible plants, used to keep healthy. The *Astragalus* genus is used medicinally, representing a major traditional Chinese medicinal material in our country, known for its effects of replenishing qi, consolidating the superficial resistance, detoxification, promoting pus discharge, diuresis and promoting tissue regeneration. In recent years, pharmacological research has discovered that the primary medicinal component in *Astragalus* is Astragaloside. Among them, Astragaloside IV is its most significant monomeric active ingredient, with effects including tumor inhibition, as well as antibacterial and anti-inflammatory properties. Although the chloroplast genomes of most plants in its family have been sequenced and annotated, there are fewer studies on comparative chloroplast genomics research and codon preference research of *Astragalus*. In this study, we collected three valuable medicinal *Astragalus* species from the northern regions and systematically compared the chloroplast genomes and codon usage characteristics of the *Astragalus* genus.

MATERIALS AND METHODS

Plant material, DNA extraction and sequencing

Fresh leaf samples of *Astragalus dahuricus* (Pall.) DC., *Astragalus melilotoides* Pall. var. *tenuis* Ledeb. and *Astragalus laxmannii* Jacq. were collected in May 2022 from National Germplasm Perennial Herbage Nursery of the Inner Mongolia Horticulture Research Institute (40.57°N, 111.93°E and altitude 1040 m above sea level), Hohhot, Inner Mongolia Autonomous Region, China. These samples had been formally identified by Juan Zhang before collection, an expert in plant taxonomy. Leaf samples were frozen immediately in liquid nitrogen, conserved in dry ice, extracted by the modified CTAB method (Yan *et al.*, 2018), used for library preparation and paired-end (PE) sequencing by the Illumina Novaseq instrument at Novogene CO., Ltd (Beijing, China).

Genome assembly and annotation

Genomic DNA was extracted from fresh leaves using a Plant DNA Isolation Kit (Tiangen, Beijing, China) and sequenced using the MiSeq PE150 platform (Illumina, San Diego, CA, United States), yielding 150 bp paired-end reads, at Novogene Co. (Tianjing, China). Chloroplast genomes of *Astragalus mongholicus* var. *dahuricus*, *Astragalus melilotoides* var. *tenuis* Ledeb. and *Astragalus laxmannii* Jacq. *de novo* assembled using NOVO Plasty (Dierckxsens *et al.*, 2020) with default parameters. Genomes were annotated using the plastid genome annotator (PGA) tool (Biomatters Limited 2018), coupled with manually edited start and stop codons using Geneious (Pánek *et al.*, 2022). The *Astragalus mongholicus* Cp genome sequence (NCBI accession

number: NC029828) was used as a reference. The annotation results were checked using the Dual Organellar GenoMe Annotator (DOGMA) (Shi *et al.*, 2019) and Cp GAVAS2 (Stephan *et al.*, 2019). OGDRAW (version 1.3.1) (Kurtz *et al.*, 2004) was used to draw the gene map of the Cp genomes.

Codon usage

Codons encoding the same amino acid are called synonymous codons and the difference in use frequency of synonymous codons is the CUB. In order to ensure the accuracy of the results, we eliminated sequences less than 300 bp before codon analysis (Ikemura, 1981; Wang *et al.*, 2018b) and the CUB was calculated using Codon W v1.4.2. We also analyzed the effective number of codons (ENC) (He *et al.*, 2016) and relative synonymous codon usage (RSCU). ENC refers to the effective number of codons and the range of its theoretical value is 20-61, representing the strength of codon bias. The larger the ENC value, the weaker the codon bias. RSCU refers to the relative probability between synonymous codons encoding corresponding amino acids for a particular codon. If there is no preference for the use of codons, the RSCU value of the codon is equal to 1.00. When the RSCU value of a codon is greater than 1.00, it indicates that the frequency of the codon use is relatively high and *vice versa*. The codon base composition and other relative parameters of three *Astragalus* were analyzed and calculated using Codon W v1.4.2 and the optimal codon usage frequency of other fungi and plants.

Codon preference influencing factor analysis

Parity rule 2 (PR2) analysis

We analyzed the distribution of the third base in the codon of amino acid, because it may affect the preference of codon usage (Wan *et al.*, 2004). If codon usage is only influenced by mutations, theoretically, the frequency of using A/T and G/C at the third position of the codon should be equal. Otherwise, codon bias may be influenced by natural selection and other factors (Zhang *et al.*, 2007). In order to analyze the composition of the third codon in CDS sequence, eight amino acids with four synonymous codons were selected for analysis, including serine (S), leucine (L), proline (P), arginine (R), threonine (T), valine (V), alanine (A) and glycine (G). The result was showed by using $G_3/(G_3+C_3)$ and $A_3/(A_3+T_3)$ as abscissa and ordinate respectively.

Neutral analysis

The chloroplast genomes of three *Astragalus* species were analyzed by neutral mapping. GC_3 content of each gene was used as abscissa, GC_{12} as ordinate and an R package was used to draw scatter plot and make straight line fitting. If the regression coefficient is close to 1, the codon preference is mainly affected by mutations. If the regression coefficient is close to 0, the codon preference is mainly affected by natural selection (Song *et al.*, 2017).

RESULTS AND DISCUSSION

Chloroplast genome characteristics of *Astragalus*

The complete chloroplast genome of the three *Astragalus* species were found to have lost one copy of the IR region, the total lengths of the chloroplast genomes were 122,815 bp (*A. dahuricus*), 123,012 bp (*A. melilotoides* Pall. var. *tenuis*) and 123,729 bp (*A. laxmannii* Jacq.), while the total GC content ranged from 33.95% to 34.10% (Fig 1). There were 107-110 genes, including 75-76 protein-coding genes (PCGs), four ribosomal RNA genes (rRNAs) and 28-30 transfer RNA genes (tRNAs). *A. dahuricus* had the fewest genes and lacked *atpE*, *tmE-UUC**, *trnM-CAU* (Table 1).

Relative synonymous codon usage analysis of the chloroplast genomes

The RSCU of the chloroplast genomes of three species of *Astragalus* was calculated using all protein-coding genes. The RSCU value is the ratio of the frequency of use of a particular codon to the expected frequency. It enables the detection of synonymous codons that do not uniformly occur in the coding sequence. Codons with no preference value are set to 1.00. The actual usage of codons with an RSCU value >1.00 is higher than expected and that of codons with an RSCU value <1.00 is lower than expected. The results showed that all species, there were 30 codons with a RSCU value greater than 1.00, of which, 29 ended with A or U codons one ended with G codon (UUG). In addition, the

RSCU value of methionine (AUG) and tryptophan (UGG) was 1.00. Isoleucine was the most amino acid encoded in the chloroplast genome, accounting for 10% on average of all amino acids. Methionine had the smallest number of codons, accounting for only 1.94% on average of all amino acids (Fig 3).

Fig 2 shows the codon contents of 20 amino acids and stop codons of all protein-coding genes in the chloroplast genomes of the three species of *Astragalus* sequenced in this study. Among the codons exhibiting usage bias, there are 53 with a relative synonymous codon usage (RSCU) greater than 1.

Optimal codons

The ENC value of the CDS of the three *Astragalus* chloroplast genome is sorted from low to high and the genetic construction library of 10% of the genes at both ends are taken. Among them, the ENC value is small as a high expression library and the large value of ENC is a low expression library. The codons of the Δ RSCU > 0.08 in the two libraries are used as high expression codons and RSCU > 1.00 codons are used as high-frequency codons. The optimal codons are both high expression and high-frequency dense codons. There are 12 optimal codons that meet RSCU greater than 1 and Δ RSCU greater than 0.08. Among them, 14 best codons are shared by three *Astragalus* species. All the best codons ended with A/U. There are 3 ending at the end of A and 8 at the end of U (Table 2).

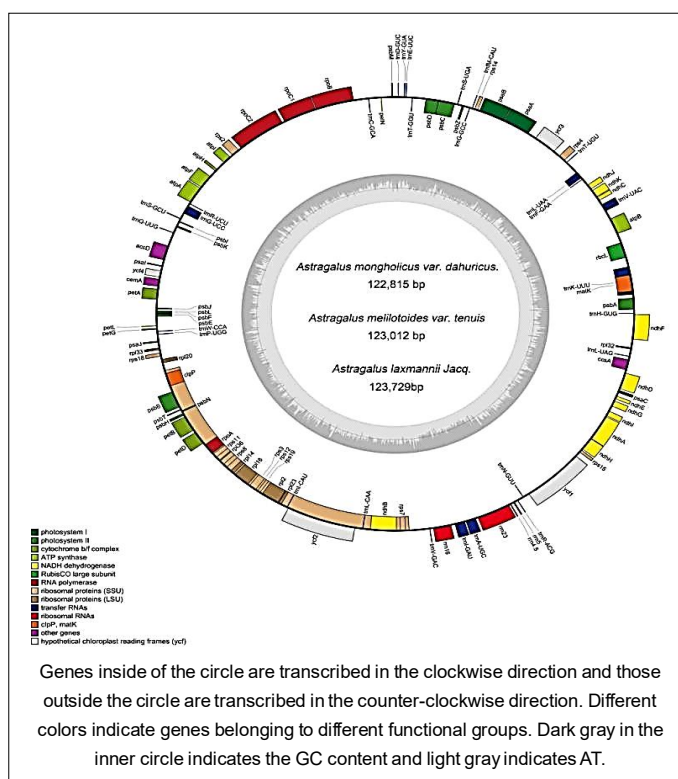


Fig 1: Structure and characteristics of the complete chloroplast genomes of three *Astragalus* species.

Table 1: List of genes encoded by three species of *Astragalus*.

Category	Gene group	Gene name
Photosynthesis	Subunits of photosystem I	<i>psaA, psaB, psaC, psal, psaJ</i>
	Subunits of photosystem II	<i>psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ</i>
	Subunits of NADH dehydrogenase	<i>ndhA*, ndhB*, ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK</i>
	Subunits of cytochrome b/f complex	<i>petA, petB*, petD*, petG, petL, petN</i>
	Subunits of ATP synthase	<i>atpA, atpB, atpF*, atpH, atpI</i>
	Large subunit of rubisco	<i>rbcL</i>
	Subunits photochlorophyllide reductase	-
Self-replication	Proteins of large ribosomal subunit	<i>rpl14, rpl16*, rpl2*, rpl20, rpl23, rpl32, rpl33, rpl36</i>
	Proteins of small ribosomal subunit	<i>rps11, rps12*, rps14, rps15, rps18, rps19, rps2, rps3, rps4, rps7, rps8</i>
	Subunits of RNA polymerase	<i>rpoA, rpoB, rpoC1*, rpoC2</i>
	Ribosomal RNAs	<i>rrn16, rrn23, rrn4.5, rrn5</i>
	Transfer RNAs	<i>trnA-UGC*, trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, trnG-GCC, trnG-UCC*, trnH-GUG, trnI-CAU, trnI-GAU*, trnK-UUU*, trnL-CAA, trnL-UAA*, trnL-UAG, trnN-GUU, trnP-UGG, trnQ-UUG, trnR-ACG, trnR-UCU, trnS-GCU, trnS-UGA, trnT-GGU, trnT-UGU, trnV-GAC, trnV-UAC*, trnW-CCA, trnY-GUA, trnM-CAU</i>
	Other genes	Maturase <i>matK</i> Protease <i>clpP*</i> Envelope membrane protein <i>cemA</i> Acetyl-CoA carboxylase <i>accD</i> c-type cytochrome synthesis gene <i>ccsA</i> Translation initiation factor - other -
Genes of unknown function	Conserved hypothetical chloroplast ORF <i>ycf1, ycf2, ycf3**, ycf4</i>	

* and ** indicate genes containing one/two introns.

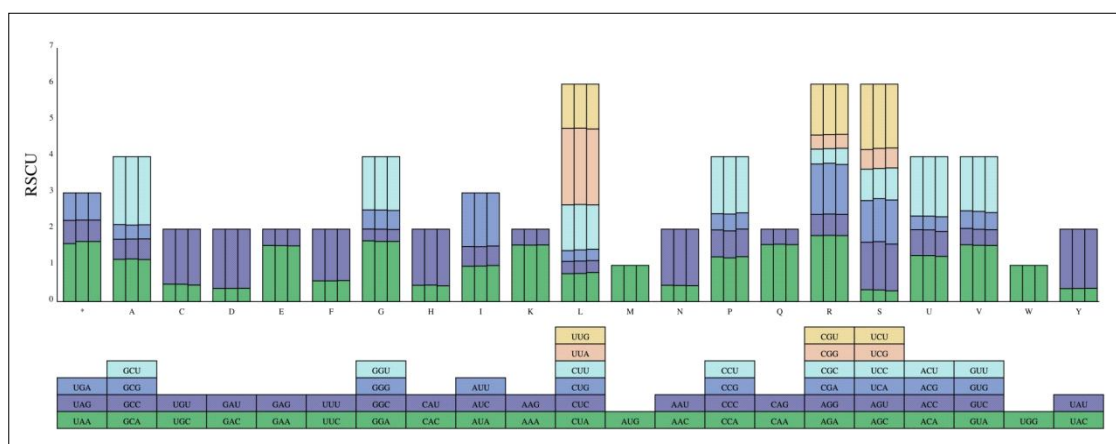


Fig 2: Codon distribution of 20 amino acid and stop codons in all protein-coding genes of the chloroplast genomes of three *Astragalus* species.

Comparison of the Complete Chloroplast Genomes in the *Astragalus*

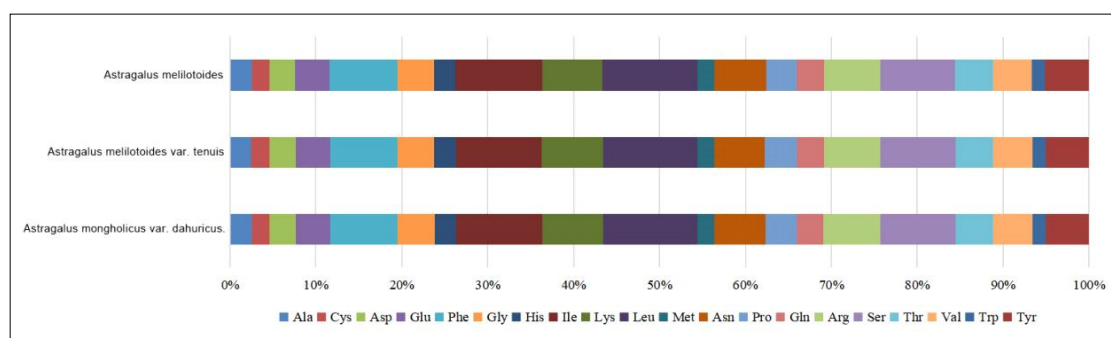


Fig 3: Amino acid proportion of protein-coding genes in three *Astragalus* chloroplast genomes.

Table 2: Optimal codons analysis of chloroplast genome in Glycyrrhiza species in *Astragalus*.

Amino acids	Codons	<i>Astragalus dahuricus</i>	<i>Astragalus melilotoides</i>	<i>Astragalus laxmannii</i>
		(Pall.) DC. ΔRSCU	Pall. var. tenuis Ledeb. ΔRSCU	Jacq. ΔRSCU
Phe	UUU	-0.48	-0.43	-0.44
	UUC	0.48	0.43	0.44
Leu	UUA	-0.29	-0.31	-0.17
	UUG	-0.14	-0.11	-0.2
	CUU	-0.3	-0.38	-0.4
	CUC	-0.02	0.06	0.11
	CUA	0.53*	0.46*	0.4*
	CUG	0.21	0.29	0.25
Ile	AUU	0.5*	0.59*	0.52*
	AUC	0.36	0.27	0.33
	AUA	-0.86	-0.86	-0.86
Val	GUU	0.18*	0.06	0.14*
	GUC	-0.03	0.03	0.08
	GUA	0.09*	0.33*	0.23*
	GUG	-0.24	-0.42	-0.45
Tyr	UAU	-0.33	-0.37	-0.27
	UAC	0.33	0.37	0.27
His	CAU	-0.02	-0.13	0
	CAC	0.02	0.13	0
Gln	CAA	-0.1	-0.12	-0.08
	CAG	0.1	0.12	0.08
Asn	AAU	-0.45	-0.46	-0.44
	AAC	0.45	0.46	0.44
Lys	AAA	0.33*	0.36*	0.36*
	AAG	-0.33	-0.36	-0.36
Asp	GAU	0.08	-0.02	0.03
	GAC	-0.08	0.02	-0.03
Glu	GAA	-0.16	-0.17	-0.13
	GAG	0.16	0.17	0.13
Ser	UCU	1.13*	1.06*	1.03*
	UCC	0.22	0.33	0.44
	UCA	-1.01	-1.05	-1.15
	UCG	-0.7	-0.68	-0.63
	AGU	0.23*	0.2*	0.22*
	AGC	0.15	0.14	0.11

Table 2: Continue...

Comparison of the Complete Chloroplast Genomes in the *Astragalus*

Table 2: Continue...

Pro	CCU	0.92*	0.85*	1.12*
	CCC	-0.61	-0.61	-0.54
	CCA	-0.25	0.02	-0.45
	CCG	-0.06	-0.26	-0.13
Thr	ACU	0.57*	0.78*	0.78*
	ACC	1.17	0.92	1.05
	ACA	-1.29	-1.23	-1.39
	ACG	-0.46	-0.49	-0.46
Ala	GCU	0.84*	0.84*	0.71*
	GCC	-0.28	-0.41	-0.26
	GCA	-0.64	-0.5	-0.63
	GCG	0.06	0.06	0.17
Cys	UGU	-0.08	0	-0.08
	UGC	0.08	0	0.08
Arg	CGU	2.24*	2.01*	1.99*
	CGC	0.46	0.67	0.63
	CGA	-0.16	-0.19	-0.1
	CGG	0.34	0.4	0.29
	AGA	-2.71	-2.68	-2.36
	AGG	-0.17	-0.21	-0.45
Gly	GGU	1.3*	1.14*	1.47*
	GGC	0.04	0.04	-0.07
	GGA	-0.97	-0.81	-1

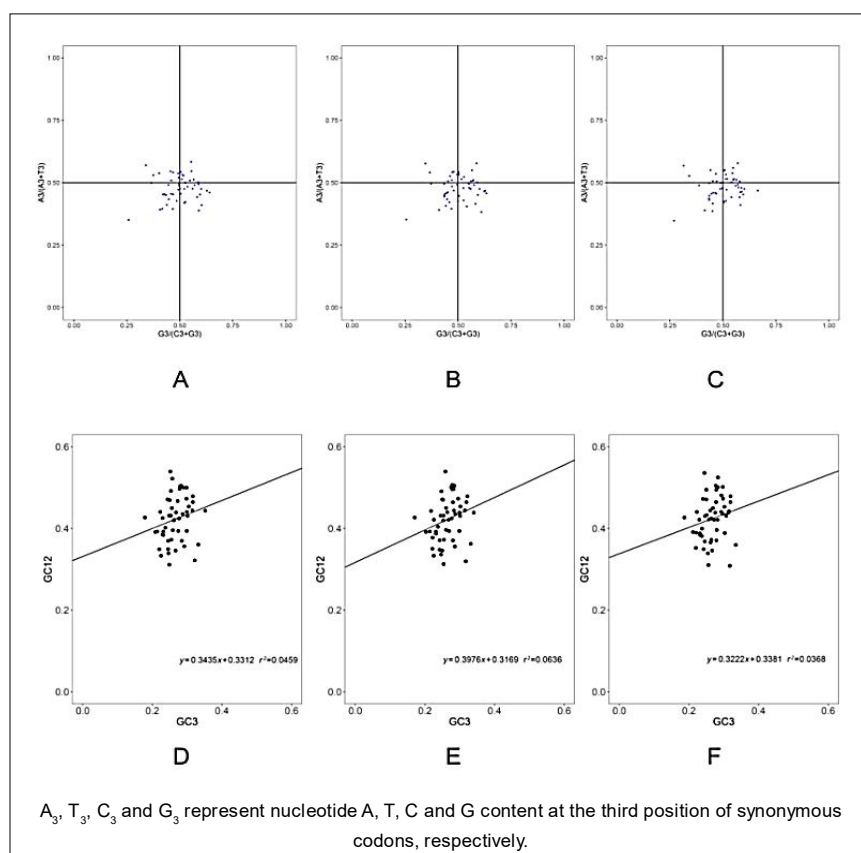


Fig 4: PR2 analysis for genes and neutral analysis in the Cp genomes of 3 *Astragalus* species.

Determining codon preference factors

In this study, the parity rule 2 (PR2) of codon A/T (A_3 and T_3) and C/G (G_3 and C_3) of three *Astragalus* chloroplast genomes were analyzed (Fig 4). The results showed that the coordinate points were unevenly distributed in four regions, mainly distributed in $G_3/(G_3+C_3) < 0.5$ and $A_3/(A_3+T_3) < 0.5$. The results further showed that the usage preference of codons affected by base. In order to further determine the main factors affecting codon preference in the chloroplast genomes of *Astragalus*, neutral mapping analysis were conducted on the chloroplast genomes of three *Astragalus*. It found that GC_{12} distributed in 0.312~0.586 and GC_3 distributed in 0.168~0.329. The regression coefficients ranged from 0.0368 to 0.0636 and GC_{12} was positively correlated with GC_3 . The results showed that the codon usage of chloroplast genome in three *Astragalus* plants were affected not only by natural selection, but also mutation.

Astragalus plants and other species of codon usage comparison

According to the calculations performed by Codon W, the usage frequencies of each codon in the chloroplasts of the three *Astragalus* species were compared with the codon usage frequencies of various organisms, including *Escherichia coli*, *Saccharomyces*, *Arabidopsis thaliana*, *Glycine max* and *Nicotiana tabacum* as published in the Codon Usage Database. The results indicate that when comparing with other organisms, for *Astragalus* plants, there is only *Arabidopsis thaliana* with a codon usage frequency ratio of 0.5 to 2.0. With a ratio greater than 2.0, there are 2 codons when compared with *Escherichia coli* and 4 codons when compared with *Saccharomyces*. When compared with *Glycine max*, there are 3 codons with a ratio greater than 2.0 and 3 codons with a ratio less than 0.5. These findings suggest that *Astragalus* plants are different from the use of 5 modes of plant codons usage.

CONCLUSION

In the present study, we sequenced and annotated the Cp genomes of *Astragalus dahuricus* (Pall.) DC., *Astragalus melilotoides* Pall. var. *tenuis* Ledeb. and *Astragalus laxmannii* Jacq.. *Astragalus* chloroplast genome ranges from 122,815 bp (*Astragalus dahuricus*) to 123,729 bp (*Astragalus melilotoides* Pall.). There were 107-110 genes, including 75-76 protein-coding genes (PCGs), four ribosomal RNA genes (rRNAs) and 28-30 transfer RNA genes (tRNAs). The chloroplast genomes of those three *Astragalus* species differ little. By analyzing the characteristics of codon usage, the study showed that the codon preference factor was affected by natural selection, the three Cp genomes from *Astragalus* had the same 11 optimal codons and ending with A/U. The codon preferences of *Astragalus* and 5 model creatures were pretty significant. By uncovering the codon usage bias and its influencing factors in the chloroplast genomes of the three

Astragalus species, it is possible to optimize and modify the *Astragalus* genome's codons, analyze and predict unknown genes and provide a theoretical basis for the improvement of *Astragalus* varieties. Additionally, codon bias influences the expression of exogenous genes in the host, making codon usage bias analysis crucial for enhancing the expression of exogenous genes in *Astragalus* species.

ACKNOWLEDGEMENT

Junjie Wang designed the experiment, Dabao Yin performed the experiments and wrote the manuscript, others contribute equally to the manuscript. This research received no external funding by "Integration of efficient breeding and processing of multi-functional grass seed" Award Number (2022YFDZ0025)

Conflict of interest

The authors declare that they have no conflicts of interest.

REFERENCES

- Archtti, M. (2004). Codon usage bias and mutation constraints reduce the level of error minimization of the genetic code. *Mol. Evol.* 59(2): 258-266.
- Biomatters Limited; VIB Accelerates. (2018). The rapeutic Biologics Discovery with Geneious Biologics. *Biotech Week*.
- Bulmer, M.G. (1991). The selection-mutation-drift theory of synonymous codon usage. *Genetics.* 129(3): 897-907.
- Dierckxsens, N., Mardulyn, P., Smits, G. (2020). Unraveling heteroplasmy patterns with NOVOPlasty. *NAR Genomics and Bioinformatics.* 2(1): 1-10.
- Fennoy, S.L., Bailey-Serres, J. (1993). Synonymous codon usage in *Zea mays* L. nuclear genes is varied by levels of C and G-ending codons. *Nucleic Acids Res.* 21(23): 5294-5300.
- He, B., Dong, H., Jiang, C., Cao, F., Tao, F., Xu, L.A. (2016). Analysis of codon usage patterns in *Ginkgo biloba* reveals codon usage tendency from A/U-ending to G/C-ending. *Scientific Reports.* 6: 35927.
- Ikemura, T. (1981) Correlation between the abundance of *Escherichiacoli* transfer RNAs and the occurrence of the respective codons in its protein genes: A proposal for a synonymous codon choicethat is optimal for the *E. coli* translational system. *Mol. Biol.* 151(3): 389-409.
- Kurtz, S, Phillippy, A., Delcher, A., Smoot, M., Shumway, M., Antonescu, C., Salzberg, S.L. (2004). Versatile and open software for comparing large genomes. *Genome Biology.* 5: R12.
- Lei, W.J., Ni, D.P., Wang, Y.J., Shao, J.J., Wang, X.C., Yang, D., Wang, J.S., Chen, H.M., Liu, C. (2016). Intraspecific and heteroplasmic variations, gene losses and inversions in the chloroplast genome of *Astragalus membranaceus*. *Scientific Reports.* 6(1): 1-13.
- Liu, X., Chang, E.M., Liu, J.F., Huang, Y.N., Wang, Y., Yao, N. (2019). Complete chloroplast genome sequence and phylogenetic analysis of *quercus bawanglingensis* huang, Li et xing, a vulnerable oak tree in China. *Forests.* 10: 0587.

- Olmstead, R. G., Palmer, J.D. (1994). Chloroplast DNA systematics: A review of methods and data analysis. *American Journal of Botany*. 81: 1205-1224.
- Pánek, T., Barcytė, D., Treitli, S.C., Záhonová, K., Sokol, M., Ševčíková, T., Zadrobílková, E., Jaške, K., Yubuki, N., Cepika, I., Eliáš, M. (2022). A new lineage of non-photosynthetic green algae with extreme organellar genomes. *BMC Biology*. 20(1): 66.
- Sharm, P., Li, W.H. (1987). The codon adaptation index-a measure of directional synonymous codon usage bias and its potential applications. *Nucleic Acids Res.* 15(3): 1281-1295.
- Shi, L.C., Chen, H.M., Jiang, M., Wang, L.Q., Wu, X., Huang, L.F., Liu, C. (2019). CPGAVAS2, an integrated plastome sequence annotator and analyzer. *Nucleic Acids Research*. 47(W1).
- Song, H., Liu, J., Song, Q., Zhang, Q., Tian, P., Nan, Z. (2017). Comprehensive analysis of codon usage bias in seven *Epicloe* species and their peramine-coding genes. *Front. Microbiol.* 8: 1419.
- Stephan, G., Pascal, L., Ralph, B. (2019). Organellar Genome DRAW (OGDRAW) version 1.3.1: Expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Research*. 47(W1).
- Tangphatsornruang, S., Sangsrakru, D., Chanprasert, J., Uthaisaisanwong, P., Yoocha, T., Jomchai N., *et al.* (2010). The chloroplast genome sequence of mungbean (*vigna radiata*) Determined by High-throughput Pyrosequencing: Structural Organization and Phylogenetic Relationships. *DNA Research*. 17: 11-22.
- Wan, S.Q., Xia, J.Y., Liu, W.X., Niu, S.L., (2009). Photosynthetic overcompensation under nocturnal warming enhances grassland carbon sequestration. *Ecology*. 90(10): 2700-2710.
- Wan, X., Xu, D., Andris, K., Zhou, J. (2004). Quantitative relationship between synonymous codon usage bias and GCC composition across unicellular genomes. *BMC Evol. Biol.* 4(1): 19.
- Wang, P.L., Yang, L.P., Wu, H.Y., Nong, Y.L., Wu, S.C., Xiao, Y.F., Zhao, Z.Y. (2018a). Codon preference of chloroplast genome in *Camellia oleifera*. *Guihaia*. 38(2): 135-144.
- Wang, Z.J., Li, B., Jiang, X.Z., Ou, Z.L., Xu, Z.D., Dai, H.H. (2018b). Comparative analysis of the codon preference patterns in two species of *Camellia sinensis* based on genome data. *Chin. J. Cell Biol.* 40(12): 62-73.
- Wolfe, K.H., Li, W.H., Sharp, P.M. (1987). Rates of Nucleotide Substitution Vary Greatly among Plant Mitochondrial, Chloroplast and Nuclear DNAs. *Proceedings of the National Academy of Sciences of the United States of America*. 84: 9054-9058.
- Yan, J.Y, Ma, C.Q., Bo, C., Fan, X.G., Li, Z., Yang, Y.Z., Zhao, Z.Y. (2018). A modified CTAB method for genomic DNA extraction from apple fruit. *Molecular Plant Breeding*. 9: 3610-3616. DOI: 10.13271/j.mpb.015.003610.
- Zhang, W.J., Zhou, J, Li, Z.F., Wang, L., Gu, X., Zhong, Y. (2007). Comparative analysis of codon usage patterns among mitochondrion, chloroplast and nuclear genes in *Triticum aestivum*. *Integr. Plant Biol.* 49: 246-254.