Kuwait J.Sci., Vol.49, No.(3),July.2022,pp(1-16)

Comparative analysis of codon usage patterns of *FUT2* from different species

Chao Xu[1], Wen B. Bao[1,2], Sheng L. Wu[1,2], Zheng C. Wu[1,2,*]

[1] *Key Laboratory for Animal Genetics, Breeding, Reproduction and Molecular Design of Jiangsu Province, College of Animal Science and Technology, Yangzhou University, Yangzhou, Jiangsu 225009, China.*
[2] *Joint International Research Laboratory of Agriculture & Agri-Product Safety, Yangzhou University, Yangzhou, Jiangsu 225009, China.*

**Corresponding author: zcwu@yzu.edu.cn*

Abstract

Enterotoxigenic *E. coli* is an important zoonotic pathogen causing diarrhea in humans and newborn animals. α - (1,2) fucosyltransferase 2 (*FUT2*) is closely associated with the formation of pathogenic receptors of Enterotoxigenic *E. coli*. Codon usage bias analysis can help to understand better the molecular mechanisms and evolutionary relationships of a particular gene. To understand the codon usage pattern of the *FUT2* gene, *FUT2* gene coding sequences of nine species were selected from the GenBank database for calculating the nucleotide composition (GC content) and genetic indices, including an adequate number of codons, relative synonymous codon usage, and relative codon usage bias using R software, to analyze codon usage bias and base composition in *FUT2* gene from different species. The results showed that the codon usage of *the FUT2* gene in other species was affected by GC bias, mainly GC frequency at the third position of codon (GC3). Most of the optimal codons were biased towards the G/C-ending types. GCC, CUG, UCC, GUG, and AUC showed the highest relative synonymous codon usage value among different species belonging to the most dominant codons. The usage characteristic of the codons for *the FUT2* gene in *Sus scrofa* was similar to that of *Bos taurus*; *Homo sapiens* was similar to *Pan troglodytes*. An adequate number of codons was significantly negatively correlated with GC3. The relatively higher frequency of optimal codons implied that *FUT2* genes from different species had a strong bias in codon usage.

Keywords: Base composition; codon usage bias; different species; optimal codons, α - (1,2) fucosyltransferase 2.

1. Introduction

Nucleic acids are templates for protein synthesis. A total of 61 codons encode 20 translated essential amino acids. In this degenerate code, most amino acids are encoded by two to six synonymous codons used at different frequencies, a phenomenon is known as codon usage bias

(Ikemura, 1985). There are mainly two kinds of theories on codon usage bias: neutral theory and selection-mutation-drift theory (Bulmer, 1991). The neutral theory explains that mutations in the third nucleotide of the codons result from unbiased selection, and mutations of synonymous codons do not affect survival fitness. The choice of codons is only associated with mutations that are not affected by natural selection pressure. "Selection-mutation-drift" theory suggests that the occurrence of mutations is directional, and codon usage bias of synonymous codons reflects the selection of the optimal codons and the effect of both mutations and drift on non-synonymous codons (Romero *et al*., 2000). In addition to the impact of selection and mutation, codon usage has been reported to be affected by a s*et* of factors, including base composition (Fedorov *et al*., 2002), gene expression level (Ikemura, 1985), tRNA abundance (Hiraoka *et al*., 2009), gene length (Duret, 2000), mRNA secondary structure (Sun *et al*., 2009), protein hydrophobicity and amino acid conservation (Gu *et al*., 2004). With the emergence of genome-wide sequencing technologies for different species and the visualization of nucleotide sequences in the NCBI GenBank database, many scientists are interested in studying codon usage bias for understanding molecular mechanisms underlying gene evolution and genome characteristics. From the research perspective, the previous studies mainly focused on the relationship b*et*ween codon bias and gene expression in prokaryotes and lower eukaryotes (Gustafsson *et al*., 2004). Current studies have begun to pay more attention to the codon bias of higher organisms, including mammals, to effectively analyze the differences in gene expression and molecular evolutionary mechanisms in animal species (Angellotti *et al*., 2007) and analyze the genetic relationship among species by comparing the degree of difference in codon bias. The study of codon bias has crucial biological significance. Based on nucleotide composition kinetics and gene transfer analysis, we can not only better understand the molecular evolutionary process but also design transgenes to enhance the expression and discovery of new genes (Carbone *et al*., 2003) and analyze the functional conservation of gene expression (Lithwick & Margalit, 2005). Inconsistence of transcriptional characteristics of synonymous codons may result in a significant change in the number of proteins produced by mRNA transcripts (Miyasaka, 2002). An earlier study confirmed a significant association of codon usage bias to the codon usage of highly expressed genes (Sueoka, 1988). In addition, studies have shown that gene sequence characteristics, mutation pressure, and natural selection affect codon usage patterns (Mandlik *et al*., 2014). The variation of codon bias may occur among species due to selection pressure on a particular gene. Therefore, there may be differences in the characteristics of codon usage patterns among gene families from different species or the same species, suggesting other pressures on genomes and genes during their evolution. Codon usage bias reflects genome or genes' origin and evolution patterns, providing significant reference value for gene expression (Pek *et al*., 2015), gene function prediction, and gene family differentiation (Wei *et al*., 2014). In addition, species' restrictions in differences in codon bias may cause gene methylation, low-level expression, and gene silencing during exogenous gene expression and genetic transformation; hence, it is necessary to optimize target gene and directional modification according to the codon bias of host genome.

Enterotoxigenic *Escherichia coli* (*E. coli*) is an important zoonotic pathogen causing diarrhea in humans and newborn young animals (Qadri *et al*., 2005; Al-Balool, 2003). Enterotoxigenic *E. coli* causes disease by adhering the pilus colonization factor to corresponding receptors in the small intestine and releasing enterotoxins to lyse cells. It was found that α-(1,2) fucosyltransferase 2 (*FUT2*) is involved in the formation of H-2 (α-fuc- (1-2) -Gal- (1-4) -GlcNAc), playing a role in the adhesion process of *E. coli* to small intestinal epithelial cells (Coddens *et al*., 2009). Pili are hair-like appendages found on the surface of *E. coli*. Studies have reported that F18, K88 (F4), and K99 (F5) receptors belong to glycoproteins and glycosaminoglycans (Mouricout, 1997). The biological function of fucosyltransferases may be significantly associated with the structure formation of *E. coli* receptors. In addition, fucosyltransferases are mainly involved in synthesizing oligosaccharide chains in the body, and the construction of glycolipids, phospholipids, and mucins, which act in various signals transmission and antigen recognition processes (Hurd *et al*., 2005). Mucin is involved in cell signaling, adhesion, growth, and immune regulation, playing an essential role in maintaining intestinal stability and inflammation (Bansil & Turner, 2018). The current studies revealed a significant association of the expression level of *the FUT2* gene with the occurrence of Crohn's disease, Norwalk virus infection, and sugar remodeling during organ transplant rejection immune response (Whyte *et al*., 2011). In addition, *FUT2* gene expression is associated with the resistance to pathogenic bacterial infection in the small intestine

(Hurd *et al*., 2005). Therefore, the adhesion ability of *E. coli* to the intestine is affected by gene expression levels of *FUT2* that may be associated with enzymatic activity. In the present study, we investigated the nucleotide composition (GC content), genetic indexes, including an adequate number of codons (ENC), relative synonymous codon usage (RSCU), and relative codon usage bias (RCBS) to analyze the kinetics and evolutionary relationships of the codons of *FUT2* gene, thereby elucidating the codon usage bias characteristics of *FUT2* gene in different species. In addition, this study provided a theoretical basis for altering the expression level of the *FUT2* gene in host cells to enhance the ability of resistance to pathogenic bacterial infection by codon optimization and transgenic engineering methods.

2. Materials and methods

2.1 Sequence data sources

The *FUT2* gene complete coding sequences of nine species (ATG is used as initiation codon, TAA, TAG, or TGA as stop codons) were downloaded from the GenBank database (http://www.ncbi.nlm.nih.gov). The primary information is shown in (Table 1).

Table 1. Nine species with accession number and length (bp) of coding sequences for the *FUT2* gene

CDS NO.	Animals	Accession NO.	Length (bp)
1	*Bos taurus*	XM_005219153	1035
2	*Cavia porcellus*	XM_003465532	1089
3	*Colobus angolensis palliates*	XM_011946446	1032
4	*Felis catus*	XM_011289738	1029
5	*Homo sapiens*	U17894.1	1032
6	*Mus musculus*	XM_011250795	1044
7	*Pan troglodytes*	AF080604	1017
8	*Rattus norvegicus*	AF131238	1065
9	*Sus scrofa*	AF136895	1023

* CDS NO: Coding sequence number.

2.2 Base composition bias analysis

The base composition of each gene sequence was calculated to determine the level of base composition bias using Mobile software (http://www.molbiol.ox.ac.uk/cu, version 1.4.2) (Peden, 2000) with criteria as follows: 1) base content of the third position in each codon (A3, U3, C3, G3); 2) GC of the first codon (GC1) and the second codon (GC2); 3) total GC of the codon (GC), and average GC of the first and second codon (GC12); 4) overall GC frequency of the third codon (GC3). In addition, AT, GC, and GC3 skewness for each coding sequence was used to assess base composition bias.

2.3 ENC analysis

The effective number of codons (ENC) is considered one of the most valuable parameters for assessing the overall codon usage bias of a gene, with a range from 20 to 61. Highly expressed genes have high levels of codon bias and lower ENC values, whereas lowly expressed genes contain more types of rare codons and higher ENC values (Fuglsang, 2004).

2.4 RSCU analysis

Relative synonymous codon usage (RSCU) is defined as the observed frequency of a codon divided by the expected frequency if all codons are used equally for any particular amino acid (Sharp, 1986). If the RSCU value of a codon is equal to one, the use of the codon is unbiased. If the RSCU value of a codon is more significant than one, the codon is more frequently used than expected, whereas if the RSCU value of a codon is less than one, the codon is less frequently used than desired. Amino acid composition effects are removed; therefore, RSCU reflects codon usage bias directly.

2.5 Frequency of optimal codon analysis

The optimal codon for an amino acid is also defined as the codon with the most significant number of tRNA genes. The frequency of optimal codon is a standard indicator of codon usage bias (Ikemura, 1985). The frequency of optimal codon value represents the ratio of the number of optimal codons to all synonymous codons (Ikemura, 1981a, b), ranging from 0.36 for a gene showing uniform codon usage bias to 1 for a gene showing strong codon usage bias (Stenico *et al*., 1994). The frequency of optimal codon value for each coding sequence was calculated according to the previous method (Lavner & Kotlar, 2005).

2.6 Gene expression evaluation analysis

Gene expression was assessed by RCBS, defined as the overall score affected by the relative codon bias of each codon in the gene (Roymondal *et al*., 2009). The RCBS value for each *FUT2* gene sequence was calculated according to the formula of Fox & Erill (2010).

2.7 Phylogenetic Analysis

To determine the evolutionary relationship between different species, phylogenetic analysis based on *FUT2* RSCU value in other species was performed using MEGA5.0 software. A neighbor-joining tree was constructed using the Kimura 2-parameter method, and the values are in the units of the number of base substitutions per site. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) is shown next to the branches. The tree is drawn to scale, with branch lengths in the same units as the evolutionary distances used to infer the phylogenetic tree. All positions containing gaps and missing data were eliminated.

2.8 Statistical analysis

Correlation analysis and significant comparison of methods were conducted using Microsoft Excel and Spearman's rank method in SPSS 18.0 (http://www.spss.com/) software. Heat maps were clustered using the hierarchical clustering method in R software. The correlation coefficient between codon and GC3 and RSCU values in different species were estimated using R software.

3. Results

3.1 Codon usage pattern analysis of FUT2 gene in different species

Codon usage pattern scanning provides the basis for revealing the mechanism of synonymous codon usage bias and has important theoretical implications for understanding molecular biology

(Hassan *et al*., 2009). To show the relationship between codon usage variation and GC condition constraints in *the FUT2* gene sequence, we analyzed the correlation between codon base composition and GC bias (GC3) using heat maps. Most of the codons with G/C-ending were positively correlated with GC3, whereas most codons ending with A/T were negatively associated with GC3 (Figure 1). The codons CAC, CGC, CTG, GGC, TAC, and TCG were strongly positively correlated with GC3 ($P < 0.05$). In contrast, the codons AAT, ACA, CGA, CCT, GAT, and TTT were strongly negatively correlated with GC3 ($P < 0.05$). The results indicated that the codon usage pattern of the *FUT2* gene is probably affected by GC3 bias.

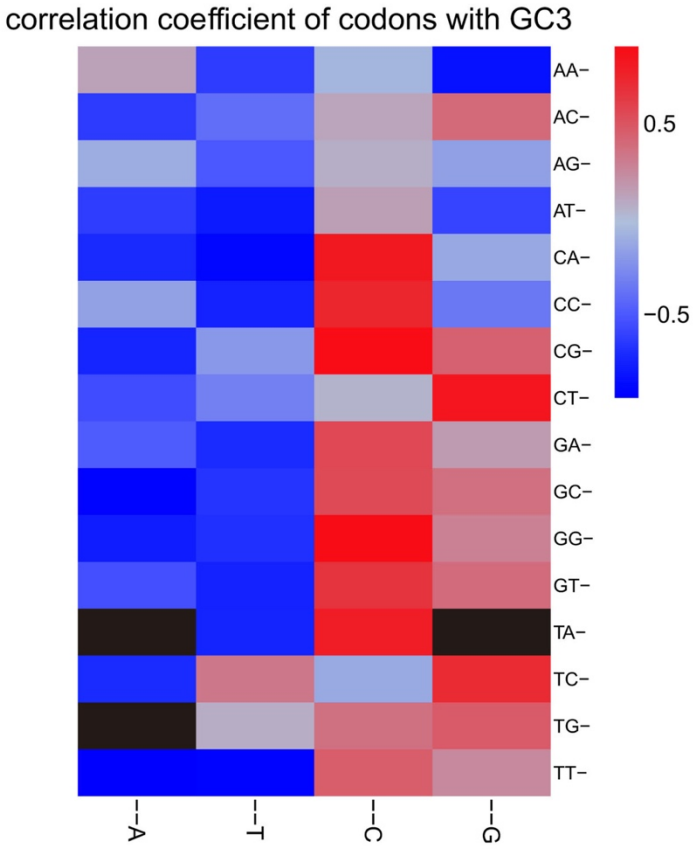

Fig. 1. Heat map of correlation coefficient of codons with GC3

3.2 Bias analysis of codon ending with G/C of FUT2 gene in different species

In the present study, the base compositions of coding sequences of *the FUT2* gene in different species were analyzed (Table 2). The average content of the third base of codons (A3, T3, C3, G3) showed the highest ratios in C3 and G3, followed by T3 and A3. GC3 values ranged from 71.0% to 93.3%, with an average of 80.6% and SD of 0.074. The average GC content (GC12) of the first and second position of codons varied from 49.3% to 55.5%, with an average of 51.2% and SD of 0.021. Base composition analysis suggested that the codons of the *FUT2* gene coding sequence

were biased with G/C-ending in the selected species. In addition, the Fop of each amino acid was calculated according to the method of (Lavner & Kotlar, 2005) to identify the codons with high and low usage frequencies. The results showed 18 optimal codons in the *FUT2* gene from different species, most of which are ended with G/C (Figure 2).

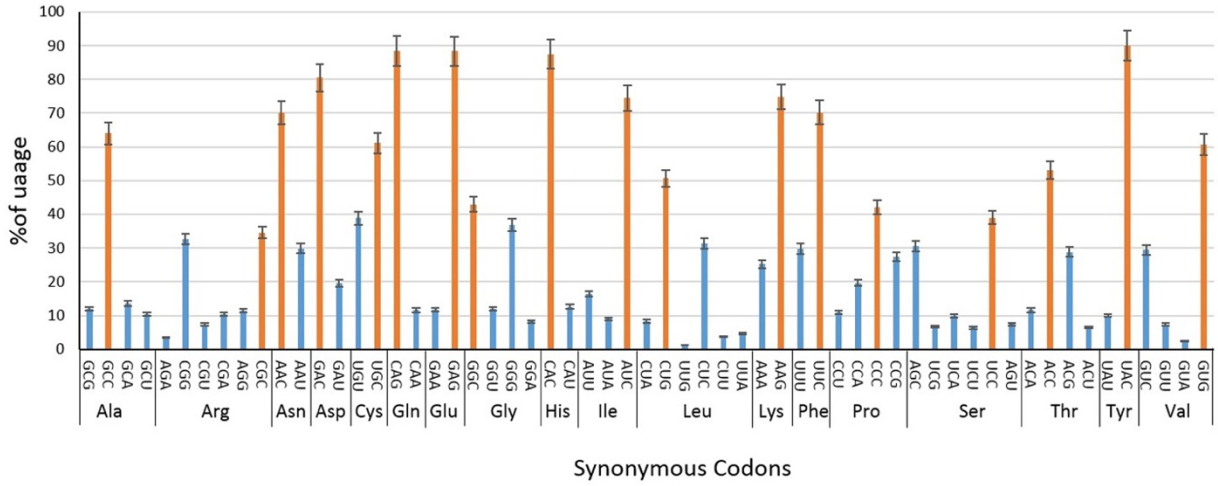

Fig. 2. The overall frequency of optimal and non-optimal codons used in *FUT2* genes among species

Table 2. Nucleotide composition analysis in the coding sequences of *FUT2* gene

NO.	A	T	G	C	A3	T3	G3	C3	AT %	GC %	GC1 %	GC2 %	CG3 %	AT3 %	GC12 %
1	203	194	280	358	27	35	114	169	0.384	0.616	0.586	0.443	0.820	0.180	0.514
2	195	190	314	390	28	23	133	179	0.354	0.646	0.623	0.457	0.860	0.140	0.540
3	213	206	279	334	35	41	116	152	0.406	0.594	0.581	0.422	0.779	0.221	0.501
4	148	169	318	383	9	11	120	200	0.308	0.681	0.630	0.481	0.933	0.058	0.555
5	213	206	281	332	37	37	116	154	0.406	0.594	0.578	0.419	0.785	0.215	0.499
6	228	226	272	318	46	55	108	139	0.435	0.565	0.566	0.420	0.710	0.290	0.493
7	208	203	279	327	35	38	115	151	0.404	0.596	0.578	0.425	0.785	0.215	0.501
8	227	231	290	317	43	60	117	135	0.430	0.570	0.577	0.423	0.710	0.290	0.500
9	187	192	301	343	14	30	133	164	0.370	0.630	0.575	0.443	0.871	0.129	0.509
Mean	202.4	201.9	290.4	344.7	30.4	36.7	119.1	160.3	0.390	0.610	0.588	0.437	0.806	0.193	0.512
SD	24.4	18.9	16.7	26.8	12.4	15	8.5	20.3	0.04	0.037	0.022	0.021	0.074	0.076	0.021

3.3 RSCU of FUT2 gene in different species

RSCU values of 59 codons of *the FUT2* gene (except for ATG methionine and TGG tryptophan) in nine species were estimated. RSCU > 1 represents relatively higher codon usage, whereas RSCU > 1.6 indicates strongly biased codons. RSCU analysis of the *FUT2* gene in the selected species showed that all 22 high-frequency codons were biased to ending with G/C (Table 3). RSCU clustering analysis showed that the RSCU values of codons including GCC, CUG, UCC, GUG, and AUC in *the FUT2* gene were more significant than 1.6 and, therefore, are considered dominant codons (Figure 3)

Table 3. Overall relative synonymous codon usage patterns (RSCU) for *FUT2* gene among different species

Amino Acid	Codon	N	RSCU[a]	Amino Acid	Codon	N	RSCU[a]
Ala	GCA	36	0.535	Leu	CUA	24	0.502
	GCC*	172	2.558		CUC*	90	1.882
	GCG	33	0.491		CUG*	145	3.031
	GCU	28	0.416		CUU	11	0.230
Arg	AGA	7	0.205		UUA	13	0.272
	AGG	23	0.673		UUG	4	0.084
	CGA	20	0.585	Lys	AAA	21	0.500
	CGC*	73	2.137		AAG*	63	1.500
	CGG*	67	1.961	Phe	UUC*	129	1.395
	CGU	15	0.439		UUU	56	0.605
Asn	AAC*	76	1.382	Pro	CCA	40	0.812
	AAU	34	0.618		CCC*	83	1.685
Asp	GAC*	85	1.619		CCG	53	1.076
	GAU	20	0.381		CCU	21	0.426
Cys	UGC*	19	1.267	Ser	AGC*	50	1.796
	UGU	11	0.733		AGU	13	0.467
Gln	CAA	13	0.211		UCA	17	0.611
	CAG*	110	1.789		UCC*	65	2.335
Glu	GAA	16	0.235		UCG	11	0.395
	GAG*	120	1.765		UCU	11	0.395
Gly	GGA	18	0.326	Thr	ACA	22	0.463
	GGC*	96	1.738		ACC*	100	2.105
	GGG*	81	1.466		ACG	56	1.179
	GGU	26	0.471		ACU	12	0.253
His	CAC*	117	1.746	Tyr	UAC*	106	1.812
	CAU	17	0.254		UAU	11	0.188
Ile	AUA	16	0.289	Val	GUA	5	0.097
	AUC*	121	2.187		GUC	61	1.179
	AUU	29	0.524		GUG*	126	2.435
					GUU	15	0.290

[a] mean value of RSCU based on the synonymous codon usage frequencies of *the FUT2* gene, N: Total number of preferred codons, *RSCU>1.

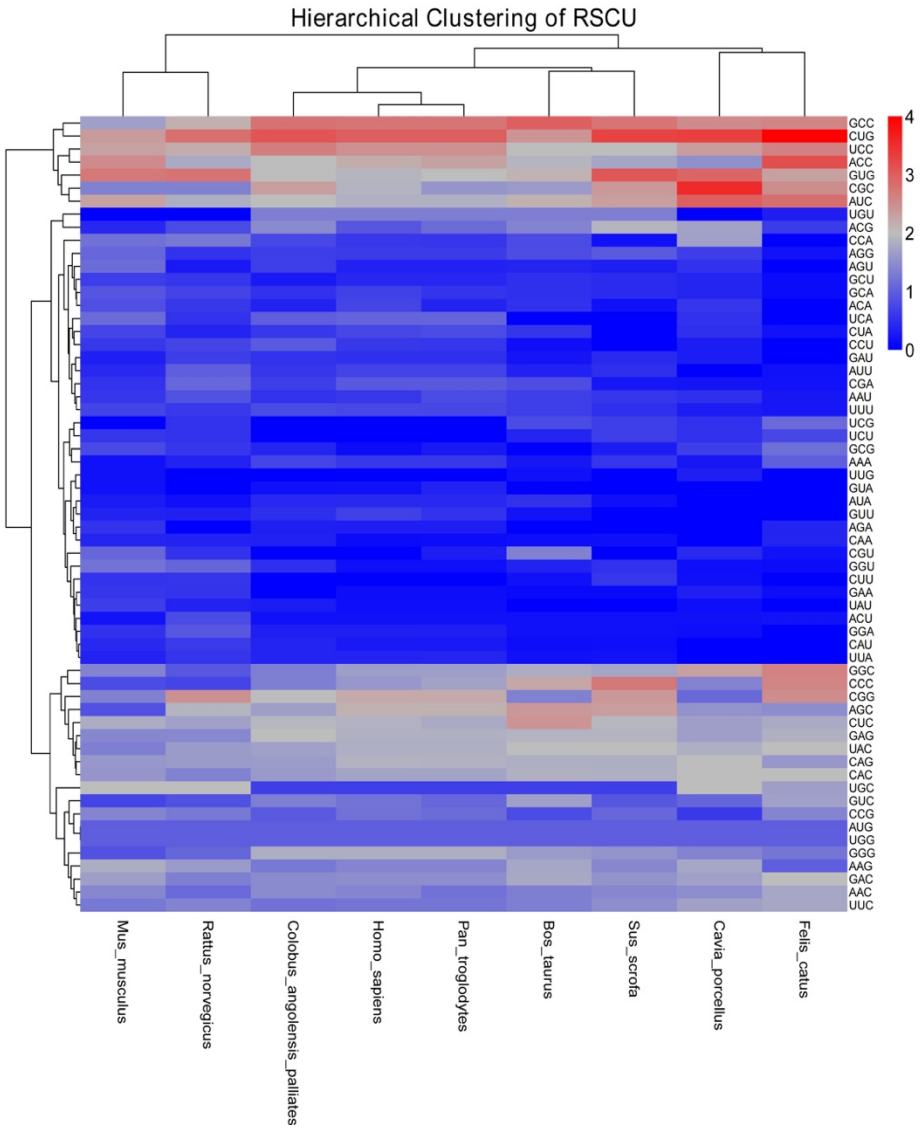

Fig. 3. Clustering of RSCU values of each codon among *FUT2* genes across species

3.4 Phylogenetic analysis based on the codon usage pattern of the FUT2 gene in different species

In the present study, a neighbor-joining tree was constructed using the Kimura 2-parameter distances of *the FUT2* gene with RSCU values of codons in different species (Figure 4). The results showed that *the FUT2* gene of the closely related species had a similar pattern of codon usage. The design of codon usage of *FUT2* genes in pigs (*Sus scrofa*) was identical to cattle (*Bos taurus*), whereas the pattern of codon usage of *FUT2* genes in humans (*Homo sapiens*) was close to chimpanzees (*Pan troglodytes*). Genes with similar functions generally have a similar pattern of codon usage (Tatarinova *et al*., 2010). The present study results suggested some differences in the codon usage pattern of the *FUT2* gene among different species.

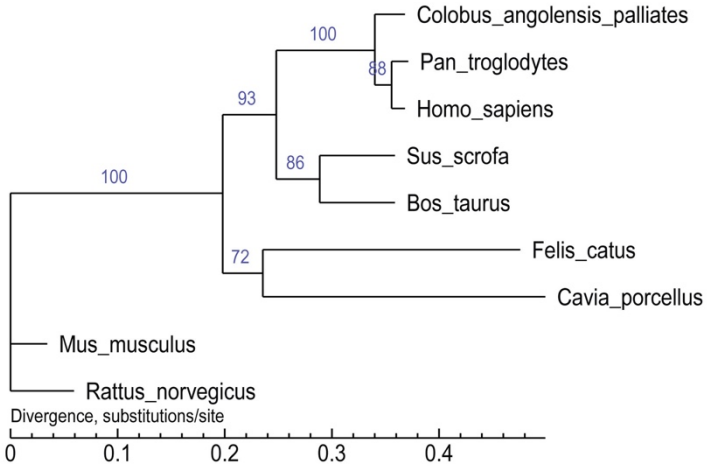

Fig. 4. Phylogenetic analysis of the Kimura 2-parameter distances of the selected RSCU value among *FUT2* genes of different species

3.5 Effect of selection pressure on FUT2 gene in different species

The ENC values ranged from 46 to 59 (mean ± SD, 54.44 ± 3.91), indicating minor variation in codon usage of the *FUT2* gene among different species. The GC3 ranged from 0.710 to 0.933 (mean ± SD, 0.806 ± 0.074), whereas ENC was significantly negatively correlated with GC3 (Pearson r = –0.922, $P < 0.01$). The ENC plot (ENC vs. GC3) showed a negative correlation of ENC values with GC3 contents, while lower ENC values had higher GC3 values (Figure 5). In addition, we calculated the GC3 skew that varied from – 0.250 to – 0.071. The relatively higher frequency of optimal codon values suggested a strong codon usage bias in *FUT2* genes of different species (Table 4). The RCBS value of a gene can estimate gene expression (Ma *et al*., 2002). The RCBS value of *the FUT2* gene ranged from 0.167 to 0.219 (mean ± SD, 0.199 ± 0.016) in different species. The lower average RCBS values place the *FUT2* gene in the class of genes with lower expression levels.

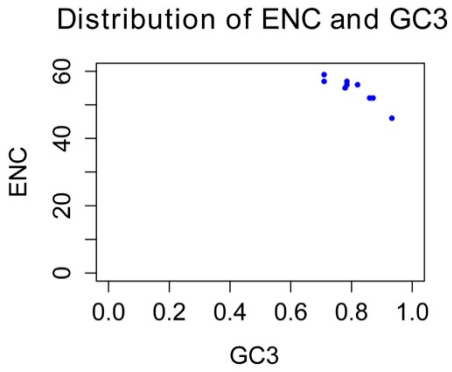

Fig. 5. ENC vs. GC3 values for *FUT2* gene

Table 4. Codon usage bias indices for *FUT2* gene across different species*

Species	RCBS	CG3%	Fop	Highest_RSCU	GC Skew	GC3 Skew
Bos_taurus	0.219	0.820	0.483	GCC(Ala)	-0.122	-0.194
Cavia_porcellus	0.206	0.860	0.566	CGC(Arg)	-0.108	-0.147
Colobus_angolensis_palliates	0.197	0.779	0.485	CUG(Leu)	-0.090	-0.134
Felis_catus	0.200	0.933	0.602	CUG(Leu)	-0.093	-0.250
Homo_sapiens	0.214	0.785	0.498	CUG(Leu)	-0.083	-0.141
Mus_musculus	0.167	0.710	0.523	GUG(Val)	-0.078	-0.126
Pan_troglodytes	0.201	0.785	0.495	CUG(Leu)	-0.079	-0.135
Rattus_norvegicus	0.181	0.710	0.494	CUG(Leu)	-0.044	-0.071
Sus_scrofa	0.206	0.871	0.516	CUG(Leu)	-0.065	-0.104

* RCBS-Relative codon usage bias, ENC-Effective number of codons, GC3-GC contents at third positions of a codon, FOP-Frequency of optimal codons, RSCU-Relative synonymous codon usage.

4. Discussion

Any species usually develops a set of genome-specific codon usage during long-term evolution history. Complex factors create codon bias, and it is not only the result of gene mutations and selection (Wong *et al*., 2002) but also related to the gene coding structure, function, and gene expression (Chiapello *et al*., 1998). Codon bias is proposed to be related to many factors in evolution. Codon usage in both prokaryotic and unicellular eukaryotes is affected primarily by mutational bias and selection (Uddin *et al*., 2016). In contrast, codon usage in multicellular eukaryotes is mainly affected by the selection of translation efficiencies in *Drosophila melanogaster* and *Caenorhabditis elegans* (Vicario *et al*., 2007). Codon usage is greatly affected by the interaction between mutation pressure and natural selection in viruses such as *Parvoviridae* (Shi *et al*., 2013). In addition, studies in multicellular and single-cell eukaryotic organisms revealed that the number of preferred codons/optimal codons containing large numbers of tRNA gene copies increased with the change in gene expression level, indicating a positive association of the optimal codons with tRNA abundance (Hiraoka *et al*., 2009). However, codon usage is mainly affected by gene base composition and gene expression level in plant species, including *Arabidopsis thaliana*, rice (Oryza sativa), and corn (*Zea mays*) (Liu *et al*., 2010). In human studies, genes' gene expression, nucleotide composition, and codon usage bias are

associated with the Y chromosome (Choudhury *et al*., 2017). In recent years many studies discovered that in the absence of natural selection pressure conditions, the base composition of gene sequence is affected by mutation pressure in a specific direction, mainly in the third base of synonymous codons (Uddin *et al*., 2019). In addition, base composition is an essential characteristic of genomic DNA. The usage frequency of different nucleotides is determined by the balance between mutation and reverse mutation. The GC content is believed to reflect the overall trend of the mutation. The alternation of the third base of the codons within a family usually does not change the encoded amino acids. Therefore, a mutation in the third base of codons has less selective pressure, and GC3 is also an essential indicator for analyzing codon usage patterns. In this study, we investigated the compositional dynamics of *the FUT2* gene in different species. Further, we found that the *FUT2* gene was biased toward G/C-ending codons, and most of the codons were significantly correlated with GC3. Therefore, the codon bias of *the FUT2* gene is strongly affected by base composition, especially GC3 content. The ENC plot (ENC vs. GC3) revealed a significant negative correlation between most ENC values and GC3 content, indicating that relatively higher ENC values were associated with lower GC3 values and smaller GC3 skew values. The results further confirmed the vital role of the GC3 composition in codon usage bias (Yang & Nielsen, 2008). In addition, higher ENC values are associated with lower gene expression. In this study, we found that the ENC value of *the FUT2* gene in all nine species was about 50, suggesting the low expression level of *the FUT2* gene in different species, which may be related to the *FUT2* gene function.

Long-term natural evolution has led to similar codon usage patterns among closely related species or genes; therefore, codon usage rules are often used as an essential reference for species or genomic classification, origin, evolution, and phylogenetic relationships. Previous studies have suggested that the clustering based on codon usage is more accurate in smaller taxonomic units. Some closely related species have similar biological functions for a particular gene (Christianson, 2005). In the present study, the codon-based clustering analysis showed that patterns of codon usage of *FUT2* genes in pig (*Sus scrofa*) and cattle (*Bos taurus*) were very close. Previous studies have shown that many factors may result in codon bias among species, among which mutation and selection are the significant factors. The optimal codons in the translation are all caused by the balance between mutation pressure, genetic drift, and weak selection within species. Therefore, mutations from unbiased codons to preferred codons are favorable, while mutations in the opposite direction are harmful. In addition, highly expressed genes generally display a higher bias toward synonymous codons (Behura *et al*., 2010). The positive correlation between codon bias and expression level is usually caused by the selection of translational efficiency (Behura & Severson, 2011), indicating that the dominant codons are selected in highly expressed genes. Therefore, predominant codons are usually clearly present in highly expressed genes (Ikemura, 1982). Our study identified the most dominant codon groups GCC, CUG, UCC, GUG, and AUC in the *FUT2* gene, providing guidance for modifying *FUT2* gene expression levels in animals to improve resistance to pathogenic infection by genetic engineering and codon optimization.

5. Conclusion

In conclusion, we found that codon usage of *the FUT2* gene in different species was affected by GC bias, especially GC3. Base composition and occurrence frequency of optimal codons indicated that the codons of *the FUT2* gene were biased toward G/C-ending types. The frequency of C was more remarkable than G. The pattern of codon usage of *FUT2* genes in pigs (*Sus scrofa*) was similar to cattle (*Bos taurus*). In contrast, the codon usage pattern of *FUT2* genes in humans (*Homo sapiens*) was close to chimpanzees (*Pan troglodytes*). The ENC values were significantly negatively associated with GC3 content. The relatively higher frequency of optimal codon value implied that *the FUT2* gene is strongly biased towards the codon usage in different species.

ACKNOWLEDGEMENTS

This study was supported by Jiangsu Agricultural Science and Technology Innovation Fund (CX (20)3011), Natural Science Foundation of Jiangsu province, China (Grant No. BK20180899), and the Priority Academic Program Development of Jiangsu Higher Education Institutions.

References

Al-Balool, F. Y. (2003). Effect of procaine and lidocaine on escherichia coli sta-induced fluid secretion across small intestine of gerbil (Gerbillus cheesmani). Kuwait journal of science & engineering, **30**: 49-58.

Angellotti, M.C., Bhuiyan, S.B., Chen, G. & Wan, X.F. (2007). Codono: codon usage bias analysis within and across genomes. Nucleic Acids Research **35**:132-136.

Bansil, R. & Turner, B.S. (2018). The biology of mucus: Composition, synthesis, and organization. Advanced Drug Delivery Reviews **124**:3-15.

Behura, S.K. & Severson, D.W. (2011). Coadaptation of isoacceptor tRNA genes and codon usage bias for translation efficiency in Aedes aegypti and Anopheles gambiae. Insect Molecular Biology **20**:177-187.

Behura, S.K., Stanke, M., Desjardins, C.A., Werren, J.H. & Severson, D.W. (2010). Comparative analysis of nuclear tRNA genes of Nasonia vitripennis and other arthropods, and relationships to codon usage bias. Insect Molecular Biology **19**:49-58.

Bulmer, M.G. (1991). The selection-mutation-drift theory of synonymous codon usage. Genetics **129**:897-907.

Carbone, A., Zinovyev, A. & Kepes, F. (2003). Codon adaptation index as a measure of dominating codon bias. Bioinformatics **19**:2005-2015.

Choudhury, M.N., Uddin, A. & Chakraborty, S. (2017). Gene Expression, Nucleotide Composition and Codon Usage Bias of Genes Associated with Human Y Chromosome. Genetica **145**:295-305.

Christianson, M.L. (2005). Codon usage patterns distort phylogenies from or of DNA sequences—American Journal of Botany **92**:1221-1233.

Chiapello, H., Lisacek, F., Caboche, M., and Hénaut, A. (1998) Codon usage and gene function are related in sequences of Arabidopsis thaliana. Gene, 209, 1-38.

Coddens, A., Diswall, M., Angstrom, J., Breimer, M.E. & Goddeeris, B. (2009). Recognition of blood group ABH type 1 determinants by the FedF adhesin of F18-fimbriated Escherichia coli. Journal of Biological Chemistry **284**:9713-9726.

Duret, L. (2000). tRNA gene number and codon usage in the C. elegans genome are co-adapted for optimal translation of highly expressed genes. Trends in Genetics **16**:287-289.

Fedorov, A., Saxonov, S. & Gilbert, W. (2002). Regularities of context-dependent codon bias in eukaryotic genes. Nucleic Acids Research **30**:1192-1197.

Fox, J. M. & Erill, I. (2010). Relative codon adaptation: a generic codon bias index for prediction of gene expression. DNA Research **17**:185-196.

Fuglsang, A. (2004). The 'effective number of codons' revisited. Biochemical and Biophysical Research Communications **317**:957–964.

Gu, W., Zhou, T., Ma, J., Sun, X. & Lu, Z. (2004). The relationship between synonymous codon usage and protein structure in Escherichia coli and Homo sapiens. Biosystems **73**:89-97.

Gustafsson, C., Govindarajan, S. & Minshull, J. (2004). Codon bias and heterologous protein expression. Trends in Biotechnology **22**:346-353.

Hassan, S., Mahalingam, V. & Kumar, V. 2009. Synonymous codon usage analysis of thirty-two mycobacteriophage genomes. Advances in bioinformatics 2009:1-11.

Hiraoka, Y., Kawamata, K., Haraguchi, T. & Chikashige, Y. (2009). Codon usage bias is correlated with gene expression levels in the fission yeast Schizosaccharomyces pombe—genes to Cells **14**:499-509.

Hurd, E.A., Holmén, J.M., Hansson, G.C. & Domino, S.E. (2005). Gastrointestinal mucins of fut2-null mice lack terminal fucosylation without affecting colonization by candida Albicans. Glycobiology **15**:1002-1007.

Ikemura, T. (1981a). 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 optimal for the E. coli translational system. Journal of molecular biology **151**:389-409.

Ikemura, T. (1981b). Correlation between the abundance of Escherichia coli transfer RNAs and the occurrence of the respective codons in its protein genes. Journal of molecular biology **146**:1-21.

Ikemura T. (1982). Correlation between the abundance of yeast transfer RNAs and the occurrence of the respective codons in protein genes. Differences in synonymous codon choice patterns of yeast and Escherichia coli with reference to the abundance of is accepting transfer RNAs. Journal of molecular biology, **158**:573–597.

Ikemura, T. (1985). Codon usage and tRNA content in unicellular and multicellular organisms. Molecular Biology and Evolution 2:13-34.

Lavner, Y. & Kotlar, D. (2005). Codon bias is a factor in regulating expression via translation rate in the human genome. Gene 345:127-138.

Lithwick, G. & Margalit, H. (2005). Relative predicted protein levels of functionally associated proteins are conserved across organisms. Nucleic Acids Research **33**:1051-1057.

Liu, H., He, R., Zhang, H., Huang, Y., Tian, M. & Zhang, J. (2010). Analysis of synonymous codon usage in Zea mays. Molecular Biology Reports **37**:677-684.

Ma, J.M., Zhou, T., Gu, W.J., Sun, X. & Lu, Z.H. (2002). Cluster analysis of the codon use frequency of MHC genes from different species. Biosystems **65**:199-207.

Mandlik, V., Shinde, S. & Singh, S. (2014). Molecular evolution of the enzymes involved in the sphingolipid metabolism of leishmania: selection pressure in relation to functional divergence and conservation. BMC Evolutionary Biology **14**:142.

Miyasaka, H. (2002). Translation initiation AUG context varies with codon usage bias and gene length in Drosophila melanogaster. Journal of Molecular Evolution **55**:52-64.

Mouricout, M. (1997). Interactions between the enteric pathogen & the host. In Mechanisms in the pathogenesis of enteric diseases. p. 109-123. Springer, Boston, MA.

Peden, J.F. (2000). Analysis of codon usage. University of Nottingham **90**:73–74.

Pek, H.B., Klement, M., Ang, K.S., Chrng, B.K. & Lee, D.Y. (2015). Exploring codon context bias for synthetic gene design of a thermostable invertase in Escherichia coli. Enzyme and Microbial Technology **75-76**:57–63.

Qadri, F., Svennerholm, A.M., Faruque, A.S. & Sack, R.B. (2005). Enterotoxigenic Escherichia coli in developing countries: epidemiology, microbiology, clinical features, treatment, & prevention. Clinical microbiology reviews **18**:465-483.

Romero, H., Zavala, A. & Musto, H. (2000). Codon usage in Chlamydia trachomatis is the result of strand-specific mutational biases and a complex pattern of selective forces. Nucleic Acids Research **28**:2084-2090.

Roymondal, U., Das, S. & Sahoo, S. (2009). Predicting gene expression level from relative codon usage bias: an application to Escherichia coli genome. DNA Research **16**:13-30.

Sharp, P.M. & Li, W.H. (1986). An evolutionary perspective on synonymous codon usage in unicellular organisms. Journal of Molecular Evolution **24**:28-38.

Shi, S.L., Jiang, Y.R., Liu, Y.Q., Xia, R.X. & Qin, L. (2013). Selective pressure dominates the synonymous codon usage inparvoviridae—virus Genes **46**:10-19.

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

Sun, Z., Ma, L., Murphy, R., Zhang, X.S. & Huang, D.W. (2009). Codon usage analysis of Wolbachia pipientis wMEL genome. Science in China Press **39**:948–953.

Stenico, M., Lloyd, A.T., Sharp, P.M. (1994） Codon usage in Caenorhabditis elegans: delineation of translational selection and mutational biases. Nucleic Acids Research, 22, 2437-2446.

Tatarinova, T.V., Alexandrov, N.N., Bouck, J.B. & Feldmann, K.A. (2010). GC 3 biology in corn, rice, sorghum, and other grasses. BMC Genomics **11**:308.

Uddin, A. & Chakraborty, S. (2016). Codon usage trend in mitochondrial CYB gene. Gene **586**:105-14.

Uddin, A., Paul, N. & Chakraborty, S. (2019). The codon usage pattern of genes involved in ovarian cancer. Annals of the New York Academy of Sciences Marrow **1440**:67-78.

Vicario, S., Moriyama, E.N. & Powell, J.R. (2007). Codon usage in twelve species of Drosophila. BMC Evolutionary Biology **7**:226.

Wei, L., He, J., Jia, X., Qi, Q. & Liang, Z.S. *et.al*. (2014). Analysis of codon usage bias of mitochondrial genome in Bombyx mori and its relation to evolution. BMC Evolutionary Biology **14**:262.

Whyte, J.J. & Prather, R.S. (2011). Genetic modifications of pigs for medicine and agriculture. Molecular Reproduction and Development **78**:879-891.

Wong, K.S., Wang, J., Tao, L., Tan, J. & Zhang, J. G. *et.al*. (2002). Compositional gradients in Gramineae genes. Genome Research **12**:851-856.

Yang, Z. & Nielsen, R. (2008). Mutation-selection models of codon substitution and their use to estimate selective strengths on codon usage. Molecular Biology and Evolution **25**:568-579.

Submitted: 24/05/2020
Revised: 24/08/2021
Accepted: 30/08/2021
DOI: 10.48129/kjs.11289