



Research Article

Codon Usage Bias and Cluster Analysis of the *MMP-2* and *MMP-9* Genes in Seven Mammals

Tanliang Ouyang ¹, Jincheng Zhong ¹, Zhixin Chai ¹, Jiabo Wang ¹, Ming Zhang ¹,
Zhijuan Wu ^{1,2} and Jinwei Xin ²

¹Key Laboratory of Qinghai-Tibetan Plateau Animal Genetic Resource Reservation and Utilization, Southwest Minzu University, Chengdu 610225, China

²State Key Laboratory of Hulled Barley and Yak Germplasm Resources and Genetic Improvement, Lhasa 850000, China

Correspondence should be addressed to Zhijuan Wu; wzjdream2005@163.com and Jinwei Xin; xinjinwei18@163.com

Received 31 March 2022; Accepted 4 August 2022; Published 5 September 2022

Academic Editor: Aldesia Provenzano

Copyright © 2022 Tanliang Ouyang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Matrix metalloproteinase (*MMP*)-2 and *MMP*-9 are a family of Zn^{2+} and Ca^{2+} -dependent gelatinase *MMP*s that regulate muscle development and disease treatment, and they are highly conservative during biological evolution. Despite increasing knowledge of *MMP* genes, their evolutionary mechanism for functional adaptation remains unclear. Moreover, analysis of codon usage bias (CUB) is reliable to understand evolutionary associations. However, the distribution of CUB of *MMP*-2 and *MMP*-9 genes in mammals has not been revealed clearly. Multiple analytical software was used to study the genetic evolution, phylogeny, and codon usage pattern of these two genes in seven species of mammals. Results showed that the *MMP*-2 and *MMP*-9 genes have CUB. By comparing the content of synonymous codon bases amongst seven mammals, we found that *MMP*-2 and *MMP*-9 were low-expression genes in mammals with high codon conservation, and their third codon preferred the G/C base. RSCU analysis revealed that these two genes preferred codons encoding delicious amino acids. Analysing what factors influence CUB showed that the third base distributors of these two genes were C/A and C/T, and GC_{3S} had a wide distribution range on the ENC plot reference curve under no selection or mutational pressure. Thus, mutational pressure is an important factor in CUB. This study revealed the usage characteristics of the *MMP*-2 and *MMP*-9 gene codons in different mammals and provided basic data for further study towards enhancing meat flavour, treating muscle disease, and optimizing codons.

1. Introduction

Codon usage bias (CUB) is defined as unequal utilisation in the frequency of synonymous codons in coding amino acids (AAs), and it has been used extensively for investigating gene phylogeny [1]. The synonymous codon characteristics include universality, degeneracy, and wobble, and they should be used randomly to encode corresponding AAs with no pressure of interference pressure. However, CUB can be affected by nucleotides composition, translation, hydrophobicity, tRNA abundance, and protein structure [2–6]. Notably, natural selection and mutational pressure, which drive the correct translation process, are the major factors associated with CUB [7, 8]. Natural selection affects the

pattern of codon usage in organisms, and mutational pressure may arise whilst the proportion of codon bases changes. CUB greatly increases the variability of genetic information and reflects the genetic drift of codons to a certain extent [9]. Therefore, CUB can reveal the evolution of genes or organisms and environmental adaptation [10].

CUB is assessed by using the effective number of codons (ENC), codon adaptation index (CAI), frequency of optimal codons (FOP), codon bias index (CBI), and relative usage of synonymous codons (RSCU). ENC is calculated by comparing the GC content of synonymous codon positions [11]. CAI is 0-1; the closer the value is to 1, the stronger the nucleotide bases prefer synonymous codons [12]. FOP and CBI are both 0-1. These two indicators are close to 1, and the

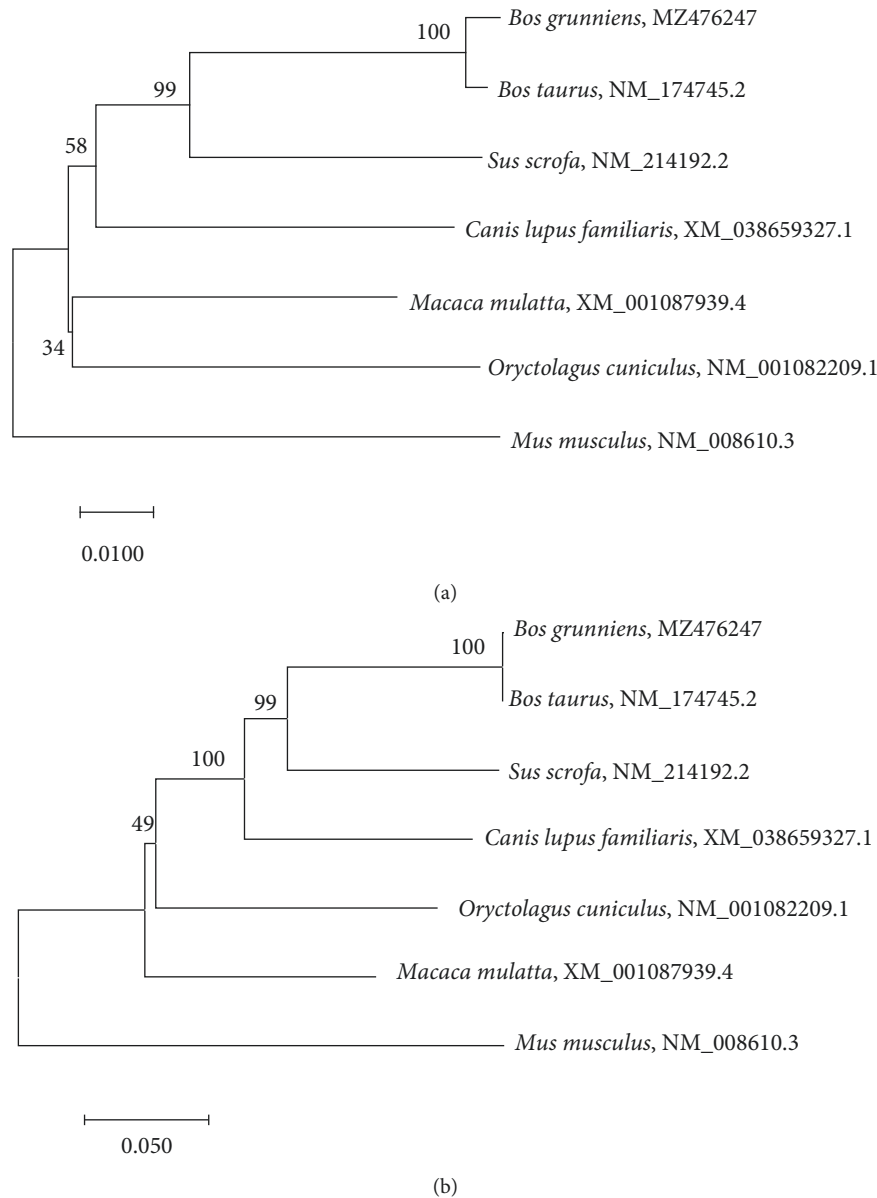


FIGURE 1: NJ trees of *MMP-2* (a) and *MMP-9* (b) genes of seven mammal species.

optimal codon for encoding amino acids is preferred. However, if CBI is negative, the optimal codon usage is less than the average number of codons used [3, 13]. RSCU is the specific value between the actual observation and theoretical observation, amongst which the theoretical observation value is the observation value when the synonymous codon usage frequency is the same, namely, there is no codon bias. If $RSCU = 1$, there is no CUB. If $RSCU > 1$, the appearance frequency of the codon is higher than the other synonymous codon. By contrast, it indicates lower genes. If $RSCU > 2$, then the frequency of CUB is extremely high [14].

Matrix metalloproteinases (MMPs) are a family of Zn^{2+} and Ca^{2+} -dependent proteolytic enzymes that are widely expressed in animal tissues and highly conservative during biological evolution [15]. *MMP-2* and *MMP-9* can regulate

muscle growth, repair, and some relative processes that affect biochemical reactions for muscle regulation [16]. Although recent research mainly focused on exploring *MMP-2* and *MMP-9* function for animal skeletal muscle development, healing diseased muscle and even meat [17–23], studies on MMP codons is rare. Therefore, there is an urgent need for exploring mammals' *MMP-2* and *MMP-9* genetic evolution and codon usage pattern regulating muscle growth.

In this study, seven mammals (*Bos grunniens*, *Bos taurus*, and *Sus scrofa* among Artiodactyla; *Macaca mulatta* in Primates; *Canis lupus familiaris* in Carnivora; *Oryctolagus cuniculus* in Lagomorpha; and *Mus musculus* in Rodentia) were chosen to analyse CUB and base pair composition dynamics. This study would give insight into the factors

TABLE 1: Nucleotide composition in the sequence of *MMP-2* gene.

Species	A/T (%)	G/C (%)	T _{3S} (%)	C _{3S} (%)	A _{3S} (%)	G _{3S} (%)	GC _{3S} (%)	AT _{3S} (%)
<i>Bos grunniens</i>	0.415	0.585	0.2093	0.372	0.2081	0.3703	0.637	0.363
<i>Bos taurus</i>	0.384	0.616	0.2638	0.3049	0.34	0.2724	0.488	0.512
<i>Macaca mulatta</i>	0.409	0.591	0.2247	0.3128	0.2741	0.3031	0.551	0.449
<i>Mus musculus</i>	0.433	0.567	0.2974	0.3234	0.3212	0.2566	0.482	0.518
<i>Oryctolagus cuniculus</i>	0.4	0.6	0.1997	0.5452	0.1304	0.3621	0.73	0.27
<i>Sus scrofa</i>	0.393	0.607	0.1871	0.5485	0.1404	0.3488	0.731	0.269
<i>Canis lupus familiaris</i>	0.417	0.583	0.284	0.3302	0.3285	0.2415	0.482	0.518

TABLE 2: Nucleotide composition in the sequence of *MMP-9* gene.

Species	A/T (%)	G/C (%)	T _{3S} (%)	C _{3S} (%)	A _{3S} (%)	G _{3S} (%)	GC _{3S} (%)	AT _{3S} (%)
<i>Bos grunniens</i>	0.366	0.634	0.1842	0.5263	0.1981	0.2891	0.685	0.315
<i>Bos taurus</i>	0.378	0.622	0.173	0.5554	0.1202	0.3781	0.759	0.241
<i>Macaca mulatta</i>	0.379	0.621	0.1766	0.5135	0.1277	0.3978	0.747	0.253
<i>Mus musculus</i>	0.433	0.567	0.2776	0.3436	0.3056	0.2458	0.503	0.497
<i>Oryctolagus cuniculus</i>	0.345	0.655	0.2237	0.3099	0.1976	0.3431	0.606	0.394
<i>Sus scrofa</i>	0.385	0.615	0.1814	0.5411	0.1207	0.3852	0.752	0.248
<i>Canis lupus familiaris</i>	0.373	0.627	0.1554	0.5702	0.1143	0.39	0.779	0.221

TABLE 3: Codon analysis of *MMP-2* gene.

Species	ENC	CAI	CBI	Fop	GC _{1S} (%)	GC _{2S} (%)	GC ₁₂ (%)
<i>Bos grunniens</i>	48.04	0.193	-0.004	0.416	0.4932	0.5961	0.54465
<i>Bos taurus</i>	53.82	0.129	-0.006	0.386	0.7471	0.576	0.66155
<i>Macaca mulatta</i>	52.31	0.157	0.006	0.413	0.4982	0.6751	0.58665
<i>Mus musculus</i>	55.09	0.163	0.034	0.416	0.6238	0.5439	0.58385
<i>Oryctolagus cuniculus</i>	43.1	0.29	0.193	0.539	0.565	0.491	0.528
<i>Sus scrofa</i>	44.62	0.261	0.17	0.524	0.5579	0.52	0.53895
<i>Canis lupus familiaris</i>	55.29	0.146	0.014	0.4	0.6706	0.5569	0.61375

TABLE 4: Codon analysis of *MMP-9* gene.

Species	ENC	CAI	CBI	Fop	GC _{1S} (%)	GC _{2S} (%)	GC ₁₂ (%)
<i>Bos grunniens</i>	44.44	0.256	0.24	0.558	0.6713	0.5295	0.6004
<i>Bos taurus</i>	42.48	0.291	0.249	0.566	0.6015	0.4955	0.5485
<i>Macaca mulatta</i>	44.99	0.252	0.191	0.53	0.5997	0.5045	0.5521
<i>Mus musculus</i>	56.96	0.16	0.042	0.419	0.6105	0.5448	0.57765
<i>Oryctolagus cuniculus</i>	54.05	0.154	0.011	0.419	0.5065	0.8191	0.6628
<i>Sus scrofa</i>	43.79	0.282	0.222	0.549	0.5958	0.4863	0.54105
<i>Canis lupus familiaris</i>	40.95	0.287	0.243	0.563	0.6025	0.4916	0.54705

affecting CUB for *MMP-2* and *MMP-9* genes and provide basic data for enhancing the meat flavour and finding a promising gene treatment for muscle disease.

2. Materials and Methods

2.1. Software. MEGA 7.0, CodonW 1.4.2, pheatmap, and ggplot packages based on R 4.4.3 software were used to complete the relevant analysis.

2.2. Base Composition of *MMP* Genes' CDS in Different Mammals. The coding sequence (CDS) of yak *MMP-2* and *MMP-9* genes were obtained in our laboratory, and the NCBI accession numbers were MZ476247 and MZ476248, respectively. The CDS of other animals' genes were from

NCBI GenBank, and their accession numbers are shown in Figure 1.

CodonW 1.4.2 software developed by J. Peden was used to analyse the *MMP-2* and *MMP-9* CDS in seven mammals for calculating A/T (A/T base content, the same below), G/C, T_{3S} (third base of the codon is T content, the same below), C_{3S}, A_{3S}, G_{3S}, GC_{3S}, AT_{3S}, ENC, CAI, CBI, FOP, and RSCU [24]. R packages pheatmap and ggplot2 were used to analyse the data.

2.3. PR2 Plot. PR2 plot could analyse the bias amongst ATCG under gene mutation [25]. If the frequency of the third base is A > T, then dots are scattered on the top of the PR2 plot. If the frequency is C > G, then dots are on the left.

TABLE 5: RSCU for *MMP-2* gene among seven species.

AA	Codon	Frequency	RSCU
Phe	UUU	66	0.675714
	UUC*	142	1.324286
Leu	UUA	28	0.268571
	UUG	71	0.732857
	CUU*	127	1.09
	CUC	101	1.011429
	CUA	53	0.468571
	CUG**	239	2.428571
Ile	AUU	34	0.801429
	AUC*	67	1.28
	AUA	34	0.918571
Met	AUG	110	1
Val	GUU	65	0.861429
	GUC	47	0.715714
	GUA	39	0.522857
	GUG*	122	1.902857
Ser	UCU*	89	1.078571
	UCC*	104	1.27
	UCA	70	0.904286
	UCG	23	0.317143
Pro	CCU	157	0.974286
	CCC*	256	1.641429
	CCA	124	0.75
Thr	CCG	102	0.635714
	ACU	74	0.795714
	ACC*	134	1.365714
	ACA*	108	1.142857
Ala	ACG	64	0.697143
	GCU*	128	1.047143
	GCC*	199	1.585714
	GCA	99	0.735714
Tyr	GCG	79	0.631429
	UAU	30	0.678571
His	UAC*	75	1.321429
	CAU	92	0.74
Gln	CAC*	148	1.26
	CAA	144	0.864286
Asn	CAG*	132	1.135714
	AAU	40	0.787143
Lys	AAC*	75	1.212857
	AAA	58	0.822857
Asp	AAG*	138	1.178571
	GAU	94	0.835714
Glu	GAC*	131	1.164286
	GAA	116	0.944286
Cys	GAG*	121	1.055714
	UGU	73	0.69
Trp	UGC*	151	1.31
	UGG	206	1
Arg	CGU	34	0.428571
	CGC*	104	1.327143
	CGA	68	0.772857
	CGG*	92	1.177143
Ser	AGU	66	0.897143
	AGC*	112	1.534286

TABLE 5: Continued.

AA	Codon	Frequency	RSCU
Arg	AGA*	109	1.212857
	AGG*	93	1.08
Gly	GGU	71	0.527143
	GGC*	207	1.371429
	GGA*	156	1.074286
	GGG*	149	1.03
TER	UAA	21	0.374286
	UAG	21	0.425714
	UGA**	132	2.201429

Note. *RSCU > 1; **RSCU > 2; AA. amino acid; TER. termination codon; the same below.

When the codon does not show usage bias, the dots are in the centre of the graph [26].

2.4. Codon Neutral Analysis. Codon neutral analysis was carried out by the correlation analysis of GC_{12} (the average of the GC content of the first and second bases) and GC_{3S} to compare the influence of natural selection pressure and mutational pressure on CUB [27]. A significant correlation between GC_{12} and GC_{3S} indicated that mutational pressure had a strong influence on codon preference; otherwise, natural selection influenced CUB [28].

2.5. ENC Plot. The relationship between ENC and GC_{3S} without environmental selection pressure could be simulated by the following formula (1). The ENC/ GC_{3S} reference curve shows the main characteristics of codon usage patterns [24]. If CUB is more affected by natural selection, it should be below the standard curve. By contrast, it should be above the standard curve if it is more affected by other factors such as gene mutation. In general, the ENC is from 35 to 61. If $ENC > 35$, CUB is weak [11].

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

3. Results

3.1. Phylogenetic Analysis. Neighbour joining (NJ) trees were established based on the *MMP-2* and *MMP-9* CDS in seven mammals. The results (Figure 1) showed that the *MMP-2* and *MMP-9* genes of *Bos grunniens* were similar to those of *B. taurus*. These two genes of *S. scrofa* were similar to those of *B. grunniens* and *B. taurus*. Interestingly, the *MMP-9* genes of *C. lupus familiaris* showed closer proximity to those of *S. scrofa* but those of the *MMP-2* gene was farther.

3.2. Nucleotide Composition of *MMP-2* and *MMP-9* Genes. Compared with the content of codon bases of the *MMP-2* and *MMP-9* genes in seven mammals, the results showed (Tables 1 and 2) that the *G/C* content was higher than the

TABLE 6: RSCU for *MMP-9* gene among seven species.

AA	Codon	Frequency	RSCU
Phe	UUU	72	0.584285714
	UUC*	197	1.415714286
Leu	UUA	17	0.17
	UUG	44	0.621428571
	CUU	76	0.848571429
	CUC*	107	1.362857143
	CUA	33	0.357142857
	CUG**	195	2.641428571
Ile	AUU	30	0.811428571
	AUC*	61	1.881428571
	AUA	11	0.308571429
Met	AUG	49	1
Val	GUU	48	0.684285714
	GUC	66	0.957142857
	GUA	29	0.375714286
	GUG*	147	1.978571429
Ser	UCU	75	0.994285714
	UCC*	118	1.862857143
	UCA	41	0.494285714
	UCG	55	0.832857143
Pro	CCU	117	0.85
	CCC*	209	1.644285714
	CCA	108	0.735714286
	CCG	102	0.774285714
Thr	ACU	74	0.688571429
	ACC**	208	2.145714286
	ACA	54	0.484285714
Ala	ACG	82	0.684285714
	GCU	84	0.792857143
	GCC*	175	1.731428571
	GCA	71	0.655714286
Tyr	GCG	87	0.815714286
	UAU	35	0.54
His	UAC	119	1.46
	CAU	39	0.441428571
Gln	CAC*	116	1.558571429
	CAA	62	0.537142857
Asn	CAG*	136	1.462857143*
	AAU	33	0.664285714
Lys	AAC*	72	1.335714286
	AAA	41	0.598571429
Asp	AAG*	105	1.401428571
	GAU	70	0.571428571
Glu	GAC*	210	1.428571429
	GAA	72	0.592857143
Cys	GAG*	159	1.407142857
	UGU	46	0.417142857
Trp	UGC	135	1.582857143
	UGG*	128	1
Arg	CGU	45	0.532857143
	CGC**	134	2.072857143
	CGA	54	0.667142857
	CGG	77	0.965714286
Ser	AGU	31	0.418571429
	AGC*	91	1.395714286

TABLE 6: Continued.

AA	Codon	Frequency	RSCU
Arg	AGA	63	0.752857143
	AGG*	79	1.011428571
Gly	GGU	77	0.615714286
	GGC*	219	1.738571429
	GGA	86	0.654285714
	GGG	124	0.991428571
TER	UAA	18	0.571428571
	UAG*	20	1.538571429
	UGA	37	0.89

A/T content. Most mammals' *MMP-2* and *MMP-9* GC_{3S} were larger than AT_{3S}, except for the *MMP-2* gene of *B. taurus* and *M. musculus*. The above findings indicated that the *MMP-2* and *MMP-9* gene codons preferred GC_{3S}.

The codon usage results (Tables 3 and 4) showed that ENC_s of the *MMP-2* and *MMP-9* genes in seven mammals were 40–56, indicating that these two genes had low expression and their codon conservation was high.

CAI showed that the preference for synonymous codons of the *MMP-9* genes in seven mammals was significantly better than that of *MMP-2*, but both were lower than 0.3, indicating that it failed to reflect the preference of synonymous codons.

FOP and CBI results of the *MMP-2* and *MMP-9* genes showed that the optimal codon usage of *MMP-2* in *B. grunniens* and *B. taurus* was inferior to the five other animals, whilst the optimal codon usage of *MMP-9* was better than that of *MMP-2*.

3.3. RSCU Analysis. The RSCU results of the *MMP-2* and *MMP-9* genes showed that these two genes had a preference for 27 and 20 codons, respectively (Tables 5 and 6). Amongst them, CUG (encoding leucine, Leu) of *MMP-2*, CUG (encoding leucine, Leu), ACC (encoding threonine, Thr), and CGC (encoding arginine, Arg) of *MMP-9* had strong high CUB (RSCU > 2).

Heat map analysis of the correlation between codon base composition and GC_{3S} (Figure 2) showed that most of the codons of the *MMP-2* and *MMP-9* genes in different mammals were positively correlated with GC_{3S} and in line with AC-, CG-, AT-, TC-, GG-, CC-, GC-, and other codons whose third base was C.

Cluster analysis of the RSCU of the *MMP-2* and *MMP-9* genes showed that the *MMP-2* gene preferred CUG, GUG, UCC, GAG, AUC, AAC, UAC, GCC, AGA, UUG, and AGG codons, which were mainly involved in encoding Leu (leucine), Val (valine), Ser (serine), Glu (glutamic acid), Iso (isoleucine) Asn (asparagine), Tyr (tyrosine), Gly (glycine), and Arg (arginine), respectively (Figures 3 and 4). In addition to *B. taurus* and *C. lupus familiaris*, the five other species had a strong preference for CUG and GUG (RSCU > 2), amongst which the RSCU of *O. cuniculus* and

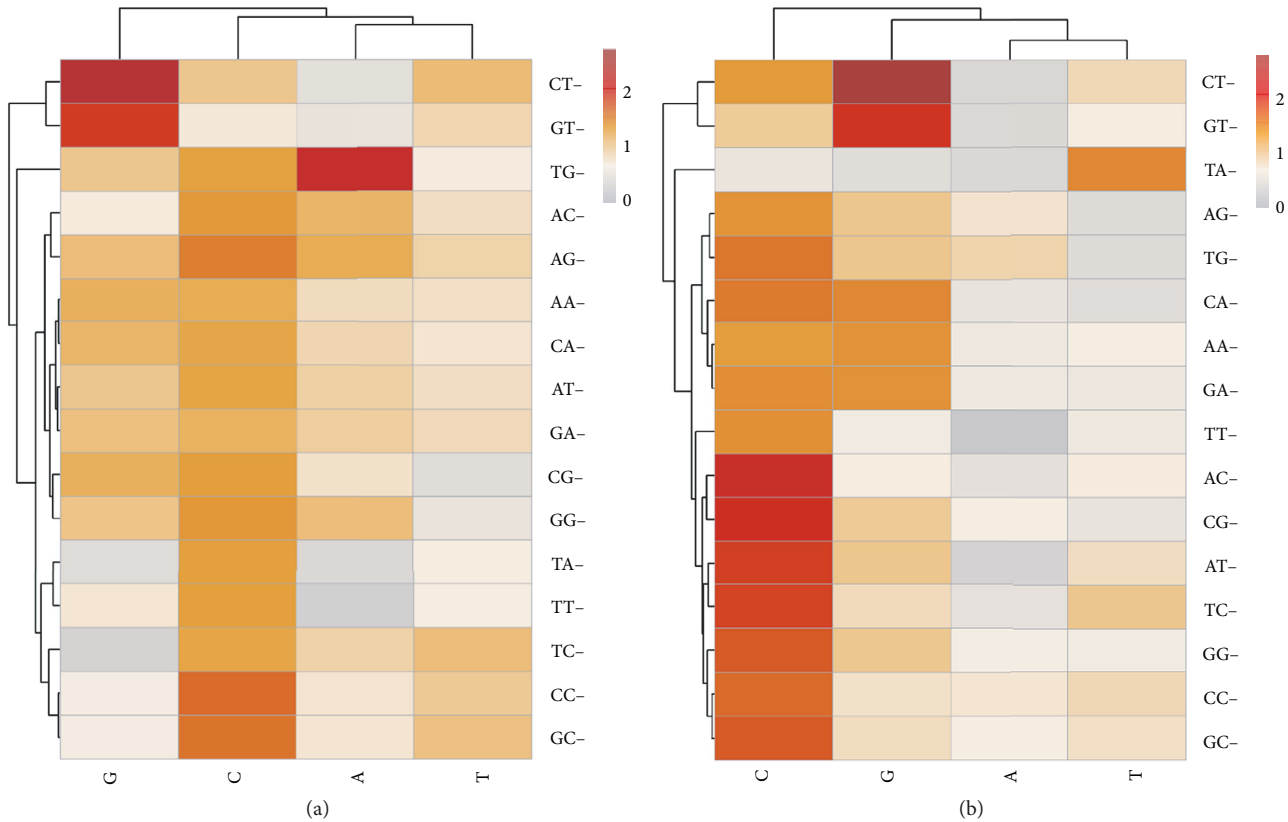


FIGURE 2: Heat maps of *MMP-2* (a) and *MMP-9*. (b) Correlation coefficient of codons with GC_{3S} .

B. grunniens > 3. The *MMP-9* gene preferred UCC, ACC, CGC, CUG, and AUC codons, which are mainly involved in Ser, Thr, Arg, Leu, and Iso, respectively. Except for *M. musculus* and *O. cuniculus*, the last five species had strong preferences similar to one another, indicating that the *MMP-9* gene was more conservative than *MMP-2*.

3.4. Factors Influenced CUB. The PR2 plot result (Figure 5) showed that the ATCG base distribution of the *MMP-2* and *MMP-9* genes amongst seven mammals was above 0.5 on the x -axis. The bases distribution of the *MMP-2* genes was mainly on the x -axis and the upper right of the y -axis and that of the *MMP-9* genes was to the x -axis and the upper right of the y -axis. The above results indicated that the contents of A_{3S} and C_{3S} for the *MMP-2* gene and the content of T_{3S} and C_{3S} for the *MMP-9* gene were high, respectively.

Neutral analysis (Figure 6 and Table 7) showed that GC_{3S} of these two genes was in the range of 0.44–0.78, whereas GC_{12} was from 0.52 to 0.67. The difference was that GC_{12} and GC_{3S} of the *MMP-2* gene were strongly negatively correlated (Pearson $r = -0.851$, p value < 0.05), whilst GC_{12} and GC_{3S} of the *MMP-9* gene were not significantly correlated, indicating that the base composition of the *MMP-2* gene codons was susceptible to mutational pressure, but the factor influencing the *MMP-9* gene was natural selection.

The ENC plot showed (Figure 7 and Table 7) that all ENC/ GC_{3S} dots of the *MMP-2* and *MMP-9* genes were distributed below the reference line. ENC and GC_{3S} had a strongly negative correlation (*MMP-2*: Pearson $r = -0.993$, p value < 0.01; *MMP-9*: Pearson $r = -0.963$, p value < 0.01), and the distribution range of GC_{3S} was large, indicating that the CUB of these two genes was affected by mutational pressure.

4. Discussion

This study found that gelatinase *MMP* genes had CUB for encoding amino acids such as Ile, Arg, Glu, and Ser related to muscle development and meat quality. Gly, Arg, and Leu can promote collagen synthesis, and animal muscle is the main way to obtain natural collagen for humans [29, 30]. Delicious amino acids (DAAs), including Glu, Gly, Ser, Asp, Arg, and Ile, are known as precursor substances that determine the flavour of meat and can improve the taste of chicken and keep the meat soft [31]. Recent research found that the quality of chicken improves and the content of DAAs increases [32]. Otherwise, Strecker amino acids (SAAs), including Phe (phenylalanine), Cys (cysteine), Ile (isoleucine), and Leu (leucine), are highly related to the production of flavour. The higher their content, the stronger the fragrance [33]. For the *MMP-2* and *MMP-9* genes, the RSCUs of AUC encoding Ile; UCC and AGC encoding Ser; CGC encoding Arg; GAC encoding Asp; GAG encoding Glu; UUC

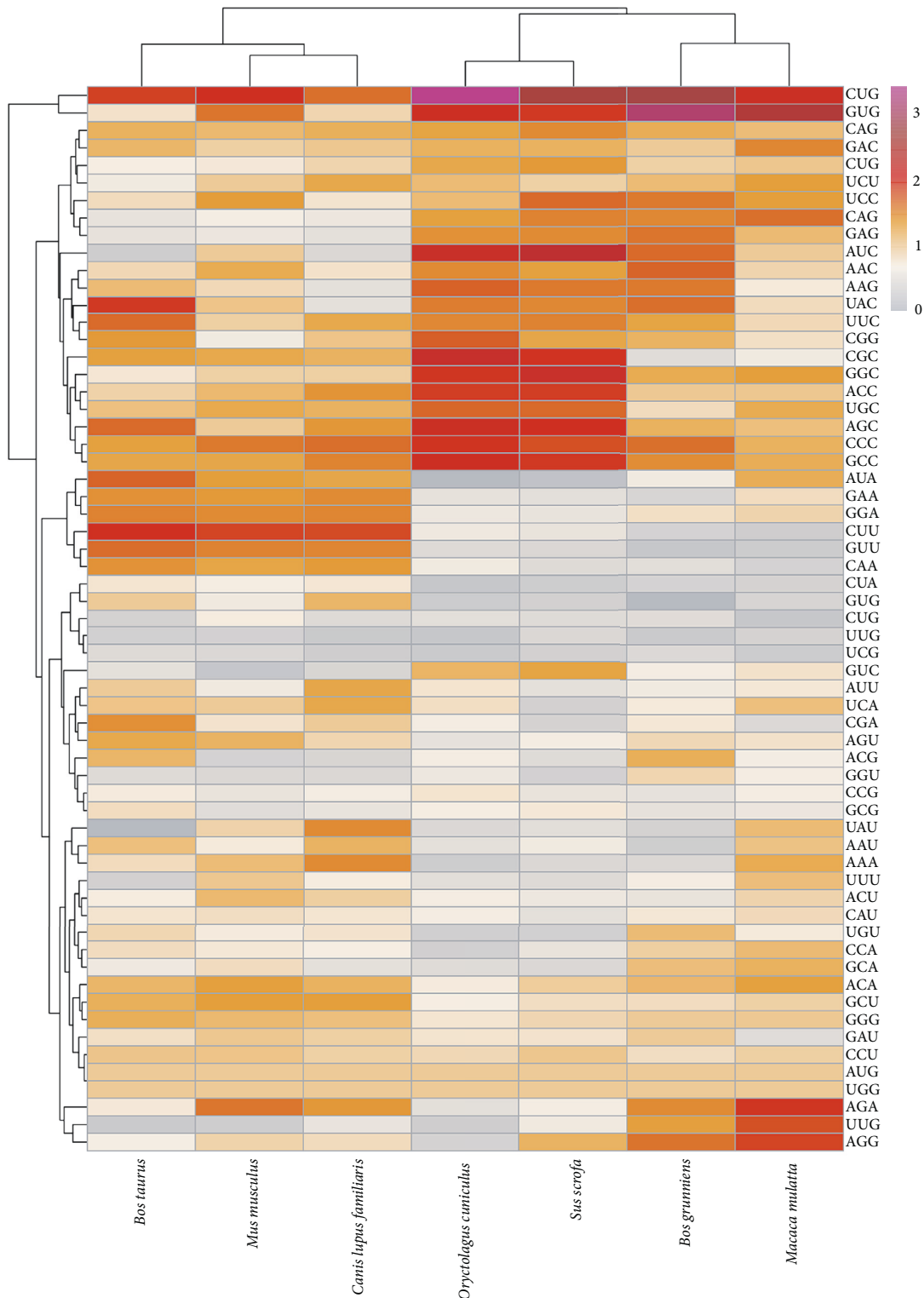


FIGURE 3: Clustering of RSCU values of each codon in *MMP-2* gene.

encoding Phe; and GGA, GGC, and GGG encoding Gly were > 1. In particular, the RSCUs of CUG encoding Leu and CGC encoding Arg > 2; this value indicated that *MMP-2* and *MMP-9* demonstrated CUB for DAAs and SAAs. Besides, Leu, Ile, and Val belong to branched-chain amino acids (BCAAs), and they are essential AAs in humans and

animals, accounting for about 35% of muscle protein. Previous studies have found that skeletal muscle, as the initial site of BCAAs catabolism, can be activated by branched-chain keto acids (BCKAs) to increase BCAAs synthesis to relieve muscle wasting disorders [34]. Also, Leu supplementation could be the prevention and treatment of

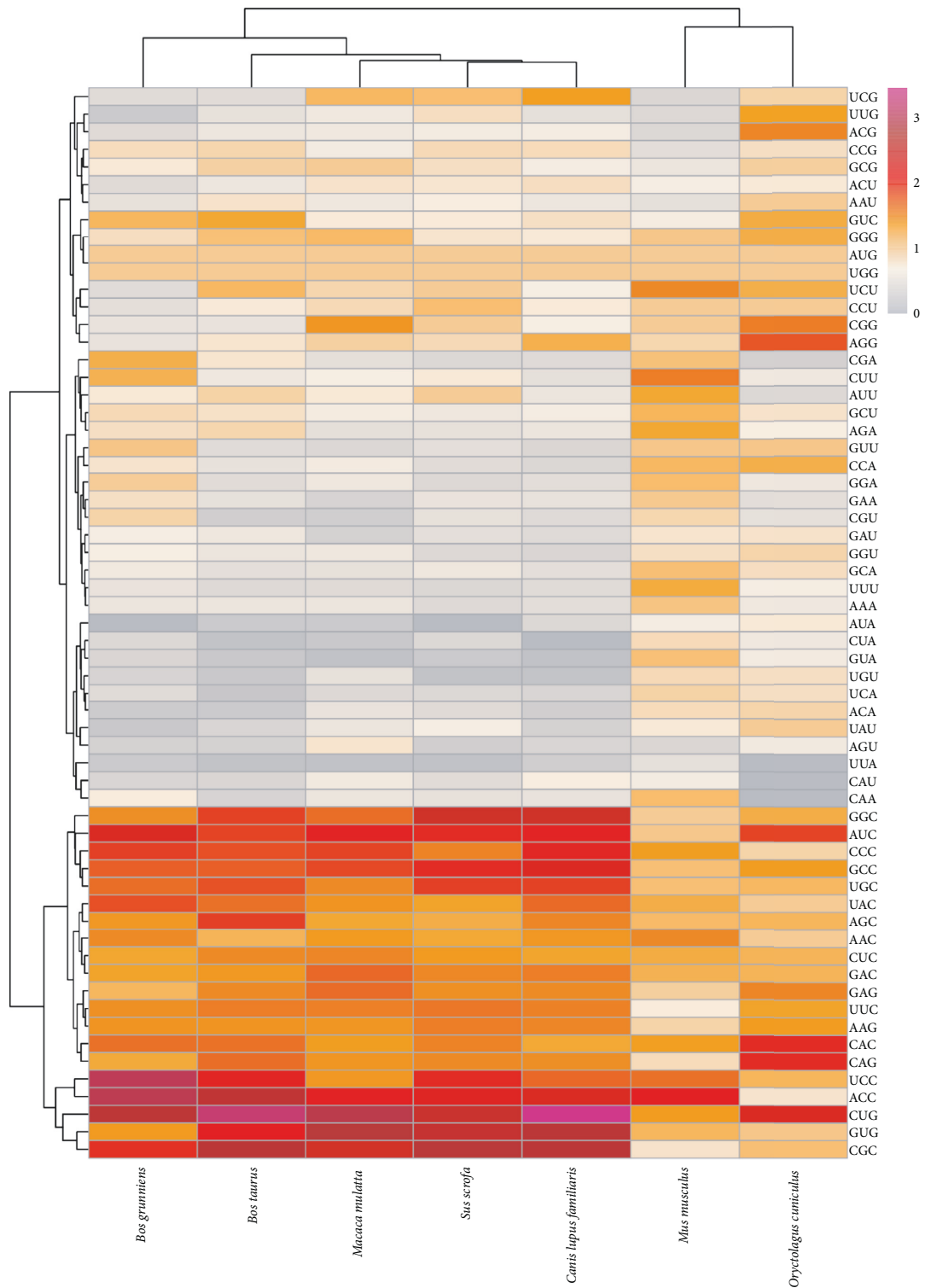


FIGURE 4: Clustering of RSCU values of each codon in *MMP-9* gene.

sarcopenia with aging [35]. Thus, BCAAs are important regulators of metabolism and metabolic health in *in vivo* [36]. The gelatinase MMP CUB associated with corresponding AAs can provide basic data for the improvement of meat quality and muscle disease of MMP molecular modification.

Mutational pressure may be the main factor influencing the CUB of MMPs. This study found that the clustering results of the RSCU were different from the NJ trees of the genes, indicating that the MMP genes were highly conserved but maybe subjected to mutations during the evolution of different species. This influence caused a decline in the

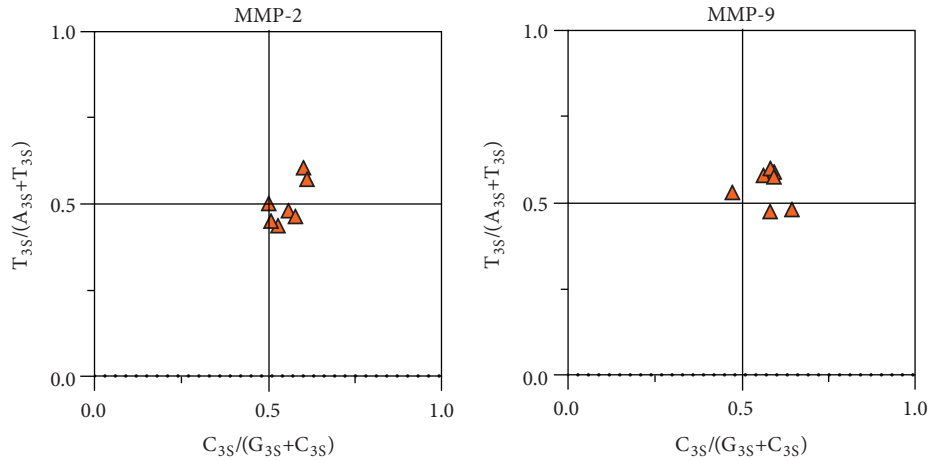


FIGURE 5: PR2 plot among seven species.

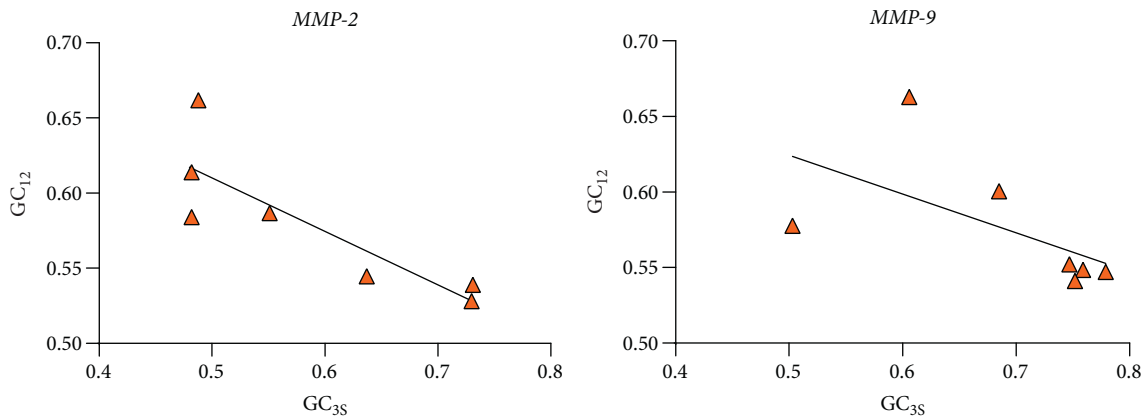


FIGURE 6: Codon neutral analysis.

TABLE 7: Pearson relative analysis with GC_{3S}.

	Pearson <i>r</i> (MMP-2)	<i>P</i> values (MMP-2)	Pearson <i>r</i> (MMP-9)	<i>P</i> values (MMP-9)
ENC	-0.993	0.000**	-0.963	0.000**
GC ₁₂	-0.851	0.015*	-0.589	0.164

Note. * *P* value < 0.05; ** *p* value < 0.01; red represents strong correlation, blue represents moderate correlation, and black represents irrelevance.

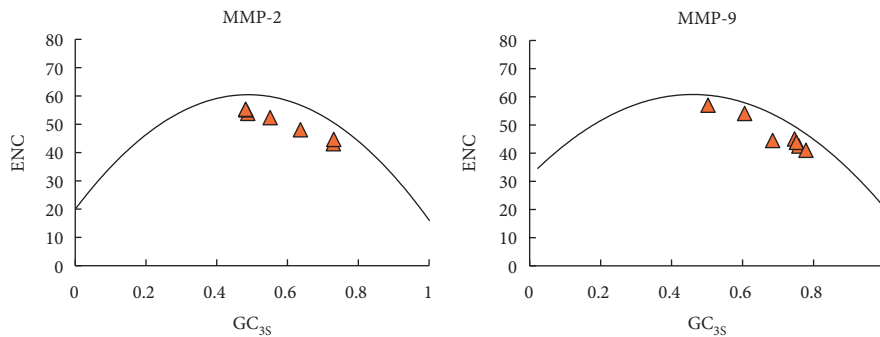


FIGURE 7: ENC plot among seven species.

accuracy of single-gene species classification. Nucleotide AT (U) CG base composition is an important feature of genes, and the GC content can reflect the overall trend of gene mutation which is a decisive factor affecting the frequency of

nucleotide use. Changes in the third base of the codon did not affect the encoded AAs, so GC_{3S} could be an important reference for analysing the codon usage pattern. The gene mutation will affect the composition of the synonymous

codon third bases with no natural selection, and the stronger the CUB, the more the codon is inclined to GC_{3S} [37, 38]. Novembre et al. also found that the third base distribution of the *MMP-2* and *MMP-9* genes is mainly AC_{3S} and CT_{3S}, respectively, and the ENC/GC_{3S} dot distribution can reach a wide range compared with the reference curve with gene mutation pressure. Thus, mutational pressure may play an important role in affecting the CUB for *MMP-2* and *MMP-9* genes, which also explains the difference in RSCU clustering in the seven mammals.

Interestingly, we also found that the clustering results based on the RSCU of the *MMP-2* gene were not completely consistent with the phylogenetic results based on the *MMP-2* gene's CDS. Given that wild yak and Tibetan antelope grow in harsh environments with low altitudes and oxygen consumption, their *EGLN1* gene has mutated changing nucleotide bases and leading to CUB changes [39, 40]. Therefore, we believe that the phylogenetic evolution of *MMP-2* genes should not only refer to gene sequence but also CUB, which could be a supplement to species classification.

5. Conclusion

MMP-2 and *MMP-9* are low-expression genes in mammals, and their codons are highly conservative. Both have a CUB at GC_{3S} and prefer codons encoding DAAs and SAAs for improving soft meat and muscle disease treatment.

Data Availability

The yak *MMP-2* and *MMP-9* genes data used to support the findings of this study are included within the article and are available from the corresponding author upon request.

Ethical Approval

The procedures for care and use of experimental animals were in accordance with the Regulations on the Administration of Experimental Animals issued by the China State Council in 2017 and the guidance on Treating Experimental Animals developed by China's Ministry of Science and Technology in 2006.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the Open Project Program of State Key Laboratory of Hulless Barley and Yak Germplasm Resources and Genetic Improvement (XZNKY-2021-C-014-K08), The Major Special Projects of Tibet Autonomous Region (XZ202101ZD0002N-01), and Program of National Beef Cattle and Yak Industrial Technology System (CARS-37). The authors thank KGSsupport (<https://www.kgsupport.com>) for its linguistic assistance during the preparation of this manuscript.

References

- [1] S. Chakraborty, T. H. Mazumder, and A. Uddin, "Compositional dynamics and codon usage pattern of *brca1* gene across nine mammalian species," *Genomics*, vol. 111, no. 2, pp. 167–176, 2019.
- [2] W. H. Li, "Models of nearly neutral mutations with particular implications for nonrandom usage of synonymous codons," *Journal of Molecular Evolution*, vol. 24, no. 4, pp. 337–345, 1987.
- [3] T. Ikemura, "Correlation between the abundance of *Escherichia coli* transfer rnas and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the *E. coli* translational system," *Journal of Molecular Biology*, vol. 151, no. 3, pp. 389–409, 1981.
- [4] L. Duret and D. Mouchiroud, "Expression pattern and, surprisingly, gene length shape codon usage in *Caenorhabditis*, *Drosophila*, and *Arabidopsis*," *Proceedings of the National Academy of Sciences*, vol. 96, no. 8, pp. 4482–4487, 1999.
- [5] X. Huang, J. Xu, L. Chen et al., "Analysis of transcriptome data reveals multifactor constraint on codon usage in *taenia multiceps*," *BMC Genomics*, vol. 18, no. 1, pp. 308–320, 2017.
- [6] H. Goodarzi, N. Tor Ab I, H. S. Najaf Ab Adi, and M. Archetti, "Amino acid and codon usage profiles: adaptive changes in the frequency of amino acids and codons," *Gene*, vol. 407, no. 1–2, pp. 30–41, 2008.
- [7] H. Musto, S. Cruveiller, G. D'Onofrio, H. Romero, and G. Bernardi, "Translational selection on codon usage in *Xenopus laevis*," *Molecular Biology and Evolution*, vol. 18, no. 9, pp. 1703–1707, 2001.
- [8] J. M. Comeron, M. Kreitman, and M. Aguade, "Natural selection on synonymous sites is correlated with gene length and recombination in *Drosophila*," *Genetics*, vol. 151, no. 1, pp. 239–249, 1999.
- [9] B. Dehlinger, J. Jurss, K. Lychuk, and C. Putonti, "The dynamic codon biaser: calculating prokaryotic codon usage biases," *Microbial Genomics*, vol. 7, no. 10, 2021.
- [10] M. C. Angellotti, S. B. Bhuiyan, G. Chen, and X. F. Wan, "Codono: codon usage bias analysis within and across genomes," *Nucleic Acids Research*, vol. 35, pp. 132–136, 2007.
- [11] F. Wright, "The 'effective number of codons' used in a gene," *Gene*, vol. 87, no. 1, pp. 23–29, 1990.
- [12] P. M. Sharp and W. H. Li, "The codon adaptation index—a measure of directional synonymous codon usage bias, and its potential applications," *Nucleic Acids Research*, vol. 15, no. 3, pp. 1281–1295, 1987.
- [13] J. L. Bennetzen and B. D. Hall, "Codon selection in yeast," *Journal of Biological Chemistry*, vol. 257, no. 6, pp. 3026–3031, 1982.
- [14] P. M. Sharp and W. H. Li, "An evolutionary perspective on synonymous codon usage in unicellular organisms," *Journal of Molecular Evolution*, vol. 24, no. 1–2, pp. 28–38, 1986.
- [15] J. C. Tian, L. Han, Q. L. Yu, X. X. Shi, and W. T. Wang, "Changes in tenderness and cathepsins, activity during post mortem ageing of yak meat," *Canadian Journal of Animal Science*, vol. 93, no. 3, pp. 321–328, 2013.
- [16] S. E. Gill and W. C. Parks, "Metalloproteinases and their inhibitors: regulators of wound healing," *The International Journal of Biochemistry & Cell Biology*, vol. 40, no. 6–7, pp. 1334–1347, 2008.
- [17] X. D. Mu, M. L. Urso, K. Murray, F. Fu, and Y. Li, "Relaxin regulates *mmp* expression and promotes satellite cell

- mobilization during muscle healing in both young and aged mice," *American Journal Of Pathology*, vol. 177, no. 5, pp. 2399–2410, 2010.
- [18] S. Christensen and P. P. Purslow, "The role of matrix metalloproteinases in muscle and adipose tissue development and meat quality: a review," *Meat Science*, vol. 119, pp. 138–146, 2016.
- [19] T. Nishimura, "Role of extracellular matrix in development of skeletal muscle and postmortem aging of meat," *Meat Science*, vol. 109, pp. 48–55, 2015.
- [20] S. M. Hindi, J. Shin, Y. Ogura, H. Li, and A. Kumar, "Matrix metalloproteinase-9 inhibition improves proliferation and engraftment of myogenic cells in dystrophic muscle of mdx mice," *PLoS One*, vol. 8, Article ID e72121, 2013.
- [21] C. Sassoli, L. Vallone, A. Tani, F. Chellini, D. Nosi, and S. Zecchi-Orlandini, "Combined use of bone marrow-derived mesenchymal stromal cells (bm-mscs) and platelet rich plasma (prp) stimulates proliferation and differentiation of myoblasts in vitro: new therapeutic perspectives for skeletal muscle repair/regeneration," *Cell and Tissue Research*, vol. 372, no. 3, pp. 549–570, 2018.
- [22] D. Miyazaki, A. Nakamura, K. Fukushima, K. Yoshida, S. Takeda, and S. Ikeda, "Matrix metalloproteinase-2 ablation in dystrophin-deficient mdx muscles reduces angiogenesis resulting in impaired growth of regenerated muscle fibers," *Human Molecular Genetics*, vol. 20, no. 9, pp. 1787–1799, 2011.
- [23] Y. X. Qi, X. H. Zhang, Y. Q. Wang, Z. B. Zhang, T. L. Zhang, and Z. Zhang, "Expression of mmp-1, -2, and -8 in longissimus dorsi muscle and their relationship with meat quality traits in cattle," *Genetics and Molecular Research: GMR*, vol. 15, no. 1, Article ID 15017593, 2016.
- [24] Y. q. Wu, Z. y. Li, D. q. Zhao, and J. Tao, "Comparative analysis of flower-meristem-identity gene *apetala2* (*ap2*) codon in different plant species," *Journal of Integrative Agriculture*, vol. 17, no. 4, pp. 867–877, 2018.
- [25] F. B. Guo and W. Wei, "Strong strand composition bias in the genome of *ehrlichia canis* revealed by multiple methods," *The Open Microbiology Journal*, vol. 4, no. 1, pp. 98–102, 2010.
- [26] N. Sueoka, "Translation-coupled violation of parity rule 2 in human genes is not the cause of heterogeneity of the DNA g + c content of third codon position," *Gene*, vol. 238, no. 1, pp. 53–58, 1999.
- [27] Y. Tamura, M. Higashi, S. Kitamoto et al., "Muc4 and *muc1* expression in adenocarcinoma of the stomach correlates with vessel invasion and lymph node metastasis: an immunohistochemical study of early gastric cancer," *PLoS One*, vol. 7, no. 11, Article ID e49251, 2012.
- [28] N. Sueoka, "Directional mutation pressure and neutral molecular evolution," *Proceedings of the National Academy of Sciences*, vol. 85, no. 8, pp. 2653–2657, 1988.
- [29] W. K. Mitchell, D. J. Wilkinson, B. E. Phillips, J. N. Lund, K. Smith, and P. J. Atherton, "Human skeletal muscle protein metabolism responds to amino acid nutrition," *Advances in Nutrition: An International Review Journal*, vol. 7, no. 4, pp. 828S–838S, 2016.
- [30] R. J. Solaro, P. Rosevear, and T. Kobayashi, "The unique functions of cardiac troponin i in the control of cardiac muscle contraction and relaxation," *Biochemical and Biophysical Research Communications*, vol. 369, no. 1, pp. 82–87, 2008.
- [31] M. A. Rabie, C. Peres, and F. X. Malcata, "Evolution of amino acids and biogenic amines throughout storage in sausages made of horse, beef and Turkey meats," *Meat Science*, vol. 96, no. 1, pp. 82–87, 2014.
- [32] J. Yan, P. Liu, L. Xu et al., "Effects of exogenous inosine monophosphate on growth performance, flavor compounds, enzyme activity, and gene expression of muscle tissues in chicken," *Poultry Science*, vol. 97, no. 4, pp. 1229–1237, 2018.
- [33] G. P. Rizzi, "The strecker degradation of amino acids: newer avenues for flavor formation," *Food Reviews International*, vol. 24, no. 4, pp. 416–435, 2008.
- [34] M. Holeccek, "Branched-chain amino acids in health and disease: metabolism, alterations in blood plasma, and as supplements," *Nutrition and Metabolism*, vol. 15, no. 1, pp. 33–44, 2018.
- [35] S. Fujita and E. Volpi, "Amino acids and muscle loss with aging," *Journal of Nutrition*, vol. 136, no. 1, pp. 277S–280S, 2006.
- [36] H. Li, X. An, L. Bao et al., "Mir-125a-3p-klf15-bcaa regulates the skeletal muscle branched-chain amino acid metabolism in Nile tilapia (*Oreochromis niloticus*) during starvation," *Frontiers in Genetics*, vol. 11, pp. 852–859, 2020.
- [37] J. A. Novembre, "Accounting for background nucleotide composition when measuring codon usage bias," *Molecular Biology and Evolution*, vol. 19, no. 8, pp. 1390–1394, 2002.
- [38] D. B. Carlini, Y. Chen, and W. Stephan, "The relationship between third-codon position nucleotide content, codon bias, mRNA secondary structure and gene expression in the *Drosophila* alcohol dehydrogenase genes *adh* and *adhr*," *Genetics*, vol. 159, no. 2, pp. 623–633, 2001.
- [39] M. L. Christianson, "Codon usage patterns distort phylogenies from or of DNA sequences," *American Journal of Botany*, vol. 92, no. 8, pp. 1221–1233, 2005.
- [40] B. B. Wang, Y. B. Zhang, F. Zhang et al., "On the origin of tibetans and their genetic basis in adapting high-altitude environments," *PLoS One*, vol. 6, no. 2, Article ID e17002, 2011.