Characterization of Synonymous Codon Usage Bias in the Duck Plague Virus UL35 Gene

Ming-Sheng Cai^a An-Chun Cheng^{a, b} Ming-Shu Wang^{a–c} Li-Chan Zhao^a
De-Kang Zhu^{a, b} Qi-Hui Luo^b Fei Liu^b Xiao-Yue Chen^{a, b}

^aAvian Disease Research Center, College of Veterinary Medicine, Sichuan Agricultural University,

^bKey Laboratory of Animal Diseases and Human Health of Sichuan Province, Yaan, and ^cCollege of Life Science and Technology, Southwest University for Nationalities, Chengdu, China

Key Words

Duck plague virus • UL35 gene • Codon usage bias

Abstract

Objective: The aim was to identify the codon usage bias between the newly identified duck plague virus (DPV) UL35 gene (GenBank accession No. EF643558) and the UL35-like genes of 27 other reference herpesviruses. **Methods:** A comparative analysis of the codon usage bias of the 28 herpesviruses was performed by using the CodonW 1.4 program and CUSP (create a codon usage table) program of EMBOSS (The European Molecular Biology Open Software Suite). **Results:** The results showed obvious differences of the synonymous codon usage bias in the 28 herpesviruses indicated by the Codon Adaptation Index, effective number of codons (ENC), and the value of G + C content at the 3rd codon position. The codon usage pattern of the DPV UL35 gene was phylogenetically conserved and similar to that of the UL35-like genes of the avian α -herpesvirus, with a strong bias towards the codons with A and T at the 3rd codon position. A cluster analysis of codon usage pattern of the DPV UL35 gene with other reference herpesviruses demonstrated that the codon usage bias of the UL35 genes of the 28 herpesviruses had a very close relation with their gene function. The ENC-plot revealed that the genetic heterogeneity in the DPV UL35 gene and the 27 reference herpesviruses were constrained

by G + C content, while gene length exerted relatively weaker influences. In addition, comparisons of the codon preferences in the UL35 gene of DPV with those of *Escherichia coli*, yeast and humans revealed that there were 33 codons showing distinct usage differences between the DPV and yeast, and 38 between the DPV and humans, but only 31 between the DPV and *E. coli*. Therefore, the *E. coli* system may be more suitable for the expression of the DPV UL35 gene. **Conclusion:** Together, these results may improve our understanding of the evolution, pathogenesis and functional studies of DPV, as well as contribute significantly to the area of herpesvirus research and possibly studies with other viruses.

Copyright © 2009 S. Karger AG, Basel

Introduction

Within the standard genetic codes used in a number of different ways, all amino acids except Met and Trp are coded by 2–6 synonymous codons. However, degenerate codons are not present at equal frequencies within and between organisms, a phenomenon termed codon usage bias [1–3]. Codon usage bias among synonymous codons

M.-S.C., A.-C.C. and M.-S.W. contributed equally to this work and should each be considered as a first author.

has been documented for many genes in many species [1, 4–8]. Studies of the synonymous codon usage can reveal information about the molecular evolution of individual genes and provide data to train genome-specific gene recognition algorithms which recognize protein coding regions in uncharacterized genomic DNA. It has been reported that synonymous codon usage biases may be associated with various biological factors, such as gene expression level, gene length, gene translation initiation signal, protein amino acid composition, protein structure, tRNA abundance, mutation frequency and patterns, and GC compositions [9–21]. It also has been reported that the codon usage bias in the genes could be related to the different functions of the encoded proteins, e.g. reports of amino acid compositions of the proteins encoded by the pseudorabies virus genes have indicated that the codon usage bias in the genes could be related to the different functions of the encoded proteins. Further analysis found that the synonymous codon usage pattern varied at different sites along a coding sequence balancing strong versus weak base pair bonding [22, 23], translational efficiency and fidelity [7], and maintenance of DNA and RNA secondary structures [24]. Furthermore, the pattern of codon usage can vary considerably among organisms, and also among genes from the same genome [25].

Codon usage bias can be driven by the neutral processes of mutation, genetic drift and/or biased gene conversion, so the relative abundance of alternative codons might reflect skews in local base composition [26, 27]. Additionally, selection for translational efficiency and/or accuracy can skew codon frequencies toward ‘optimal’ codons [3, 28]. Although it is an ongoing problem to quantify the relative importance of selective and neutral forces as causes of codon usage bias within and between species, selection in codon usage can be inferred from genomic correlations with the relative abundance of alternative tRNA molecules or gene copies, gene expression levels, synonymous substitution rates or skewed levels of polymorphism at synonymous sites [29–32]. Generally, differences in patterns of codon usage between species are principally thought to be due to mutational processes in base composition [33, 34]. Brownian motion models may capture the predominant dynamics in the divergence of genomic base composition [35] and, therefore, may also describe interspecific dynamics of overall codon usage trends. However, intraspecific variation does not fit neutral mutational models as well, suggesting that deviations in the effectiveness of selection among loci is likely to be an important force for shaping

patterns of intragenomic codon usage variation across all domains of life [33]. Recently, analyses of the patterns of codon usage bias of herpesviruses are primarily focused on the pseudorabies virus [20], herpes simplex virus type 1 (HSV-1) [36], Epstein-Barr virus [37], UL24 gene [38] and dUTPase gene [39] of the duck plague virus (DPV).

Duck plague, caused by a DPV of Herpesviridae, is an acute, febrile, contagious and septic disease that occurs frequently in domestic waterfowl such as ducks, geese and swans. It can result in a marked decrease of egg yield and death of commercial waterfowl [40]. Most of the previous research has focused on the epidemiology and prevention of this disease. However, the molecular biological information about the DPV genome is limited. Recently, a DPV CHv strain genomic library was constructed successfully for the first time [41] and the UL35 gene was isolated and identified in our laboratory. However, data in respect to codon usage bias in the newly identified DPV UL35 gene is not yet available. Given this background, it has become important to analyze the codon preference used in the DPV UL35 gene. In this study, we analyzed the synonymous codon usage data of the DPV UL35 gene and compared it with the UL35-like genes of 27 reference herpesviruses. Then, we examined how other factors may affect codon usage variation in the DPV UL35 gene and its reference species. Moreover, we compared the codon usage preference of the DPV UL35 gene with those of *Escherichia coli*, yeast and humans. Taken together, analysis of codon usage patterns of the DPV UL35 gene and a comparison of codon preference between the DPV UL35 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. It may also provide some insights into the features of the DPV genome and improve the understanding of factors shaping codon usage patterns as well as contributing significantly to the area of herpesvirus research or even studies with other viruses.

Materials and Methods

Virus Species and Gene Sequences

The DPV CHv strain, which is a highly virulent field strain of DPV, was obtained from the Key Laboratory of Animal Disease and Human Health of Sichuan Province. The UL35 gene (GenBank accession No. EF643558) of the DPV CHv strain was first identified by constructing a DPV genomic library in our labora-

tory. The nucleotide sequences (table 1) of the UL35 gene and the UL35-like genes of 27 reference herpesviruses were obtained from GenBank (<http://www.ncbi.nlm.nih.gov/>; Bethesda, Md., USA) [42].

Molecular Phylogenetic Tree

To compare with those of the UL35-like proteins of the 27 reference herpesviruses, for which nucleotide sequences are available at GenBank (table 1), the nucleotide sequences of the DPV UL35 gene and its reference herpesviruses were translated into amino acid sequences. After this, multiple sequence alignment and phylogenetic analysis were performed by employing the Clustal V in MegAlign program of DNASTar (Version 7.0; DNASTAR Inc., Madison, Wisc., USA).

Codon Usage Analysis of the DPV UL35 Gene and the Other 27 Reference Herpesviruses

For each gene, codon usage was estimated by using the program CodonW 1.4 (written by John Peden, obtained from <http://codonw.sourceforge.net/>). A number of indices of codon usage bias including the effective number of codons (ENC), relative synonymous codon usage (RSCU), G + C content at the 3rd codon position (GC3s) and the Codon Adaptation Index (CAI) were calculated. ENC is the effective number of codons used in a gene which is used to quantify how far the codon usage of a gene departs from equal usage of synonymous codons, and without dependence on sequence length or specific knowledge of preferred codons, although it is affected by base composition [43–45]. Values of ENC can range between 20 (when only 1 codon is used per amino acid) and 61 (when all synonyms are used with equal frequency). ENC appears to be a good measure of general codon usage bias. The lower the ENC, the higher the codon bias is. RSCU was used to study the overall synonymous codon usage variation among the genes without the confounding influence of the amino acid composition of different gene samples. 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 >1.0 indicate that the corresponding codon is more frequently used than expected, whereas the reverse is true for RSCU values <1.0 [46]. Heat maps of RSCU were constructed with CIMMiner (<http://discover.nci.nih.gov/cimminer>). GC3s is a good indicator of the extent of base composition bias, which represents the frequency of the nucleotide G + C at the synonymous 3rd codon position, excluding Met, Trp and the stop codons. CAI uses a reference set of highly expressed genes from a species to assess the relative merits of each codon, and a score for a gene is calculated from the frequency of use of all codons in that gene. The index assesses the extent to which selection has been effective in molding the pattern of codon usage. In that respect it is useful for predicting the level of expression of a gene. The codon usage pattern across genes was examined by the ENC-plot, which is a plot of ENC versus GC3s and length or GC3s versus length.

Comparison of Codon Preferences of DPV UL35 Gene with those of E. coli, Yeast and Humans

To examine whether different species follow with the same codon usage rule, we compared the codon preferences among the DPV UL35 gene with those of *E. coli*, yeast and humans. The codon usage analysis of these species was carried out by using the codon usage database (<http://www.kazusa.or.jp/codon>) and the

CUSP (create a codon usage table) program of EMBOSS (<http://bioinfo.pbi.nrc.ca:8090/EMBOSS/>; The European Molecular Biology Open Software Suite, Cambridge, UK).

Statistical Methods

The correlations between codon usage variations among the DPV UL35 gene and 27 reference herpesviruses and 4 parameters (CAI, ENC, GC3s and gene length) were analyzed by using SPSS 12.0 software.

Results

Molecular Phylogenetic Tree

A phylogenetic tree based on the deduced amino acid sequences of the DPV UL35 gene and its UL35-like genes in the reference herpesviruses is shown in figure 1. From figure 1 we can see that the general branching pattern coincides with other previously published phylogenetic analyses and that the UL35-like genes within the same herpesvirus subfamily (including α -herpesvirus, β -herpesvirus, γ -herpesvirus) or in the same microorganism are clustered together. Simultaneously, it is shown that DPV first clusters with some avian herpesviruses, such as GaHV-2, GaHV-3 and MeHV-1, in a monophyletic clade and then clusters with other α -herpesvirus members such as EHV-1, BoHV-1, CeHV-1, HHV-1 and SuHV-1. Therefore, it can be concluded from the tree and the identified high amino acid sequence similarity that the DPV UL35 protein has a close evolutionary relationship with other avian α -herpesviruses and at the same time differs from them in certain ways [47].

Codon Usage Analysis of the UL35 Gene in DPV and the Reference Herpesviruses

The results obtained by CodonW analysis of 28 herpesviruses species are shown in table 2. Codon usage in the DPV CHv UL35 gene and its homologous genes is highly nonrandom in all the herpesviruses and the overall base composition of the UL35 gene and its homologous genes in these species also differs dramatically. However, it is interesting that there is no distinction of the codon usage bias parameters of the UL35 gene in both the DPV CHv strain and DPV clone-03 strain. It is shown in table 2 that the CAI values of different herpesviruses vary from 0.109 to 0.298, with a mean value of 0.215 and a standard deviation (SD) of 0.049, and the ENC values range from 30.45 to 61, with a mean value of 47.55 and SD of 10.354. Since most ENC values of the 28 herpesviruses are much higher (ENC >40), the codon usage bias in the herpesvirus UL35-like gene is slightly lower. There is also

Table 1. Nucleotide sequences of the UL35 gene of DPV and 27 reference herpesviruses

Rank	Species	Virus name (abbreviation)	Description	Natural host	GenBank accession No.	Sequence length, bp
1	Un-classified	Duck plague virus CHv strain (DPV CHv)	UL35 gene, its product is a small capsid protein, similar to HSV-1 VP26 protein, herpesvirus UL35 family. UL35 represents a true late gene which encodes a 12-kDa capsid protein.	Duck	EF643558	354
2		Duck plague virus clone 03 strain (DPV clone-03)	UL35 gene, its product is a capsid protein, similar to HSV-1 VP26 protein, herpesvirus UL35 family. UL35 represents a true late gene which encodes a 12-kDa capsid protein.	Duck	EF203708	354
3	α -Herpesvirinae	Bovine herpesvirus 1 (BoHV-1)	UL35 gene, its product is a small capsid protein, possibly capsid transport, located externally on capsid hexons.	Bovine	NC_001847	375
4		Bovine herpesvirus 5 (BoHV-5)	UL35 gene, its product is a small capsid protein, possibly capsid transport, located externally on capsid hexons.	Bovine	NC_005261	378
5		Cercopithecine herpesvirus 1 (CeHV-1)	UL35 gene, its product is a small capsid protein.	Cercopithecine	AF533768	345
6		Cercopithecine herpesvirus 2 (CeHV-2)	UL35 gene, its product is a small capsid protein.	Cercopithecine	AY714813	345
7		Cercopithecine herpesvirus 16 (CeHV-16)	UL35 gene, its product is a small capsid protein VP26, interacts with dynein, possibly capsid transport, located externally on capsid hexons.	Cercopithecine	NC_007653	345
8		Equid herpesvirus 1 (EHV-1)	ORF25, similar to HSV-1 UL35, its product is a capsid protein, located on tips of hexons.	Equid	AY464052	360
9		Equid herpesvirus 4 (EHV-4)	25 gene, its product is a capsid protein, located on tips of hexons, counterpart of HSV-1 gene UL35 and VZV gene 23.	Equid	AF030027	360
10		Gallid herpesvirus 1 (GaHV-1)	UL35 gene, its product is a small capsid protein, possibly capsid transport, located externally on capsid hexons.	Avian	NC_006623	375
11		Gallid herpesvirus 2 (GaHV-2)	UL35 gene, its product is a VP26 capsid protein.	Avian	AY510475	396
12		Gallid herpesvirus 3 (GaHV-3)	UL35 gene, its product is a small capsid protein, possibly capsid transport, located externally on capsid hexons.	Avian	NC_002577	384
13		Human herpesvirus 1 (HHV-1)	UL35 gene, its product is a capsid protein VP26, herpesvirus UL35 family. UL35 represents a true late gene which encodes a 12-kDa capsid protein.	Human	AB252868	339
14		Human herpesvirus 2 (HHV-2)	UL35 gene, its product is a small capsid protein VP26, possibly capsid transport, located externally on capsid hexon.	Human	NC_001798	339
15		Human herpesvirus 3 (HHV-3)	ORF23, its product is a small capsid protein, similar to HHV-1 UL35, located on tips of hexons.	Human	EU154348	708
16		Meleagrid herpesvirus 1 (MeHV-1)	UL35 gene, its product is a capsid protein VP26, herpesvirus UL35 family. UL35 represents a true late gene which encodes a 12-kDa capsid protein.	Meleagrid	AF282130	714
17		Psittacid herpesvirus 1 (PsHV-1)	UL35 gene, its product is VP26.	Psittacid	AY372243	426
18		Suid herpesvirus 1 (SuHV-1)	UL35 gene, its product is a capsid protein, similar to VP26, herpesvirus UL35 family. UL35 represents a true late gene which encodes a 12-kDa capsid protein.	Swine	BK001744	312

Table 1 (continued)

Rank	Species	Virus name (abbreviation)	Description	Natural host	GenBank accession No.	Sequence length, bp
19	β-Herpesvirinae	Cercopithecine herpesvirus 8 (CeHV-8)	Similar to the smallest capsid protein hCMV UL48a/UL48.5, its product is rhUL48a.	Cercopithecine	DQ120516	222
20		Human herpesvirus 5 (HHV-5)	Its product is UL48A capsid protein located at tips of hexons, HSV-1 UL35 counterpart.	Human	NC_006273	228
21		Human herpesvirus 6 (HHV-6)	U33 gene, its product is a capsid protein, old name XILF3, homolog HCMV UL48 (capsid protein), HVS GENE 66, EBV BFRF2, EHV1 25, VZV 23, HSV1 UL35.	Human	X83413	1,413
22		Human herpesvirus 7 (HHV-7)	U32 gene, its product is U32, small capsid protein on hexon tips, herpesvirus core gene UL35 family, similar to HHV-5 UL48A.	Human	NC_001716	273
23	γ-Herpesvirinae	Pongine herpesvirus 4 (PoHV-4)	Its product is capsid protein UL48A, located at tips of hexons, similar to HSV-1 UL35.	Chimpanzee	NC_003521	255
24		Alcelaphine herpesvirus 1 (AIHV-1)	ORF65, its product is a capsid protein, positional similar to EBV BFRF3 and HSV UL35.	Alcelaphine	NC_002531	759
25		Human herpesvirus 4 (HHV-4)	ORF65, its product is BFRF3, small capsid protein VP26 on hexon tips, γ-herpesvirus capsid protein, core gene UL35 family.	Human	AJ507799	531
26		Human herpesvirus 8 (HHV-8)	VAC, its product is a small viral capsid antigen, γ-herpesvirus capsid protein, homolog of HSV VP26, EBV BFRF3 or HVS ORF65.	Human	KSU50141	513
27		Murid herpesvirus 4 (MuHV-4)	65 gene, its product is a capsid protein, located on tips of hexons.	Murine	AF105037	561
28		Ovine herpesvirus 2 (OvHV-2)	ORF65, its product is a capsid protein on hexon tips, similar to HSV UL35.	Ovine	NC_007646	636

a slight variation in codon usage patterns which is the same among different herpesviral UL35-like genes (SD = 10.354). Similarly, the GC3s content of each UL35-like gene also confirms the homogeneity of synonymous codon usage among the different herpesviruses, which varies from 30.77 to 98.26%, with a mean of 65.01% and SD of 21.217%.

It has been reported that a plot of ENc against GC3s can be effectively used to explore the heterogeneity of codon usage among genes [43]. If a particular gene is subject to G + C compositional constraint for shaping the codon usage pattern, it will lie on a continuous curve, which represents random codon usage [48]. If a gene is subject to selection for translationally optimal codons, it will lie considerably below the expected curve. The ENc values of each UL35-like gene in the reference herpesviruses are plotted against their corresponding GC3s in figure 2a. From figure 2a, it can be seen that although a few genes lay on the expected curve, a large number of points lie near the solid curve of this distribution, suggesting that these genes are subject to GC compositional constraints.

The relationship between gene length and synonymous codon usage bias has been reported for *Drosophila melanogaster*, *E. coli*, *Saccharomyces cerevisiae*, *Pseudomonas aeruginosa* and *Yersinia pestis* [49–51]. Here, the plot of gene length against ENc or against GC3s (fig. 2b, c) shows the distribution of each gene. It appears that shorter genes have a much wider variance in ENc values, vice versa for longer genes. We analyzed the relationship between ENc value and gene length, and the relationship between GC3s and gene length of the 28 herpesviruses. However, none of the correlations were statistically significant. Evidently, the gene length may play a minor role in shaping the codon usage bias of the 28 herpesviruses. Similar results were also found in *S. pneumoniae*, *P. aeruginosa* and SARS coronavirus [51–53].

Variation in the DPV UL35 Gene Codon Usage and Amino Acid Composition

While the CAI, ENc and related measures indicate the overall codon bias of the DPV UL35 gene, it is also instructive to look more closely at the pattern of codon bias.

Fig. 1. Evolutionary relationships of the putative DPV CHv UL35 protein with its 27 reference herpesviruses from different species (table 1). Phylogenetic tree of these proteins was generated by using the MegAlign (DNA-STAR Inc.) program with Clustal V multiple alignment and sequence distance indicated by the scale calculated using PAM250 matrix in Lasergene.

Table 3 shows the overall codon preference of the DPV UL35 gene. From the values of RSCU calculated for the DPV UL35 gene, we can see that the amino acids, excluding Met, Trp, and the termination codons in the polypeptide, Arg, Leu, Ser, Ala, Gly, Pro, Thr and Val, have a high level of diversity in codon usage biases because they have a 6-fold and 4-fold coding degeneracy, while the preferred codons of amino acids that have 2-fold or 3-fold coding degeneracy are more uniform. As a whole, although the most and the least used frequencies of codons of all the amino acids are different, the analyzed DPV UL35 gene shows significant preference for one or more postulate codons for each amino acid. Meanwhile, the RSCU values of these codons, whose 3rd positions are A or T, are much more than those codons whose 3rd positions are C or G, indicating that the DPV UL35 gene shows a high bias of codon usage toward the codons with a T and/or A ending rather than a C and/or G ending for all degenerate codons.

Phylogenetic Persistence in Codon Usage Bias of the DPV UL35 Gene

In order to show whether gene function was correlated with the observed variation in codon bias, a cluster analysis of the codon usage pattern based on the DPV CHv UL35 gene and its reference herpesviruses (clustered by the RSCU values, data not shown) was conducted (fig. 3). From figure 3 we can see that DPV CHv and DPV clone-03 are different from other herpesviruses as they first cluster together and form a separate branch, then cluster with the avian herpesviruses of GaHV-3 and GaHV-1 and afterwards cluster with other herpesviruses such as MuHV-4, EHV-4 and HHV-3. This result fully indicates the internal relations of the codon usage pattern between DPV and other herpesviruses, suggesting that the codon usage pattern of DPV has differences with other herpesviruses: the more distant the relationship, the bigger the variation of the codon usage bias, and vice versa. Thus, we can make a conclusion that the codon usage pattern of DPV is fairly close to that of avian herpesviruses and is largely different from other types of herpesviruses, and

Table 2. Summary analysis of the UL35 gene in different herpesvirus species

Rank	Virus name	CAI	ENc	Coding GC, %	GC3s %
1	DPV CHv	0.139	59.88	49.72	43.22
2	DPV clone-03	0.139	59.88	49.72	43.22
3	BoHV-1	0.219	36.20	70.67	86.40
4	BoHV-5	0.210	33.11	73.28	93.65
5	CeHV-1	0.252	32.04	73.04	95.65
6	CeHV-2	0.264	30.45	76.23	96.52
7	CeHV-8	0.298	44.52	54.05	70.27
8	CeHV-16	0.253	30.62	75.65	98.26
9	EHV-1	0.294	48.83	52.50	61.67
10	EHV-4	0.269	48.21	45.83	45.83
11	GaHV-1	0.109	54.87	47.20	41.60
12	GaHV-2	0.179	58.90	44.95	37.12
13	GaHV-3	0.186	50.56	51.82	57.03
14	HHV-1	0.233	46.25	65.78	76.99
15	HHV-2	0.253	40.78	66.96	84.07
16	HHV-3	0.153	50.25	55.51	48.31
17	MeHV-1	0.197	58.18	46.78	45.80
18	PsHV-1	0.214	54.30	62.21	68.31
19	SuHV-1	0.239	32.08	70.19	95.19
20	HHV-5	0.266	32.20	59.21	81.58
21	HHV-6	0.207	57.31	42.04	50.96
22	HHV-7	0.250	61.00	37.36	30.77
23	PoHV-4	0.269	42.91	61.96	85.88
24	AlHV-1	0.189	59.45	55.20	40.32
25	HHV-4	0.190	50.36	63.28	63.84
26	HHV-8	0.179	52.52	58.48	66.67
27	MuHV-4	0.189	50.09	48.66	47.06
28	OvHV-2	0.173	55.56	62.89	64.15

All indices were calculated by CodonW 1.4.

the UL35 gene’s function contributes to the codon usage pattern.

Comparison of UL35 Gene Codon Usage in DPV with those of E. coli, Yeast and Humans

Generally, the codon usage bias in genes remains at a certain level across species. The DPV UL35 gene was compared with those of *E. coli*, yeast and humans to see which would be the suitable host for optimal expression. Data from table 4 reveals that there are 33 codons showing a DPV-to-yeast ratio higher than 2 or lower than 0.50 and 38 codons showing a DPV-to-human ratio higher than 2 or lower than 0.50, but only 31 codons showing a DPV-to-*E. coli* ratio higher than 2 or lower than 0.50, indicating that the codon preferences are significantly different among them. That DPV has the lowest distinct us-

Fig. 2. Relationship between ENc, GC3s and gene length of the UL35 genes in the DPV CHv strain and 27 reference herpesviruses. **a** Plot of ENc vs. GC3s for the UL35 genes in the DPV CHv strain and 27 reference herpesviruses. The solid curve shows the expected position of genes whose codon usage is only determined by the variation in GC3s. **b** Plot of ENc vs. gene length. **c** Plot of GC3s vs. gene length.

Table 3. Codon preferences in the DPV CHv UL35 gene analyzed with the CUSP program

AA	Codon	Fract ^a	Frequency ^b (1/1,000)	Number	RSCU ^c	AA	Codon	Fract ^a	Frequency ^b (1/1,000)	Number	RSCU ^c
R(Arg)	AGA	0.286	33.898	4	1.717	T	ACG	0.000	0.000	0	0.000
R	AGG	0.286	33.898	4	1.717	T	<u>ACT</u>	0.222	16.949	2	0.889
R	<u>CGA</u>	0.143	16.949	2	0.858	V(Val)	<u>GTA</u>	0.600	25.424	3	2.400
R	<u>CGC</u>	0.286	33.898	4	1.717	V	GTC	0.000	0.000	0	0.000
R	CGG	0.000	0.000	0	0.000	V	GTG	0.000	0.000	0	0.000
R	CGT	0.000	0.000	0	0.000	V	GTT	0.400	16.949	2	1.600
L(Leu)	CTA	0.300	25.424	3	1.800	I(Ile)	ATA	0.500	16.949	2	1.500
L	CTC	0.200	16.949	2	1.200	I	ATC	0.500	16.949	2	1.500
L	CTG	0.000	0.000	0	0.000	I	ATT	0.000	0.000	0	0.000
L	<u>CTT</u>	0.100	8.475	1	0.600	N(Asn)	<u>AAC</u>	0.200	8.475	1	0.40
L	<u>TTA</u>	0.100	8.475	1	0.600	N	AAT	0.800	33.898	4	1.600
L	<u>TTG</u>	0.300	25.424	3	1.800	D(Asp)	<u>GAC</u>	0.429	25.424	3	0.857
S(Ser)	<u>AGC</u>	0.077	8.475	1	0.461	D	GAT	0.571	33.898	4	1.143
S	AGT	0.154	16.949	2	0.922	C(Cys)	TGC	0.000	0.000	0	0.000
S	TCA	0.154	16.949	2	0.922	C	<u>TGT</u>	1.000	8.475	1	2.000
S	TCC	0.000	0.000	0	0.000	H(His)	CAC	0.000	0.000	0	0.000
S	<u>TCG</u>	0.308	33.898	4	1.843	H	CAT	1.000	8.475	1	2.000
S	<u>TCT</u>	0.308	33.898	4	1.843	F(Phe)	<u>TTC</u>	0.167	8.475	1	0.333
A(Ala)	<u>GCA</u>	0.214	25.424	3	0.857	F	<u>TTT</u>	0.833	42.373	5	1.667
A	GCC	0.000	0.000	0	0.000	Y(Tyr)	TAC	0.667	16.949	2	1.333
A	<u>GCG</u>	0.429	50.847	6	1.714	Y	<u>TAT</u>	0.333	8.475	1	0.667
A	GCT	0.357	42.373	5	1.429	Q(Gln)	<u>CAA</u>	0.333	8.475	1	0.667
G(Gly)	GGA	0.429	25.424	3	1.714	Q	CAG	0.667	16.949	2	1.333
G	GGC	0.286	16.949	2	1.143	E(Glu)	<u>GAA</u>	0.333	8.475	1	0.667
G	<u>GGG</u>	0.143	8.475	1	0.571	E	GAG	0.667	16.949	2	1.333
G	<u>GGT</u>	0.143	8.475	1	0.571	K(Lys)	AAA	0.000	0.000	0	0.000
P(Pro)	<u>CCA</u>	0.429	25.424	3	1.714	K	AAG	1.000	8.475	1	2.000
P	<u>CCC</u>	0.143	8.475	1	0.571	M(Met)	ATG	1.000	25.424	3	1.000
P	CCG	0.286	16.949	2	1.143	W(Trp)	TGG	1.000	8.475	1	1.000
P	<u>CCT</u>	0.143	8.475	1	0.571	^d	TAA	1.000	8.475	1	3.000
T(Thr)	ACA	0.444	33.898	4	1.778	^d	TAG	0.000	0.000	0	0.000
T	ACC	0.333	25.424	3	1.333	^d	TGA	0.000	0.000	0	0.000

Italicized codons indicate the highest frequency in coding the amino acid. Underlined codons appear during the lower frequency coding of the amino acid. Triplets in bold face indicate the lowest frequency (frequency is zero) in coding the amino acid.

^a Proportion of all synonymous codons encoding the same amino acid.

^b The frequency of each codon that appears in the coding sequence of the individual gene is 1/1,000.

^c The RSCU refers to the relative synonymous codon usage 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, whereas the reverse is true for RSCU values less than 1.0.

^d Refers to stop codon.

age differences with *E. coli* suggests that codon usage of the DPV UL35 gene more closely resembles that of *E. coli* than that of yeast and humans. Thus, to express the DPV UL35 gene efficiently in the yeast or human cell system, codon optimization of the DPV UL35 gene may be required. At the same time, we can speculate that the DPV UL35 gene may be more efficiently expressed in the prokaryotic expression system such as the *E. coli* system.

Discussion

Previous analyses of codon usage have suggested that both a huge interspecific variation and a clear intra-genomic variability, even for functionally homologous genes, and remarkable differences in codon usage exist across species [32, 46, 55]. Codon usage bias is found to be related to different biological factors. However, direc-

Fig. 3. Heat map of RSCU values for 28 reference herpesvirus species (clustered by the RSCU values, data not shown). Each column represents a different codon, with the corresponding amino acid abbreviations and codon identity. Different species are represented in each row (identifiers as in table 1). Cluster is shown to the left based on euclidean distance and average method.

tional mutation pressure on DNA sequences and natural selection affecting gene translation are the 2 major factors that have been widely accepted to account for both interspecific codon usage variation and intragenomic codon usage variability. In *E. coli* and yeast, synonymous codon usage patterns are related to the abundance of isoaccepting tRNAs [7, 56], where highly expressed genes have a strong selective preference for the codons complementary to the most abundant tRNA species, whereas lowly expressed genes display more uniform codon usage patterns largely compatible with the mutational bias in the absence of translational selection [57, 58]. In mammals and birds, the diverse patterns of codon usage may arise from the compositional constraints of the genomes [19, 59, 60].

In our study, a comprehensive analysis of codon usage including ENC, CAI value, GC content and the RSCU values of the DPV CHv strain UL35 gene was performed by using CodonW 1.4 and EMBOSS CUSP programs, and these values were subsequently compared with those of the 27 reference herpesvirus species. The data of synonymous codon usage bias showed certain disparity of each herpesvirus from the different organisms with the result revealing that: (1) the DPV CHv UL35 gene and its 27 reference herpesviruses adopt relatively similar codon usage patterns, although the DPV CHv UL35 gene shows a few differences of codon usage bias with its reference herpesvirus species; and (2) the DPV UL35 gene prefers to use the codons with A and T at the 3rd codon position. Furthermore, the biased trend towards A and T coincides

Table 4. Comparison of codon preferences between the DPV UL35 gene and *E. coli*, yeast and humans

Codon	Amino acid	<i>E. coli</i> (1/1,000)	Yeast (1/1,000)	Human (1/1,000)	DPV (1/1,000)	DPV/ <i>E. coli</i>	DPV/ yeast	DPV/ human
GCA	A(Ala)	20.6	16.1	16.1	25.424	1.23	1.58	1.58
GCC	A	25.5	12.5	28.4	0.000	0.00	0.00	0.00
GCG	A	31.7	6.1	7.5	59.322	1.87	9.72	7.91
GCT	A	15.6	21.1	18.6	33.898	2.17	1.61	1.82
TGC	C(Cys)	6.9	4.7	12.2	0.000	0.00	0.00	0.00
TGT	C	5.5	8.0	10.0	8.475	1.54	1.06	0.85
GAC	D(Asp)	18.6	20.2	25.6	25.424	1.37	1.26	0.99
GAT	D	32.1	37.8	21.9	33.898	1.06	0.90	1.55
GAA	E(Glu)	38.2	48.5	29.0	8.475	0.22	0.17	0.29
GAG	E	17.7	19.1	39.9	16.949	0.96	0.89	0.42
TTC	F(Phe)	16.9	18.2	20.6	8.475	0.50	0.47	0.41
TTT	F	23.2	26.1	17.1	42.373	1.83	1.62	2.48
GGA	G(Gly)	9.0	10.9	16.4	25.424	2.82	2.33	1.55
GGC	G	27.9	9.7	22.5	16.949	0.61	1.75	0.75
GGG	G	11.3	6.0	16.3	8.475	0.75	1.41	0.52
GGT	G	24.4	24.0	10.8	8.475	0.35	0.35	0.78
CAC	H(His)	9.8	7.7	15.0	0.000	0.00	0.00	0.00
CAT	H	13.6	13.7	10.5	8.475	0.62	0.62	0.81
ATA	I(Ile)	5.4	17.8	7.7	16.949	3.14	0.95	2.20
ATC	I	24.2	17.0	21.6	16.949	0.70	1.00	0.78
ATT	I	29.8	30.4	16.1	0.000	0.00	0.00	0.00
AAA	K(Lys)	33.2	42.2	24.1	0.000	0.00	0.00	0.00
AAG	K	10.7	30.7	32.2	8.475	0.79	0.28	0.26
CTA	L(Leu)	4.0	13.3	7.8	25.424	6.36	1.91	3.26
CTC	L	11.0	5.4	19.8	16.949	1.54	3.14	0.86
CTG	L	50.9	10.4	39.8	0.000	0.00	0.00	0.00
CTT	L	11.7	12.1	13.0	8.475	0.72	0.70	0.65
TTA	L	13.9	26.7	7.5	8.475	0.61	0.32	1.13
TTG	L	14.0	27.0	12.6	25.424	1.82	0.94	2.02
ATG	M(Met)	27.0	20.9	22.2	25.424	0.94	1.22	1.15
AAC	N(Asn)	21.4	24.9	19.5	8.475	0.40	0.34	0.43
AAT	N	18.6	36.3	16.7	33.898	1.82	0.93	2.03
CCA	P(Pro)	8.5	18.2	16.7	25.424	2.99	1.40	1.52
CCC	P	5.8	6.8	20.1	8.475	1.46	1.25	0.42
CCG	P	21.8	5.3	6.9	16.949	0.78	3.20	2.46
CCT	P	7.3	13.6	17.3	8.475	1.16	0.62	0.49
CAA	Q(Gln)	15.0	27.5	12.0	8.475	0.57	0.31	0.71
CAG	Q	29.5	12.1	34.1	16.949	0.57	1.40	0.50
AGA	R(Arg)	2.9	21.3	11.5	33.898	11.69	1.59	2.95
AGG	R	1.9	9.2	11.4	33.898	17.84	3.68	2.97
CGA	R	3.9	3.0	6.3	16.949	4.35	5.65	2.69
CGC	R	21.0	2.6	10.7	33.898	1.61	13.04	3.17
CGG	R	6.3	1.7	11.6	0.000	0.00	0.00	0.00
CGT	R	20.3	6.5	4.6	0.000	0.00	0.00	0.00
AGC	S(Ser)	16.0	9.7	19.3	8.475	0.53	0.87	0.44
AGT	S	9.5	14.2	11.9	16.949	1.78	1.19	1.42
TCA	S	7.8	18.8	12.0	16.949	2.17	0.90	1.41
TCC	S	8.9	14.2	11.9	0.000	0.00	0.00	0.00
TCG	S	8.7	8.5	4.4	33.898	3.90	3.99	7.70
TCT	S	8.7	23.5	14.7	33.898	3.90	1.44	2.31
ACA	T(Thr)	8.2	17.8	15.1	33.898	4.13	1.90	2.24
ACC	T	22.8	12.6	19.4	25.424	1.12	2.02	1.31
ACG	T	14.8	7.9	6.1	0.000	0.00	0.00	0.00

Table 4 (continued)

Codon	Amino acid	<i>E. coli</i> (1/1,000)	Yeast (1/1,000)	Human (1/1,000)	DPV (1/1,000)	DPV/ <i>E. coli</i>	DPV/ yeast	DPV/ human
ACT	T	9.1	20.3	13.0	16.949	1.86	0.83	1.30
GTA	V(Val)	11.1	11.8	7.2	25.424	2.29	2.15	3.53
GTC	V	15.1	11.6	14.6	0.000	0.00	0.00	0.00
GTG	V	25.5	10.6	28.4	0.000	0.00	0.00	0.00
GTT	V	18.5	22.0	11.0	16.949	0.92	0.77	1.54
TGG	W(Trp)	15.2	10.3	12.7	8.475	0.56	0.82	0.67
TAC	Y(Tyr)	12.1	14.6	15.5	16.949	1.40	1.16	1.09
TAT	Y	16.5	18.9	12.1	8.475	0.51	0.45	0.70
TAA	^a	2.0	1.0	0.7	8.475	4.24	8.48	12.11
TAG	^a	0.3	0.5	0.6	0.000	0.00	0.00	0.00
TGA	^a	1.1	0.7	1.5	0.000	0.00	0.00	0.00

DPV/*E. coli*, DPV/yeast and DPV/human indicate the ratio of codon usage frequency in DPV to that in *E. coli*, yeast and humans, respectively. A ratio higher than 2 or lower than 0.5 (marked in bold) indicates that the codon preference differs greatly, and vice versa [54].

^a Refers to stop codon.

with high A + T content in the DPV CHv UL35 gene. Since the DPV CHv UL35 gene is an AT-rich gene, it is reasonable that codons ending in A and/or T are predominant in the gene. In order to show the codon usage variation, we used the ENc-plot to analyze the factors affecting codon usage variation among genes. Here, genetic heterogeneity in the DPV and its reference herpesviruses was observed to be constrained by GC content, while the gene length had only a minor impact on the codon choice.

Comparative analysis of UL35 genes in DPV and the reference herpesviruses indicated that synonymous codon usage in these genes are phylogenetically conserved. Data in table 2 show that the UL35 genes in DPV, MeHV-1, GaHV-1, GaHV-2 and GaHV-3, whose natural host is avian, have a stronger correlation than the UL35 genes of herpesviruses with mammalian hosts, such as CeHV-1, EHV-1, BoHV-1 and SuHV-1. This indicates that the UL35 genes of herpesviruses belonging to the same host have a similar sequence length and CAI value. Although the codon usage pattern among different species is a complex phenomenon, it is important to elucidate the underlying mechanisms of codon usage pattern in order to understand the evolution of the species. From the phylogenetic tree and cluster analysis results (fig. 1, 3), we can see that DPV is evolutionarily closer to GaHV-3 than GaHV-2 and MeHV-1. Simultaneously, its codon usage pattern is also closer to GaHV-3 than GaHV-1 and MuHV-4. Therefore, we can draw a conclusion that the species has a certain

influence to the preference of codon usage, but is less important than the influence of gene function, and the codon usage bias of the DPV UL35 gene has a very close relation with its gene function. A previous study has shown that the codon usage pattern of the DPV dUTPase gene is phylogenetically conserved and similar to that of the dUTPase genes of the avian α -herpesvirus [39]. Furthermore, analysis of codon usage bias in the UL24 gene of DPV also indicates that the synonymous codon usage pattern of the UL24 gene in DPV is similar to other avian α -herpesviruses and the UL24 gene's function also contributes to the codon usage [38]. So, we can infer from these studies that the codon usage pattern of DPV is similar to that of avian α -herpesvirus; however, the conclusive evidence may require further analysis of other DPV genes, such as viral genes for maintenance or replication.

With respect to the functions of the UL35 gene product (VP26) in the herpesvirus life cycle, it is well documented in studies on the HSV UL35 protein of the subfamily α -herpesvirus that the UL35 gene, acting as a true late γ 2 gene, encodes for a small nucleocapsid protein and its encoding protein, located on the hexon of the nucleocapsid, and can interact with viral DNA of nucleocapsid. It is nonessential for the formation of viral capsids and viral propagation in cell culture. However, it is essential for viral replication production in the nervous system, the absence of which will result in a 100-fold yield reduction of the infected-virus in trigeminal ganglion [61–63]. It is located at the exterior of the icosahedral capsid, in-

dicating that VP26 might serve as a core for connecting the capsid with the external tegument and envelope in the late stage of viral assembly [62]. Its exposed sites outside the capsid also reveal that VP26 is in a favorable position, coupling the capsid and its functional ligand, such as tegument [64]. Similarly, research on the homolog of VP26 also points out that the smallest capsid protein, which is located at the most exterior region of the capsid, except for interacting with viral DNA, may relate to tegumentation and/or uncoating in the course of infection and plays an important role in regulating the interaction between the capsid and tegument as well as cytoskeletal proteins [65]. Recent studies have also revealed that VP26 of HSV-1 correlates with ribosome and may regulate the translation of the host protein [66]. Therefore, because of the important role played by HSV VP26 in the process of infection, VP26 of DPV may also play a similar role to that of HSV in the course of infection. However, further studies are required to confirm this hypothesis.

Among the codon usage bias patterns in *E. coli*, yeast and humans, the codon usage bias pattern in the DPV UL35 gene is similar to that of *E. coli* (table 4), hence the assumption was made in the present study that the DPV UL35 gene may express more efficiently in the *E. coli* system. This may serve as a guide for manipulating the ex-

pression of the targeted gene. The optimization of the DPV UL35 gene with host-preferred codons is likely to improve the expression level of the DPV UL35 gene in a given host. Therefore, the *E. coli* expression system may be better applied to the production of the DPV UL35 protein. Our hypothesis was verified in our recent study, in which the DPV UL35 protein was successfully expressed in the *E. coli* expression system [67]. In summary, our work has provided a basic understanding of the evolution and pathogenesis of the DPV UL35 gene and offered some new insights into the mechanisms for codon usage bias, as well as contributing significantly to the area of herpesvirus research and possibly studies with other viruses.

Acknowledgements

Research was supported by grants from the National Natural Science Foundation of China (30771598), Changjiang Scholars and Innovative Research Team in University (PCSIRT0848), the earmarked fund for Modern Agro-Industry Technology Research System (2009-2013), New Century Excellent Talents Program in University (NCET-06-0818), Scientific and Technological Innovation Major Project Funds in University (706050), Cultivation Fund of the Key Scientific and Technical Innovation Project, Department of Education of Sichuan Province (07ZZ028) and Sichuan Province Basic Research Program (07JY029-016/17).

References

- 1 Grantham R, Gautier C, Gouy M, Mercier R, Pavé A: Codon catalog usage and the genome hypothesis. *Nucleic Acids Res* 1980;8:R49–R62.
- 2 Sharp PM, Averof M, Lloyd AT, Matassi G, Peden JF: DNA sequence evolution: the sounds of silence. *Philos Trans R Soc Lond B Biol Sci* 1995;349:241–247.
- 3 Duret L: Evolution of synonymous codon usage in metazoans. *Curr Opin Genet Dev* 2002;12:640–649.
- 4 Maruyama T, Gojobori T, Aota S, Ikemura T: Codon usage tabulated from the GenBank genetic sequence data. *Nucleic Acids Res* 1986;14:r151–r197.
- 5 Wain-Hobson S, Nussinov R, Brown RJ, Sussman JL: Preferential codon usage in genes. *Gene* 1981;13:355–364.
- 6 Blake RD, Hinds PW: Analysis of the codon bias in *E. coli* sequences. *J Biomol Struct Dynam* 1984;2:593–606.
- 7 Ikemura T: Codon usage and tRNA content in unicellular and multicellular organisms. *Mol Biol Evol* 1985;2:13–34.
- 8 Grantham R, Gautier C, Gouy M, Jacobzone M, Mercier R: Codon catalog usage is a genome strategy modulated for gene expressivity. *Nucleic Acids Res* 1981;9:r43–r74.
- 9 Bernardi G: Compositional constraints and genome evolution. *J Mol Evol* 1986;24:1–11.
- 10 Gu W, Zhou T, Ma J, Sun X, Lu Z: The relationship between synonymous codon usage and protein structure in *Escherichia coli* and *Homo sapiens*. *Biosystems* 2004;73:89–97.
- 11 Ikemura T: Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the *E. coli* translational system. *J Mol Biol* 1981;151:389–409.
- 12 Lobry JR, Gautier C: Hydrophobicity, expressivity and aromaticity are the major trends of amino-acid usage in 999 *Escherichia coli* chromosome-encoded genes. *Nucleic Acids Res* 1994;22:3174–3180.
- 13 Ma J, Campbell A, Karlin S: Correlations between Shine-Dalgarno sequences and gene features such as predicted expression levels and operon structures. *J Bacteriol* 2002;184:5733–5745.
- 14 Wan XF, Xu D, Zhou J: Intelligent Engineering Systems Through Artificial Neural Networks. New York, ASME Press, 2003, pp 1101–1118.
- 15 Wan XF, Xu D, Kleinhofs A, Zhou J: Quantitative relationship between synonymous codon usage bias and GC composition across unicellular genomes. *BMC Evol Biol* 2004;4:19.
- 16 Bains W: Codon distribution in vertebrate genes may be used to predict gene length. *J Mol Biol* 1987;197:379–388.
- 17 Gouy M, Gautier C: Codon usage in bacteria: correlation with gene expressivity. *Nucleic Acids Res* 1982;10:7055–7074.
- 18 D'Onofrio G, Ghosh TC, Bernardi G: The base composition of the genes is correlated with the secondary structures of the encoded proteins. *Gene* 2002;300:179–187.
- 19 Karlin S, Mrazek J: What drives codon choices in human genes? *J Mol Biol* 1996;262:459–472.
- 20 Ma XR, Xiao SB, Fang LR, Chen HC: Bias of base composition and codon usage in pseudorabies virus genes. *J Genet Genomics* 2005;32:616–624.
- 21 Hooper SD, Berg OG: Gradients in nucleotide and codon usage along *Escherichia coli* genes. *Nucleic Acids Res* 2000;28:3517–3523.

- 22 Blaisdeli BE: Choice of base at silent codon site 3 is not selectively neutral in eucaryotic structural genes: it maintains excess short runs of weak and strong hydrogen bonding bases. *J Mol Evol* 1983;19:226–236.
- 23 Grosjean H, Fiers W: Preferential codon usage in prokaryotic genes: the optimal codon-anticodon interaction energy and the selective codon usage in efficiently expressed genes. *Gene* 1982;18:199–209.
- 24 Wada A, Sujama A: Local stability of DNA and RNA secondary structure and its relation to biological functions. *Prog Biophys Mol Biol* 1986;47:113–157.
- 25 Grocock RJ, Sharp PM: Synonymous codon usage in *Cryptosporidium parvum* identification of two distinct trends among genes. *Int J Parasitol* 2001;31:402–412.
- 26 Sueoka N: Directional mutation pressure and neutral molecular evolution. *Proc Natl Acad Sci USA* 1988;85:2653–2657.
- 27 Marais G: Biased gene conversion: implications for genome and sex evolution. *Trends Genet* 2003;19:330–338.
- 28 Ikemura T: Correlation between the abundance of yeast transfer RNAs and the occurrence of the respective codons in protein genes. *J Mol Biol* 1982;158:573–597.
- 29 Akashi H: Inferring weak selection from patterns of polymorphism and divergence at silent sites in *Drosophila* DNA. *Genetics* 1995;139:1067–1076.
- 30 Duret L, Mouchiroud D: Expression pattern and, surprisingly, gene length shape codon usage in *Caenorhabditis*, *Drosophila*, *Arabidopsis*. *Proc Natl Acad Sci USA* 1999;96:4482–4487.
- 31 Bennetzen JL, Hall BD: Codon selection in yeast. *J Biol Chem* 1982;257:3026–3031.
- 32 Sharp PM, Li WH: The rate of synonymous substitution in enterobacterial genes is inversely related to codon usage bias. *Mol Biol Evol* 1987;4:222–230.
- 33 Knight RD, Freeland SJ, Landweber LF: A simple model based on mutation and selection explains trends in codon and amino-acid usage and GC composition within and across genomes. *Genome Biol* 2001;2:r0010.1–r0010.13.
- 34 Chen SL, Lee W, Hottes AK, Shapiro L, McAdams HH: Codon usage between genomes is constrained by genome-wide mutational processes. *Proc Natl Acad Sci USA* 2004;101:3480–3485.
- 35 Haywood-Farmer E, Otto SP: The evolution of genomic base composition in bacteria. *Evolution* 2003;57:1783–1792.
- 36 Hall JD, Gibbs JS, Coen DM, Mount DW: Structural organization and unusual codon usage in the DNA polymerase gene from herpes simplex virus type 1. *DNA* 1986;5:281–288.
- 37 Karlin S, Blaisdel BE, Schachtel GA: Contrasts in codon usage of latent versus productive genes of Epstein-Barr virus: data and hypotheses. *J Virol* 1990;64:4264–4273.
- 38 Jia R, Cheng A, Wang M, Xin H, Guo Y, Zhu D, Qi X, Zhao L, Ge H, Chen X: Analysis of synonymous codon usage in the UL24 gene of duck enteritis virus. *Virus Genes* 2009;38:96–103.
- 39 Zhao LC, Cheng AC, Wang MS, Yuan GP, Cai MS: Characterization of codon usage bias in the dUTPase gene of duck enteritis virus. *Prog Nat Sci* 2008;18:1069–1076.
- 40 Sandhu TS, Shawky SA: Duck virus enteritis (duck plague); in Saif YM, Barnes HJ, Glisson JR (eds): *Diseases of Poultry*, ed 11. Ames, Iowa State University Press, 2003, pp 354–361.
- 41 Cheng AC, Wang MS, Wen M, Zhou WG, Guo YF, Jia RY, Xu C, Yuan GP, Liu YC: Construction of duck enteritis virus gene libraries and discovery, cloning and identification of viral nucleocapsid protein gene. *High Technol Lett* 2006;16:948–953.
- 42 Burland TG: DNASTAR's Lasergene sequence analysis software. *Methods Mol Biol* 2000;132:71–91.
- 43 Wright F: The 'effective number of codons' used in a gene. *Gene* 1990;87:23–29.
- 44 Comerón JM, Aguade M: An evaluation of measures of synonymous codon usage bias. *J Mol Evol* 1998;47:268–274.
- 45 Novembre JA: Accounting for background nucleotide composition when measuring codon usage bias. *Mol Biol Evol* 2002;19:1390–1394.
- 46 Sharp PM, Li WH: The codon Adaptation Index – a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res* 1987;15:1281–1295.
- 47 Weinstein JN, Myers TG, O'Connor PM, Friend SH, Fornace AJ, Kohn KW, Fojo T, Bates SE, Rubinstein LV, Anderson NL, Buolamwini JK, van Ossdl WW, Monks AP, Scudiero DA, Sausville EA, Zaharevitz DW, Bunow B, Viswanadhan VN, Johnson GS, Wittes RE, Paull KD: An information-intensive approach to the molecular pharmacology of cancer. *Science* 1997;275:343–349.
- 48 Jiang P, Sun X, Lu Z: Analysis of synonymous codon usage in *Aeropyrum pernix* K1 and other *Crenarchaeota* microorganisms. *J Genet Genomics* 2007;34:275–284.
- 49 Moriyama EN, Powell JR: Gene length and codon usage bias in *Drosophila melanogaster*, *Saccharomyces cerevisiae* and *Escherichia coli*. *Nucleic Acids Res* 1998;26:3188–3193.
- 50 Hou ZC, Yang N: Factors affecting codon usage in *Yersinia pestis*. *Acta Biochim Biophys Sin* 2003;35:580–586.
- 51 Gupta SK, Ghosh TC: Gene expressivity is the main factor in dictating the codon usage variation among the genes in *Pseudomonas aeruginosa*. *Gene* 2001;273:63–70.
- 52 Hou ZC, Yang N: Analysis of factors shaping *S. pneumoniae* codon usage. *J Genet Genomics* 2002;29:747–752.
- 53 Gu W, Zhou T, Ma J, Sun X, Lu Z: Analysis of synonymous codon usage in SARS Coronavirus and other viruses in the Nidovirales. *Virus Res* 2004;101:155–161.
- 54 Lu H, Zhao WM, Zheng Y, Wang H, Qi M, Yu XP: Analysis of synonymous codon usage bias in *Chlamydia*. *Acta Biochim Biophys Sin (Shanghai)* 2005;37:1–10.
- 55 Karlin S, Ladunga I, Blaisdel BE: Heterogeneity of genomes: measures and values. *Proc Natl Acad Sci USA* 1994;91:12837–12841.
- 56 Hoekema A, Kastelein RA, Vasser M, de Boer HA: Codon replacement in the PGK1 gene of *Saccharomyces cerevisiae*: experimental approach to study the role of biased codon usage in gene expression. *Mol Cell Biol* 1987;7:2914–2924.
- 57 Lesnik T, Solomovici J, Deana A, Ehrlich R, Reiss C: Ribosome traffic in *E. coli* and regulation of gene expression. *J Theor Biol* 2000;202:175–185.
- 58 Sharp PM, Tuohy TM, Mosurski KR: Codon usage in yeast: cluster analysis clearly differentiates highly and lowly expressed genes. *Nucleic Acids Res* 1986;14:5125–5143.
- 59 Ghosh TC, Gupta SK, Majumdar S: Studies on codon usage in *Entamoeba histolytica*. *Int J Parasitol* 2000;30:715–722.
- 60 Romero H, Zavala A, Musto H, Bernardi G: The influence of translational selection on codon usage in fishes from the family Cyprinidae. *Gene* 2003;317:141–147.
- 61 McNabb DS, Courtney RJ: Identification and characterization of the herpes simplex virus type 1 virion protein encoded by the UL35 open reading frame. *J Virol* 1992;66:2653–2663.
- 62 Booy FP, Trus BL, Newcomb WW, Brown JC, Conway JF, Steven AC: Finding a needle in a haystack: detection of a small protein (the 12-kDa VP26) in a large complex (the 200-MDa capsid of herpes simplex virus). *Proc Natl Acad Sci USA* 1994;91:5652–5656.
- 63 Desai P, DeLuca NA, Person S: Herpes simplex virus type 1 VP26 is not essential for replication in cell culture but influences production of infectious virus in the nervous system of infected mice. *Virology* 1998;247:115–124.
- 64 Booy FP, Trus BL, Davison AJ, Steven AC: The capsid architecture of channel catfish virus, an evolutionarily distant herpesvirus, is largely conserved in the absence of discernible sequence homology with herpes simplex virus. *Virology* 1996;215:134–141.
- 65 Lo P, Yu X, Atanasov I, Chandran B, Zhou ZH: Three-dimensional localization of pORF65 in Kaposi's sarcoma-associated herpesvirus capsid. *J Virol* 2003;77:4291–4297.
- 66 Greco A, Bienvenut W, Sanchez JC, Kindbeiter K, Hochstrasser D, Madjar JJ, Diaz JJ: Identification of ribosomal viral and cellular basic proteins during the course of infection with herpes simplex virus type 1. *Proteomics* 2001;1:545–549.
- 67 Cai MS, Cheng AC, Wang MS, Zhao LC, Zhu DK, Luo QH, Liu F, Chen XY: His6-tagged UL35 protein of duck plague virus expression, purification, and production of polyclonal antibody. *Intervirology* 2009;52:141–151.