



Full Length Article

Analysis of Synonymous Codon usage of the Complete Chloroplast Genome in *Phleum pratense* cv. Minshan

Guangxin Cui¹, Chunmei Wang¹, Xiaoxing Wei², Hongshan Yang¹, Yuan Lu¹, Xiaoli Wang¹, Xinqiang Zhu¹, Qian Zhang¹, Yaqin Gao^{3*} and Huirong Duan^{1*}

¹Lanzhou Institute of Husbandry and Pharmaceutical Science, Chinese Academy of Agricultural Sciences, Lanzhou, China

²Academy of Animal and Veterinary Sciences, Qinghai University, Xining, China

³Laboratory of Quality & Safety Risk Assessment for Livestock Products, Ministry of Agriculture and Rural Affairs, Lanzhou Institute of Husbandry and Pharmaceutical Science, Chinese Academy of Agricultural Sciences, Lanzhou, China

*For Correspondence: gaoyaqin@caas.cn; duanhuirong@caas.cn

Received 20 January 2020; Accepted 14 March 2020; Published 31 May 2020

Abstract

Timothy (*Phleum pratense* L.) is extraordinary roughage for racing horses and dairy cows. *Phleum pratense* L. cv. Minshan is the only cultivar in China, and severe variety degeneration restricts its productivity and promotion. Little information on its genetic characteristics is revealed up to now. To fully display the architectures of Minshan chloroplast genome and lay a theory foundation for future gene modification, synonymous codon usage pattern in Minshan including codon nucleotide base preference, optimal codons and the factors affecting synonymous codon usage was inferred. The tools of Codon W, CUSP of EMBOSS were used to calculate nucleotide compositions and codon usage bias parameters. The average ENC value of 49 genes was 49.44, and all CAI values were less than 0.35, suggesting weak preference of synonymous usage in Minshan. Of 10 identified optimal codons, 6 ended with T and 4 ended with A, implying synonymous codons preferred to use A/T ended codons, and these codons were closer to Axis 1 than others. Most points of the coding genes were under the expected curve of ENC plot analysis and except for *InfA*, a weak correlation between GC₃ and GC₁₂ in other 48 genes indicated natural selection was the chief contributor in synonymous codon usage bias in Minshan. The results could lay a theory foundation for codon modification of exogenous genes and molecular assisted breeding in the future. © 2020 Friends Science Publishers

Keywords: Synonymous codon usage; Chloroplast genome; Natural selection; *Phleum pratense* L. cv. Minshan

Introduction

Timothy (*Phleum pratense* L.) is extraordinary roughage because it is rich in fiber promoting performance of racing horses and helping dairy cows maintain high milk production. As the only cultivated variety registered in 1990 in China, *Phleum pratense* L. cv. Minshan (Minshan) is highly adapted to the climate of Minshan County, Gansu Province (Cao 2003). The specific cold and moist climate in turn limits its seed production and breeding. In the last twenty years, the production and promotion of Minshan has been hindered due to variety degeneration (Du 2003). Variety improvement and new cultivars of high quality and yield are eagerly demanded to meet the needs of the emerging racing horse industry (Wang *et al.* 2018a).

Information carried on chloroplast genome (cp genome) has been widely applied to gene mapping, variety identification, plant barcode sequences screening, population genetics, gene diversity study and molecular assisted breeding (Parks *et al.* 2009). Codon is the key link connecting nucleic acids and proteins and plays a vital role

in biological genetic information transmission. Among 20 amino acids forming proteins in organisms, except for methionine and tryptophan which are encoded by unique codons, other 18 amino acids correspond to 2–6 synonymous codons. Synonymous codons are used differently in organisms even in different genes in one genome and different parts of one gene, which is called synonymous codon usage bias (SCUB) (Li *et al.* 2012). SCUB is an important feature of organism evolution and exists in numerous living organisms (Sau *et al.* 2006; Parks *et al.* 2009; Chen *et al.* 2014; Li *et al.* 2019; Zhou *et al.* 2019). SCUB analyses enable the scientific community to increase target gene expression genetically, make the exogenous genes more efficient and stable as well as variety improvement (Li *et al.* 2019).

The cp genome of Minshan was assembled and reported using Illumina pair-end sequencing data (Cui *et al.* 2019). To comprehensively understand the architectures of Minshan cp genomes and provide useful information for molecular assisted breeding, synonymous codon usage of Minshan was studied in this study.

Materials and Methods

Sequence data

Healthy fresh leaves were sampled from Lanzhou Scientific Observation and Experiment Field Station of the Ministry of Agriculture for Ecological System in the Loess Plateau Area (36°01'N, 103°45'E, altitude 1700 m), Gansu, China, on July 22th 2019. The complete cp genome sequencing of Minshan was determined (Genbank accession number: MN551180) based on Illumina NovaSeq platform at Benagen Tech Solution Co., Ltd. (Wuhan, China). GeSeq was employed to annotate the assembled genome (Tillich *et al.* 2017). After filtering the repeated sequences and the sequences length less than 300 bp, 49 sequences with the start codon of ATG, TTG, CTG, ATT, ATC, GTG and ATA, also the end codon of TGA, TAG, and TAA, were used to carry on the subsequent analysis.

Relative synonymous codon usage

A great deal of codon usage indices were acquired via the program Codon W (version 1.3, <https://sourceforge.net/projects/codonw/>), including the relative synonymous codon usage (RSCU) value, the codon adaptation index (CAI), the effective number of codons (ENC), the nucleotides G and C content of all 49 coding sequences (GC), the frequency of G + C at the third position of synonymous codons (GC_{3s}), and the silent base compositions (A_{3s}, T_{3s}, G_{3s}, and C_{3s}) (Li *et al.* 2019). The G+C content at the first, second, third positions of codons (GC₁, GC₂, GC₃) and the average G+C content of the first and second positions (GC₁₂) were calculated by the online CUSP function from EMBOSS (<http://imed.med.ucm.es/EMBOSS/>) (Wang *et al.* 2018b).

Identification of the optimal codon

According to RSCU value of each codon, the highest frequency synonymous codons with the largest RSCU value were identified (Li *et al.* 2019). Using ENC analysis as preference standard, the 49 sequences of Minshan were ordered from high to low, and the highest 5% sequences and the lowest 5% sequences were taken to form the high and low expression gene group, separately. ΔRSCU was subtracted the RSCU value of each codons in the low expressed gene group from the high expressed gene group. The codons with ΔRSCU value larger than 0.08 were recognized as high expressed. The optimal codons were identified as the ones which were high frequency and high expressed (Wang *et al.* 2019).

Correspondence analysis

Correspondence analysis is a vital multivariate tool to explore codon usage change trends (Choudhury *et al.* 2017).

The corresponding analysis of genes and codon bias was carried out by Codon W based on the RSCU values. RSCU values of 49 coding sequences in Minshan were spread into a 40-dimensional vector space. The axes were related to the influencing factors on SCUB and the data of different axes were obtained according to codon base bias and genes. The correlation analysis among GC, ENC, CAI, G_{3s}, GC₃, Axis 1, Axis 2, Axis 3 and Axis 4 was accomplished through SPSS 16.0 based on the Spearman's rank correlation method ($P < 0.05$ or $P < 0.01$). The graphs were depicted using EXCEL 2016.

ENC-plot analysis

ENC displays the degree of codons deviated from random selection (Li *et al.* 2019). ENC value ranges from 20 to 61 with the boundary value of 35 and the ENC value less than 35 means strong codon preference, otherwise, weak codon preference takes place. There is extreme preference when the ENC value is 20. In contrast, there is no preference with the ENC value of 61, indicating random selection of codons (Wang *et al.* 2018b). ENC-plot mapping analysis is used to explore the dominating reason affecting the SCUB. The ENC plot of ENC versus GC_{3s} was drawn by EXCEL 2016. The ENC formula of expected curve is as follows:

$$ENC = 1 + GC_{3s} + \frac{29}{GC_{3s}^2 + (1 - GC_{3s})^2}$$

The genes would be distributed alongside or next to the expected curve when SCUB is merely involved by mutation, however if the points of genes are far away from the expected curve, it illustrates SCUB is primarily affected by natural selection other than mutation pressure (Wang *et al.* 2018b).

Neutrality plot analysis

GC content of the cp genome is highly conserved, while under the condition of various evolution pressures, different bases preferences would happen, and synonymous codon mutation usually occurs at the third position. In the neutral graph, the GC₁₂ value of each gene is used as vertical coordinate, and the corresponding GC₃ value is used as horizontal axis (Wei *et al.* 2014). If the point distributes alongside or nearby the diagonal line which means that GC₁₂ is equal to GC₃, it implies that other evolution pressure excluding mutation pressure is weak. Oppositely, if GC₁₂ and GC₃ correlates weakly, the regression coefficient is approximately to 0, it shows that the base composition of the 3 positions are significantly different, revealing that natural selection is the dominating factor affecting SCUB (He *et al.* 2013).

Statistical analysis

Correlation analysis was conducted by SPSS 16.0 employing the correlation method of Spearman's rank ($P < 0.05$ or $P < 0.01$). The graphs were depicted using EXCEL 2016.

Results

The codon usage pattern of Minshan

Nucleotide A and T were abundant in Minshan cp genome. The average GC, GC₁, GC₂ and GC₃ content of Minshan cp genome were 38, 47, 39 and 30%. The frequencies of A_{3s}, T_{3s}, G_{3s}, C_{3s}, and GC_{3s} were 43, 46, 17, 17 and 27%, respectively. The length of 49 amino acids was between 101 and 1473 with the average of 338. The value of ENC was between 39.91 and 60.65 with an average of 49.44, implying weak preference of synonymous usage. All CAI values of these 49 sequences were less than 0.35, which additionally demonstrated the weak preference of synonymous usage (Table 1).

According to the RSCU values of 61 codons in Minshan cp genome, 18 high frequency synonymous codons with the largest RSCU value, were observed (Table 2). 26 codons were identified as the high expressed codons (Table 3). 10 codons including AGA, TGT, TTT, TAT, TTA, CAA, CAT, GCT, GAA and GTT were located in the intersection set of high frequency codons and high expressed codons and they were identified as the optimal codons, of which, 6 ended with T and 4 ended with A, implying the synonymous codons is biased in using A and T ended codons in Minshan cp genome.

Correspondence analysis of SCUB

Synonymous codons and the cluster of 49 coding genes in Minshan were characterized by different color points in 40 dimensional axes and went through the correspondence analysis (Fig. 1, 2 and 3). The first four axes comprised 33.35% of the whole variation and Axis 1 and Axis 2 were the two chief contributors to SCUB in Minshan, which accounted for 9.18 and 8.85% of the total variation, respectively (Fig. 1). The codons ending with G or C distributed dispersedly and away from Axis 1 and Axis 2 while A or T ended codons were closer to Axis 1 (Fig. 2), indicating the nucleotide constitution for mutation pressure may associate with SCUB. Different gene types showed different distribution patterns (Fig. 3). The genes of photosystem I and photosystem II distributed in the third and fourth quadrant, and rubisCO large subunit distributed next to Axis 1. For the genes of Cytochrome b/f complex, the *petA* and *petD* distributed in the second and third quadrant, *petB* was close to Axis 1. Moreover, *rpoB* and *ropCI* points were also close to Axis 1. However, ATP synthase, NADH dehydrogenase, Ribosomal proteins (LSU) and Ribosomal proteins (SSU) distributed discretely, indicating that other SCUB influencing factors such as natural selection may work.

Correlation analysis

To fully explore the SCUB affected by evolution pressure of

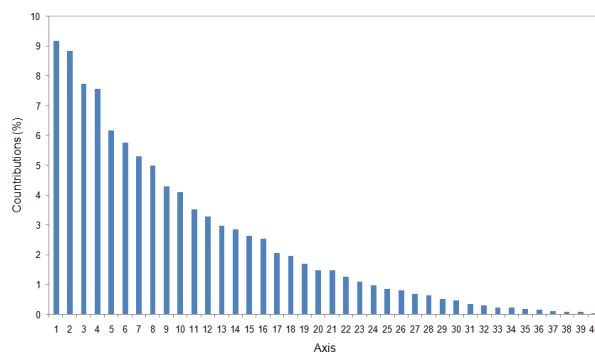


Fig. 1: Contributions of 40 axes from a correspondence analysis

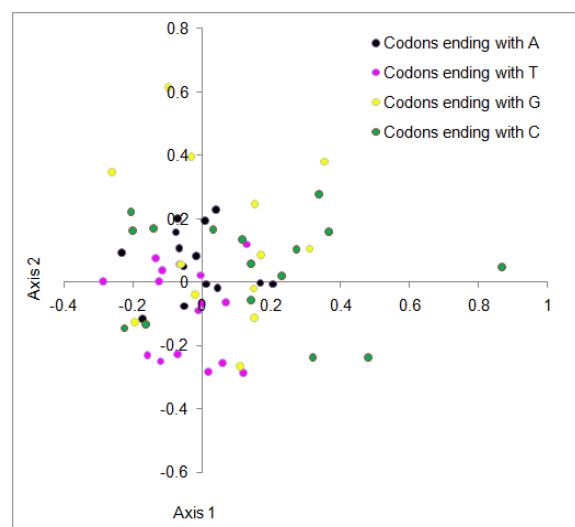


Fig. 2: Correspondence analysis of synonymous codon usage towards the codons in Minshan cp genome

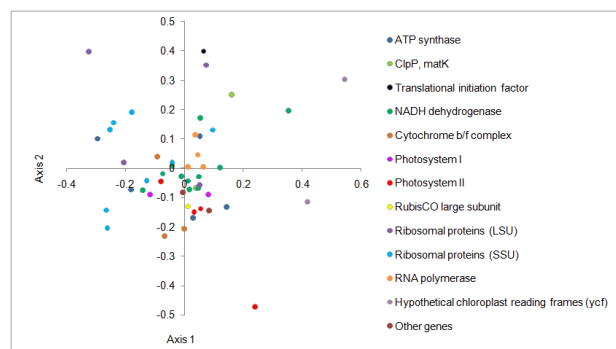


Fig. 3: Correspondence analysis of synonymous codon usage towards the coding genes in Minshan cp genome

mutation or natural selection, correlation among GC, ENC, CAI, GC_{3s}, GC₃, Axis 1, Axis 2, Axis 3, and Axis 4 were calculated (Table 4). Axis 1 showed significant positive correlation with ENC, GC_{3s} and GC₃ ($r = 0.415^{**}$, $P < 0.01$; $r = 0.556^{**}$, $P < 0.01$; $r = 0.599^{**}$, $P < 0.01$) while Axis 2 exhibited significant positive correlation with ENC ($r =$

Table 1: Codon usage indices of 49 coding genes in Minshan cp genome

Gene	GC	GC ₃	GC _{3s}	CAI	ENC	Gene	GC	GC ₃	GC _{3s}	CAI	ENC
<i>atpA</i>	0.41	0.29	0.28	0.18	48.20	<i>psaB</i>	0.41	0.32	0.28	0.17	47.90
<i>atpB</i>	0.42	0.30	0.29	0.20	46.20	<i>psbA</i>	0.42	0.33	0.29	0.31	39.91
<i>atpE</i>	0.41	0.34	0.31	0.18	57.47	<i>psbB</i>	0.43	0.28	0.25	0.20	47.04
<i>atpF</i>	0.38	0.31	0.28	0.15	49.63	<i>psbC</i>	0.44	0.34	0.30	0.19	49.17
<i>atpI</i>	0.39	0.32	0.29	0.16	48.40	<i>psbD</i>	0.44	0.34	0.30	0.24	49.69
<i>ccsA</i>	0.35	0.28	0.23	0.14	47.33	<i>rbcL</i>	0.44	0.30	0.27	0.29	48.10
<i>cemA</i>	0.34	0.30	0.27	0.16	54.91	<i>rpl14</i>	0.39	0.23	0.22	0.18	48.68
<i>clpP</i>	0.42	0.36	0.30	0.18	52.22	<i>rpl16</i>	0.45	0.27	0.22	0.13	44.40
<i>infA</i>	0.40	0.39	0.37	0.19	61.00	<i>rpl20</i>	0.36	0.31	0.27	0.12	52.92
<i>matK</i>	0.33	0.28	0.25	0.17	48.09	<i>rpl22</i>	0.36	0.33	0.30	0.17	49.89
<i>ndhA</i>	0.34	0.33	0.20	0.13	44.95	<i>rpoA</i>	0.36	0.27	0.26	0.14	50.83
<i>ndhB</i>	0.38	0.33	0.30	0.16	46.76	<i>rpoB</i>	0.40	0.31	0.29	0.15	50.32
<i>ndhC</i>	0.41	0.36	0.30	0.17	59.32	<i>rpoC1</i>	0.40	0.31	0.29	0.16	51.21
<i>ndhD</i>	0.36	0.30	0.26	0.13	47.86	<i>rpoC2</i>	0.39	0.31	0.30	0.16	52.51
<i>ndhE</i>	0.33	0.28	0.25	0.13	57.52	<i>rps11</i>	0.44	0.24	0.22	0.18	43.39
<i>ndhF</i>	0.34	0.26	0.21	0.15	46.66	<i>rps14</i>	0.39	0.31	0.28	0.14	49.68
<i>ndhG</i>	0.35	0.25	0.21	0.13	48.32	<i>rps18</i>	0.32	0.26	0.24	0.15	46.59
<i>ndhH</i>	0.38	0.29	0.24	0.15	49.59	<i>rps2</i>	0.37	0.31	0.27	0.17	47.85
<i>ndhI</i>	0.35	0.27	0.25	0.17	44.98	<i>rps3</i>	0.34	0.27	0.26	0.20	48.62
<i>ndhJ</i>	0.39	0.30	0.26	0.16	50.46	<i>rps4</i>	0.37	0.25	0.23	0.16	48.07
<i>ndhK</i>	0.38	0.31	0.27	0.16	52.06	<i>rps7</i>	0.40	0.24	0.21	0.16	48.31
<i>petA</i>	0.41	0.32	0.31	0.17	49.09	<i>rps8</i>	0.36	0.26	0.23	0.11	44.14
<i>petB</i>	0.40	0.30	0.23	0.20	43.25	<i>ycf3</i>	0.41	0.48	0.44	0.14	60.65
<i>petD</i>	0.40	0.31	0.28	0.16	48.83	<i>ycf4</i>	0.41	0.34	0.30	0.17	47.57
<i>psaA</i>	0.43	0.34	0.30	0.19	51.85	Average	0.39	0.30	0.27	0.17	49.44

Table 2: Codon usage in Minshan cp genome

Amino acid	Codon	Number	RSCU	Amino acid	Codon	Number	RSCU
Ala (A)	GCT	460	1.75	Asn (N)	AAT	485	1.49
	GCC	146	0.56		AAC	165	0.51
	GCA	323	1.23		Pro (P)	CCT	274
GCG	120	0.46	CCC	165		0.92	
Cys (C)	TGT	128	1.48	CCA		192	1.07
	TGC	45	0.52	CCG	86	0.48	
Asp (D)	GAT	485	1.53	Gln (Q)	CAA	442	1.52
	GAC	150	0.47		CAG	140	0.48
Glu (E)	GAA	654	1.45	Arg (R)	CGT	227	1.39
	GAG	245	0.55		CGC	94	0.58
Phe (F)	TTT	634	1.35		CGA	208	1.28
	TTC	307	0.65	CGG	65	0.4	
Gly (G)	GGT	388	1.27	AGA	275	1.69	
	GGC	131	0.43	AGG	108	0.66	
	GGA	486	1.59	Ser (S)	TCT	330	1.66
	GGG	221	0.72		TCC	223	1.12
His (H)	CAT	273	1.46	TCA	201	1.01	
	CAC	102	0.54	TCG	101	0.51	
Ile (I)	ATT	696	1.51	AGT	255	1.28	
	ATC	253	0.55	AGC	85	0.43	
Lys (K)	ATA	438	0.95	Thr (T)	ACT	382	1.74
	AAA	605	1.48		ACC	162	0.74
Leu (L)	AAG	212	0.52	ACA	241	1.1	
	TTA	638	2.11	ACG	93	0.42	
	TTG	333	1.1	Val (V)	GTT	364	1.52
	CTT	389	1.29		GTC	115	0.48
	CTC	119	0.39		GTA	356	1.49
	CTA	251	0.83	GTG	120	0.5	
CTG	86	0.28	Tyr (Y)	TAT	473	1.55	
Met (M)	ATG	386		1	TAC	139	0.45
Trp (W)	TGG	311	1				

The highest frequency used synonymous codons (the largest RSCU value) are in bold RSCU, relative synonymous codon usage

0.496**, $P < 0.01$) but significant negative correlation with CAI ($r = -0.424$ **, $P < 0.01$) and Axis 3 negatively correlated with GC and CAI significantly ($r = -0.410$ **, $P <$

0.01; $r = -0.362$ *, $P < 0.05$), moreover Axis 4 showed no significant correlation with other indices, suggesting Axis 1 and Axis 3 were the major contributors for codon nucleotide

Table 3: The codon statistics within high and low expressed genes and ΔRSCU value for each codon in Minshan cp genome

Amino acid	Codon	High expressed gene		Low expressed gene		ΔRSCU	Amino acid	Codon	High expressed gene		Low expressed gene		ΔRSCU
		Frequency	RSCU	Frequency	RSCU				Frequency	RSCU	Frequency	RSCU	
Ala (A)	GCT*	11	1.63	0	0.80	0.83	Asn (N)	AAT	8	1.33	9	1.50	-0.17
	GCC	4	0.59	5	2.00	-1.41		AAC*	4	0.67	3	0.50	0.17
	GCA	8	1.19	3	1.20	-0.01		Pro (P)	CCT	3	0.80	5	1.43
GCG*	4	0.59	0	0.00	0.59	CCC*	5		1.33	1	0.29	1.04	
Cys (C)	TGT*	2	1.33	3	1.20	0.13	CCA		3	0.80	5	1.43	-0.63
	TGC	1	0.67	2	0.80	-0.13	CCG*	4	1.07	3	0.86	0.21	
Asp (D)	GAT	5	1.43	5	1.43	0.00	Gln (Q)	CAA*	6	1.50	7	1.40	0.10
	GAC	2	0.57	2	0.57	0.00		CAG	2	0.50	3	0.60	-0.10
Glu (E)	GAA*	15	1.67	11	1.29	0.38	Arg (R)	CGT*	6	1.24	5	1.07	0.17
	GAG	3	0.33	6	0.71	-0.38		CGC	2	0.41	5	1.07	-0.66
Phe (F)	TTT*	5	2.00	9	0.90	1.10	CGA*	9	1.86	5	1.07	0.79	
	TTC	0	0.00	11	1.10	-1.10		CGG	0	0.00	5	1.07	-1.07
Gly (G)	GGT	7	1.40	9	1.50	-0.10	AGA*	11	2.28	4	0.86	1.42	
	GGC*	4	0.80	2	0.33	0.47		AGG	1	0.21	4	0.86	-0.65
	GGA	7	1.40	9	1.50	-0.10		Ser (S)	TCT	0	0.00	8	1.92
GGG	2	0.40	4	0.67	-0.27	TCC*	6		2.57	5	1.20	1.37	
His (H)	CAT*	2	2.00	0	0.00	2.00	TCA		1	0.43	5	1.20	-0.77
	CAC	0	0.00	2	2.00	-2.00	TCG	0	0.00	2	0.48	-0.48	
Ile (I)	ATT	11	1.18	22	1.69	-0.51	AGT*	6	2.57	3	0.72	1.85	
	ATC*	5	0.54	6	0.46	0.08		AGC	1	0.43	2	0.48	-0.05
	ATA*	12	1.29	11	0.85	0.44		Thr (T)	ACT	3	1.00	5	1.54
Lys (K)	AAA*	13	1.63	10	0.95	0.68	ACC		1	0.33	4	1.23	-0.90
	AAG	3	0.38	11	1.05	-0.67	ACA*		4	1.33	3	0.92	0.41
Leu (L)	TTA*	6	1.57	4	0.71	0.86	ACG*	4	1.33	1	0.31	1.02	
	TTG	3	0.78	8	1.41	-0.63		Val (V)	GTT*	6	1.60	4	1.00
	CTT	6	1.57	13	2.29	-0.72	GTC		1	0.27	3	0.75	-0.48
	CTC*	2	0.52	1	0.18	0.34	GTA*		7	1.87	6	1.50	0.37
	CTA	2	0.52	6	1.06	-0.54	GTG		1	0.27	3	0.75	-0.48
	CTG*	4	1.04	2	0.35	0.69	Tyr (Y)	TAT*	5	2.00	6	1.00	1.00
Met (M)	ATG	7	1.00	10	1.00	0.00		TAC	0	0.00	6	1.00	-1.00
Trp (W)	TGG	7	1.00	9	1.00	0.00							

RSCU, relative synonymous codon usage

* indicates the high expression codons (ΔRSCU>0.08)

Table 4: Correlation coefficients of the indices influencing codon bias in Minshan cp genome

Indices	GC	ENC	CAI	GC _{3s}	GC ₃	Axis 1	Axis 2	Axis 3	Axis 4
GC	1								
ENC	-0.017	1							
CAI	0.515**	-0.227	1						
GC _{3s}	0.356*	0.614**	0.176	1					
GC ₃	0.334*	0.565**	0.127	0.896**	1				
Axis 1	0.025	0.415**	0.148	0.556**	0.599**	1			
Axis 2	0.086	0.496**	-0.424**	0.199	0.178	0.007	1		
Axis 3	-0.410**	-0.066	-0.362*	-0.03	-0.061	-0.006	0.001	1	
Axis 4	0.055	0.251	-0.064	0.247	0.28	0.011	0.008	-0.009	1

**correlation is significant at the 0.01 level.

*correlation is significant at the 0.05 level.

constitution variation, and Axis 1, Axis 2 and Axis 3 all contributed to SCUB. GC_{3s} correlated with GC₃, ENC and Axis 1 significantly ($r = 0.896^{**}$, $P < 0.01$; $r = 0.614^{**}$, $P < 0.01$; $r = 0.556^{**}$, $P < 0.01$), implying that the codon nucleotide base constitution for the pressure of mutation may affect SCUB. CAI positively correlated with GC ($r = 0.515^{**}$, $P < 0.01$) and negatively correlated with Axis 2 and Axis 3 ($r = 0.424^{**}$, $P < 0.01$; $r = 0.362^*$, $P < 0.01$), indicating natural selection may play a considerable role in SCUB.

ENC plot analysis

Most points of the total 49 genes in Minshan cp genome distributed discretely (Fig. 4). The points of *clpP*, *ndhE*,

rp114, *rps18*, *rps4* and *yef* located on and the points of *ndhH*, *ndhI*, *ndhK* and *ccsA* were close to the expected curve, indicating mutation pressure was the major factor affecting their SCUB. Meanwhile, the rest of genes could be divided into two groups, the points of *infA*, *ndhJ* and *clpP* located above the expected curve and other genes were below the expected curve, both of which were apart from the expected curve implying natural selection affected their SCUB momentarily in the terms of warranting the most effective use of codons.

Neutrality plot analysis

It is an effective way to study the degree of mutation pressure against natural selection in SCUB in cp genome

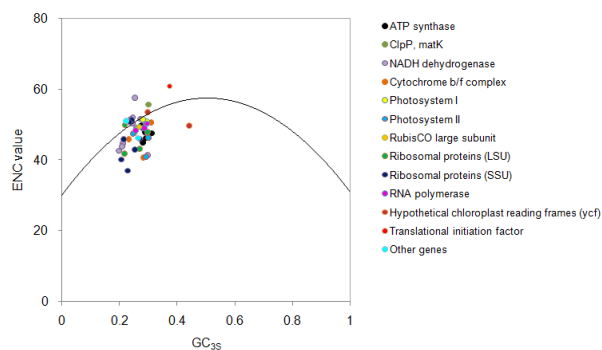


Fig. 4: ENC-plot analysis of Minshan cp genome. ENC, effective number of codons. GC_{3s} , the frequencies of nucleotide G + C at the third position of synonymous codons. The curve shows the expected relationship between ENC values and GC_{3s} under random codon usage assumption

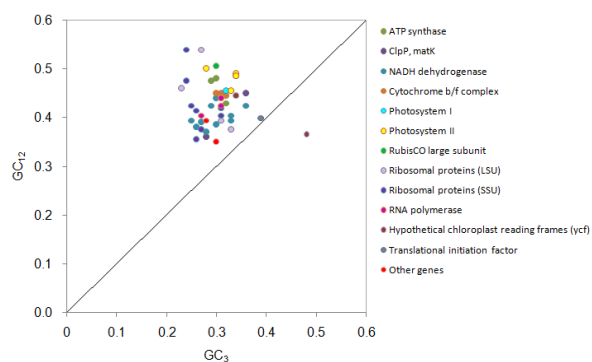


Fig. 5: Neutrality plot analysis of Minshan cp genome. GC_{12} , the average frequencies of nucleotide G + C at the first and second positions of synonymous codons. GC_3 , the frequencies of nucleotide G + C at the third position of synonymous codons. The curve shows that GC_{12} is equal to GC_3

employing the method of neutrality plot analysis. The point of *infA* was diagonally distributed (Fig. 5), suggesting no significant difference existed among GC_1 , GC_2 and GC_3 . Besides, GC_3 correlated negatively with GC_{12} in other 48 coding sequences of Minshan cp genome, and the correlation was very little ($r = -0.1133$). The results showed that natural selection influenced the SCUB for 48 coding sequence in Minshan expect for *infA* which was mainly affected by the pressure of mutation.

Discussion

Minshan is the only timothy cultivar in China, and severe variety degradation has restricted its promotion and production. Research on SCUB in Minshan cp genome could help reveal its biological architectures, gene evolution and assist molecular breeding in further study. SCUB affects the speed and efficiency of mRNA translation and the folding characteristics of polypeptide chain (Brule and Grayhack 2017; Hu *et al.* 2019). SCUB is quite different

according to different species, tissues and genes (Qiu *et al.* 2011; He *et al.* 2013; Chakraborty *et al.* 2017; Paulet *et al.* 2017; Zhang *et al.* 2018; Cai *et al.* 2019; Hu *et al.* 2019). Among numerous affecting factors, mutation pressure and natural selection are of great importance (Prabha *et al.* 2017). SCUB is the result of long-term competition between the nucleotide constitution for mutation pressure and natural selection. Research on SCUB in Minshan also could conduce to find out the main influencing factor in its evolution and advance the understanding of the balance between them (Sharp and Li 1987; Olejniczak and Uhlenbeck 2006; Zalucki *et al.* 2007; Wang *et al.* 2018b).

In this study, nucleotide composition in Minshan cp genome was abundant in A or T, showing A or T bias. 10 optimal synonymous codons were identified and of which, 6 ended with A and 4 ended with T. Moreover, the correspondence analysis reflected that A or T ended codons were closer to Axis 1 than others. The nucleotide preference in Minshan may be related to the relative evolutionary conservation of cp genome (Hu *et al.* 2019). Nucleotide A or T preference in SCUB was in line with earlier studies in *Oncidium gower ramsey* (Chen *et al.* 2011), seed plants (*Meng et al.* 2008), *Lonicera japonica* (He *et al.* 2017), solanum (Zhang *et al.* 2018) and *Elaeagnus angustifolia* (Wang *et al.* 2019), indicating that the codon bias may correlate to the base composition for mutation pressure.

ENC is an important index to reflect the preference degree of unequal use of synonymous codons (Gupta *et al.* 2004). In Minshan, ENC values ranged from 39.91 to 60.65 with the average of 49.44, implying the synonymous codon usage bias was weak. On ENC plot, only the points of *clpP*, *ndhE*, *rp114*, *rps18*, *rps4* and *ycf* were located on and the points of *ndhH*, *ndhI*, *ndhK* and *ccsA* were close to the expected curve, others presented a discrete distribution. Additionally, except for *infA* which diagonally distributed in neutrality plot, GC_3 and GC_{12} correlated weakly in other 48 genes in Minshan cp genome. All of the results suggest SCUB in Minshan is predominantly influenced by natural selection and different genes have different evolutionary pressure (Mukhopadhyay *et al.* 2008; Li *et al.* 2019; Wang *et al.* 2019).

Conclusion

The research is the first one which systematically analyzes codon usage pattern in Minshan cp genome and comprehensively explores the influencing factors on SCUB. Weak preference of synonymous usage in Minshan exists, and nucleotide constitution, mutation stress and natural selection all have an effect on SCUB, of which natural selection is the major contributor. It still needs further study to clarify whether natural selection has effect on the evolution of functional genes in Minshan cp genome. The results exhibit the architectures of Minshan cp genomes and afford useful information for codon modification and molecular assisted breeding for further study in the future.

Acknowledgements

This work was supported by Quality evaluation and pollution-free production standard system construction of Minshan Timothy (GSZYTC-ZCJC-18027), the Key Laboratory of Superior Forage Germplasm in the Qinghai-Tibetan Plateau (2020-ZJ-Y12), and the National Science Foundation of China (31700338).

References

- Brule CE, EJ Grayhack (2017). Synonymous codons: choose wisely for expression. *Trends Genet* 33:283–297
- Cai MS, AC Cheng, MS Wang, LC Zhao, DK Zhu, QH Luo, F Liu, XY Chen (2019). Characterization of synonymous codon usage bias in the duck plague virus UL35 gene. *Intervirology* 52:266–278
- Cao ZZ (2003). The cultivation and production of *Phleum pretense* in China. *Grassl Chin* 6:72–74
- Chakraborty S, D Nag, TH Mazumder, A Uddin (2017). Codon usage pattern and prediction of gene expression level in *Bungarus* species. *Gene* 604:48–60
- Chen X, XN Cai, QZ Chen, HX Zhou, Y Cai, A Ben (2011). Factors affecting synonymous codon usage bias in chloroplast genome of *Oncidium Gower Ramsey*. *Evol Bioinform* 7:271–278
- Chen Y, YZ Shi, HJ Deng, T Gu, J Xu, JX Ou, ZG Jiang, YR Jiao, T Zou, C Wang (2014). Characterization of the porcine epidemic diarrhea virus codon usage bias. *Infect Genet Evol* 28:95–100
- Choudhury MN, A Uddin, S Chakraborty (2017). Codon usage bias and its influencing factors for Y-linked genes in human. *Comput Biol Chem* 69:77–86
- Cui GX, Y Lu, XX Wei, XL Wang, CM Wang, YQ Gao, HR Duan (2019). Characterization of the complete chloroplast genome of *Phleum pretense* L. cv. Minshan. *Mitochondr DNA* 4:4180–4181
- Du WH (2003). Research advance in nutritive value, cultivation and utilization of Timothy. *Grassl Turf* 4:7–11
- Gupta SK, TK Bhattacharyya, TC Ghosh (2004). Synonymous codon usage in *Lactococcus lactis*: mutational bias versus translational selection. *J Biomol Str Dyn* 21:527–536
- He L, J Qian, X Li, Z Sun, X Xu, S Chen (2017). Complete chloroplast genome of medicinal plant *Lonicera japonica*: genome rearrangement, intron gain and loss, and implications for phylogenetic studies. *Molecules* 22:249–261
- He S, SJ Zhang, LS Xiu, ZB Xing (2013). Codon usage analysis in squalene synthase gene. *Genom Appl Biol* 32:232–239
- Hu XY, YQ Xu, YZ Han, SH Du (2019). Codon usage bias analysis of the chloroplast genome of *Ziziphus jujube* var *spinosa*. *J For Environ* 39:621–628
- Li GL, ZL Pan, SC Gao, YY He, QY Xia, J Yan, HP Yao (2019). Analysis of synonymous codon usage of chloroplast genome in *Porphyra umbilicalis*. *Genes Genomics* 41:1173–1181
- Li ML, ZY Zhao, JH Chen, BY Wang, Z Li, J Li, MS Cai (2012). Characterization of synonymous codon usage bias in the pseudorabies virus *US1* gene. *Virolog Sin* 27:303–315
- Meng Z, L Wei, L Xia (2008). Patterns of synonymous codon usage bias in chloroplast genomes of seed plants. *For Stud Chin* 10:235–242
- Mukhopadhyay P, S Basak, TC Ghosh (2008). Differential selective constraints shaping codon usage pattern of housekeeping and tissue-specific homologous genes of rice and *Arabidopsis*. *DNA Res* 15:347–356
- Olejniczak M, OC Uhlenbeck (2006). tRNA residues that have coevolved with their anticodon to ensure uniform and accurate codon recognition. *Biochimie* 88:943–950
- Parks M, R Cronn, A Liston (2009). Increasing phylogenetic resolution at low taxonomic levels using massively parallel sequencing of chloroplast genomes. *BMC Biol* 7: Article 84
- Paulet D, A David, E Rivals (2017). Ribo-seq enlightens codon usage bias. *DNA Res* 24:303–310
- Prabha R, DP Singh, S Sinha, K Ahmad, A Rai (2017). Genome-wide comparative analysis of codon usage bias and codon context patterns among cyanobacterial genomes. *Mar Genom* 32:31–39
- Qiu S, K Zeng, T Slotte, S Wright, D Charlesworth (2011). Reduced efficacy of natural selection on codon usage bias in selfing *Arabidopsis* and *Capsella* species. *Genome Biol Evol* 3:868–880
- Sau K, SK Gupta, S Sau, SC Mandal, TC Ghosh (2006). Factors influencing synonymous codon and amino acid usage biases in Mimivirus. *Biosystems* 85:107–113
- Sharp PM, WH Li (1987). The rate of synonymous substitution in enterobacterial genes is inversely related to codon usage bias. *Mol Biol Evol* 4:222–230
- Tillich M, P Lehwark, T Pellizzer, ES Ulbricht-Jones, A Fischer, R Bock S Greiner (2017). GeSeq-versatile and accurate annotation of organelle genomes. *Nucl Acids Res* 45: 6–11
- Wang HX, ZD Cao, JC Xiang, P Fu (2018a). Present situation of grass industry and countermeasures of transformational development in Dingxi. *Pratac Sci* 7:1811–1817
- Wang J, TY Wang, LY Wang, JG Zhang, YF Zeng (2019). Assembling and Analysis of the whole chloroplast genome sequence of *Elaeagnus angustifolia* and its codon usage bias. *Acta Bot Sin* 9:1559–1572
- Wang L, H Xing, Y Yuan, X Wang, M Saeed, J Tao, W Feng, G Zhang, X Song, X Sun (2018b). Genome-wide analysis of codon usage bias in four sequenced cotton species. *PLoS One* 13; Article e0194372
- Wei L, J He, X Jia, Q Qi, ZS Liang, H Zheng, Y Ping, SY Liu, JC Sun (2014). Analysis of codon usage bias of mitochondrial genome in *Bombyx mori* and its relation to evolution. *BMC Evol Biol* 14:262–273
- Zalucki YM, PM Power, MP Jennings (2007). Selection for efficient translation initiation biases codon usage at second amino acid position in secretory proteins. *Nucl Acids Res* 35:5748–5754
- Zhang R, L Zhang, W Wang, Z Zhang (2018). Differences in codon usage bias between photosynthesis-related genes and genetic system-related genes of chloroplast genomes in cultivated and wild *solanum* species. *Intl J Mol Sci* 19:3124–3148
- Zhou CL, LY Peng, X Wang, JL Chen, L Wang, H Chen, ZX Lai, SC Liu (2019). Codon bias and evolution analysis of *AtGAI* in *Amaranthus tricolor* L. *J Chin Agric Univ* 24:10–22