Advanced Journal of Microbiology Research Vol. 2015
Available online at http://internationalscholarsjournals.org/journal/ajmr

Full Length Research Paper

Factors affecting synonymous codon usage bias in the gC gene of DPV CHv strain

Fu Xiao Shen[1], An Chun Cheng[1,2,3]*, Ming Shu Wang[1,2], Hai Bin Huang[1], Chuan Feng Li[1], Jin Feng Jiang[1], Ren Yong Jia[2], De Kang Zhu[1,2] and Xiao Yue Chen[3]

[1] Avian Disease Research Center, College of Veterinary Medicine, Sichuan Agricultural University, Ya‟an, Sichuan,
[2] Key Laboratory of Animal Disease and Human Health of Sichuan Province, Sichuan Agricultural University, Ya‟an, Sichuan, 625014, PR China.
[3] Epizootic Diseases Institute of Sichuan Agricultural University, Sichuan Agricultural University, Ya‟an, Sichuan, 625014, PR China.

Accepted 27 January, 2014

The aim was to identify the factors affecting synonymous codon usage bias in the newly discovered gC gene of duck plague virus (DPV) CHv strain and a comparative analysis of the codon usage bias in the gC genes of 27 other reference herpesviruses was performed by using CAI, CHIPS and CUSP program of EMBOSS. The results showed that base composition, mutational bias and natural selection were the major determinants of the codon usage variation in the DPV gC gene. The primary codon usage trends in the DPV gC gene showed obvious difference with a strong bias towards the synonymous codons with A and T at the third codon position. Characterization of nucleotide composition in the DPV gC gene was related to dinucleotide usage bias characterized by the variation of CpG islands. The ENc-plot and GC12s-GC3s revealed that the genetic heterogeneity in the gC genes of the 28 herpesviruses were constrained by G + C content. Moreover, the phylogenetic analysis demonstrated that codon usage patterns of the DPV gC gene were phylogenetically conserved and similar to the gC genes of the avian alphaherpesvirinae. Furthermore, comparisons of the codon preferences in the DPV gC gene with those of Escherichia coli, yeast and human revealed that there was a statistically positive correlation between DPV and yeast (r = 0.646, P < 0.05) . The above results could provide useful information for the synonymous codons usage bias of the DPV gC gene and promote the relevant mechanism for evolution, pathogenesis and functional studies in the area of DPV research and possibly studies with other herpesvirus viruses.

Key words: Duck plague virus (DPV), gC gene, codon usage bias.

INTRODUCTION

Despite the gene degeneration, synonymous codons are not used equally intra and inter-genomic variations in the wide range of living nature, giving rise to the phenomenon termed „codon usage bias‟ (Duret, 2002; Fuglsang, 2006). It has been reported that multiplicity factors are accounting for the biased usage of synonymous codons, such as gene length (Marais and Duret, 2001; Miyasaka, 2002), CpG islands (Scaiewicz et al., 2006; Woo et al., 2007), gene expression level (Ma et al., 2002; Comeron, 2004), proteins secondary structure (Gu et al., 2004; Kahali et al., 2007) and gene density (Versteeg et al., 2003) and so on. Many literatures were recorded about the extent and origin in the synonymous codon bias of many genes in many species. In unicellular organisms, the codon usage pattern is the result of the equilibrium between natural selection and compositional mutation bias, but mutation bias is the major factor in accounting for the variation of codon usage in some prokaryotes organisms (Ikemura, 1985; Gupta et al., 2004) . This paradigm has been successfully attributable to many viruses (Meintjes and Rodrigo, 2005; Pinto et al., 2007). Most retroviruses also exhibited a profound discrimination against the CpG dinucleotide sequence and the genome

*Corresponding author. E-mail: chenganchun@vip. Tel: +86 835 2885774

combination characteristics strongly influenced codon usage patterns (Berkhout et al., 2002). In human RNA viruses, most important determinants were mutational pressure rather than translational selection (Jenkins and Holmes, 2003). Besides this typical preference for a particular nucleotide sequence, codon usage bias have sufficient evidences that molecular weight, hydropathy, aromaticity and cys-teine contents were responsible for the variation of amino acid usage in adenovirus genes (Das et al., 2006). Studies of the synonymous codon usage could display information about the molecular evolution of individual genes and prepare data to train genome-specific gene recognition algorithms which recognize protein coding regions in uncharacterized genomic DNA. At the last few years, the synonymous codon preference patterns in herpesvirus were primarily centralized on the Pseudorabies Virus (PRV) (Ma et al., 2005), Herpes Simplex Virus 1 (HSV-1) (Hall et al., 1986), Epstein-Barr Virus (EBV) (Karlin et al., 1990), UL24 gene, UL35 gene and dUTPase gene of DPV (Zhao et al., 2008; Cai et al., 2009; Jia et al., 2009).

Duck Plague (DP), which is caused by DPV, a member of the Herpesviridae family, is an acute and contagious herpesvirus infection of *Anseriformes* (ducks, geese and swans) with high morbidity and mortality in the worldwide waterfowl production (Converse and Kidd, 2001; Shawky and Schat, 2002). Most of the previous researches had focused on the epidemiology and prevention of this disease (Hansen et al., 2000; Yang et al., 2009). However, the molecular biological information about the DPV genome was limited. Recently, a DPV genomic library was constructed successfully in time and the gC gene was discovered in our laboratory (Cheng et al., 2006; Xu et al., 2008). Unfortunately, the delineation of the DPV gC gene could not to be known, although glycoprotein C has several important biological functions in other herpesvirus (Gupta et al., 2001; Rux et al., 2002). It has become important to analyze the codon usage patterns in the DPV gC gene, which may provide more information on the features of the DPV complete genome.

In this communication, the analysis of codon usage bias in the gC gene of the DPV CHv strain are engaged and compared it with the gC genes of 27 other reference herpesvirus. Then, we examine how other factors may affect codon usage variation in the DPV gC gene and its reference species. Meanwhile, the codon usage patterns in the DPV gC gene were compared with *E.coli*, yeast and Human by multivariate statistical analysis. On the whole, a comparison analysis of codon preference between the DPV gC gene and other species can provide a basis for understanding the relevant mechanism for biased usage of synonymous codons and for selecting an appropriate host expression system to improve the expression of a target gene. Such information‟s can not only provide new insight into the features of the DPV genome and improve the understanding of factors shaping codon usage patterns in the gC gene of DPV, but also may increase the studies efficiency in the area of other herpesvirus.

MATERIALS AND METHODS

Virus and gene sequences

The DPV CHv strain, which is a high- virulence field strain of DPV, was obtained from Key Laboratory of Animal Disease and Human Health of Sichuan Province. The gC gene of DPV CHv strain (GenBank Accession No: EU076811) was discovered by constructing a DPV genomic library in our laboratory. The nucleotide sequences of the gC genes of 27 other reference herpesviruses are extracted from the NCBI GenBank nucleotide database. To keep the statistical significance of codon usage bias, only sequences with length above 300 bps are indagated.

Characterization of synonymous codon usage in the gC genes of DPV and 27 reference herpesviruses

For each gene, codon usage was estimated by using CAI, CHIPS and CUSP program of EMBOSS. Generally, the effective number of codons (ENC) was used to measure the codon bias magnitude for an individual gene, without dependence on gene length or specific knowledge of preferred codons (Novembre, 2002). Values of ENc could range from 20 (for a gene with extreme bias used only one codon per amino acid) to 61 (for a gene with no bias using synonymous codons equally) (Comeron and Aguade, 1998). The codon adaptation index (CAI) value was regarded as a reference set of highly expressed genes from a species to assess the relative merits of each codon. Higher CAI value expected stronger codon usage bias and higher expression level, whereas the reverse was true for lower CAI value. The relative synonymous codon usage (RSCU) value was used to examine the codon usage variation among the genes without the confounding influence of amino acid composition. 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(Liu et al., 2005). The GC3s is a good indicator in the extent of base composition bias, which represents the frequency of the nucleotide G + C at the third synonymous codon position (excluding Met, Trp and the termination codons). The GC12s is the average frequency of the nucleotide G + C at the once and twice synonymous codon position. Since the parameters have been effective in investigating the codon usage bias, it is useful for examining the codon usage pattern by the ENc-plot and GC12s-GC3s.

Molecular characterization and phylogenetic analysis of the DPV gC gene

The nucleotide sequences of the DPV gC gene and 27 reference herpesviruses were translated into amino acid sequences by using DNASTAR software. After this, multiple sequence alignment and phylogenetic analysis were performed for the gC genes of 28 herpesviruses with CLUSTAL-X and TREEVIEW software.

Comparison of codon preferences of DPV gC gene with those of *E. coli,* Yeast and Human

Correlation analysis was performed by the SPSS statistics 17.0 software, which is used to compare codon usage bias among DPV, *E.coli*, yeast and Human. The database of the codon usage in *E. coli*, yeast and Human is available at http: www.kazusa. or.jp/condon.

RESULTS AND DISCUSSIONS

Characterization of synonymous codon usage in the gC gene of DPV CHv strain

While the CAI, ENc and related measures indicate the overall DPV gC gene codon bias, it is also important to look more closely at the patterns of codon bias. The details of the coding-gC gene of 59 nucleotide sense codons (excluding Met, Trp and the termination codons) in the DPV CHv strain are shown in Table 1. It is evident that codon usage bias in the predicted polypeptide is strong bias towards the synonymous codons with A and T in the wobble position, but not all codons display the preference service as expected. “Optimal codons” are existed a high level of diversity for coding Leu, Pro, Arg, Ser, Thr and Val amino acids at codon usage bias because they have a 6-fold and 4-fold coding degeneracy, while the preferred codons of that have 2-fold or 3-fold coding degeneracy are more uniform. On the other hand, the frequencies of codons usage in coding the same amino acids are different; For instance, the frequencies of triplet CGC is more than 13.889 (‰), which is the preferred codon coding Arg amino acids, Interestingly, the triplet AGG has become an avoidance of particular codon in coding Arg amino acid.

Codon[a] = The preferentially used codons for each amino acid are displayed in bold. Fractb = The “Fract” column shows the proportion of a given codon usage in its redundant set. RSCUc = The “RSCU” column shows the proportion of relative synonymous codon usage Frequency/1000d = The “Frequency” column lists the number of codons present per 1000 bases in the input sequence. * shows that a strong bias towards the codons with A and T at the third codon position Place the bottom of the Table 1, because these explaintation are the explanation of Table 1

Characterization of synonymous codon usage in the gC genes of the 28 herpesviruses

The results obtained by EMBOSS analysis in the gC genes of 28 herpesviruses are shown in Table 2. Codon usage in the DPV gC gene and its homologous genes is highly nonrandom in all the herpesviruses and characterization of synonymous codon usage in the gC genes of the 28 herpesviruses also differs dramatically. Up to date, the CAI values range from 0.58 to 0.77, with a mean value of 0.69 and standard deviation (SD) of 0.05 and the ENc values range from 27.54 to 61.00, with a mean value of 44.73 and standard deviation (SD) of 11.21. Because most ENc values of the 28 herpesviruses are much higher (ENc> 40), the codon usage bias in the herpesvirus gC gene is slightly lower. There is also a slight variation in codon usage patterns which is the same among different herpesvirus gC genes (SD = 11.21).

Similarly, the GC3s contents of each gC gene also confirm the homogeneity of synonymous codon usage among the different herpesviruses, which varies from 14.77% to 99.22%, with a mean of 62.50% and a standard deviation (SD) of 26.20% and the GC12s contents also range from 31.20 to 71.17%, with a mean of 56.10% and a standard deviation (SD) of 12.04%. This observation indicates that there is a significant heterogeneity in compositional bias as well as in the codon usage pattern within and among the members of herpesvirus. Taken together, there are marked variations of codon usage in base composition, other factors might also have some influences in shaping the codon usage variation in the gC genes of 28 herpesvirus strains.

A comparative analysis of the codon usage bias in the DPV gC gene and its homologous genes of 27 other herpesviruses was performed. The data of synonymous codon usage bias show the DPV gC gene and its 27 reference herpesviruses adopt relatively similar codon usage patterns, although the DPV gC gene shows a few variations of codon usage bias with its reference herpesvirus species and the DPV gC gene prefers to use the codons with A and T at the third codon position. It is reasonable that the nucleotide „„C and/or G“‟ seems to face the strongest selection force to make it change to „„A and/or T” in silent sites during DPV CHv strain evolutionary procession (Shields et al., 1988). At the same time, random mutations increase the population of AT rather than CG as a result, undergo spontaneous deaminations (Lindahl, 1993).

Phylogenetic analysis of the DPV gC gene and the 27 other reference herpesviruses

Although the codon usage pattern among different species is a complex phenomenon, it plays an important role in illuminating the underlying mechanisms of codon usage patterns in order to understand the evolution of the species. After the deduced amino acid sequence is encoded by the 1296 bp open reading frame (ORF) of the DPV gC gene, a phylogenetic tree based on the amino acid sequences of the gC genes of the 28 herpesviruses is shown in Figure 1. The major clades in the herpesviruses family are corresponding to three accepted subfamily “A”, “B” and “C”. Multiple sequence alignment of the amino acid sequences demonstrates that DPV is evolutionarily closer to some avian alphaherpesvirinae, such as CaHV-2, MeHV-1 and MDV- 2 clustered within a “D” monophyletic clade and its codon usage pattern have a relatively correlation with MuHV-1, HHV-5 and HHV-7. Comparative analysis in the gC genes of 28 herpesviruses indicates that synonymous codon usage in these genes is phylogenetically conserved. The data show that the gC genes of DPV, CaHV-2, MeHV-1 and MDV-2, whose natural host is avian, have a stronger correlation than the gC genes of herpesviruses with

Table 1. Characterization of synonymous codon usage in the gC gene of DPV CHv strain.

AA	Codon[a]	Fract[b]	RSCU[c]	Frequency/1000[d]
Ala	GCA*	0.379	1.52	25.463
	GCC	0.207	0.83	13.889
	GCG	0.138	0.55	9.259
	GCT*	0.276	1.10	18.519
Cys	TGC	0.333	0.67	6.944
	TGT*	0.667	1.33	13.889
Asp	GAC	0.333	0.67	13.889
	GAT*	0.667	1.33	27.778
Glu	GAA*	0.778	**1.56**	48.611
	GAG	0.222	0.44	13.889
Phe	TTC	0.417	0.83	11.574
	TTT*	0.583	1.17	16.204
Gly	GGA*	0.375	1.50	27.778
	GGC*	0.375	1.50	27.778
	GGG	0.188	0.75	13.889
	GGT	0.062	0.25	4.630
His	CAC	0.143	0.29	2.315
	CAT*	0.857	1.71	13.889
Ile	ATA*	0.353	1.06	27.778
	ATC	0.324	0.97	25.463
	ATT	0.324	0.97	25.463
Lys	AAA*	0.550	1.10	25.463
	AAG	0.450	0.90	20.833
Leu	CTA	0.133	0.80	9.259
	CTC	0.033	0.20	2.315
	CTG	0.167	1.00	11.574
	CTT	0.167	1.00	11.574
	TTA*	0.267	1.60	18.519
	TTG*	0.233	1.40	16.204
Asn	AAC	0.278	0.56	11.574
	AAT*	0.722	1.44	30.093
Pro	CCA*	0.308	1.23	18.519
	CCC	0.231	0.92	13.889
	CCG	0.231	0.92	13.889
	CCT	0.231	0.92	13.889
Gln	CAA*	0.700	1.48	16.204

Table 1. Contd.

	CAG	0.300	0.60	6.944
Arg	AGA*	0.400	2.40	13.889
	AGG	0.000	0.00	0.000
	CGA*	0.200	1.20	6.944
	CGC	0.067	0.40	2.315
	CGG*	0.200	1.20	6.944
	CGT	0.133	0.80	4.630
Ser	AGC	0.030	0.18	2.315
	AGT	0.152	0.91	11.574
	TCA	0.121	0.73	9.259
	TCC	0.091	0.55	6.944
	TCG	0.273	0.64	20.833
	TCT*	0.333	2.00	25.463
Thr	ACA*	0.500	2.00	43.981
	ACC	0.132	0.53	11.574
	ACG	0.158	0.63	13.889
	ACT	0.211	0.84	18.519
Val	GTA	0.152	0.61	11.574
	GTC	0.212	0.85	16.204
	GTG	0.182	0.73	13.889
	GTT*	0.455	1.82	34.722
Tyr	TAC	0.211	0.42	9.259
	TAT*	0.789	1.58	34.722
Pro	CCA*	0.308	1.23	18.519

Table 2. Characterization of synonymous codon usage in the gC genes of 28 herpesviruses.

Species	Virus name	Access No	CAI	ENC	GC_{3S}(%)	GC_{12S}(%)
Alpha herpesvirinae	Duck enteritis virus CHv strain (DPV CHv)	EU076811	0.65	54.41	38.89	45.84
	Gallid herpesvirus 1 (GaHV-1)	DQ862824	0.66	57.04	40.96	48.56
	Gallid herpesvirus 2 (GaHV-2)	AF147806	0.65	55.93	39.84	46.42
	Human herpesvirus 3 (HHV-3)	EU154348	0.68	53.57	50.62	45.64
	Human herpesvirus 2 (HHV-2)	EU018123	0.74	35.70	87.53	62.17
	Bovine herpesvirus 5 (BoHV-5)	U35883	0.77	33.26	88.14	68.33
	Equid herpesvirus 4 (EHV-4)	A21044	0.68	58.59	48.35	48.36
	Canid herpesvirus 1 (CnHV-1)	AF361074	0.61	37.86	16.52	31.20
	Caprine herpesvirus 1 (CpHV-1)	AY821804	0.72	32.54	93.04	69.25
	Cervid herpesvirus 1 (CeHV-1)	DQ333390	0.7727.54		99.01	69.65
	Cervid herpesvirus 2 (CeHV-2)	DQ333391	0.7628.61		98.61	69.42
	Cervid herpesvirus 16 (CeHV-16)	DQ149153	0.7729.83		96.98	71.17
	Elk herp Elk herpesvirus (ElHV)	DQ333389	0.7430.83		96.85	69.67
	Bovine herpesvirus 1 (BoHV-1)	DQ184913	0.62	33.27	90.48	62.20
	Meleagridherpesvirus 1 (MeHV-1)	AF291866	0.67	57.80	45.71	48.06
	Human herpesvirus 1 (HHV-1)	NC001806	0.69	55.52	61.34	70.11
	Marek's disease virus 2 (MDV-2)	AB024414	0.67	61.00	54.07	49.17

Table 2. Contd.

	Felid herpesvirus 1 (FeHV-1)	D86616	0.65	54.29	38.32	45.89
	Bubaline herpesvirus 1 (BuHV-1)	EU723234	0.75	27.80	99.22	66.80
	Psittacid herpesvirus 1 (PsHV-1)	AY372243	0.69	47.59	73.00	55.51
	Equine herpesvirus 1 (EHV-1)	AY464052	0.69	54.53	50.11	55.44
	Chimpanzee alpha-1 herpesvirus (ChaHV-1)	AB218905	0.74	39.51	82.95	61.13
	Human herpesvirus 5 (HHV-5)	NC001347	0.67	48.28	55.53	59.91
	Human herpesvirus 7 (HHV-7)	NC001716	0.66	53.85	33.47	33.48
Beta herpesvirinae	Murid herpesvirus 1 (MuHV-1)	NC004065	0.64	47.41	51.94	64.20
	Panine herpesvirus 2 (PaHV-2)	NC003521	0.65	42.06	57.51	67.59
Gamma herpesvirina	Alcelaphine herpesvirus 1 (AlHV-1)	NC002531	0.70	52.47	46.36	48.43
	Saimiriine herpesvirus 2 (SaHV-2)	AJ410492	0.58	41.45	14.77	37.08

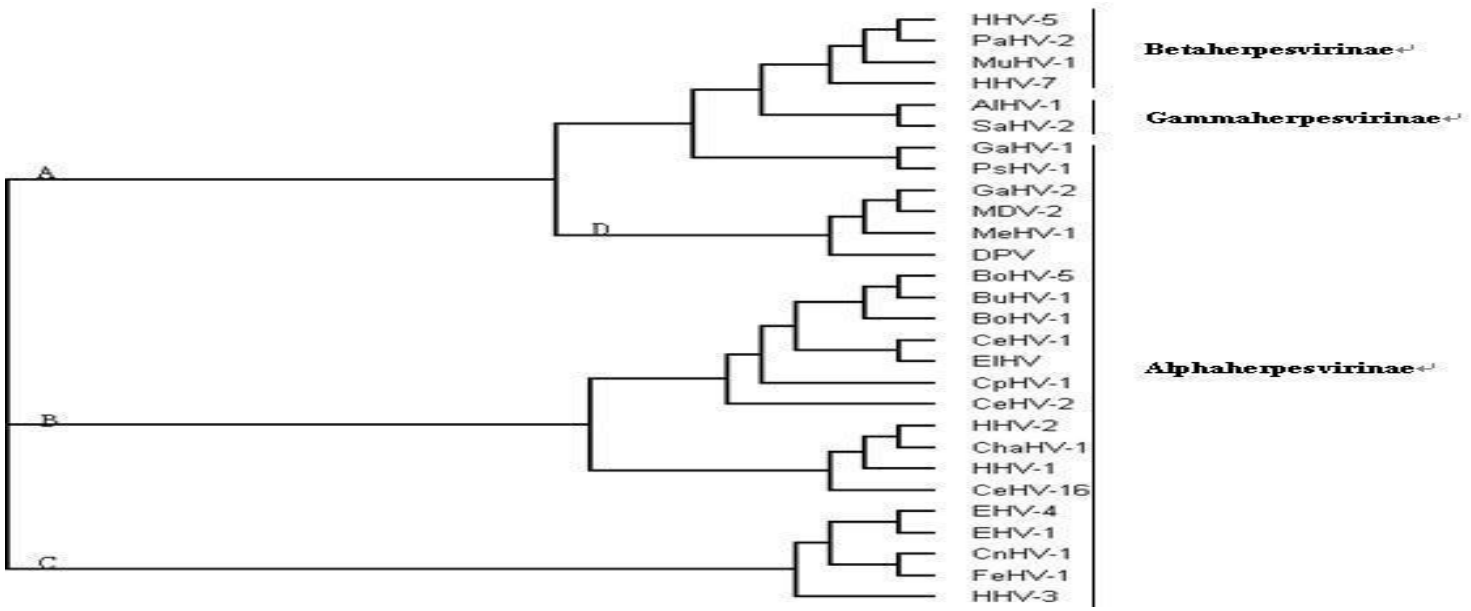

Figure 1. Phylogenetic analysis in the gC genes of the 28 herpesviruses strains.

mammalian hosts, such as MuHV-1, HHV-5 and HHV-7. The results suggest that codon usage patterns are complicated by distinguished virus-specific mutational bias and these do not exhibit any signs of host-specificity in the gC genes of 28 herpesviruses. It can be concluded that the DPV gC gene has a close evolutionary relationship with other avian alphaherpesvirinae and is different from other types of herpesviruses in the certain ways. This type distribution of codon usage pattern is consistent with the classified reports in DPV dUTPase, UL24 and UL35 (Zhao et al., 2008; Cai et al., 2009; Jia et al., 2009). The results suggest that the codon usage pattern of DPV is similar to that of avian alphaherpesvirinae, however, the conclusive evidence may require further analysis of other DPV genes, such as viral genes for maintenance or replication.

Factors affecting synonymous codon usage bias in the gC gene of DPV CHv Strain

Codon usage bias is one of the most important indicators of the selective forces shaping genome evolution. The classical explanation for unequal usage of synonymous codons are accepted as the most reasonable hypothesis of neutral theory and the “selection-mutation-drift” model, which are natural selection, mutational bias, or both in different species (Ghosh et al., 2000). According to neutral theory, mutations at degenerate coding positions resulting in random synonymous codon choice should be selectively neutral, while in the “selection–mutation–drift” model, synonymous codon usage bias simply reflects the balance between selection favoring optimal codons and mutation–drift allowing persistence of non-optimal codons.

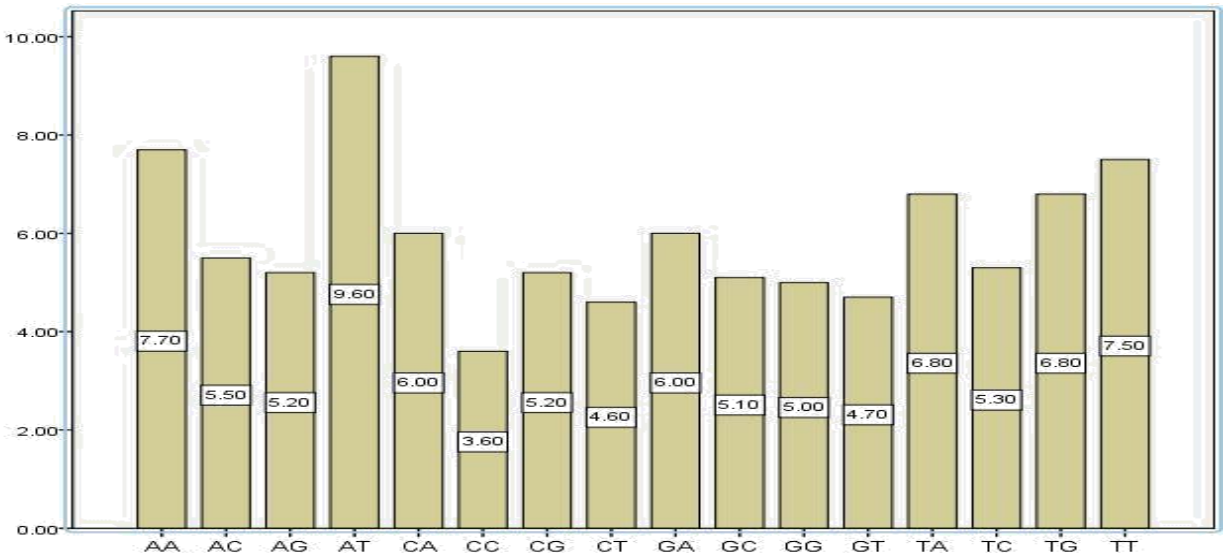

Figure 2. Compositional distribution of dinucleotides in the gC gene of DPV CHv strain.

However, these hypotheses were not fit to explain codon usage variation with the development of genome project of many organisms (Zavala et al., 2002). For this reason, several measures of the degree of codon usage bias in a certain gene have been developed.

While codon usage bias influenced with mutation pressure, the driving force may result in CpG suppression that have also been proposed to be responsible for shaping the dinucleotides group contents. Characterization of nucleotide composition in the DPV gC gene reflects dinucleotide usage bias characterized by the variation of CpG islands (Figure 2). The ApT and ApA dinucleotides have the higher proportion that the ratio is more than 9.6 and 7.7%, respectively. Nevertheless, the great majority of containing G and C dinucleotides was suppressed at a low level. The variation of CpG islands are two possibilities about dinucleotide usage bias in the DPV gC gene. Firstly, this depletion is believed to be a result of accidental mutations by deamination of 5-methylcytosine to thymine. Since the product of this mutation is indistinguishable from endogenous nucleotides, it cannot be recognized by DNA repair systems (Jones et al., 1992). DNA methylation mainly elevates the mutation rate of C to T through spontaneous deamination of the resultant 5- methylcytosine (Sved and Bird, 1990). Consequently, the C to T transition in one strand leads to G to A transition on the opposite strand, which in turn increases the AT content (Cooper and Krawczak, 1989). Secondly, DNA methylation has encouraged targeted disruption of the gene for cytosine methyltransferase (Li et al., 1992). It is believed that this phenotype is related to the fact that DNA methylation has been found to have a repressing effect on the binding of a specific methyl-CpG binding protein (Boyes and Bird, 1992; Eden and Cedar, 1994). Based on these observations it has been suggested that general DNA methylation may have evolved as a way of suppressing CpG islands. To reveal the factors influencing architecture of the protein-coding DPV gC gene, mutational bias appears to be the main determinant accounting for the codon usage variation in dinucleotides compositional distribution.

It has been reported that a plot of ENC against GC3s can be used to explore the heterogeneity codon usage among the recommending genes (Wright, 1990). If codon usage bias has some other influences than mutational bias, the actual distribution of genes could not be expected to be indicative. In other words, if a particular gene is only subject to G+C on the third codon position, the values of ENC would lie on or just below the continuous curve (Jiang et al., 2007). The ENc-plot of the gC genes of 28 herpesviruses shows that a few genes lie on or just below the expected curve towards originates GC poor regions from extreme mutational bias (Figure 3). Oppositely, the bulk of points fairly spread out the theoretical curve, that indicating that the codon usage is highly influenced by the G+C compositional constraints. Hence, other than mutational bias, some other factors have primary influences in detecting the codon usage variation among the gC gene, which are independent of compositional constraints.

A plot of GC12s against GC3s is another effective way to investigate the heterogeneity of codon usage in the gC genes of 28 difference herpesviruses. As to the gene itself, natural pressure always act as an effective factor in promoting the gene to adapt to the change in the external environment. If natural pressure determined codon usage in each codon group, the level of divergence between the species compared is lower in the GC12s-GC3S plot. The result indicates that most of the gene points could not be localized as a fundamental confine in the plot (Figure 4).

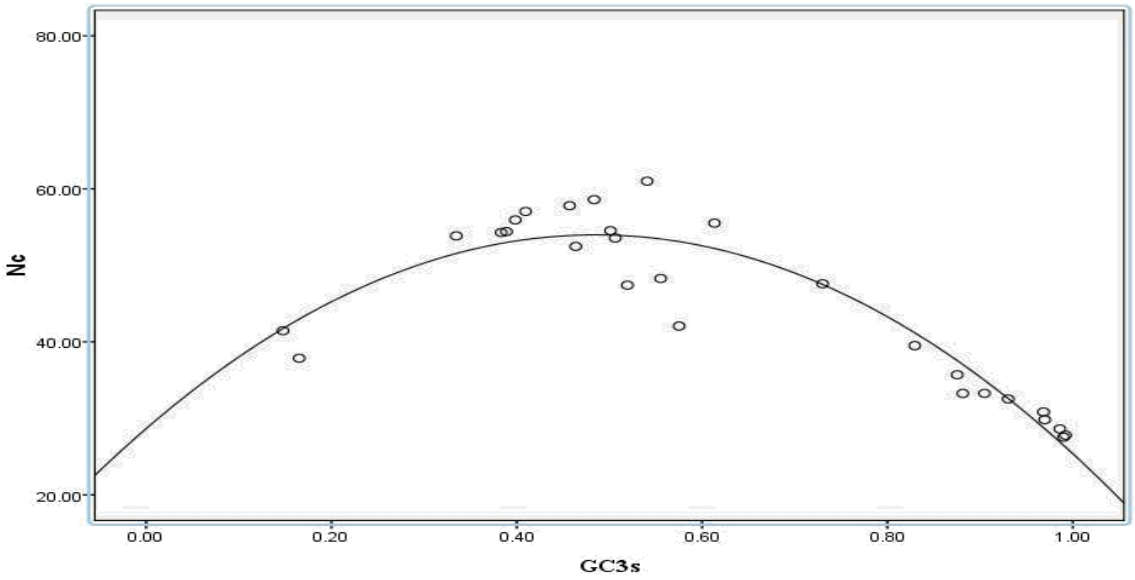

Figure 3. The plot of ENC and C3s of the gC gene in the 28 different herpesviruses strains.

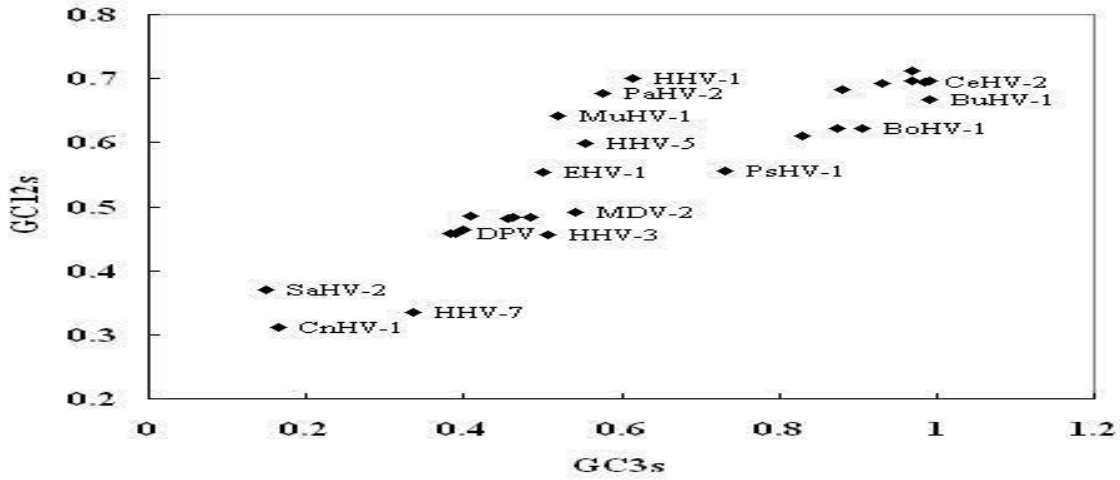

Figure 4. The plot of C12s against GC3s of the gC genes in the 28 different herpesviruses strains.

The GC12s value widespread change when the GC3s value has little fluctuation in the gC genes of 28 herpesvirus, especially CnHV-1, HHV-7, PsHV-1 and SaHV- 2. It can be speculated that mutational bias might play a crucial role, natural selection bias in a minor way, in shaping codon usage among the gC genes from the GC12s-GC3S plot. The result also suggested that, in addition to mutational bias, natural selection may also be the secondary evolutionary forces to shape codon usage in the gC gene of DPV CHv strain. It is known to all that synonymous codon choice patterns are related to the abundance of isoaccepting tRNAs (Higgs and Ran, 2008), moreover, the extent of the bias in codon usage is postively correlated to the level of gene expression level (Zouridis and Hatzimanikatis, 2008). In unicellular organisms, highly expressed genes have a strong selective preference for the codons complementary to the most abundant tRNA species, whereas lower expressed genes display more uniform codon usage patterns largely compatible with the mutational bias in the absence of translational selection (Lesnik et al., 2000). In multicellular organisms, the diverse patterns of codon usage may arise from compositional constraints of the genomes (Ghosh et al., 2000; Romero et al., 2003) . In this paper, we show the base composition of the DPV gC gene and its pattern of codon bias contributing to the existing variation in codon usage bias within and between other species. Characterization of nucleotide composition

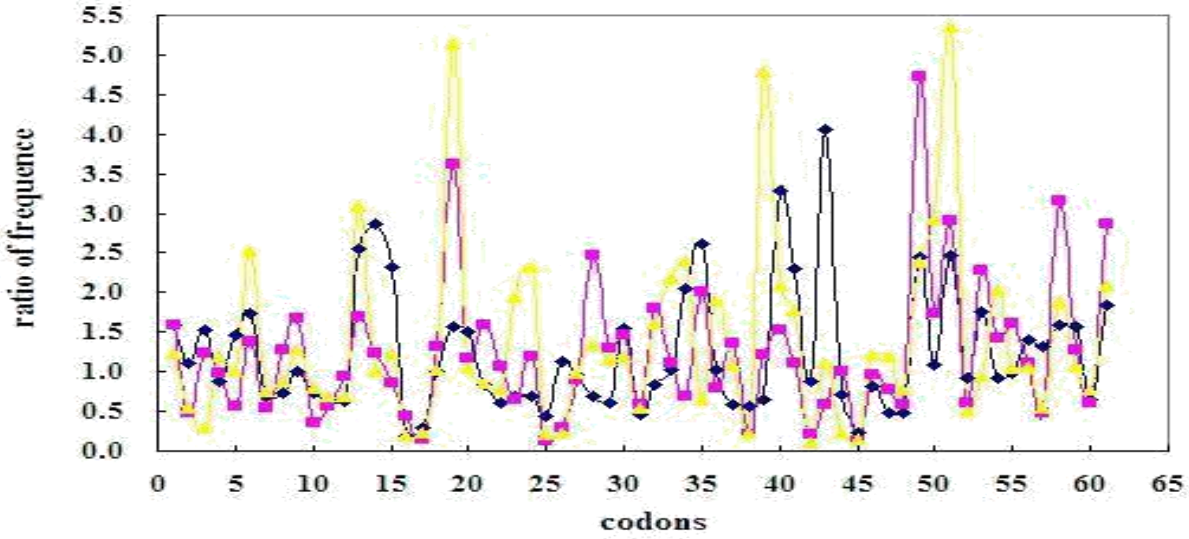

Figure 5. The comparisons in the ratio of codon usage frequency (1/1000) of DPV to *E. coli*, yeast, Human.

Table 3. Correlation analysis among DPV, *E.coli*, yeast and Human.

r	DPV	Yeast	Human	E. coli
DPV	1	0.646*	0.266*	0.331**
Yeast		1	0.384**	0.418**
Human			1	0.604**
E. coli				1

*Significant correlation at the 0.05 level. **Significant correlation at the 0.05 level.

in the DPV gC gene was related to dinucleotide usage bias characterized by the variation of CpG islands. The ENc-plot and GC12s-GC3s revealed that the genetic heterogeneity in the gC genes of the 28 herpesviruses was constrained by G + C content. Base composition, mutational bias and natural selection were the major determinants of the codon usage variation in the DPV gC gene that have been accepted to account for codon usage variation.

Comparison of codon usage bias among DPV, *E. coli*, yeast and humans

In general, the codon usage bias in genes remains at a certain level across species. The DPV gC gene compared with those of *E. coli*, yeast and Human to see which will be the suitable host for the optimal expression of DPV genes (Figure 5) . The data revealed that there are 22 codons showing a DPV-to-*E. coli* ratio higher than 2 or lower than 0.50, with 19 codons between DPV-to-Human, but only 17 codons showing a DPV- to- yeast ratio higher than 2 or lower than 0.50. According to the lowest codon usage differences between DPV and yeast, the codon usage of the DPV gC gene more closely resembles that of yeast than that of *E. coli* and Human. The yeast expression system may be better applied to the production of the DPV gC protein. Correlation analysis, a commonly used multivariate statistical approach, reveals that the codon usage frequencies of the DPV gC gene and three mode organisms (*E. coli*, yeast and Human) show enormous variation (Table 3). The correlation analysis was made in this study that high familiar relationship was observed between the DPV and yeast ($r = 0.646$, $P < 0.05$). The optimization of the DPV gC gene with host-preferred codons is likely to improve the expression level of the DPV gC gene in a given host. This may serve as a guide for manipulating the expression of the targeted gene. Codon optimization of the DPV gC gene may be required to express the DPV gC gene efficiently in the *E. coli* or Human.

In summary, our work has provided a basic understanding of the evolution and pathogenesis of the DPV gC gene and offered some new insights into the mechanisms for codon usage bias, as well as contributing significantly to the area of DPV research and possibly studies with other herpesvirus virus.

The ratio higher than 2 or lower 0.5 indicates that the

codon usage preference differs.

ACKNOWLEDGMENTS

The research was supported by Changjiang Scholars and Innovative Research Team in University (IRT0848), National Natural Science Foundation of China (30771598), the Cultivation Fund of the Key Scientific and Technical Innovation Project, Ministry of Education of China (706050), and Supported by the earmarked fund for Modern Agro-industry Technology Research System (nycytx-45-12).

REFERENCES

Berkhout B, Grigoriev A, Bakker M, Lukashov VV (2002). Codon and amino acid usage in retroviral genomes is consistent with virus-specific nucleotide pressure. Aids. Res. Hum. Retrovir. 18: 133-141.

Boyes J, Bird A (1992). Repression of genes by DNA methylation depends on CpG density and promoter strength: evidence for involvement of a methyl-CpG binding protein. EMBO J. 11: 327-333.

Cai MS, Cheng AC, Wang MS, Zhao LC, Zhu DK, Luo QH, Liu F, Chen XY (2009). Characterization of synonymous codon usage bias in the duck plague virus UL35 gene. Intervirology 52: 266-278.

Cheng AC, Wang MS, Wen M, Zhou WG, Guo YF, Jia RY, XU C, Yuan GP, Liu YC (2006). Construction of duck enteritis virus gene libraies and discovery, cloning and identification of viral nucleocapsid protein gene. High. Technol. Lett. 16: 948-953.

Comeron JM (2004). Selective and mutational patterns associated with gene expression in humans: influences on synonymous composition and intron presence. Genetics. 167: 1293-1304.

Comeron JM, Aguade M (1998). An evaluation of measures of synonymous codon usage bias. J. Mol. Evol. 47: 268-274.

Converse KA, Kidd GA (2001). Duck plague epizootics in the United States, 1967-1995. J. Wildl. Dis. 37: 347-357.

Cooper DN, Krawczak M (1989). Cytosine methylation and the fate of CpG dinucleotides in vertebrate genomes. Hum. Genet. 83: 181-188.

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

Duret L (2002). Evolution of synonymous codon usage in metazoans. Curr. Opin. Genet. Dev. 12: 640-649.

Eden S, Cedar H (1994). Role of DNA methylation in the regulation of transcription. Curr. Opin. Genet. Dev. 4: 255-259.

Fuglsang A (2006). Accounting for background nucleotide composition when measuring codon usage bias: brilliant idea, difficult in practice. Mol. Biol. Evol. 23: 1345-1347.

Ghosh TC, Gupta SK, Majumdar S (2000). Studies on codon usage in Entamoeba histolytica. Int. J. Parasitol. 30: 715-722.

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

Gupta PK, Saini M, Gupta LK, Rao VD, Bandyopadhyay SK, Butchaiah G, Garg GK, Garg SK (2001). Induction of immune responses in cattle with a DNA vaccine encoding glycoprotein C of bovine herpesvirus-1. Vet. Microbiol. 78: 293-305.

Gupta SK, Bhattacharyya TK, Ghosh TC (2004). Synonymous codon usage in Lactococcus lactis: mutational bias versus translational selection. J. Biomol. Struct. Dyn. 21: 527-536.

Hall JD, Gibbs JS, Coen DM, Mount DW (1986). Structural organization and unusual codon usage in the DNA polymerase gene from herpes simplex virus type 1. DNA 5: 281-288.

Hansen WR, Nashold SW, Docherty DE, Brown SE, Knudson DL (2000). Diagnosis of duck plague in waterfowl by polymerase chain reaction. Avian. Dis. 44: 266-274.

Higgs PG, Ran W (2008). Coevolution of codon usage and tRNA genes leads to alternative stable states of biased codon usage. Mol. Biol. Evol. 25: 2279-2291.

Ikemura T (1985). Codon usage and tRNA content in unicellular and multicellular organisms. Mol. Biol. Evol. 2: 13-34.

Jenkins GM, Holmes EC (2003). The extent of codon usage bias in human RNA viruses and its evolutionary origin. Virus. Res. 92: 1-7.

Jia RY, Cheng AC, Wang MS, Xin HY, Guo YF, Zhu DK, Qi XF, Zhao LC, Ge H, Chen XY (2009). Analysis of synonymous codon usage in the UL24 gene of duck enteritis virus. Virus Genes. 38: 96-103.

Jiang P, Sun X, Lu Z (2007). Analysis of synonymous codon usage in Aeropyrum pernix K1 and other Crenarchaeota microorganisms. J. Genet. Genomics. 34: 275-284.

Jones PA, Rideout WM, Shen JC, Spruck CH, Tsai YC (1992). Methylation, mutation and cancer. Bioessays 14: 33-36.

Kahali B, Basak S, Ghosh TC (2007). Reinvestigating the codon and amino acid usage of S. cerevisiae genome: a new insight from protein secondary structure analysis. Biochem. Biophys. Res. Commun. 354: 693-699.

Karlin S, Blaisdell BE, Schachtel GA (1990). Contrasts in codon usage of latent versus productive genes of Epstein-Barr virus: data and hypotheses. J. Virol. 64: 4264-4273.

Lesnik T, Solomovici J, Deana A, Ehrlich R, Reiss C (2000). Ribosome traffic in E. coli and regulation of gene expression. J. Theor. Biol. 202: 175-185.

Li E, Bestor TH, Jaenisch R (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell 69: 915-926.

Lindahl T (1993). Instability and decay of the primary structure of DNA. Nature. 362: 709-715.

Liu Q, Dou S, Ji Z, Xue Q (2005). Synonymous codon usage and gene function are strongly related in Oryza sativa. Biosystems 80: 123-131.

Ma J, Campbell A, Karlin S (2002). Correlations between Shine-Dalgarno sequences and gene features such as predicted expression levels and operon structures. J. Bacteriol. 184: 5733-5745.

Ma XR, Xiao SB, Fang LR, Chen HC (2005). Bias of base composition and codon usage in pseudorabies virus genes. J. Genet. Genomics 32: 616-624.

Marais G, Duret L (2001). Synonymous codon usage, accuracy of translation and gene length in Caenorhabditis elegans. J. Mol. Evol. 52: 275-280.

Meintjes PL, Rodrigo AG (2005). Evolution of relative synonymous codon usage in Human Immunodeficiency Virus type-1. J. Bioinform. Comput. Biol. 3:157-168.

Miyasaka H (2002). Translation initiation AUG context varies with codon usage bias and gene length in Drosophila melanogaster. J. Mol. Evol. 55: 52-64.

Novembre JA (2002). Accounting for background nucleotide composition when measuring codon usage bias. Mol. Biol. Evol. 19: 1390-1394.

Pinto RM, Aragones L, Costafreda MI, Ribes E, Bosch A (2007). Codon usage and replicative strategies of hepatitis A virus. Virus Res. 127: 158-163.

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

Rux AH, Lou H, Lambris JD, Friedman HM, Eisenberg RJ, Cohen GH (2002). Kinetic analysis of glycoprotein C of herpes simplex virus types 1 and 2 binding to heparin, heparan sulfate and complement component C3b. Virology 294: 324-332.

Scaiewicz V, Sabbia V, Piovani R, Musto H (2006). CpG islands are the second main factor shaping codon usage in human genes. Biochem. Biophys. Res. Commun. 343: 1257-1261.

Versteeg R, van Schaik BD, Van Batenburg MF, Roos M, Monajemi R, Caron H, Bussemaker HJ, van Kampen AH (2003). The human transcriptome map reveals extremes in gene density, intron length, GC content and repeat pattern for domains of highly and weakly expressed genes. Genome Res. 13: 1998-2004.

Woo PC, Wong BH, Huang Y, Lau SK, Yuen KY (2007). Cytosine deamination and selection of CpG suppressed clones are the two major independent biological forces that shape codon usage bias in coronaviruses. Virology 369: 431-442.

Wright F (1990). The 'effective number of codons' used in a gene.

Genetics 87: 23-29.
Xu C, Li XR, Xin HY, Lian B, Cheng AC, Wang Ms, Zhu DK, Jia RY, Luo QH, Chen XY (2008). Cloning and molecular characterization of gC gene of duck plague virus. Chin. Vet. Sci. 38: 1038-1044.
Yang XY, Qi XF, Cheng AC, Wang MS, Zhu DK, Jia RY, Chen XY (2009). Intestinal mucosal immune response in ducklings following oral immunisation with an attenuated duck enteritis virus vaccine. Vet. J. 4: 1-5.
Zavala A, Naya H, Romero H, Musto H (2002). Trends in codon and amino acid usage in Thermotoga maritima. J. Mol. Evol. 54: 563-568.
Zhao LC, Cheng AC, Wang MS, Yuan GP, Cai MS (2008). Characterization of codon usage bias in the dUTPase gene of duck enteritis virus. Prog. Nat. Sci. 18: 1069-1076.
Zouridis H, Hatzimanikatis V (2008). Effects of codon distributions and tRNA competition on protein translation. Biophys. J. 95: 1018-1033.