

Available online at <http://www.mecs-press.net/ijem>

Analysis of Synonymous Codon Usage in the newly identified DPV UL43 Gene

Qin HE^a, Mingshu WANG^{a,b}, Anchun CHENG^{a,b,c}, Dekang ZHU^{a,b}, Xiaoyue CHEN^{a,b,c}
Renyong JIA^b, Qihui LUO^b, Yi ZHOU^b, Zhengli CHEN^b

^a Avian Disease Research Center, College of Veterinary Medicine of Sichuan Agricultural University, Ya'an, Sichuan, 625014, China

^b Key Laboratory of Animal Diseases and Human Health of Sichuan Province, Sichuan Agricultural University, Ya'an, Sichuan, 625014, China

^c Key Laboratory of Animal Diseases and Human Health of Sichuan Province, Sichuan Agricultural University, Ya'an, Sichuan, 625014, China

Abstract

In this paper, we analyzed the UL43 gene of duck plague virus (DPV) of codon usage bias. The results may provide a basis for understanding the evolution and pathogenesis of DPV and for selecting appropriate host expression systems to improve the expression of target gene in vitro. In this study, the synonymous codon usage bias of UL43 gene in the 24 herpesviruses have been analyzed and the results showed obvious differences by the Codon Adaptation Index (CAI), effective number of codons (ENC) and the value of G + C content at the 3rd codon position (GC_{3s}). In addition, the results revealed that the synonymous codons with A and T at the third codon position have widely usage in the codon of UL43 gene of DPV. G + C compositional constraint is the main factor that determines the codon usage bias in UL43 gene. The phylogenetic analysis suggested that DPV was evolutionarily closer to fowl herpesviruses which further clustered into Alphaherpesvirinae. Furthermore the ORF of UL43 gene has sequential rare codons. There were 25 codons showing distinct usage differences between DPV with *Escherichia coli*, 24 codons showing distinct usage differences between DPV with yeast, and 32 between DPV and *H. sapiens*. Therefore the yeast and *E. coli* expression system may be suitable for the expression of DPV UL43 gene if some codons could be optimized.

Index Terms: Duck Plague Virus; UL43 Gene; Codon Usage Bias

© 2011 Published by MECS Publisher. Selection and/or peer review under responsibility of the Research Association of Modern Education and Computer Science.

* Corresponding author:

E-mail address: chenganchun@vip.163.com

1. Introduction

Codon Usage Bias was defined as deviation from equal usage of synonymous codons [1]. There are 64 codons to code 20 amino acids and the start and stop signals. Most amino acids were coded by more than one codon (synonymous codons) [2], but the synonymous codon usage are not used equally both within and between genomes [3-4]. Previous research have been showed that codon usage bias may be very complicated and associated with various biological factors, such as gene expression level [5], gene length [6], gene translation initiation signal [7], protein amino acid composition [8], protein structure [9], tRNA abundance [10,11], mutation frequency and patterns [8,12], GC composition [13,14], and environmental factors [15]. For a number of different organisms, it was suggested that codon usage is best explained by selection for tRNA abundance, gene expression levels, and translational optimization [16]. Recently, it was also suggested that codon usage is related to gene function [17] and the the evolutionary history of an organism in metazoan genomes [18]. Codon usage bias is widely studied in particular organisms to achieve high expression of heterologous proteins in vitro and to improve the design of oligonucleotide probes and primers in order to provide a general understanding of the molecular evolution of species. Codon usage information in different organisms has also been analyzed. Several evolutionary processes have been provided an opportunity to inspect the relationship between codon usage and the evolutionary age of genes [18], and all synonymous codons might be integrated parts of the Genetic Code with equal importance in maintaining its functional integrity [1]. Studies of the synonymous codon usage in viruses can reveal information about the molecular evolution of individual genes and such information would be relevant to understanding the regulation of viral gene expression and also to vaccine design where the efficient expression of viral proteins may be required to generate immunity [19,20]. Recently, analyses of the patterns of codon usage bias of herpesviruses are primarily focused on the pseudorabies virus (PRV) [21], herpes simplex virus type 1 (HSV-1) [22], Epstein-Barr virus [23]. However, except for UL24, UL26.5, UL35, gE, dUTPase, the codon usage bias in DPV genome was known little [2, 24-27].

Duck plague, known as duck viral enteritis (DVE), is an acute, lethal and contagious disease caused by Herpesviridae, duck plague virus (DPV) that occurs worldwide among domestic and wild ducks, geese, swans, and other water fowl, with migratory waterfowl contributing to spread between continents [28,29]. Morbidity and mortality vary from 5 to 100%, most ducks that develop clinical signs die. Strains of virus vary in virulence, but only a single antigenic type has been recognized. Now most of the previous research work has focused on the epidemiology and prevention of this disease. However, the molecular biology information about the DPV genome is limited. Recently, the UL43 gene was isolated and identified from DPV CHv strain in our laboratory [30]. The UL43 gene encodes nonglycosylated membrane-associated protein and is conserved only within the Alpha- and Gammaherpesvirinae subfamilies [31]. Little is known about the molecular informations and function of DPV UL43 protein at present. In this study, we first analyzed the synonymous codon usage in the UL43 gene of DPV and compared with those of 23 other species of herpesviruses. Moreover, the codon usage bias in the DPV gene was compared with those of *Escherichia coli*, yeast and *H. sapiens*. In addition, we also investigated the rare codons of UL43 gene. All these datas might provide some insights into the features of the DPV genome, the possible function of DPV UL43 gene as well as the suitable expression system in in vitro.

2. Materials And Methods

2.1. Virus Species and Gene Sequences

The DPV CHv strain, a high-virulence strain of DPV, was obtained from Key Laboratory of Animal Disease and Human Health of Sichuan Province. The UL43 gene of the DPV CHv strain (GenBank accession no.EU071037) was isolated and identified by our laboratory. The nucleotide sequences of the UL43 genes of 23 reference herpesviruses were obtained from the NCBI GenBank nucleotide database (Table 2).

2.2. Analysis on Codon Usage in UL43 Gene of DPV and 23 Reference Herpesviruses

The 'Effective Number of Codons' (ENC) was often used to quantify the codon usage bias of an ORF in an individual gene. The values range from 20 to 61. In an extremely biased gene where only one codon is used for each amino acid, this value would be 20; in an unbiased gene, it would be 61. The ENC value of the UL43 gene in each reference herpesvirus was computed with The European Molecular Biology Open Software Suite (EMBOSS) CHIPS online service program. The peculiarity in codon usage frequency and the G + C content of the gene sequences were also calculated with the CUSP program of EMBOSS.

2.3. Molecular Characterization and Phylogenetic Analysis of the DPV UL43 Gene

SHOWORF, an EMBOSS nucleotide translation program, was used to read and transfer the nucleotide sequence data to a computer file. Phylogenetic analysis was performed for the UL43 genes of 24 herpesviruses with CLUSTAL-X and TREEVIEW software. Relative synonymous codon usage values (RSCU) of each codon in a gene was used to examine the synonymous codon usage without the confounding influence of amino acid composition [32]. RSCU value >1.0 indicated that the corresponding codon was more frequently used than expected, whereas the reverse was true for RSCU value <1.0 [33]. The RSCU values of UL43 genes were analyzed with the CodonW.

2.4. Analysis the Rare Codons of DPV UL43 Gene

The proteins in heterologous hosts are often difficult to express or at very low levels. They might contain codons that are rarely used in the desired host. Log to <http://nihserver.mbi.ucla.edu/RACC/> to analyze the rare codons of the DPV UL43 gene.

2.5. Comparison of Codon Preferences of DPV UL43 Gene with Those of *E. coli*, Yeast and *H. sapiens*

To examine whether different species follow with the same codon usage rule, Codon usage bias in the DPV UL43 gene was determined with the SPSS 13.0 software, and we compare the UL43 codon usage bias among DPV, *E. coli*, yeast and *H. sapiens* (create a codon usage table). The database of the codon usage in *E. coli*, yeast and *H. sapiens* is available at <http://www.kazusa.or.jp/codon>.

3. Results

3.1. Variation in DPV UL43 Codon Usage and Amino Acid Composition

While the relative synonymous codon usage values (RSCU) and the related measures indicate the overall DPV UL43 codon bias, it is also important to closely investigate the pattern of codon bias. Table 1 shows the codon preferences of DPV UL43 gene. Sixty-one codons (excepting Met and the termination codons) in the polypeptide, with eleven synonymous codons strong bias toward A-ended and eight toward T-ended at the third codon position, were used. A high level of diversity in codon usage bias existed for coding the Ala, Gly, Leu, Pro, Arg, Ser, Thr and Val amino acids because they have a 6-fold and 4-fold coding degeneracy.

3.2. Codon Usage Analysis of the UL43 genes of DPV and Reference Herpesviruses

The results obtained by CodonW and EMBOSS analysis of the ENC, CAI, coding G + C content (GC%)

and the G + C contents at the third codon position content (GC_{3S}%) of 24 herpesviruses species are shown in Table 2 . Codon usages in the UL43 genes are highly nonrandom in all the herpesviruses, and the overall base compositions of the UL43 genes in these species also differ dramatically. From the Table 2, the ENC values of different UL43 genes vary from 29.375 to 61, with a mean value of 47.366 and standard deviation (S.D.) of 9.367. The GC_{3S} contents of UL43 genes range from 33.18 to 95.99% with a mean of 60.214% and S.D. of 20.973. The CAI values of different UL43 genes vary from 0.588 to 0.78, with a mean value of 0.689 and standard deviation (S.D.) of 0.055.

Table 1. Synonymous Codon Usage Of DEV UL43 Gene Analyzed With Cusp Program

Rank	AA	Codon ^a	Fraction ^b	Frequency ^c	Number ^d	RSCU ^e	
1	A(Ala)	GCA ^f	0.388	38	19	1.55	
2		GCC	0.204	20	10	0.82	
3		GCG	0.265	26	13	1.06	
4		GCT	0.143	14	7	0.57	
5	C(Cys)	TGC	0.200	6	3	0.40	
6		TGT ^f	0.800	24	12	1.60	
7	D(Asp)	GAC	0.500	10	5	1.00	
8		GAT	0.500	10	5	1.00	
9	E(Glu)	GAA	0.400	12	6	0.80	
10		GAG	0.600	18	9	1.20	
11	F(Phe)	TTC	0.238	10	5	0.48	
12		TTT ^f	0.762	32	16	1.52	
13	G(Gly)	GGA ^f	0.333	22	11	1.33	
14		GGC	0.182	12	6	0.73	
15		GGG	0.242	16	8	0.97	
16		GGT	0.242	16	8	0.97	
17	H(His)	CAC	0.250	6	3	0.50	
18		CAT ^f	0.750	18	9	1.50	
19	I(Ile)	ATA ^f	0.388	38	19	1.16	
20		ATC	0.184	18	9	0.55	
21	K(Lys)	ATT ^f	0.429	42	21	1.29	
22		AAA	0.462	12	6	0.92	
23		AAG	0.538	14	7	1.08	
24		CTA ^f	0.222	24	12	1.33	
25	L(Leu)	CTC	0.037	4	2	0.22	
26		CTG	0.074	8	4	0.44	
27		CTT	0.148	16	8	0.89	
28		TTA ^f	0.241	26	13	1.44	
29	M(Met)	TTG	0.278	30	15	1.67	
30		ATG	1.000	40	20	1.00	
31		N(Asn)	AAC	0.231	6	3	0.46
32			AAT ^f	0.796	20	10	1.54
33	P(Pro)	CCA ^f	0.407	22	11	1.63	
34		CCC	0.037	2	1	0.15	
35		CCG	0.333	18	9	1.33	
36	Q(Gln)	CCT	0.222	12	6	0.89	
37		CAA	0.500	10	5	1.00	
38		CAG	0.500	10	5	1.00	
39		R(Arg)	AGA ^f	0.308	16	8	1.85
40	AGG		0.038	2	1	0.23	
41	CGA ^f		0.192	10	5	1.15	

42		CGC	0.077	4	2	0.46
43		CGG	0.115	6	3	0.69
44		CGT ^f	0.269	14	7	1.62
45	S(Ser)	AGC	0.111	8	4	0.67
46		AGT	0.139	10	5	0.83
47		TCA ^f	0.250	18	9	1.50
48		TCC	0.111	8	4	0.67
49		TCG	0.111	8	4	0.67
50		TCT ^f	0.278	10	10	1.67
51	T(Thr)	ACA ^f	0.409	36	18	1.64
52		ACC	0.182	16	8	0.73
53		ACG	0.273	24	12	1.09
54		ACT	0.136	12	6	0.55
56	V(Val)	GTA ^f	0.438	28	14	1.75
57		GTC	0.125	8	4	0.50
58		GTG	0.188	12	6	0.75
59		GTT	0.250	16	8	1.00
60	W(Trp)	TGG	1.000	2	1	1.00
61	Y(Tyr)	TAC	0.550	22	11	1.10
62		TAT	0.450	18	9	0.90
63	*	TAA	1.000	2	1	3.00
64	*	TAG	0.000	0.000	0	0.00
65	*	TGA	0.000	0.000	0	0.00

a The preferentially used codons for each amino acid are displayed in bold

b The ‘‘Fract’’ shows the proportion of all synonymous codons encoding the same amino acid

c The ‘‘Frequency’’ lists the number of codons present per 1000 bases in the input sequence(s)

d The ‘‘Number’’ lists the number of codons

e The ‘‘RSCU’’ shows the proportion of relative synonymous codon usage

f Shows a strong bias towards the codons with A and T at the third codon position

In general speaking, if the ENC value of a gene is 35 or less, that gene is thought to possess strong codon bias [34]. Analyzing the ENC values of all the UL43 genes, the results showed the majority of them do not have a strong codon bias. The plot of ENC and GC_{3S} content is another effective way to explore codon usage variation among different genes [34]. In Fig.1, the solid line represents the curve if codon usage is only determined by GC_{3S} content. If GC_{3S} is the only determinant factor shaping the codon usage pattern, the values of ENC would fall on a continuous curve, which represents random codon usage [35]. For UL43 genes, only a few ENC values were lined on the curve and others were plotted against both GC_{3S} content and the expected ENC value. Result suggested that there are other factors that contributed to the codon usage pattern in the UL43 genes besides the genomic composition.

Table 2. Summary Analysis Of UL43 Gene In Different Herpesvirus Species

	Virus name	GenBank	L(bp) ^a	ENC ^b	GC(%)	CAI ^c	GC _{3s} (%) ^d
Alphaherpesvirinae	Duck plague virus (DPV)	EU071037	1503	55.903	43.85	0.588	39.32
	Meleagrid herpesvirus 1 (MeHV-1)	NC 002641	1269	60.597	45.00	0.603	39.72
	Bovine herpesvirus 1 (BoHV-1)	NC 001847	1137	36.571	78.63	0.779	79.42
	Bovine herpesvirus 5 (BoHV-5)	NC 005261	1143	33.894	81.98	0.780	82.41

	Equid herpesvirus 1 (EHV-1)	NC 00149	1206	53.195	51.99	0.712	53.73
	Equid herpesvirus 4 (EHV-4)	NC 001844	1212	51.633	46.62	0.731	49.75
	Gallid herpesvirus 2 (GaHV-2)	NC 002229	1263	54.273	39.59	0.597	34.20
	Gallid herpesvirus 3 (GaHV-3)	NC 002577	1245	61.000	51.57	0.607	54.94
	Human herpesvirus 1 (HHV-1)	NC 001806	1254	43.870	72.41	0.694	83.01
	Human herpesvirus 2 (HHV-2)	NC 001798	1245	37.501	74.06	0.707	86.99
	Suid herpesvirus 1 (SuHV-1)	BK 001744	1122	29.375	74.42	0.746	95.99
	pseudorabies virus (PRV)	AF 158096	1356	56.604	73.45	0.684	57.74
	Psittacid herpesvirus 1 (PsHV-1)	NC 005264	1194	52.828	58.63	0.618	71.11
	Ceropithecine herpesvirus 1 (CerHV-1)	NC 005264	1140	33.754	77.98	0.695	93.16
	Ceropithecine herpesvirus 2 (CerHV-2)	NC 006560	1146	32.484	79.93	0.694	95.55
	Ceropithecine herpesvirus 9 (CerHV-9)	NC 002686	1266	47.159	32.15	0.676	33.18
Betaherpesvirinae	Human herpesvirus 7 (HHV-7)	NC 001716	963	44.125	33.75	0.705	40.50
	Human herpesvirus 6 (HHV-6)	NC 001664	951	59.265	43.53	0.673	47.95
	Human herpesvirus 5 (HHV-5)	NC 001347	1272	53.859	62.89	0.650	61.79
Gammaherpesvirinae	Human herpesvirus 8 (HHV-8)	NC 009333	1074	52.390	53.17	0.726	52.51
	Human herpesvirus 4 (HHV-4)	NC 009334	1074	48.537	54.84	0.751	69.83
	Ovine herpesvirus 2 (OvHV-2)	NC 007646	1056	46.334	47.63	0.694	43.18
	Murid herpesvirus 4 (MuHV-4)	NC 001826	1044	47.323	40.13	0.705	40.52
	Alcelaphine herpesvirus 1 (AIHV-1)	NC 00253	1056	44.300	41.48	0.715	38.64

a Represents the length of identified ORF

c Codon Adaptation Index

b Effective number of codons

d G + C frequency at the synonymous third position of codons

3.3. Characterization of the DPV UL43 Gene

Using CLUSTAL-X and TREEVIEW software, A phylogenetic tree was established from the deduced amino acids encoded by the 1503 bp ORF of the UL43 gene of DPV and the 23 reference herpesviruses (Fig. 2). It shows that there are mainly three branches for the 24 herpesviruses. DPV has been clustered in one branch Alphaherpesvirinae. The DPV, MaHV-2, GaHV-3, and MeHV-1 are clustered in a distinct subbranch. The identified high amino acid sequence similarity suggests that the UL43 protein of DPV is evolutionary closer related to MaHV-2, GaHV-3, and MeHV-1, and that the amino acid sequences of UL43 gene are higher similarity.

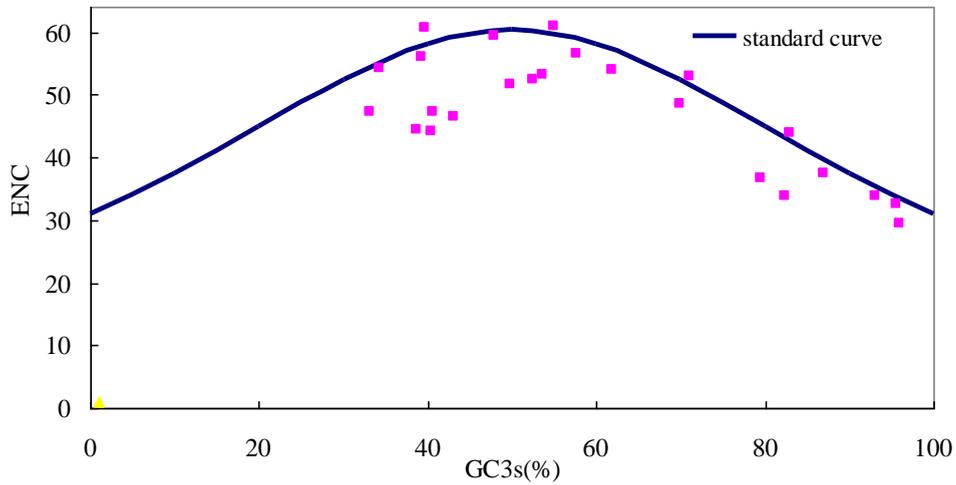


Fig. 1. The plot of ENC and guanine (G) + cytosine (C) frequency at the synonymous third position of codons (GC3S) of the UL43 gene in the DEV CHv strain and those of 23 reference herpesviruses

The curve indicates the expected codon usage if GC compositional constraints alone account for codon usage bias.

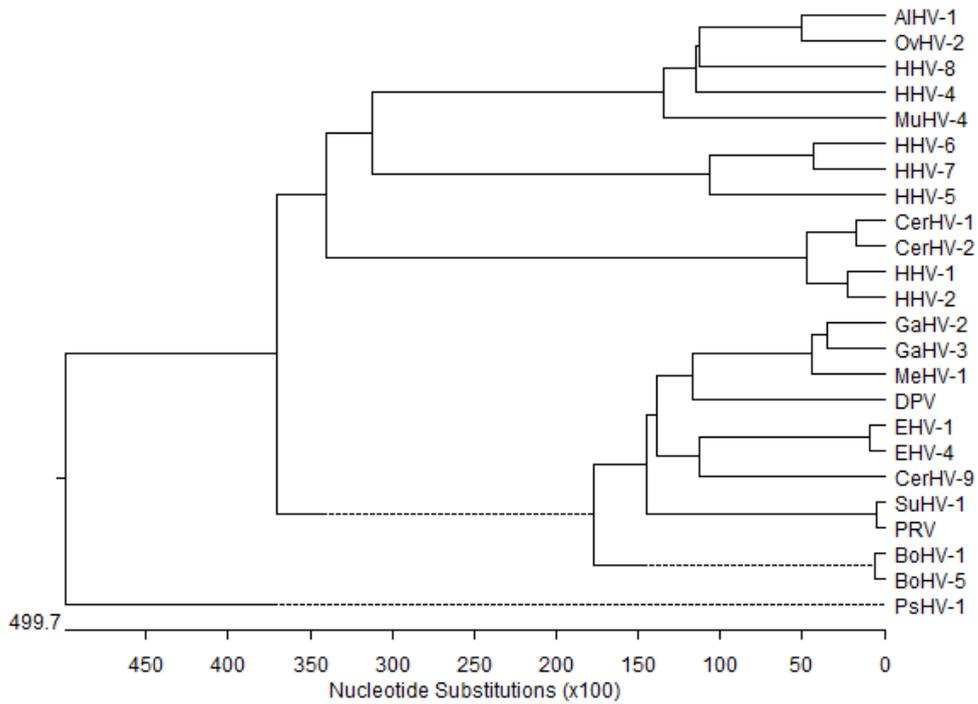


Fig. 2. Phylogenetic tree based on the UL43 amino acid sequences in 24 herpesviruses (Table 2), and constructed with CLUSTAL-X and MEGA software

Rare condons analysis showed that there are 46 rare condons (9.318%) in the ORF of the DPV UL43 gene by using codon usage database on line (<http://nihserver.mbi.ucla.edu/RACC/>) (shown in Fig.3). The result revealed that there have sequential more than two rare codons in the in the ORF of the DPV UL43 gene.

atg agc gag cag gcg aac gcg acg gag cgt cct gtc gtg gcc gaa gtg cag cct acg gtt acg gac acg acc ggc gag aca ccg
CGA acg cca ccg cct cct tac gaa gaa gcg agc caa cat ggc aat tct gca tgc acc **CTA** cct cca tac agc atg aat tta gag
AGA aat ggg cat act gct **ATA** ccg caa gca ccg atc gcc gca aat aca ggg atg gcg ccg **CTA** ttg acc gat ggg tcc
AGA cag ttg atg aca tct agt tct **ATA** tgc **AGA** tgt **AGA** gca tgc **CGA** gat ggt tgt ggg cca tgt aca atc gca tct act
ATA tct gta ttc atg atc ggg ttt cat gca gcc atg att gct gca tgt aca aca aca att ctt atg cca acg aat cgc att tcg ttg gcc
gga gcc aca att gcg **CTA** ttg gct **ATA** gca **ATA** ctt aat att tta **CGA** tac tca tca aaa ttt atg aaa atg atg tgc **CTA** acg
ttt aag ttg ttg caa att ttg gca tgt att tcc gcc ttg gtt **ATA** gga ctg aca **AGA** acg gag gtc aag aca gaa tta ttg ccg aat
aag **ATA** ccg atc gac tcg aac atg cat gcg ttt aac **ATA** gca tat gta acg gcg **CTA** gta tta tct gtc gta ttc ggt acg cca
gtt ttc gca tac tac att tca tgt gcc gcg acc gga gca ccg ccg cat atg **ATA** gct aca ttt att agt gcc tct ctt gcc att tct ttg
gga att gta acc cca tta att **AGA** ggt aat gtg tgg **ATA** gcg att ggg ttt gga gcc gct **ATA** atg atc ttg gcc tgt ttg aag
gac tat ggc gca aaa atg cgt gac aca tgt cat tac aaa tta gcg cgt ttt gct acg atg **AGG** aca tat gcg gat atg ggt ttc gga
gta gca ttt cag **CCC** gct tca att cca cca aat ggc gat gga **CTA** cct **CGA** atg cac att gga aca cac gaa gag gac gtg tct
att ttt gat gtc ctt aaa ccg ccg aaa **AGA** cat tca tgt tat aca **CTA** ttt tca atc **CTA** aca att ccg ttt tta tac gga gta ctt
acc ttc cca tat ggt ggt acg **ATA** cca atc att aag tta act gag act act gca tta gca gtt ctg ttg ggg cat ctc gta aat gtg ttt
ATA tta cca cat aag aca tgt tcc atg gcc att tat gta gag cgt gta ctt **ATA** **ATA** aca tat **ATA** **CTA** **CTA** cag gtt atc
tct acc **ATA** tta gtg act **AGA** ggt tat gag gaa **CTA** **ATA** tat agt tac gta ttt tcc gtt agt tca caa gta gcg ttg tgt **ATA**
tta ttg ctg cac cgt **CGA** tgc gtt gga ctc aag ggg ctg gca ttt tca gta gta gca cgt agc atg ttt gca tta ctt ttt tgt tca atc
gcg **CTA** ggt ctt gga att acc tac gtt gcg cgt att tac caa atg agt tac taa

Red = rare Arg codons **AGG, AGA, CGA** ; Green = rare Leu codon **CTA** ;
Blue = rare Ile codon **ATA** ; Orange = rare Pro codon **CCC**

Fig. 3. Rare condons analysis of the DPV UL43 gene

3.4. Comparison of Codon Usage between DPV and *E. coli*, Yeast and *H. sapiens*

The DPV UL43 gene was compared with those of *E. coli*, yeast and *H. sapiens* to see which will be the suitable host for the optimal expression of DPV genes. From Table 3, there are 25 codons distinct usage differences between DPV UL43 gene and *E. coli* (a DPV-to-*E. coli* ratio higher than 2 or lower than 0.50), 24 between DPV UL43 gene and yeast (a DPV -to-yeast ratio higher than 2 or lower than 0.50), 32 between DPV UL43 gene and *H. sapiens* (a DPV -to- *H. sapiens* ratio higher than 2 or lower than 0.50). Codons usage analysis datas (Fig.4) shows variation between DPV UL43 gene, *E. coli*, yeast and *H. sapiens*. All these might suggest that expressing DPV genes more efficiently in *E. coli* or yeast systems.

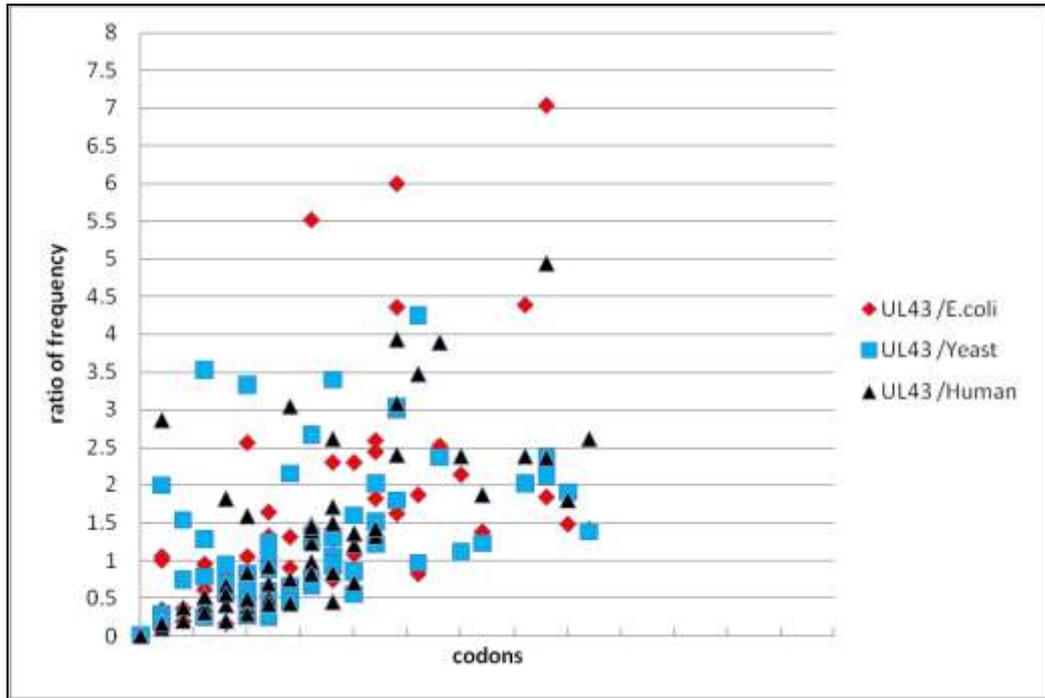


Fig. 4. The comparisons in the ratio of codon usage frequency (1/1000) of DPV to E. coli, yeast and H.sapiens

The ratio higher than 2 or lower 0.5 indicates that the codon usage preference differs, and vice versa.

Table 3. Comparison Of Codon Preferences Between The DPV UL43 Gene And E. Coli, Yeast And H. Sapiens

Condon	AA	E.coli (1/1000)	Yeast (1/1000)	H.sapiens (1/1000)	UL43 (1/1000)	UL43/E.coli /E.coli	UL43/Yeast /Yeast	UL43/H.sapiens /Human
GCA	A(Ala)	20.6	16.1	16.1	38	1.84	2.36	2.36
GCC	A	25.5	12.5	28.4	20	0.78	1.60	0.70
GCG	A	31.7	6.1	7.5	26	0.82	4.26	3.47
GCT	A	15.6	21.1	18.6	14	0.90	0.66	0.75
TGC	C(Cys)	6.9	4.7	12.2	6	0.87	1.28	0.49
TGT	C	5.5	8.0	10.0	24	4.36	3.00	2.4
GAC	D(Asp)	18.6	20.2	25.6	10	0.54	0.50	0.39
GAT	D	32.1	37.8	21.9	10	0.31	0.26	0.46
GAA	E(Glu)	38.2	48.5	29.0	12	0.31	0.25	0.41
GAG	E	17.7	19.1	39.9	18	1.02	0.94	0.45
TTC	F(Phe)	16.9	18.2	20.6	10	0.59	0.55	0.49
TTT	F	23.2	26.1	17.1	32	1.38	1.23	1.87
GGA	G(Gly)	9.0	10.9	16.4	22	2.44	2.02	1.34
GGC	G	27.9	9.7	22.5	12	0.43	1.24	0.53
GGG	G	11.3	6.0	16.3	16	1.42	2.67	0.98
GGT	G	24.4	24.0	10.8	16	0.66	0.67	1.48
CAC	H(His)	9.8	7.7	15.0	6	0.61	0.78	0.40
CAT	H	13.6	13.7	10.5	18	1.32	1.31	1.71
ATA	I(Ile)	5.4	17.8	7.7	38	7.04	2.13	4.94
ATC	I	24.2	17.0	21.6	18	0.74	1.06	0.83
ATT	I	29.8	30.4	16.1	42	1.41	1.38	2.61

AAA	K(Lys)	33.2	42.2	24.1	12	0.36	0.28	0.50
AAG	K	10.7	30.7	32.2	14	1.31	0.46	0.43
CTA	L(Leu)	4.0	13.3	7.8	24	6.00	1.80	3.08
CTC	L	11.0	5.4	19.8	4	0.36	0.74	0.20
CTG	L	50.9	10.4	39.8	8	0.16	0.77	0.20
CTT	L	11.7	12.1	13.0	16	1.37	1.32	1.23
TTA	L	13.9	26.7	7.5	26	1.87	0.97	3.47
TTG	L	14.0	27.0	12.6	30	2.14	1.11	2.38
ATG	M(Met)	27.0	20.9	22.2	40	1.48	1.91	1.80
AAC	N(Asn)	21.4	24.9	19.5	6	0.28	0.24	0.31
AAT	N	18.6	36.3	16.7	20	1.08	0.55	1.20
CCA	P(Pro)	8.5	18.2	16.7	22	2.59	1.21	1.32
CCC	P	5.8	6.8	20.1	2	0.34	0.29	0.10
CCG	P	21.8	5.3	6.9	18	0.83	3.40	2.61
CCT	P	7.3	13.6	17.3	12	1.64	0.88	0.69
CAA	Q(Gln)	15.0	27.5	12.0	10	0.67	0.36	0.83
CAG	Q	29.5	12.1	34.1	10	0.34	0.83	0.29
AGA	R(Arg)	2.9	21.3	11.5	16	5.52	0.75	1.39
AGG	R	1.9	9.2	11.4	2	1.05	0.22	0.18
CGA	R	3.9	3.0	6.3	10	2.56	3.33	1.59
CGC	R	21.0	2.6	10.7	4	0.19	1.54	0.37
CGG	R	6.3	1.7	11.6	6	0.95	3.53	0.52
CGT	R	20.3	6.5	4.6	14	0.69	2.15	3.04
AGC	S(Ser)	16.0	9.7	19.3	8	0.50	0.82	0.41
AGT	S	9.5	14.2	11.9	10	1.05	0.70	0.84
TCA	S	7.8	18.8	12.0	18	2.30	0.96	1.50
TCC	S	8.9	14.2	11.9	8	0.90	0.56	0.67
TCG	S	8.7	8.5	4.4	8	0.92	0.94	1.82
TCT	S	8.7	23.5	14.7	20	2.30	0.85	1.36
ACA	T(Thr)	8.2	17.8	15.1	36	4.39	2.02	2.38
ACC	T	22.8	12.6	19.4	16	0.70	1.27	0.82
ACG	T	14.8	7.9	6.1	24	1.62	3.04	3.93
ACT	T	9.1	20.3	13.0	12	1.32	0.59	0.92
GTA	V(Val)	11.1	11.8	7.2	28	2.52	2.37	3.89
GTC	V	15.1	11.6	14.6	8	0.53	0.69	0.55
GTG	V	25.5	10.6	28.4	12	0.47	1.13	0.42
GTT	V	18.5	22.0	11.0	16	0.86	0.72	1.45
TGG	W(Trp)	15.2	10.3	12.7	2	0.13	0.19	0.16
TAC	Y(Tyr)	12.1	14.6	15.5	22	1.82	1.51	1.42
TAT	Y	16.5	18.9	12.1	18	1.09	0.95	1.49
TAA	*	2.0	1.0	0.7	2	1.00	2.00	2.86
TAG	*	0.3	0.5	0.6	0.00	0.00	0.00	0.00
TGA	*	1.1	0.7	1.5	0.00	0.00	0.00	0.00

4. Discussions

The degeneracy of the genetic code implies that multiple triplets codes for the same amino acid . The frequencies with which different codons are used vary significantly between organisms and between proteins within the same organism [36]. Nonetheless, several evolutionary processes have been postulated as the major factors that determine codon usage: selection, mutation, and genetic drift . However, the relative contribution of each of these factors in different species remains debatable [37-40]. In this paper, we used the EMBOSS CUSP program and CHIPS program to deduce the DPV UL43 gene's RSCU, ENC values, GC and GC3s content and CAI, comparing with 23 reference herpesvirus UL43 gene nucleotide sequences. RSCU was used to study the overall synonymous codon usage variation among the genes without the confounding influence of the amino acids composition of different genes, it is defined as the ratio of the observed frequency of codons to the expected frequency if all the synonymous codons for those amino acids are used equally. RSCU values greater

than 1.0 indicate that the corresponding codon is more frequently used than expected [33]. From the Table 1, eleven synonymous codons strong bias toward A-ended and eight toward T-ended at the third codon position were used. A high level of diversity in codon usage bias existed for coding the Ala, Gly, Leu, Pro, Arg, Ser, Thr and Val amino acids. Codons of hydrophobic amino acids, highly represented in integral membrane proteins, are composed of 50% uracils (U) [41]. So the content of A or T ending codons is an important factor. The ENC value and GC_{3S} content, two important codon usage indices, have been widely used to explore the codon usage variation among different genes [42-44]. Values of ENC range from 20 (when only one codon is used per amino acid) to 61 (when all synonyms are used with equal frequency). The ENC values of herpesvirus UL43 genes are dramatically different (from 29 to 61 shown in Table 2) but mostly the codon usage bias is lower stronger. The ENC value of DPV UL43 gene is high (ENC>50), so that the codon usage bias is a little stronger. If G+C compositional constraint influences the codon usage, then the GC_{3S} and ENC correlated spots would lie on [35] or just below the expected curve. If a gene is subject to selection for translationally optimal codons, it will lie considerably below the expected curve [45]. In Fig.1, a large number of points do not follow the theoretical curve suggesting that other factors other than gene composition contribute to the codon usage pattern in the reference herpesviruses, which maybe mutational bias and natural selection, such as translational selection, the tRNA abundance, leading to the codon usage variation among genes in different organisms.

CAI (Codon Adaptation Index), is a species-dependent codon bias measure, and used as a measure for gene expressivity in studies investigating mutational and selectional [46]. The CAI value is much closer to 1, the codon usage is much stronger and the gene expressing level is much higher. Comparative analysis of UL43 genes in DPV and the reference herpesviruses indicated that synonymous codon usage in these genes is phylogenetically conserved. Datas in Table 2 show that the UL43 genes in DPV, MeHV-1, GaHV-2 and GaHV-3, whose natural host is avian, have a stronger correlation than the UL43 genes of herpesviruses with other hosts. In the Table 3, the CAI value of DPV UL43 gene is 0.588, which is a little slight lower. We can infer UL43 gene is lowly expressed gene in DPV genome. Simultaneously, the phylogenetic tree analysis based on the UL43 gene products revealed that UL43 protein of the DPV CHv strain and some avian herpesviruses such as MeHV-1, GaHV-2 and GaHV-3, were clustered within a monophyletic clade and grouped within alphaherpesvirinae. We speculate that the codon usage bias of DPV UL43 gene has a very close relation with its gene function and gene type. DPV UL43 gene is similar to the UL43 gene function of the herpesviruses. Studies on the herpes simplex virus (HSV-1) and pseudorabies virus (PRV) UL43 proteins have well documented that the envelope UL43 protein did not impair viral replication either in cell culture or in a mouse infection model, including no effect on neuroinvasion and transneuronal spread [47,48]. If it is deleted, plaque size is reduced by approximately 10% and viral titers are decreased by ca. 2- to 5-fold [49]. Furthermore, The DPV UL43 protein has several membrane domains, potential ability to cross a membrane several times suggesting a role in signal transduction or as ion channels, just like VZV gene 15, which corresponds to HSV-1 gene UL43, contains a periodic charge pattern similar to that found in voltage-gated ion channels [50].

The most plausible selection-based explanation for codon usage bias is the selection for efficient translation related to the relative abundance of isoaccepting tRNAs [42,43]. In order to show the codon usage variation among genes from different organisms, the codon usage bias of DPV was compared with that of *E. coli*, yeast, and *H. sapiens*. There are 25 codons distinct usage differences between DPV UL43 gene and *E.coli*, 24 between DPV UL43 gene and yeast, 32 between DPV UL43 gene and *H. sapiens*. Thus, we can assume that the *E.coli* and yeast expression system is suitable for heterologous expression of the DPV gene. In addition, we analyzed the rare condons of the DPV UL43 gene. There were 46 rare codons and 3 consecutive rare codons in UL43 gene ORF, which may influence the expression of the UL43 gene in vitro. So if we choose the prokaryotic expression system, we should choose the host bacteria *Rosseta*, which should improve the expression of the exogenous genes. All of these may be of great importance for gene characterization, for gene classification and for assessing the possible role of UL43 protein in viral pathogenesis.

Acknowledgements

The research was supported by grants from Changjiang Scholars and Innovative Research Team in University (PCSIRT0848) and the earmarked fund for Modern Agro-industry Technology Research System (nycytx-45-12). The first two authors contribute equally to this work and should be considered as first author. An-chun Cheng and Ming-shu Wang are the corresponding authors. Key Laboratory of Animal Disease and Human Health of Sichuan Province & Avian Disease Research Center, College of Veterinary Medicine of Sichuan Agricultural University, 46# Xinkang Road, Yucheng district, Yaan 625014, Sichuan province of China. Tel.: +86 835 2885774; fax: +86 835 2885774. E-mail address: chenganchun@vip.163.com (A. Cheng); mshwang@163.com (M. Wang).

References

- [1] B. Jan, "Does codon bias have an evolutionary origin?," *Theor Biol Med Model*, BioMed Central, England, vol. 5, pp. 1-16, July 2008.
- [2] R. Y. Jia, A. C. Cheng, M. S. Wang, H. Y. Xin, Y. F. Guo, D. K. Zhu, et al, "Analysis of Synonymous Codon Usage in the UL24 Gene of Duck Enteritis Virus," *Virus Genes*, Kluwer Academic, United States, vol.38, pp. 96-103, October 2009.
- [3] R. Grantham, C. Gautier, M. Gouy, R. Mercier and A. Pavé, "Codon catalog usage and the genome hypothesis," *Nucleic Acids Res*, Oxford University Press, England, vol. 8, pp. 49-62, January 1980.
- [4] A. Martin, J. Bertranpetit, J. L. Oliver and J. R. Medina, "Variation in G+C content and codon choice: differences among synonymous codon groups in vertebrate genes," *Nucleic Acids Res*, Oxford University Press, England, vol.15, pp.6181-6189, June 1989.
- [5] M. Nakamura, and M. Sugiura, "Translation efficiencies of synonymous codons are not always correlated with codon usage in tobacco chloroplasts," *Plant J*, Blackwell Scientific Publishers and BIOS Scientific Publishers in association with the Society for Experimental Biology, England, vol. 49, pp. 128-134, January 2007.
- [6] A. Eyre-Walker, "Synonymous codon bias is related to gene length in *Escherichia coli*: selection for translational accuracy?" *Mol. Bio. Evol.*, Oxford University Press, United States, vol. 13, pp. 864-872, July 1996.
- [7] J. Ma, A. Campbell, and S. Karlin, "Correlations between shine-dalgarno sequences and gene features such as predicted expression levels and operon structures," *J. Bacteriol*, American Society for Microbiology, United States, vol. 184, pp. 5733-5745, July 2002.
- [8] S. Noboru, "Two aspects of DNA base composition: G+C content and translation-coupled deviation from intra-strand rule of A = T and G = C," *J. Mol. Evol.*, Springer-Verlag, Germany, vol.49, pp. 49-62, July 1999.
- [9] J. B. Plotkin, J. Dushoff, M. M. Desai, and H. B. Fraser, "Codon usage and selection on proteins," *J. Mol. Evol.*, Springer-Verlag, Germany, vol. 63, pp. 635-653, October 2006.
- [10] S. Kanaya, Y. Yamada, Y. Kudo, and T. Ikemura, "Studies of codon usage and tRNA genes of 18 unicellular organisms and quantification of *Bacillus subtilis* tRNAs: gene expression level and species-specific diversity of codon usage based on multivariate analysis," *Gene*, Elsevier/North-Holland, Netherlands, vol.238, pp. 143-155, September 1999.
- [11] S. Noguchi, and Y. Satow, "Purification of human β 2-adrenergic receptor expressed in methylotrophic yeast *Pichia pastoris*," *J. Biochem*, Oxford University Press, England, vol. 140, pp. 799-804, October 2006.
- [12] K. Sau, S. K. Gupta, S. Sau, and T. C. Ghosh, "Synonymous codon usage bias in 16 *Staphylococcus*

- aureus phages: implication in phage therapy,” *Virus Res*, Elsevier Science, Netherlands, vol. 113, pp. 123–231, November 2005.
- [13] N. Sueoka, and Y. Kawanishi, “DNA G+C content of the third codon position and codon usage biases of human genes,” *Gene*, Elsevier/North-Holland, Netherlands, vol. 261, pp. 53–62, December 2000.
- [14] X. F. Wan, D. Xu, A. Kleinhofs, and J. Zhou, “Quantitative relationship between synonymous codon usage bias and GC composition across unicellular genomes,” *BMC Evol. Biol.*, BioMed Central, England, vol. 4, June 2004.
- [15] D. J. Lynn, G. A. Singer, and D. A. Hicker, “Synonymous codon usage is subject to selection in thermophilic bacteria,” *Nucleic Acids Res*, Oxford University Press, England, vol. 30, pp. 4272–4277, October 2002.
- [16] L. Duret, “tRNA gene number and codon usage in the *C.elegans* genome are co-adapted for optimal translation of highly expressed genes,” *Trends Genet*, Elsevier Trends Journals, vol. 16, pp. 287–289, July 2000.
- [17] J. Ma, T. Zhou, W. Gu, X. Sun, and Z. Lu, “Cluster analysis of the codon use frequency of MHC genes from different species,” *Biosystems*, Elsevier Science Ireland, Ireland, vol. 65, pp. 199–207, March–May 2002.
- [18] Y. Prat, M. Fromer, N. Linial, and M. Linial, “Codon usage is associated with the evolutionary age of genes in metazoan genomes,” *BMC Evol. Biol.*, BioMed Central, England, vol. 9, pp. 285–299, December 2009.
- [19] S. Hassard and G. Ward, “Efficient creation of sequencing libraries from blunt-ended restriction enzyme fragments,” *Biotechniques*, Informa Healthcare USA, Inc., England, vol. 18, pp. 396–398, March 1995.
- [20] G. M. Jenkins and E. C. Holmes, “The extent of codon usage bias in human RNA viruses and its evolutionary origin,” *Virus Res*, Elsevier Science, Netherlands, vol. 92, pp. 1–7, March 2003.
- [21] X. R. Ma, S. B. Xiao, L. R. Fang, and H. C. Chen, “Bias of base composition and codon usage in pseudorabies virus genes,” *Yi Chuan Xue Bao*, Ke xue chu ban she, China, vol. 32, pp. 616–624, June 2005.
- [22] J. D. Hall, J. S. Gibbs, D.M. Coen, and D. W. Mount, “Structural organization and unusual codon usage in the DNA polymerase gene from herpes simplex virus type 1,” *DNA*, Mary Ann Liebert, United States, vol. 5, pp. 281–288, August 1986.
- [23] S. Karlin, B. E. Blaisdell, and G. A. Schachtel, “Contrasts in codon usage of latent versus productive genes of Epstein-Barr virus: data and hypotheses,” *J. Virol*, American Society for Microbiology, United States, vol. 64, pp. 4264–4273, September 1990.
- [24] Y. Zhang, A. C. Cheng, M. S. Wang, D. K. Zhu, R. Y. Jia, F. Liu, et al, “Analysis of Synonymous Codon Usage in the UL26.5 Gene of Duck Enteritis Virus,” *IEEE*, the 2nd International Conference on BioMedical Engineering and Informatics, Tianjing, China, vol. 4, pp. 1829–1835, October 2009.
- [25] M. S. Cai, A. C. Cheng, M. S. Wang, L. C. Zhao, D. K. Zhu, Q. H. Luo, et al, “Characterization of synonymous codon usage bias in the duck plague virus UL35 gene,” *Intervirology*, Karger, Switzerland, Vol. 52, pp. 266–278, August 2009.
- [26] H. Chang, A. C. Cheng, M. S. Wang, X. Wei, K. P. Lou, M. S. Cai, et al, “Characterization of codon usage bias in the newly identified DPV gE gene,” *IEEE*, the 2nd International Conference on BioMedical Engineering and Informatics, Tianjing, China, vol. 4, pp. 1836–1841/1874, October 2009.
- [27] L. C. Zhao, A. C. Cheng, M. S. Wang, G. P. Yuan, and M. S. Cai, “Characterization of codon usage bias in the dUTPase gene of duck enteritis virus,” *Prog. Nat. Sci.*, Science Press, China, Vol. 18, pp. 1069–1076, March 2008.
- [28] S. Davison, K. A. Converse, A. N. Hamir, and R. J. Eckroade, “Duck viral enteritis in domestic Muscovy ducks in Pennsylvania,” *Avian Dis*, American Association Of Avian Pathologists, Ithaca, Vol. 37, pp. 1142–1146, October–December 1993.
- [29] Y. M. Saif, H. J. Barnes, J. R. Glisson, A. M. Fadly, L. R. McDougald, and D. E. Swayne, *Diseases of Poultry*[M]. 11th ed. Ames: Iowa State University Press, 2003, pp. 354–363.

- [30] A. C. Cheng, M. S. Wang, and M. Wen, "Construction of duck enteritis virus gene libraries and discovery, cloning and identification of viral nucleocapsid protein gene," *High Technology Letters*, Institute of Scientific and Technical Information of China, China, vol. 16, pp. 948–953, 2006.
- [31] K. L. Carter, P. L. Ward, and B. Roizman, "Characterization of the Products of the UL43 Gene of Herpes Simplex Virus 1: Potential Implications for Regulation of Gene Expression by Antisense Transcription," *J. Virol*, American Society for Microbiology, United States, vol. 70, pp. 7663–7668, July 1996.
- [32] P. M. Sharp, W. H. Li, "Codon usage in regulatory genes in *Escherichia coli* does not reflect selection for 'rare' codons," *Nucleic Acids Res*, Oxford University Press, England, Vol. 2, pp. 7737–7749, February 1986.
- [33] 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 Res*, Oxford University Press, England, vol. 15, pp. 1281–1295, February 1987.
- [34] F. Wright, "The 'effective number of codons' used in a gene," *Gene*, Elsevier/North-Holland, Netherlands, vol. 87, pp. 23–29, March 1990.
- [35] P. Jiang, X. Sun, and Z. Lu, "Analysis of Synonymous codon usage in *Aeropyrum pernix* K1 and other Crenarchaeota Microorganism," *J Genet Genomics*, Science press, China, vol. 34, pp. 275–284, March 2007.
- [36] H. Akashi, "Gene expression and molecular evolution," *Curr. Opin. Genet. Dev.*, Elsevier, England, vol. 11, pp. 660–666, December 2001.
- [37] R. Hershberg and D. A. Petrov, "Selection on codon bias," *Annu. Rev. Genet.*, Annual Reviews, United States, vol. 42, pp. 287–299, December 2008.
- [38] H. Akashi, "Codon bias evolution in *Drosophila*. Population genetics of mutation-selection drift," *Gene*, Elsevier/North-Holland, Netherlands, vol. 205, pp. 269–278, December 1997.
- [39] L. Duret, "Evolution of synonymous codon usage in metazoans," *Curr. Opin. Genet. Dev.*, Elsevier, England, vol. 12, pp. 640–649, December 2002.
- [40] Z. Yang and R. Nielsen, "Mutation-selection models of codon substitution and their use to estimate selective strengths on codon usage," *Mol. Biol. Evol.*, Oxford University Press, United States, vol. 25, pp. 568–579, January 2008.
- [41] J. Prilusky and E. Bibi, "Studying membrane proteins through the eyes of the genetic code revealed a strong uracil bias in their coding mRNAs," *Proc. Natl. Acad. Sci. U.S.A.*, National Acad Sciences, United States, vol. 106, pp. 6662–6666, April 2009.
- [42] H. Sakai, T. Washio, R. Saito, A. Shinagawa, M. Itoh, K. Shibata, et al. "Correlation between sequence conservation of the 5' untranslated region and codon usage bias in *Mus musculus* genes," *Gene*, Elsevier/North-Holland, Netherlands, vol. 276, pp. 101–105, October 2001.
- [43] W. Gu, T. Zhou, J. Ma, X. Sun, and Z. Lu, "Analysis of synonymous codon usage in SARS Coronavirus and other viruses in the Nidovirales," *Virus Res*, Elsevier Science, Netherlands, vol. 101, pp. 155–161, May 2004.
- [44] T. Zhou, W. Gu, J. Ma, X. Sun, and Z. Lu, "Analysis of synonymous codon usage in H5N1 virus and other influenza A viruses," *Biosystems*, Elsevier Science Ireland, Ireland, vol. 81, pp. 77–86, April 2005.
- [45] T. Zhou, X. Sun, and Z. Lu, "Synonymous codon usage in environmental chlamydia UWE25 reflects an evolutionary divergence from pathogenic chlamydiae," *Gene*, Elsevier/North-Holland, Netherlands, vol. 368, pp. 117–125, March 2006.
- [46] H. S. Najafabadi, J. Lehmann, and M. Omid. "Error minimization explains the codon usage of highly expressed genes in *Escherichia coli*," *Gene*, Elsevier/North-Holland, Netherlands, vol. 387, pp. 150–155, January 2007.
- [47] C. A. MacLean, S. Efstathiou, M. L. Elliott, F. E. Jamieson, D. J. McGeoch, "Investigation of herpes simplex virus type 1 genes encoding multiply inserted membrane proteins," *J. Gen. Virol*, Society for General Microbiology, England and Wales, Vol. 72, pp. 897–906, 1991.

- [48] L. P. Powers, K. S. Wilkinson, and P. Ryan, "Characterization of the prv43 gene of pseudorabies virus and demonstration that it is not required for virus growth in cell culture," *Virology*, Academic Press, United States, vol. 199, pp. 81–88, February 1994.
- [49] B. G. Klupp, J. Altmann, H. Granzow, W. Fuchs, and T. C. Mettenleiter, "Identification and characterization of the pseudorabies virus UL43 protein," *Virology*, Academic Press, United States, vol. 334, pp. 224–233, January 2005.
- [50] S. Karlin, B. E. Blaisdell, E. S. Mocarski, and V. Brendel, "A method to identify distinctive charge configurations in protein sequences, with application to human herpesvirus polypeptides," *J. Mol. Biol.*, Academic Press, England, vol. 205, pp. 165-177, January 1989.