African Journal of Microbiology Research Vol. 5(26), pp. 4592-4600, 16 November, 2011
Available online at http://www.academicjournals.org/AJMR
ISSN 1996-0808 ©2011 Academic Journals
DOI: 10.5897/AJMR11.668

Full Length Research Paper

Analysis of codon usage in peste des petits ruminant's virus

Xin-sheng LIU, Yong-lu Wang*, Yong-guang Zhang, Yu-zhen Fang, Li Pan, Jian-liang LÜ, Peng Zhou, Zhong-wang Zhang, Shou-tian Jiang

State Key Laboratory of Veterinary Etiological Biology, National Foot and Mouth Disease Reference Laboratory, Key Laboratory of Animal Virology of Agriculture, Lanzhou Veterinary Research Institute, Chinese Academy of Agricultural Sciences, Lanzhou 730046, China.

Accepted 8 September, 2011

In this study, it is indicated that the bias of synonymous codon usage of Peste des Petits Ruminants virus (PPRV) gene is lower without intra-specific performance through analysis on PPRV gene synonymous codon usage. The virus has interspecies codon usage specific performance in Morbillivirus. PPRV synonymous codon usage pattern is mainly caused by genetic base, the changes of codon usage bias are mainly caused by pressure of gene mutations. Meanwhile, the gene function also determines PPRV codon usage to some extent. However, the gene length and translation stress have no influence on codon usage patterns in genes of the virus. The codon usage bias pattern of PPRV that we analyzed plays an important role for understanding the evolution process of the virus, in particular to understanding the role of mutation and selection in this process.

Key words: Peste des petits ruminants virus, synonymous codon usage, mutational bias, base composition.

INTRODUCTION

Genetic code refers to corresponding relation between nucleotide sequence in deoxyribonucleic acid (DNA) or m ribonucleic acid (RNA) and the amino acid sequences of its encoded protein, including sense codons of 61 kinds of amino acids and 3 kinds of stop codons which generally do not encode any amino acids (UAA, UAG and UGA). Generally each kind of amino acid corresponds to one or more (up to 6) codons. Codons encoding the same amino acid are called synonymous codons. In the process of protein synthesis and the synonymous codons encoding amino acids are not randomly used (Dittmar et al., 2006; Lloyd et al., 1992; Xie et al., 1998). Many studies showed that different species or different genes of the same species have obvious preference on the codon usage (Chiapello et al., 1998; Dams and Antoniw, 2003; Zhou et al., 2005). Codon usage bias is mainly affected by muta-tion preference, translation choice, protein secondary structure, replication and transcription choice, protein hydrophobic and hydrophilic nature, external environment, and other factors (Levin et al., 2000; Gupta and Ghosh, 2001; Onofrio et al., 2002; Gu et al., 2004; Romero et al., 2000; 2003; Vander and Farias, 2006). Peste des Petits Ruminants virus (PPRV) is the member of measles virus category in the paramyxovirus family, other members of the same category also include Rinderpest virus, Canine distemper virus, Porpoise distemper virus, Measles virus and Dolphin distemper virus (Amjad et al., 1996). PPRV is single strand negative chain ribonucleic acid (RNA) virus without segment with six genes of N, P, M, F, H and L from the RNA 3 'end to 5' end in turn, six types of structure proteins with corresponding codes are respectively core clothing capsid protein (N), phosphoprotein (P), membrane protein (M), fusion protein (F), hemagglutinin (H) and large protein (L)

*Corresponding author. E-mail: wangyonglumd@yahoo.cn. Tel: 09318343796. Fax: 09318343796.

Abbreviations: PPRV, Peste des Petits Ruminants virus; **ENC,** effective number of codons; **CA,** correspondence analysis; **ORF,** open reading frame; **RSCU,** relative synonymous codon usage; **NDV,** Newcastle disease virus; **RNA,** ribonucleic acid; **DNA,** deoxyribonucleic acid.

Table 1. PPRV and other viruses of Morbollivirus genome sequences included in this study.

No	Virus	Length (bp)	Accession no
1	PPRV	15,948	EU267273
2	PPRV	15,948	FJ905304
3	PPRV	15,948	NC_006383
4	PPRV	15,948	EU267274
5	PPRV	15,948	AJ849636
6	PPRV	15,948	X74443
7	CDV	15,690	AY445077
8	RPV	15,882	GU168576
9	MV	15,894	NC_001498

in turn, P gene also encodes two nonstructural proteins C and V (Contzer et al., 1994; Bailey et al., 2005) at the same time. PPRV is currently divided into four systems, wherein I, II and III systems are generated from Africa, and IV system is generated from Asia (Kwiatek et al., 2007; Shaila et al., 2007). The virus only has one serotype. Disease was firstly generated in Côte d'Ivoire of Africa in 1942 for the first time (Cargadennec and Lawman, 1942), Peste des petits ruminants almost spread over all countries from Sahara to the equator in the next few years. It was early believed that the hosts of PPRV were limited to goats and sheep, but it was found that wild small ruminants were also infected by PRRV in 1987 (Furley et al., 1987), and camel respiratory disease caused by PPRV was discovered in 2001 (Abraham et al., 2005; Haroun et al., 2002). However, the research on the aspect of PRRV codon usage has not been reported in all studies on PRRV. Therefore, this article analyzes the PRRV codon usage patterns and the factors that affect codon usage, and compares the codon usage differences between PRRV and other virus of the same family. We hope that the research can play an active role in describing PRRV codon usage rules, and future research on the PRRV.

MATERIALS AND METHODS

Sequence data

Full genome sequences of 9 complete virus samples include six PRRV genome sequences, a CDV genome sequence, an RPV genome sequence and an MV genome sequence, all sequence are downloaded from the National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/Genbank/), and detailed information of these strains is listed in Table 1. Open reading frame (ORF) (> 300 bp) contained in each strain virus and the nucleotide composition of each ORF are analyzed using BioEdit 7.0.9 software.

Synonymous codon usage measures

In order to eliminate the influence of amino acid composition on codon usage and directly reflect the usage characteristics of codon, the study evaluates synonymous codon usage bias through statistical estimation on relative synonymous codon usage (RSCU) frequency. The relative synonymous codon usage frequency of the No. j codon of No. i amino acid is calculated according to the published calculation formula (Sharp et al., 1986). RSCU:

$$\mathrm{RSCU}_{ij} = \frac{X_{ij}}{\frac{1}{n_i}\sum_{j-1}^{n_i} X_{ij}}$$

In the Formula, X_{ij} is the occurrence frequency of No. j codon encoding No. i amino acid, n_i is the quantity of synonymous codon family encoding No. i amino acid ($l=n_i=6$). RSCU value refers to the ratio between the usage frequency of one codon in gene sample and expected frequency in the synonymous codon family. If the synonymous codon usage of one amino acid has no preferences, that is, codon usage frequency is close to expected frequency, the RSCU values of codons are equal to 1; if a codon RSCU value is greater than 1, it is indicated that the codon use frequency is higher than expected frequency, whereas it is less than expected value.

The definition on a single gene codon bias is mainly based on effective number of codons (ENC). ENC values can reflect the preference degree of synonymous codon non-equilibrium use in codon family. ENC values are calculated through using the following calculating formula (Wright et al., 1990). ENC:

$$\mathrm{ENC} = 2 + \frac{9}{\mathrm{F}_2} + \frac{1}{F_3} + \frac{5}{F_4} + \frac{3}{F_6}$$

In the formula, F_K (K = 2, 3, 4, 6) refers to the codon use equilibrium with k synonymous codon families in genes. The range of ENC values is from 20 (each amino acid only uses one codon) to 61 (all synonymous codons are equivalently used). ENC value is closer to 20, the degree of being used non-randomly is higher, and the bias is stronger. It is generally believed that the genes are provided with significant codon bias when ENC ≤ 35.

GC% refers to the percentage of base held by G and C in encoding genes. GC 3% refers to the occurrence frequency of G and C in the position of the third codon in addition to methionine, tryptophan, and stop codon, and the contents often reflect the strength of directional mutation pressure, and are closely related to codon preference. A_3, T_3, G_3 and C_3%, respectively refer to the occurrence frequency of nucleotide A, T, G and C in the third codon

Table 2. Synonymous codon usage in PPRV[a, b].

AA[a]	Codon	RSCU [b]	AA[a]	Codon	RSCU [b]
Leu	UUA	0.976	Tyr	UAU	1.191
	UUG	1.600		UAC	0.808
	CUU	0.610	His	CAU	0.945
	CUC	0.946		CAC	1.055
	CUA	0.830	Gln	CAA	0.885
	CUG	1.028		CAG	1.115
Ile	AUU	0.840	Asn	AAU	0.933
	AUC	1.080		AAC	1.066
	AUA	1.073	Lys	AAA	0.925
Val	GUU	0.831		AAG	1.075
	GUC	1.113	Asp	GAU	1.060
	GUA	0.735		GAC	0.940
	GUG	1.321	Glu	GAA	0.735
Gly	GGU	0.820		GAG	1.265
	GGC	0.776	Cys	UGU	1.095
	GGA	1.265		UGC	0.905
	GGG	1.140	Arg	CGU	0.341
Pro	CCU	0.720		CGC	0.328
	CCC	1.033		CGA	0.711
	CCA	1.405		CGG	0.683
	CCG	0.848		AGA	2.175
Thr	ACU	0.701		AGG	1.756
	ACC	1.183	Ser	AGU	0.690
	ACA	1.536		AGC	0.715
	ACG	0.578		UCU	0.921
Ala	GCU	0.838		UCC	1.070
	GCC	1.231		UCA	1.878
	GCA	1.368		UCG	0.726
	GCG	0.560	Phe	UUU	0.901
				UUC	1.098

a AA is the abbreviation of Amino Acid.; b The preferentially used codons for each amino acid are displayed in bold.

position.

Correspondence analysis

Correspondence analysis (CA) is mainly used for detecting the changes of codon RSCU values in genes (Naya et al., 2001). It is an effective multivariate statistical method of studying the internal relation between the variables and samples, and it is successfully applied to the study of codon. In correspondence analysis, all genes in samples are distributed in a 59-dimensional (59 justice codons, in addition to the stop codon, Met, and Trp) vector space, each gene is described with 59 (f'_1, f'_2,…, f'_{59}) variables, the results can be applied for finding out the major factors affecting codon usage bias in genes (Hao et al., 2008). We can judge major factors affecting the gene codon usage according to the correlation and significance between distribution positions of these genes on f'_1 and other parameters (Sau et al., 2006).

Correlation analysis

Correlation analysis of Newcastle disease virus (NDV) was used to identify the relationship between nucleotide composition and synonymous codon usage pattern (Ewens et al., 2001).This analysis was implemented based on the Spearman's rank correlation analysis way.

All statistical processes were carried out with statistical software SPSS 11.5 for windows.

RESULTS

Synonymous codon usage in NDV

RSCU values of 59 synonymous codons in PPRV and nucleotide composition information of each ORF in PPRV genome are listed in Tables 2 and 3. It was discovered that the average GC content is 47.98% and SD is 0.14 through analysis on GC contents of six strain PPRV genomes selected by the experiment. Meanwhile, the A% + U% value is greater than G% + C% value (Table 3). This shows that PPRV is a genome with low GC content. However, the codons with the most preferred usage tend to use A or G at the end in PPRV genome, a total of 12

Table 3. Identified ORFs (length>300 bps) in the PPRV (6 isolates) genome.

ORF	No	A%	U%	G%	C%	A_3%	U_3%	G_3%	C_3%	GC%	GC3%	ENC
N	1	27.88	22.05	27.57	22.50	25.50	21.67	30.01	25.10	50.06	55.11	54.50
	2	27.76	21.17	27.76	23.32	25.90	19.39	30.00	27.00	51.08	57.00	54.66
	3	27.63	21.48	27.69	23.19	25.30	19.39	30.11	27.80	50.89	57.91	53.85
	4	27.69	21.80	27.31	23.19	24.50	20.53	30.02	27.20	50.51	57.22	57.45
	5	27.56	21.46	27.75	23.24	25.30	19.39	30.03	27.80	50.98	57.83	53.85
	6	27.95	21.86	27.12	23.07	25.10	20.91	30.01	26.80	50.19	56.81	57.40
P	1	31.31	21.24	24.12	23.33	27.80	25.29	20.00	26.30	47.45	46.30	56.23
	2	31.11	20.69	24.23	23.97	26.91	24.16	20.62	28.29	48.20	48.91	54.41
	3	30.84	20.63	24.49	24.03	26.52	23.96	20.82	28.68	48.53	49.50	55.03
	4	30.78	21.41	24.17	23.64	26.71	26.91	20.03	26.32	47.81	46.35	58.37
	5	30.84	20.63	24.49	24.03	28.48	23.96	20.82	28.68	48.53	49.50	55.03
	6	30.84	21.94	22.92	24.30	26.32	27.30	20.82	25.54	47.22	46.36	58.13
M	1	29.96	24.31	23.71	22.02	28.27	22.02	21.13	28.57	45.73	49.70	53.15
	2	29.55	23.58	24.28	22.59	26.86	20.59	22.08	30.44	46.87	52.52	53.84
	3	29.95	23.78	23.98	22.29	28.95	22.38	20.59	25.07	46.27	45.66	53.90
	4	29.35	23.98	24.48	22.19	26.86	22.08	22.68	28.65	46.67	51.33	52.76
	5	29.95	23.78	23.98	22.29	28.95	22.38	20.59	28.05	46.27	48.64	53.90
	6	29.35	24.38	24.18	22.09	26.86	23.28	21.79	28.05	46.27	49.84	52.86
F	1	29.98	24.86	23.83	21.33	28.33	25.22	24.49	21.93	45.16	46.42	56.08
	2	30.16	23.69	23.93	22.22	29.30	22.71	23.80	24.17	46.15	47.97	54.03
	3	29.91	23.57	24.05	22.47	29.12	22.52	23.80	24.54	46.52	48.34	52.89
	4	30.04	24.11	23.81	22.04	28.75	23.99	23.99	23.26	45.85	47.25	54.56
	5	29.91	23.57	24.05	22.47	29.12	22.52	23.80	24.54	46.52	48.34	52.89
	6	30.04	23.99	23.69	22.28	29.12	23.26	23.26	24.35	45.97	47.61	54.35
H	1	27.26	26.00	24.36	22.39	23.15	27.58	23.64	25.61	46.74	49.25	53.80
	2	26.98	26.49	24.19	22.33	22.16	28.24	24.30	25.28	46.52	49.58	54.81
	3	27.09	26.60	24.14	22.17	23.31	28.24	23.15	25.28	46.13	48.43	55.81
	4	26.98	26.00	24.52	22.50	22.00	26.43	24.95	26.60	47.02	51.55	55.87
	5	27.09	26.60	24.14	22.17	23.31	28.24	23.15	25.28	46.13	48.43	55.81
	6	27.20	25.94	24.47	22.39	22.98	27.09	24.13	25.78	46.85	49.91	55.92
L	1	28.91	25.74	23.15	22.20	23.40	25.69	23.27	27.62	45.35	50.89	56.30
	2	29.26	25.91	22.84	21.99	24.69	26.52	22.67	26.11	44.83	48.78	56.66
	3	29.01	25.90	23.00	22.09	24.18	26.24	22.95	26.61	45.09	49.56	56.36
	4	28.92	25.67	23.04	22.37	23.91	25.56	22.90	27.62	45.41	50.52	56.63
	5	29.01	25.90	23.00	22.09	24.18	26.24	22.95	26.61	45.09	49.56	56.36
	6	29.03	25.84	22.98	22.16	23.95	26.06	22.90	27.07	45.14	49.97	56.35

most preferred codons are ended with A and G (Table 2). UCA, ACA, CCA, and GUG are codons with the highest usage frequency in 36 ORFs. In addition, GC3% maximum value is 57.91%, the minimum value is 46.35% with the mean value of 50.25%, and the standard deviation is 3.45 in PPRV genes. This implies that the GC content of the third position of the codon affects the PPRV codon usage patterns. But it is worth noting that there are zones with local high GC contents in the PPRV genome mainly in N gene (Table 3).

We also calculated the ENC values and the GC3% value of each gene in order to study the different codon usages of different genes in PPRV genome, the calculation results are listed in Table 3. Data show that the ENC values of different genes are different in PPRV genome. ENC values range from 52.76 to 58.37 with the

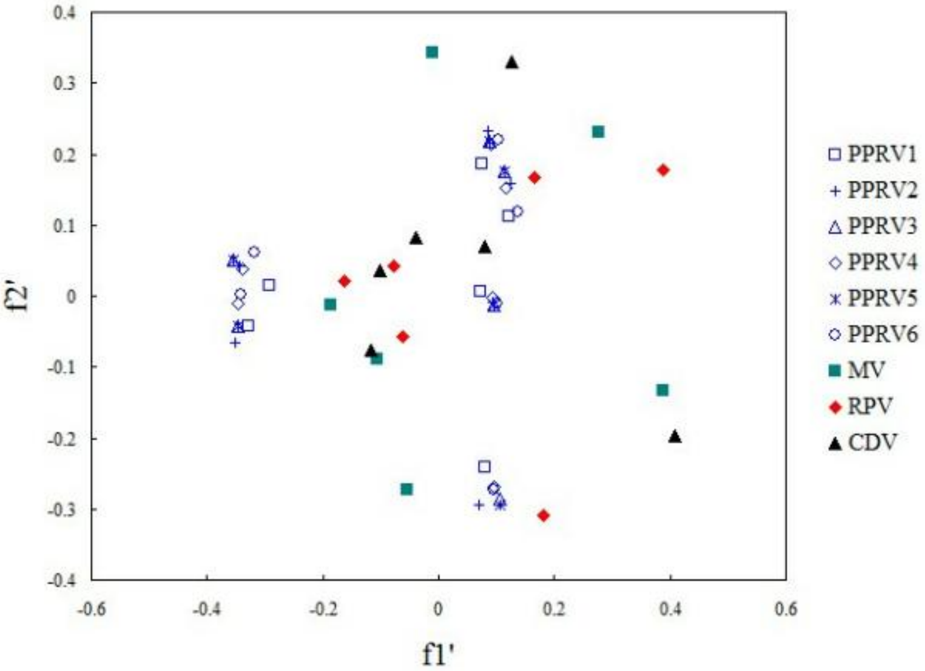

Figure 1. A plot of the values of the first axis and the second axis of each gene of different virus of Morbillivirus in CA (f'1 and f'2, respectively, represent the values of the first and the second axis of each gene in CA). CA has detected one major trend in the first axis, which accounted for 26.03%of the total variation, and none of the other axes have individually accounted for more than 11.94% of the total variation. The two major axes in this plot contributed to the codon usage bias.

mean of 55.13 and standard deviation of 1.53. ENC values of all PPRV genes are larger (ENC> 50), therefore, it can be said that PPRV genome synonymous codon bias is generally low.

It can be seen from the above analysis that PPRV gene synonymous codon usage bias is relatively low collectively, the usage bias is mostly caused by base composition of the genome. These results are similar to reports in some literatures, the report believed that RNA virus overall codon usage bias is relatively weak, and the differences between the genomes are relatively small (Drake and Holland, 1999).

Synonymous codon usage in different viruses of Morbillivirus is virus specific

We divided genes in the same type of virus in Morbillivirus into a group and used principal component analysis to analyze all selected 54 coding genes of four different types of virus in Morbillivirus in order to compare synonymous codon usage patterns of different viral genomes in Morbillivirus. The first dimension and second dimension variables were selected to analyze the difference of synonymous codon usage among different strain PPRV genes. The first dimensional variable that we obtained can reflect 26.03% of synonymous codon usage variation among these genes, and the second dimensional variable can reflect 11.94% of the variation in principal component analysis.

Figure 1 shows bitmap decided by the first and second dimension variables of each gene, it can be seen from the figure that each same gene in different virus strain genomes is basically collected together without significant differences (the first axis: r = -0.205, p>0.05, the second axis: r = 0.065, p>0.1). It is indicated that codon usage patterns of PPRV different strains are similar. However, each gene of PPRV and genes of other virus in Morbillivirus are prominently and dispersedly located in different positions, thereby indicating that their codon usage patterns have differences. Therefore, we can consider that synonymous codon usage patterns in PPRV genome are the same without specificity among the various strains, but the synonymous codon usage patterns among various viruses in Morbillivirus are different with interspecies specificity.

Mutational bias is the main factor that determines the codon usage variation

We used linear regression analysis to respectively

Table 4. The correlation analysis between the A, U, C, G contents and A3, U3, C3, G3 contents in all ORF of PPRV[a].

	A_3%	**U3%**	**G3%**	**C3%**	**GC3%**
A%	0.770***	0.074 ^{NS}	0.659***	0.050 ^{NS}	0.575***
U%	0.489**	0.617***	0.181 ^{NS}	0.290 ^{NS}	0.313**
G%	0.084 ^{NS}	0.677***	0.845***	0.157 ^{NS}	0.850***
C%	0.081 ^{NS}	0.182 ^{NS}	0.005 ^{NS}	0.393**	0.207 ^{NS}
GC%	0.029 ^{NS}	0.619***	0.678 ***	0.280**	0.761***

a Value in this table is the R value of each linear regression analysis.; NS in superscript represent non-significant; *** P-value <0.001; ** P-value <0.01.

Table 5. Linear regression analysis between the first two axes in CA and the nucleotide contents on the third codon position in all ORF of PPRV[a].

Base composition	**f'_1**	**f'_2**
A3%	0.089 ^{NS}	0.465 ***
U3%	0.392**	0.049 ^{NS}
G3%	0.321*	0.040 ^{NS}
C3%	0.292 ^{NS}	0.57 2***
GC3%	0.442***	0.257 ^{NS}

a Value in this table is the R value of each linear regression analysis; b *f'_1*and *f'_2* , respectively, represent the values of the first and the second axis of each gene in CA, NS in superscript represent non-significant, *** P-value <0.001, ** P-value <0.01, * 0.01<P-value<0.05.

compare the pertinence among A3%, U3%, G3%, C3%, GC3% and A%, U%, G%, C% and GC% in order to explore whether the determinants of PPRV codon usage variation is mutation pressure or natural selection. It is discovered that GC3% and A%, U%, G% and GC% were significantly correlated except irrelevance with C% (Table 4). This shows that the GC content of the third position of codon affects the interaction between mutation pressure and natural selection to a certain extent. Meanwhile, GC% and U3%, G3%, C3% and GC3% form significant correlation except the irrelevance with A3% (Table 4), which showed that nucleotide composition restriction affects PPRV codon usage pattern variation.

In addition, we also applied linear regression analysis to discover the correlation between synonymous codon usage and nucleotide composition (Table 5). The results showed that the first axis of each gene in the principal component analysis is closely related to GC 3% (r = 0.442, P <0.001), thereby suggesting that codon usage variation of these genes has pertinence with mutation bias. Meanwhile, it can be seen from the corresponding relationship distribution diagram (Figure 2) between the ENC values and GC3% that most points are well located near the theoretical curve or on the lower part of theoretical curve, thereby indicating that the codon usage of these PPRV genes are closely related with the GC content of the third position of codon. In addition, mutation bias is the major influence factor for deciding the mutation of these gene codons, meanwhile, there may be some factors affecting the codon usage variations of the genes besides mutation bias. Other literatures about RNA virus also had similar reports (Drake and Holland, 1999). It is generally believed that the mutation rate of RNA virus genome is much higher than the DNA viruses (Gareth et al., 1999), so it is not hard to understand why mutation bias becomes the decisive factor of PPRV codon usage bias.

Gene function also drives the codon usage variation

Although gene base composition in the virus genes mainly determines their codon usage bias, gene codon usage is also related with influence factors such as gene translation choice, the function of genes, genetic length and the like. We tested these factors to see whether these factors affect the virus gene codon usage bias. It can be seen from (Figure 3) that genes with the same function types tend to gather together in PPRV genes. It can be found by T test that the first and second dimensional components of PPRV genes with different functions in principal component analysis have significant differences (r = 0.605, p<0.001). These results indicate that the gene function has certain selection function on PPRV gene codon usage bias.

Generally speaking, if translation choice is one of the

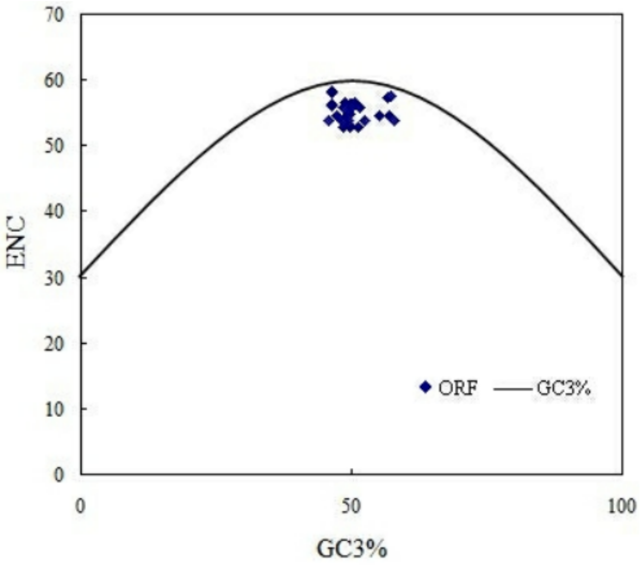

Figure 2. The relationship between the effective number of codons (ENC) and the GC content of the third codon position (GC 3%).

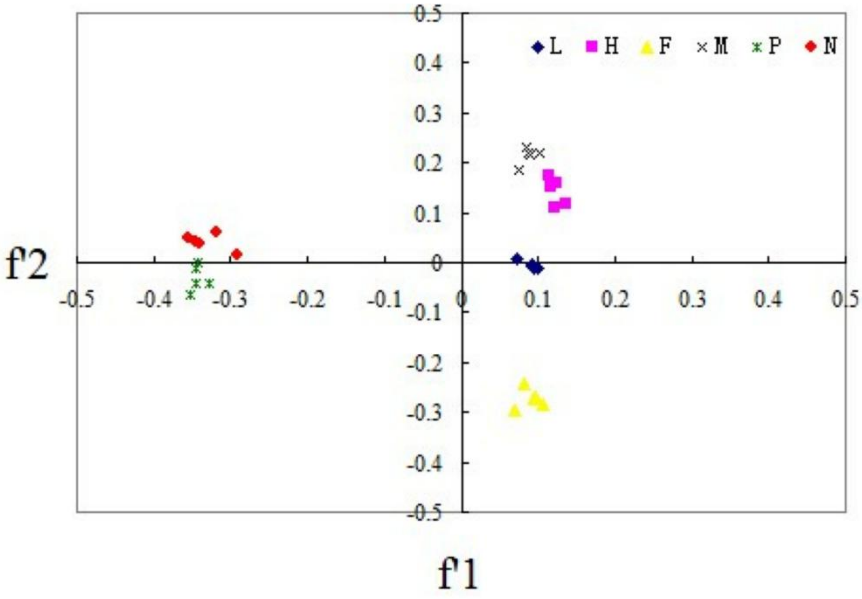

Figure 3. A plot of the values of the first axis and the second axis of each gene of different virus of PPRV in CA (f'_1 and f'_2, respectively, represent the values of the first and the second axis of each gene in CA).

factors affecting the gene codon usage patterns, the genes in genome encoding structural proteins should have more prominent codon usage bias than the genes encoding non-structural protein. Six types of proteins encoded in PPRV genomes are structural proteins. We calculated the average ENC value of the same structure protein genes in different PPRV virus strains and their corresponding standard difference in order to explore whether the translation choice is one of the factors affecting gene codon usage patterns or not. The average ENC value change range of six structure protein genes of PPRV is smaller from 53.40 to 56.44 with standard deviation of 1.01. This shows that the codon bias of different function genes in the translation process of PPRV is not greatly changed due to the influence of translation selection. Therefore, we believe that gene codon usage of these viruses have no relationship with gene translation choice.

At the same time, we used correlation analysis to check the gene codon usage bias and gene length, and found that gene codon usage bias and gene length have no significant correlation (r = 0.175, p>0.05) in the virus genes, it can be seen from the above analysis that gene function also may affect the synonymous codon usage of these virus genes in addition to the influence of gene base composition on codon usage, and other factors such as gene length and gene translation rate basically have no influence on codon usage of these virus genes.

DISCUSSION

Some studies that have been reported displayed that viruses including the influenza A virus subtype H5N1, severe acute respiratory syndrome (SARS) Coronavirus, human bocavirus and the like give priority to codons ended with A or U (Tong et al., 2005; Wanjun et al., 2004; Zhao et al., 2008). But so far, related studies on Peste des Petits Ruminants virus (PPRV) codon usage patterns have not yet been carried out, and this research fills up this gap. The study results indicated that PPRV gives priority to use codons ended with A or G, 12 most preferred codons are ended with A and G, wherein UCA, ACA, CCA and GUG are codons with the highest used frequency in 36 ORFs. Since PPRV genome GC content is low, the GC3% mean is 50.25%, PPRV gives priority to use codons ended with A or G, which is decided by base composition of the third position of codons. Meanwhile, we also found that there are zones with local high GC contents in PPRV genome mainly in the N gene, which suggests that N gene may give priority to use codons ended with G and C compared with other PPRV genes.

Although different PPRV strains have great difference in genome level, and have no prominent difference on codon usage patterns encoding protein without intra-species specific performance. This shows that the codon usage patterns among all PPRV strains are consistent. However, the codon usage among PPRV, CDV, RPV and MV in Morbillivirus of PPRV has difference, and thereby the codon usage of all viruses in Morbillivirus has intra-species differences.

As for RNA viruses, codon usage pattern formation was mainly caused by the mutation pressure rather than natural selection (Shackelton et al., 2006). In this study, GC 3% in each PPRV gene is prominently related with A%, U%, G% and GC%, and the GC% also forms prominent correlation with U 3%, G 3%, C 3% and GC 3% (Table 4). This shows that GC content of the third position of codon affects the interaction between mutation pressure and natural selection to a certain extent, and the nucleotide composition constraints affect the codon use pattern mutation of PPRV. Therefore, mutation pressure in the whole PPRV genome is greater than the influence of natural selection, and is the main determinant factor of codon usage variation. The first axis of each PPRV gene in the principal component analysis is closely related with GC 3% (r = 0.442, P <0.001), thereby suggesting that codon usage variations of these genes have correlation with mutation bias. Correspondence map between ENC values and GC 3 content are also strong supports for this conclusion. The analytical methods which are the same as the study A have also been successfully applied in some reports (Liu et al., 2006; Zhou et al., 2009; Renyong et al., 2009). Therefore, it can be said that mutation bias played a decisive role in the evolution process of PPRV codon usage patterns.

In general, natural selection such as translation selection, gene function, gene length and other influence factors are also related with codon usage variation (Tong et al., 2005). Some published results showed that the same genes with different virus functions tend to gather together in corresponding analysis (Das et al., 2006). In the present study, genes with the same function types in PPRV genes tend to gather together, and t tests showed that the first and second dimensional components of PPRV genes with different functions have prominent differences (P<0.001) in principal component analysis. This shows that the gene function also plays a role for codon usage variation. Therefore, gene function is another influencing factor compared with mutation bias.

The gene length and codon usage patterns have certain correlation in some research reports, and the gene length has no influence on synonymous codon usage variation as in some viruses (Hou and Yang, 2002; Wanjun et al., 2004). We have tested the gene codon usage bias and gene length through the correlation analysis and found that the codon usage bias and gene length have no significant correlation (P>0.1) in the virus genes in the PPRV. The results showed that PPRV gene length has no effect on variation of synonymous codon usage. In this study, we revealed PPRV codon usage patterns, and analyzed all factors affecting PPRV codon usage, thereby providing effective information for future PPRV research.

REFERENCES

Abraham G, Sintayehu A, Libeau G (2005). Antibody seroprevalences against peste des petit rum-inants (PPR) virus in camels, cattle, goats and sheep in Ethiopia. Prev. Vet. Med., 70(1-2):51-57.

Amjad H, Qamar-ul-Islam, Forsyth M, Barrett T, Rossiter PB (1996). Peste des Petits Ruminants in goats in Pakistan. Vet. Rec., 139(5):118-119.

Bailey D, Banyard A, Dash P, Ozkul A, Barrett T (2005). Full genome sequence of Peste des Petits Ruminants virus, a member of the Morbillivirus genus. Virus Res., 110(1-2):119-124.

Cargadennec L, Lawman A (1942). Lapestedes petits ruminants. Bull. Serv. Zoot. Epizoot. AOF, 5(1):16-21.

Chiapello H, Lisacek F, Caboche M (1998). Codon usage and gene function are related in sequences of Arabidopsis tha liana. Gene, 209(1-2):GC1- GC38.

Contzer J, Thomson C, R Tustin (1994). Infectious disease of livestock with special reference to southern Africa. Oxford University Press, South Africa.

Dams MJ, Antoniw JF (2003). Codon usage bias amongst plant viruses.

Arch. Virol., 149(1):113 - 135.

Das S, Paul S, Dutta C (2006). Synonymous codon usage in adenoviruses: influence of mutation, selection and protein hydropathy. Virus Res., 117(2):227–236.

Dittmar KA, Goodenbour JM, Pan J (2006).Tissue-specific differences in human transfer RNA expression. PLoS. Genet., 2, 2107–2115.

Drake JW, Holland JJ (1999). Mutation rates among RNA viruses. Proc. Natl. Acad. Sci. USA, 96:13910-13913

Ewens WJ, Grant GR (2001). Statistical Methods in Bioinformatics. Springer, New York.

Furley CW, Taylor WP, Obi TU (1987). An outbreak of peste des petits ruminants in a zoological collection. Virus Res., 121(19):443-447.

Gareth MJ, Edward CH (2003). The extent of codon usage bias in human RNA viruses and its evolutionary origin. Virus Res., 92:1-7

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(2):89-97

Gupta SK, Ghosh TC (2001).Gene expressivity is the main factor in dictating the codon usage variation among the genes in *Pseudomonas aeruginosa*. Gene, 273(1): 63 - 70.

Hao S, Zhang Q, Liu X, Wang X, Zhang H, Wu Y, Jiang F (2008). Analysis of synonymous codon usage in 11 Human Bocavirus isolates. Biosystems, (92):207-214.

Haroun M, Hajer I, Mukhtar M (2002). Detection of antibodies against peste des petits ruminants virus in sera of cattle, camels, sheep and goat in Sudan. Vet. Res. Commun., 26(7):537-541.

Hou ZC, Yang N (2002). Analysis of factors shaping Spneumoniae codon usage. Yi. Chuan Xue. Bao 29(8): 747–752.

Kwiatek Q, Minet C, Grillet C (2007). este des petits ruminants (PPR) outbreak in Tajikistan. Comp. Pathol., 136(2-3):111-9.

Levin DB, Whittome B (2000). Codon usage in nucleopolyhedroviruses. Gen. Virol., 81(Pt 9): 2313 - 2325.

Liu Qingpo (2006). Analysis of codon usage pattern in the radioresistant bacterium *Deinococcus radiodurans*. BioSystems, 85:99–106.

Lloyd AT, Sharp PM (1992). Evolution of codon usage patterns: the extent and nature of divergence between *Candida albicans* and *Saccharomyces cerevisiae*. Nucleic Acids Res., 20: 5289–5295.

Naya H, Romero H, Carels N, Zavala A, Musto H (2001).Translational selection shapes codon usage in the GC-rich genome of *Chlamydomonas reinhardtii*. FEBS Lett., (501):127-130.

Onofrio GD, Ghosh TC, Bernardi G (2002). The base composition of the genes is correlated with the secondary structures of the encoded Proteins. Gene, 300(1):179-187.

Renyong Jia, Anchun Cheng, Mingshu Wang (2009). Analysis of synonymous codon usage in the UL24 gene of duck enteritis virus. Virus Genes, 38:96–103

Romero H, Zavala A, Musto H (2000). Compositional Pressure and translational selection determine codon usage in the extremely GC poor unicellular eukaryote *Entamoeba histolytica*. Gene, 242(1-2):307-311

Romero H, Zavala A, Musto H, Bernardi G (2003).The influence of translational selection on codon usage in fishes from the family cyprinidae. Gene, 317(1-2):141-147

Sau K, Gupta SK, Sau S, Mandal SC, Ghosh TC (2006). Factors influencing synonymous codon and amino acid usage biases in Mimivirus. Biosystems, (85):107-113.

Shackelton LA, Parrish CR, Holmes EC (2006).Evolutionary basis of codon usage and nucleotide composition bias in vertebrate DNA viruses. J. Mol. Evol., 62 (5), 551–563.

Shaila MS, Shamaki D, Forsyth MA (2007). Geographic distribution and epidemiology of peste des petits ruminants virus. Virus Res.,43(2):149-153.

Sharp PM, Tuohy TM, Mosurski KR (1986). Codon usage in yeast: cluster analysis clearly differentiates highly and lowly expressed genes. Nucleic Acids Res., 14, 5125–5143.

Tong Zhou, Wanjun Gu, Jianmin Ma, Xiao Sun, Zuhong Lu (2005). Analysis of synonymous codon usage in H5N1 virus and other influenza A viruses. BioSystems, 81:77–86

Vander Linden MG, de Farias ST (2006).Correlation between codon usage and thermostability. Extremophiles, 10(5):479-481.

Wanjun Gu, Tong Zhou, Jianmin Ma, Xiao Sun, Zuhong Lu (2004). Analysis of synonymous codon usage in SARS Coronavirus and other viruses in the Nidovirales. Virus Res., 101:155–161

Wright F (1990). The ‘effective number of codons’ used in a gene. Gene 87: 23–29.

Xie T, Ding D, Tao X, Dafu D (1998). The relationship between synonymous codon usage and protein structure. FEBS Lett., 434, 93–96.

Zhao S, Zhang Q, Liu X, Wang X, Zhang H, Yan W, Jiang F (2008). Analysis of synonymous codon usage in 11 Human Bocavirus isolates. BioSystems, 92:207–214

Zhou H, Wang H, Huang LF (2005). Heterogeneity in codon usages of sobem ovirus genes. Archives Virol., 150(8): 1591 - 1605.

Zhou Meng, Li Xia (2009). Analysis of synonymous codon usage patterns in different plant mitochondrial genomes. Mol. Biol. Rep., 36:2039–2046.