

口蹄疫病毒密码子使用偏性的位点差异

马晓霞 冯玉萍 赵永清 陈莉 刘俊林 郭鹏辉 陈红 杨具田,
卢建雄 陈世恩 马忠仁

(西北民族大学 生命科学与工程学院,甘肃 兰州 730010)

摘要: 分析了 FMDV 各血清型病毒基因编码起始区与终止区的密码子使用偏性。结果表明, FMDV 各血清型中的稀有密码子倾向出现于编码起始区, 而终止区附近稀有密码子出现的倾向较弱。这一有趣的现象, 可利用“稀有密码子调控假说”来解释为 FMDV 起始编码区的稀有密码子的使用对其编码区表达具有负调控作用。而终止区附近的这种较弱的倾向性表明, 终止区密码子的使用对编码区基因表达的影响可能没有起始区那样强烈。同时说明, “稀有密码子调控假说”不仅适用于细菌, 而且也适用于一些病毒基因组。

关键词: FMDV; 血清型; 密码子; 调控

中图分类号: Q78 文献标识码: A 文章编号: 1000-7091(2013)02-0032-06

Site Discrepancy of Synonymous Codon Usage in FMDV

MA Xiao-xia, FENG Yu-ping, ZHAO Yong-qing, CHEN Li, LIU Jun-lin, GUO Peng-hui,
CHEN Hong, YANG Ju-tian, LU JIAN-xiong, CHEN Shi-en, MA Zhong-ren

(College of Life Science and Engineering, Northwest University for nationalities, Lanzhou 730010, China)

Abstract: The synonymous codon usage in the translational initiation and termination regions of genes of seven different serotypes belonging to foot-and-mouth disease virus (FMDV) was particularly analyzed. The results indicate that the pattern of codon usage with low propensity at some sites in the initial region is preferentially existed, and the interesting phenomenon may imply a negative effect on gene expression, which can be explained by the "minor codon modulator hypothesis". Nevertheless, this interesting phenomenon fails to exist obviously in the terminal region. Obviously, some information about the pattern of codon usage in the initial translation region may indicate that the sites of codon usage with low propensity may have an more obvious effect on FMDV gene transcription or translation than the terminal region. The proposed results imply that the "minor codon modulator hypothesis" may be applied to both bacteria and some RNA viruses (e. g. FMDV).

Key words: Foot-and-mouth disease virus; Serotype; Codon usage; Modulator

Because of the degeneracy of genetic code, many amino acids are coded by more than one codon. The genetic code is redundant: 20 amino acids plus start and stop signals are coded by 64 codons. This redundancy increases the resistance of genes to mutation: the third codon letters, which are termed wobble bases, can often be interchanged without affecting the primary sequence of the protein product. Nevertheless, wobble base usage is highly conserved in mRNA sequences (there is no

very little individual or intra-species variation) and interestingly, some wobble mutations, which are called silent mutations, are known to contribute to genetic disease with no change in the amino acid sequences^[1].

Search of the synonymous codon usage can cast slight on the molecular evolution of individual genes and obtain data to train genome specific gene recognition algorithms which recognize protein coding regions in uncharacterized and non-understanding genomic

收稿日期: 2013-01-28

基金项目: 校级中青年科研基金 (12XB37)

作者简介: 马晓霞 (1982-), 女, 甘肃兰州人, 助理实验师, 硕士, 主要从事分子生物学与生物化学研究。

通讯作者: 马忠仁 (1962-), 男, 甘肃兰州人, 教授, 博士, 主要从事基因工程疫苗研究。

DNA or RNA. It has also been reported that synonymous codons are not used equally both within and between genomes^[2]. Codon usage bias may rise from some factors. Base composition constraints and translation selection are thought to be the main factors accounting for codon usage variation among genes in different organisms^[3]. The diverse patterns of codon usage in mammals may arise from compositional constraints of the genomes. In contrast, in some unicellular organisms i. e. *Escherichia coli* and *Saccharomyces cerevisiae* high expressed genes have a strong selective preference for codons which are recognized by most abundant tRNAs, whereas low expressed genes appear a more uniform pattern of codon usage. Ma et al. pointed out that codon usage was related to gene functional protein secondary structure^[2,4]. Further analysis addressed that the pattern of synonymous codon usage varied at sites along a coding sequence^[5]. In many bacterial species, *Deinococcus radiodurans* and *Methanobacterium thermoautotrophicum*, some minor codons are preferentially used near the translational initiation site^[6]. These minor codons are thought to play an important role in gene expression.

Foot-and-mouth disease virus (FMDV) is the causative agent of the economically most important animal viral disease world-wide. FMDV is a non-enveloped virus with icosahedral symmetry that belongs to the *aphthovirus* genus of the *picornavirus* family. The virus exists in the form of seven different serotypes: A, O, C, Asia 1 and South African Territories 1 (SAT 1), SAT 2 and SAT 3. Its genome is a single stranded RNA molecule of about 8 500 nucleotides of positive polarity. Purified genomic FMDV RNA can act as messenger RNA *in vitro* and *in vivo*. The RNA extracted from virions or transcribed from full-length cDNA copies of the viral genome is infectious. This permits manipulation of DNA copies of specific FMDV genomic segments to study the phenotypic effects of mutations and other genomic alterations^[7]. The genome have a single long open reading frame (ORF) and encode all of its proteins in form of a polyprotein.

Although genome sequence of FMDV has been published and many studies have been performed on FMDV, little genomic analysis is available on this virus. In this search, we have used the available complete

gene sequences and analyzed the codon usage patterns in the translational initiation and termination regions of FMDV. Codon usage bias data of FMDV and the comparison results might give some clues to the features of FMDV genome.

1 Materials and Methods

1.1 Nucleic acid data set

For some analyses, complete genome or the whole polyprotein FMDV sequences currently available in GenBank release seven were considered and included: Serotype A: AY593769, AY593789, AY593767, AY593770; Serotype O: EF552697, EF552696, EF552695, EF552694; Serotype Asia 1: DQ989322, DQ533483, AY687334, AY687333; Serotype C: AF274010, FJ824812, AM409325, DQ409191; Serotype SAT 1: AY593838, AY593839, AY593840, AY593841; Serotype SAT 2: AF540910, AY593847, AY593848, AY593849; Serotype SAT 3: AY593850, AY593851, AY593852, AY593853.

1.2 Relative synonymous codon usage (RSCU)

To examine synonymous codon usage without the confounding influence of amino acid composition of different gene samples, the values of Relative Synonymous Codon Usage (RSCU) of different codons in per genome have been calculated. The RSCU value of the i -th codon for the j -th amino acid is calculated^[8].

$$RSCU = \frac{g_{ij}}{\sum_j g_{ij}} \cdot n_i$$

Where g_{ij} is the observed number of the j -th codon for the i -th amino acid which has n_i type of synonymous codons. It is obvious that RSCU values close to 1.0 demonstrate lack of bias for the corresponding codon.

1.3 Analyzing the tendency of codon usage in given regions

Due to the mutational phenomenon at each site of the target region, we define a variant of the Tendency of Codon Usage (TCU) to analyze the tendency of codon usage of given regions in coding sequences. TCU values are calculated followed by: $\Delta RSCU = RSCU_{ij} - RSCU_0$. More simply, TCU is the cumulative value of $\Delta RSCU$ values in each position of the interesting region, namely, the initiation site. $RSCU_{ij}$ means that one codon usage value corresponding a particular amino acid in a

position of the interesting region; n means all samples in this work. When all RSCU values according to a particular position in the region are $RSCU_0$, TCB is zero. It means that there are little dominant codons or minor codons existing in this position. In contrast, when the absolute value of TCB is much more than $RSCU_0$, it means that optimizing codons or minor codons are used in a particular position.

We focused on the initial region from the translational initiation site to the 30th downstream codon and the terminal region from the translational termination site to the 30th upstream codon. The TCU values of the corresponding genome were calculated. And the RSCU values of the corresponding genome were calculated to compare with the TCU values in order to address the relationship between certain codons in the given region and one corresponding genome.

2 Results and Analysis

2.1 Synonymous codon usage and codon usage bias in FMDV

The overall RSCU values of 59 sense codons in FMDV were respectively shown in Tab. 1. Both the codons with strong propensity and the ones with relative strong propensity have been respectively tagged with asterisk and plus sign. As for them, there are codons with very strong propensity and ones with weak propensity existing in the same amino acid.

2.2 Codon usage near translation initiation site

Fig. 1 represents the relationship between each particular site (from the start codon to 30th codon) and TCU for seven serotypes of FMDV. Each bar represents a tendency of codon usage in a particular site. A high TCU value demonstrates the corresponding site inclines to appear codons with strong propensity in the translational initiation region.

Our study presents that there are several sites with negative codon usage propensity in the all interesting region. This means that some minor codons trend to appear in the translational initiation region, while they are low preferentially used in the whole genome of these serotypes. These results are partially consistent with the previous report where several bacteria genomes are studied^[6].

Tab. 1 Synonymous codon usage and codon usage bias in FMDV genome

AA ^a	Codon	RSCU	AA ^a	Codon	RSCU
Ala ⁺	GCA	1.02	Leu [*]	CUA	0.17
	GCC	1.45		CUC	1.95
	GCG	0.59		CUG	1.65
	GCU	0.94		CUU	1.10
Arg [*]	AGA	1.46	Lys	UUA	0.05
	AGG	0.80		UUG	1.05
	CGA	0.33		AAA	0.86
	CGC	1.77		AAG	1.14
Asn [*]	CGG	0.66	Phe	UUC	1.23
	CGU	0.97		UUU	0.77
	AAC	1.72		CCA	0.82
	AAU	0.28		CCC	1.29
Asp [*]	GAC	1.50	Pro	CCG	0.78
	GAU	0.50		CCU	1.10
Cys	UGC	1.17	Ser ⁺	AGC	1.23
	UGU	0.83		AGU	0.65
Gln	CAA	0.89		UCA	1.02
	CAG	1.15		UCC	1.56
Glu ⁺	GAA	0.71	Thr ⁺	UCG	0.89
	GAG	1.29		UCU	0.76
Gly	GGA	1.04		ACA	0.90
	GGC	1.20		ACC	1.56
His [*]	GGG	0.82	Tyr [*]	ACG	0.60
	GGU	0.94		ACU	0.94
	CAC	1.75		UAC	1.71
	CAU	0.25		UAU	0.29
Ile [*]	AUA	0.17	Val [*]	GUA	0.27
	AUC	1.80		GUC	1.13
	AUU	1.10		GUG	1.62
				GUU	0.98

Note: ^aAA is the abbreviation of amino acid; ^bRSCU values are mean values; ^cThe preferentially used codons for each amino acid are described in bold; ^{*} The amino acids which include the very discrepancy among synonymous codons usage; ⁺ The amino acids which include the relative discrepancy among synonymous codons usage.

Codons with minus TCU values in seven serotypes are showed in Fig. 1. This phenomenon is especially evident at the 2nd, 3rd, 4th, 5th, 6th, 9th, 18th, 19th, 22th, 23th, 24th, 26th, 30th sites in serotype A, the 3rd, 4th, 6th, 9th, 13th, 18th, 19th, 23th, 26th sites in serotype O, the 3rd, 6th, 7th, 9th, 21th, 23th, 24th, 25th, 26th, 28th, 30th sites in serotype Asia 1, the 2nd, 3rd, 4th, 6th, 7th, 9th, 11th, 17th, 18th, 22th, 23th, 26th, 28th, 30th site in serotype C, the 3rd, 4th, 6th, 14th, 16th, 20th, 23th, 28th sites in serotype SAT 1, the 2nd, 3rd, 4th, 6th, 7th, 9th, 20th, 22nd, 23rd, 26th, 26th, 29th sites in serotype SAT-2, and the 3rd, 4th, 6th, 7th, 9th, 14th, 16th, 20th, 23th, 28th, 30th sites in

serotype SAT 3. To some extent, these sites with low codon usage propensity often trend to gather in both

ends and in the middle frame, many sites with high codon usage propensity exist.

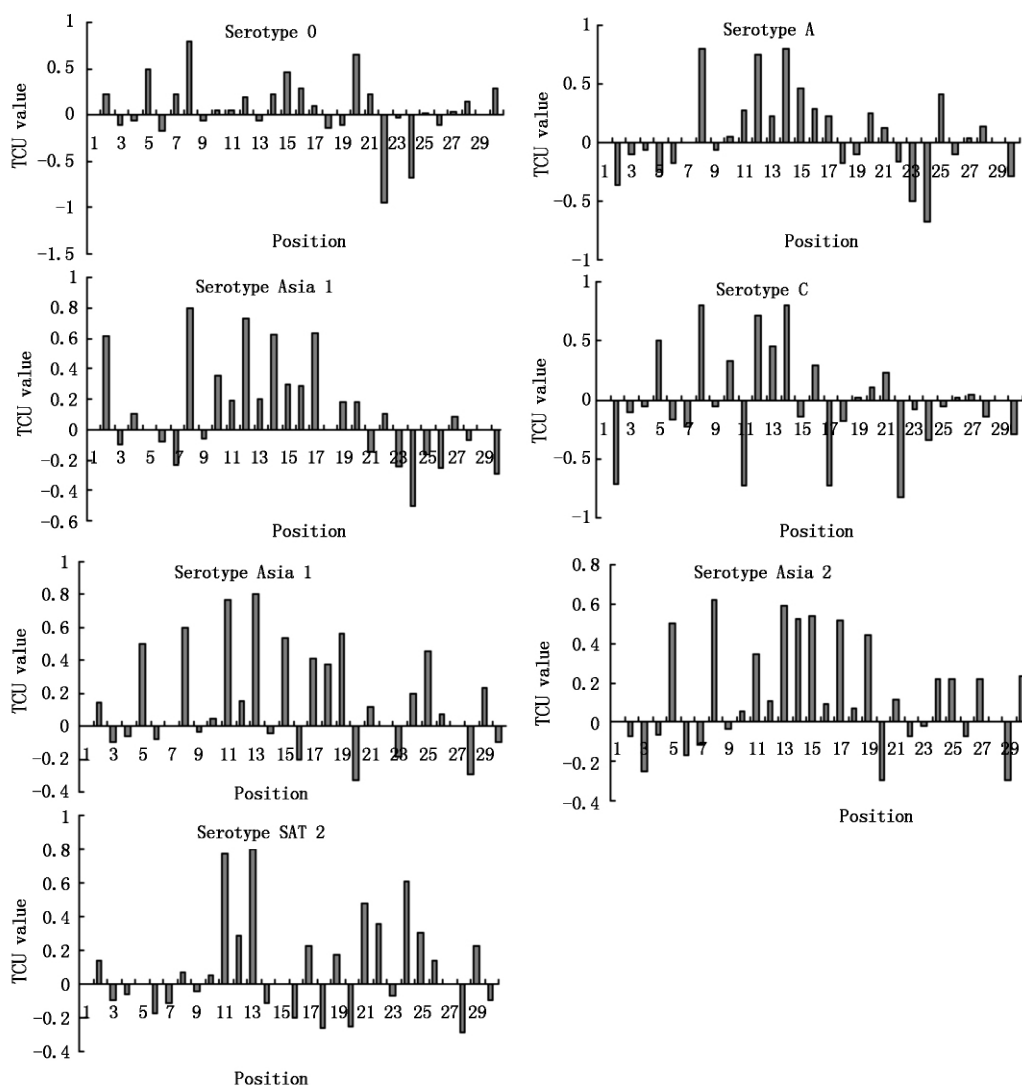


Fig. 1 The tendency of codon usage at each site in the initial translation region

2.3 Codon usage near translational termination site

Fig. 2 describes TCU values for sites which derive from the terminal translation region. Compared Fig. 2 to Fig. 1, like the sites with codon usage near the initial translation region, these translational termination sites fail to be preferentially used. These results are inconsistent with the previous report^[9]. In our study, only a few sites with negative propensity, while a large part of sites trend to use codons which own high RSCU values. The phenomenon may reflect that the translational termination sites with low codon usage propensity may fail to regulate the gene expression.

The two analytic results are only partially corre-

spondent with other RNA virus^[10]. It may imply that RNA viruses' genomes have specific characteristics for the initial translation region and the terminal translation region respectively.

3 Discussion

By calculating RSCU values for certain codons, the method is one of the approaches to the codon usage bias in the whole genome of some organisms^[8]. RSCU is the observed number of codon occurrences divided by the number expected if synonymous codons were used uniformly. This approach made it possible to avoid subjectivity and serotypes limitations in choosing the reference set of codons.

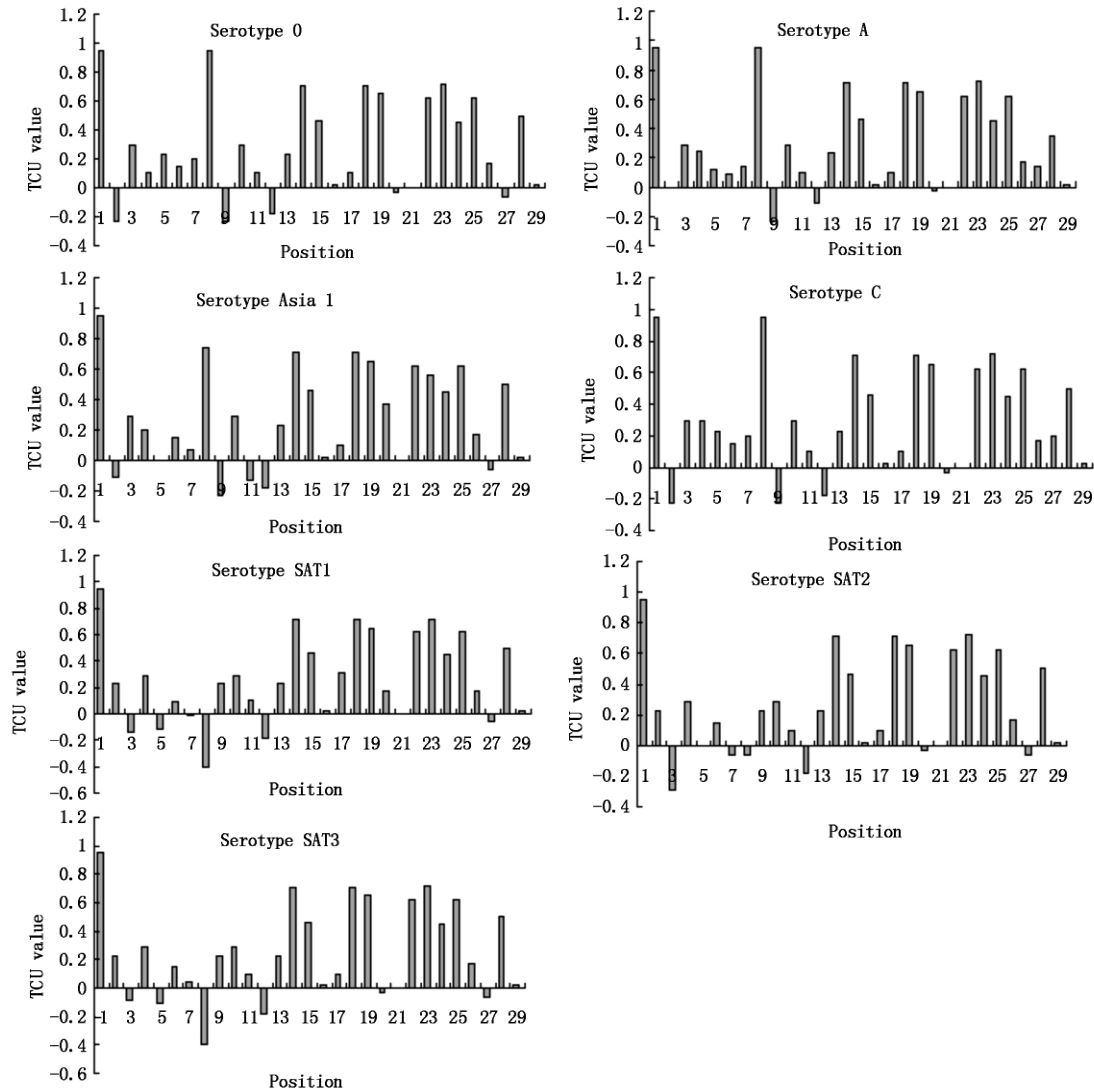


Fig. 2 The tendency of codon usage at each site in the terminal translation region

Among many organisms, synonymous codons are not used equally. Often the pattern of codon usage can be attributed to a general A + T composition or G + C composition preference that pervades the entire genome. As for the nucleic composition of FMDV genome, due to rich in G + C, all optimal codons are G-end and C-end, namely high RSCU values for these codons. In our study, it is generally believed that minor codons often are T-end or A-end (Tab. 1). It means that in the initial translation region, if a series of sites choose minor codons, this structure may have a significant effect on the rate of translation by means of the secondary structure in mRNA.

By means of calculating RSCU values and TCU values, the two parameters demonstrate some interesting bio-informatics about both the initial translation region and the terminal translation region of FMDV. The information about the initial translation region may confirm

that FMDV polyprotein from the initial translation region containing minor codons or non-optimal codons is less efficient than that from the terminal translation region containing non-minor codons. Because even a single minor codon in the initiation site can reduce gene expression as a result of limited availability of tRNAs depending on the host cell, the area comprised of a series of sites with low codon usage propensity in the initial translation region probably have a negative effect on gene expression. This can be explained by the "minor codon modulator hypothesis"^[11]. When the concentration of tRNA for minor codons becomes extremely limited, ribosomes of the host cell stall at minor codon sites, inhibiting the effective entry of a ribosome at the initiation site, thereby resulting in a decrease in the rate of translation.

Comparing Fig. 2 to Fig. 1, the observation indicates that many sites in the translational termination re-

gion with high TCU values address that sites in the terminal translation region trend to choose non-minor codons. As for codon composition prior to achievement of polyprotein production expression, the terminal translation region fail to limit translation.

In inclusion, the codon usage in the translational initiation site is more connected with the regulation of gene expression than in the translational termination site. Many studies confirm the existence of codon bias and significant correlations have been found between codon bias and various biological parameters such as gene expression level^[12-14], gene length^[15-16], gene translation initiation signal^[17], protein amino acid composition^[18], protein structure^[19-20], tRNA abundance^[21], mutation frequency and pattern^[22-23], and GC composition^[24]. There is a interesting phenomenon that the tendency of codon choice in the interesting region reflects a general biological function because of the universal nature of the genetic code and the structure and function of nucleic acids and proteins.

参考文献:

- [1] Sauna Z E, Kimchi-Sarfaty C, Ambudkar S V, *et al.* The sounds of silence: synonymous mutations affect function [J]. *Pharmacogenomics* 2007 8: 527 - 532.
- [2] Gu W, Zhou T, Ma J *et al.* Analysis of synonymous codon usage in SARS coronavirus and other viruses in the Nidovirales [J]. *Virus Research* 2004 101 (2) : 155 - 161.
- [3] Francino H P, Ochman H. Isochores result from mutation not selection [J]. *Nature* 1999 400(6739) : 30 - 31.
- [4] Ma J, Zhou T, Gu W *et al.* Cluster analysis of the codon use frequency of MHC genes from different species [J]. *Biosystems* 2002 65(2) : 199 - 207.
- [5] Hooper S D, Berg O G. Gradients in nucleotide and codon usage along *Escherichia coli* genes [J]. *Nucleic Acids Research* 2000 28(18) : 3517 - 3523.
- [6] Ohno H, Sakai H, Washio T, *et al.* Preferential usage of some minor codons in bacteria [J]. *Gene* 2001 276: 107 - 115.
- [7] Belsham G J. Distinctive features of foot-and-mouth disease virus a member of the picornavirus family; aspects of virus protein synthesis, protein processing and structure [J]. *Prog Biophys Mol Biol* 1993 60: 241 - 60.
- [8] Sharp P M, Touhy T, Mosurski K. Codon usage in yeast cluster analysis clearly differentiates highly and lowly expressed genes [J]. *Nucleic Acids Research* 1986 14 (13) : 5125 - 5143.
- [9] Li W. Position dependence of synonymous codon usage [J]. *Acta Biophysica Sinica* 2001 17(3) : 529 - 534.
- [10] Zhou T, Gu W J, Ma J M *et al.* Site discrepancy of synonymous codon usage in SARS Coronavirus and other viruses in Coronaviridae [J]. *Journal of Southeast University* 2005 21: 203 - 206.
- [11] Chen G T, Inouye M. Suppression of the negative effect of minor arginine codons on gene expression; preferential usage of minor codons within the first 25 codons of the *Escherichia coli* genes [J]. *Nucleic Acids Research* 1990 18 1465 - 1437.
- [12] Gouy M, Gautier C. Codon usage in bacteria: correlation with gene expressivity [J]. *Nucleic Acids Res* 1982 10: 7055 - 7074.
- [13] Sharp P M, Li W H. An evolutionary perspective on synonymous codon usage in unicellular organisms [J]. *J Mol Evol* 1986 24: 28 - 38.
- [14] Sharp P M, Li W H. The codon Adaptation Index—a measure of directional synonymous codon usage bias, and its potential applications [J]. *Nucleic Acids Res* 1987 15: 1281 - 1295.
- [15] Bains W. Codon distribution in vertebrate genes may be used to predict gene length [J]. *J Mol Biol* 1987 197: 397 - 388.
- [16] Eyre-Walker A. Synonymous codon bias is related to gene length in *Escherichia coli*: selection for translational accuracy [J]. *Mol Biol Evol* 1996 13: 864 - 872.
- [17] Ma J, Campbell A, Karlin S. Correlations between Shine-Dalgarno sequences and gene features [J]. *J Bacteriol* 2002 184: 5733 - 7545.
- [18] Lobry J R, Gautier C. Hydrophobicity, expressivity and aromaticity are the major trends of amino-acid usage in 999 *Escherichia coli* chromosome-encoded genes [J]. *Nucleic Acids Res* 1994 22: 3174 - 3180.
- [19] D'Onofrio G, Ghosh T C, Bernardi G. The base composition of the genes is correlated with the secondary structures of the encoded proteins [J]. *Gene* 2002 300: 179 - 187.
- [20] Xie T, Ding D, Tax X, *et al.* The relationship between synonymous codon usage and protein structure [J]. *FEBS Lett* 1998 434: 93 - 96.
- [21] Bulmer M. Coevolution of codon usage and transfer RNA abundance [J]. *Nature* 1987 325: 728 - 730.
- [22] Sueoka N. Directional mutation pressure, selective constraints and genetic equilibria [J]. *J Mol Evol* 1999 34: 95 - 114.
- [23] Sueoka N. Two aspects of DNA base composition: G + C content and translation-coupled deviation from intrastrand rule of A = T and G = C [J]. *J Mol Evol* 1999 49: 49 - 62.
- [24] Karlin S, Mrazek J. What drives codon choice in human genes? [J]. *J Mol Biol* 1996 262: 459 - 472.