Indian Journal of Biochemistry & Biophysics
Vol. 53, June-August 2016, pp. 81-94

Characterization of synonymous codon usage bias in pseudorabies virus *EP0* gene

Zongmin Liao[2#], Tao Chen[2#], Ping Wang[2#], Yuanfang Wang[2], Xingmei Zou[2], Zuo Xu[2], Jingying Mai[2], Jinlu Huang[5], Gengde Hong[6], Yao Wang[3], Yun Bian[3], Haifan Li[4], Qiusan Chen[6], Kenie Wang[1], Delong Liu[3], Hao Peng[3], Ruiyi Luo[3], Shuxuan Deng[7], Mingsheng Cai[2]*, Daixiong Chen[2]* & Meili Li[1,2]*

[1]Guangdong Provincial Key Laboratory of Allergy and Clinical Immunology, Second Affiliated Hospital of Guangzhou Medical University, No.250 Changgang Dong Road, Haizhu District, Guangzhou-510 260, Guangdong, China

[2]Department of Pathogenic Biology and Immunology, Sino-French Hoffmann Institute, School of Basic Medical Science; [3]GMU-GIBH Joint School of Life Sciences; [4]School of Public Health, Guangzhou Medical University, Xinzao Town, Panyu, Guangzhou-511 436, Guangdong, China

[5]Guangdong Haid Group Co. Ltd., Guangzhou-511 400, Guandong Province, China

[6]The Third Clinical School of Guangzhou Medical University, No. 63 Duobao Road, Liwan District, Guangzhou-510 150, Guangdong, China

[7]Agricultural Technology Service Center of Dalingshan Town, Dongguan-523 830, P.R. China

Received 20 October 2014; *revised 19 May 2016*

Codon usage bias among synonymous codons is not an uncommon phenomenon and it is known to involve various biological factors, such as GC compositions, gene length, mutation frequency and patterns, gene expression level, etc. Knowledge on synonymous codon usage may help in understanding the molecular evolution of the individual gene better. In the present study, we examined the codon usage bias between pseudorabies virus (PRV) *EP0* gene and the *EP0*-like genes of 24 reference alpha herpesviruses. Comparative analysis showed noticeable disparities of the synonymous codon usage bias in the 25 alpha herpesviruses, indicated by codon adaptation index, an effective number of codons (ENc) and GC3s value. The codon usage pattern of PRV *EP0* gene was phylogenetically conserved and similar to that of the *EP0*-like genes of the genus *Mardivirus* of alphaherpesvirus, with a strong bias towards the codons with C and G at the third codon position. Cluster analysis of codon usage pattern of PRV *EP0* gene with its reference alpha herpesviruses demonstrated that the codon usage bias of *EP0*-like genes of 25 alpha herpesviruses had close relation with their gene functions. ENc-plot revealed that the genetic heterogeneity in PRV *EP0* gene and the 24 reference alpha herpesviruses was constrained by G+C content, but not gene length. In addition, comparison of codon preferences in the *EP0* gene of PRV with those of *E. coli*, yeast and human revealed that there were 40 codons showing distinct usage differences between PRV and yeast, 27 between PRV and *E. coli*, but only 22 between PRV and human. Therefore, the human expression system may be more suitable for expression of PRV *EP0* gene. In conclusion, these results may improve our understanding of the evolution, pathogenesis and functional studies of PRV.

Keywords: Alpha herpesvirus, Codon usage bias, *EP0* gene, Pseudorabies virus

Within the standard genetic codes utilized in diverse ways, all amino acids (aa) are coded by two to six synonymous codons, except Met and Trp. However, degenerate codons are not used at equal frequencies within an organism a phenomenon called codon usage bias[1-3]. Codon usage bias among synonymous codons has been described for many genes in various species[3-10]. Researches of the synonymous codon usage can uncover knowledge concerning the molecular evolution of individual genes. It is reported that synonymous codon usage bias is associated with various biological factors, such as GC compositions, gene length, mutation frequency and patterns, gene expression level, tRNA abundance, gene translation initiation signal and protein structure[6,10-13]. Further analysis discovered that synonymous codon usage pattern varied at different sites along a coding sequence[11], balances of strong versus weak base pair bonding[12,13], maintenance of DNA and RNA secondary structure[14], and translational efficiency and fidelity[6].

*Correspondence:
Telefax : +86 20 37103216
E-mail: caimingsheng@163.com (MC), daixiongchen@hotmail.com (DC), meili_2011@hotmail.com (ML)
#These authors contributed equally

Aujeszky's disease, which is revoked by the causative agent of pseudorabies virus (PRV) also known as suid herpesvirus 1, SuHV-1, is a frequently fatal disease with a global distribution that affects swine primarily and other domestic wild animals incidentally[15-20]. PRV belongs to the genus *Varicellovirus*, subfamily *Alpha herpesvirinae*, which is a swine alpha herpesvirus. Most of the previous works have focused on the epidemiology and prevention of this disease[17,20-24]. However, specific molecular biological characteristics about the PRV genome are still not well understood.

PRV *EP0* gene, a 1104-base pair sequence encodes a putative polypeptide of 367 aa residues designated EP0. Concerning the function of *EP0* gene product (contains the conserved C3HC4 RING-finger domain in the amino-terminal region) in the herpesvirus life cycle, herpes simplex virus 1 (HSV-1) ICP0[25], varicella-zoster virus (VZV) ORF61[26], BoHV-1 p135 (BICP0)[27] and EHV-1 gene 63 (ORF63)[28], the homologue of EP0, have been extensively studied. However, the functional characteristics of PRV *EP0* gene as well as its codon usage bias is poorly understood. In this study, we analyzed the synonymous codon usage data of PRV *EP0* gene and compared it with the *EP0*-like genes of 24 reference alpha herpesviruses. Further, we investigated how other factors may impact codon usage variation in the PRV *EP0* gene and its reference species. Moreover, we compared the codon usage preference of PRV *EP0* gene with those of *E. coli*, yeast and human.

Materials and Methods

Virus species and gene sequences

The nucleotide sequences (Table 1) of PRV Becker strain *EP0* gene (GenBank accession no. JF797219) and the *EP0*-like genes of 24 reference alpha herpesviruses were obtained from the GenBank (Bethesda, Maryland, USA; http://www.ncbi.nlm.nih.gov/).

Molecular phylogenetic tree of EP0-like proteins of the 25 reference alpha herpesviruses

To compare with those of EP0-like proteins of the 25 reference alpha herpesviruses, for which nucleotide sequences are available in GenBank (listed in Table 1), the nucleotide sequences of PRV *EP0* gene and its reference alpha herpesviruses were translated into aa sequence, then multiple sequence alignment and phylogenetic analysis (rooted tree) were performed by employing the Jotun Hein in MegAlign program of DNAStar (version 7.0, DNAStar, Inc.)[29].

Codon usage analysis of the PRV Becker strain *EP0* gene and other 24 reference alpha herpesviruses

For each gene, codon usage was assessed by using CAI, CHIPS and CUPS programs of EMBOSS (The European Molecular Biology Open Software Suite, http://emboss.bioinformatics.nl/). Some indices of codon usage bias including CAI (codon adaptation index), ENc (effective number of codons), GC3s (G+C content at the third positions of codons) and RSCU (relative synonymous codon usage) were calculated. CAI uses a reference set of highly expressed genes from a species to estimate the relative virtues of each codon (a full gene list is available at http://helixweb.nih.gov/emboss/html/cai.htm), and a score for a gene is calculated from the frequency of use of all codons in that gene. The index assesses the level to which selection has been effective in shaping codon usage[30]. ENc is the effective number of codons used in a gene and can be used to quantify how far the codon usage of a gene deviates from the equal usage of synonymous codons without reliance on sequence length or a given knowledge of preferred codons, although it is affected by base composition[31-33]. Values of ENc can range from 20 (when only one codon is used per aa) to 61 (when all synonyms are used with equal frequency). Thus, ENc can be a useful measure of general codon usage bias. The lower the ENc, the higher the codon bias. GC3s is a useful parameter of the degree of base composition bias, and represents the frequency of the nucleotide G+C at the synonymous third position of codons, excluding Met, Trp and the stop codons. The relative synonymous codon usage (RSCU) was employed to investigate the overall synonymous codon usage variation among the genes without the confounding influence of the aa 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 aa are used equally. An RSCU value greater than 1.0 indicates that the corresponding codon is more frequently used than expected, whereas the reverse is true for a RSCU value less than 1.0[30]. A heat map to represent the clustering of RSCU values was constructed with the CIMMiner software tool (http://discover.nci.nih.gov/cimminer)[34] with each row representing a specific codon and each column

Table 1 — Nucleotide sequences of the PRV Becker strain *EP0* gene and the *EP0*-like genes of 24 reference alpha herpesviruses from different species (*contd.*)

Rank	Virus name (Abbreviation)	Genus	Strain	Description	Natural host	GeneBank accession no.	Sequence length, bp
1	Suid herpesvirus 1(SuHV-1) Pseudorabies virus (PRV)	Varicellovirus	Becker	EP0 gene (early protein 0), its product is ICP0, a ubiquitin E3 ligase (functions in protein degradation and gene regulation), which contains RING-finger (Really Interesting New Gene) domain, a specialized type of Zn-finger of 40 to 60 residues that binds two atoms of zinc, defined by the 'cross-brace' motif C-X2-C-X(9-39)-C-X(1-3)-H-X(2-3)-(N/C/H)-X2-C-X(4-48) C-X2-C.	Sus scrofa (Pig)	JF797219	1104
2			Bartha			JF797217	1104
3			Kaplan			JQ809328	1233
4			Fa			EU333164	1230
5			PRV-FZ			FJ477294	1230
6			Ea			AF298586	1230
7			Indiana-Funkhauser			M57504	1233
8	Equid herpesvirus 1(EHV-1) Equine abortion virus (EAV)		T-529	ORF63 gene, its product is ICP0, a ubiquitin E3 ligase (functions in proteasome-dependent degradation of several cellular proteins, disrupting ND10, gene regulation and latency), which contains RING-finger (Really Interesting New Gene) domain, a specialized type of Zn-finger of 40 to 60 residues that binds two atoms of zinc, defined by the 'cross-brace' motif C-X2-C-X(9-39)-C-X(1-3)-H-X(2-3)-(N/C/H)-X2-C-X(4-48) C-X2-C.	Equus caballus (Horse)	AB194266	1596
9	Equine herpesvirus 4 (EHV-4) Equine rhinopneumonitis virus (ERV)		NS80567			NC_001844	1611
10	Equid herpesvirus 8 (EHV-8)		wh			NC_017826	1623
11	Equid herpesvirus 9 (EHV-9) Gazelle herpesvirus 1 (GHV-1)		P19			NC_011644	1602
12	Cercopithecine herpesvirus 9 (CeHV-9) Simian varicella virus (SVV)		Delta	ORF61 gene, its product is ICP0, a ubiquitin E3 ligase (functions in proteasome-dependent degradation of several cellular proteins, disrupting ND10, gene regulation and latency), which contains RING-finger (Really Interesting New Gene) domain, a specialized type of Zn-finger of 40 to 60 residues that binds two atoms of zinc, defined by the 'cross-brace' motif C-X2-C-X(9-39)-C-X(1-3)-H-X(2-3)-(N/C/H)-X2-C-X(4-48) C-X2-C.	Erythrocebus patas (Monkey)	NC_002686	1512
13	Felid herpesvirus 1 (FeHV-1)		C-27	ICP0 gene, its product is a ubiquitin E3 ligase, which contains RING-finger (Really Interesting New Gene) domain, a specialized type of Zn-finger of 40 to 60 residues that binds two atoms of zinc, defined by the 'cross-brace' motif C-X2-C-X(9-39)-C-X(1-3)-H-X(2-3)-(N/C/H)-X2-C-X(4-48) C-X2-C.	Felidae (Cat)	NC_013590	1497
14	Bovine herpesvirus 1 (BoHV-1) Infectious bovine rhinotracheitis virus (IBRV)		Composite of 5 strains	BICP0 gene, its product is ICP0, a ubiquitin E3 ligase (functions in proteasome-dependent degradation of several cellular proteins, disrupting ND10, gene regulation and latency), which contains RING-finger (Really Interesting New Gene) domain, a specialized type of Zn-finger of 40 to 60 residues that binds two atoms of zinc, defined by the 'cross-brace' motif C-X2-C-X(9-39)-C-X(1-3)-H-X(2-3)-(N/C/H)-X2-C-X(4-48) C-X2-C.	Bos taurus (Cattle)	NC_001847	2031
15	Bovine herpesvirus 5 (BoHV-5) Bovine encephalitis herpesvirus (BEHV)		SV507/99			NC_005261	2163

(*contd.*)

Table 1 — Nucleotide sequences of the PRV Becker strain *EP0* gene and the *EP0*-like genes of 24 reference alpha herpesviruses from different species

Rank	Virus name (Abbreviation)	Genus	Strain	Description	Natural host	GeneBank accession no.	Sequence length, bp
16	Canid herpesvirus 1 (CaHV-1)			CICP0 gene, its product is infected cell protein 0, which contains RING-finger (Really Interesting New Gene) domain, a specialized type of Zn-finger of 40 to 60 residues that binds two atoms of zinc, defined by the 'cross-brace' motif C-X2-C-X(9-39)-C-X(1-3)-H-X(2-3)-(N/C/H)-X2-C-X(4-48) C-X2-C.	Greyhound (Dog)	AB042275	1002
17	Human herpesvirus 3 (HHV-3) Varicella-zoster virus (VZV)		11	ORF61 gene (similar to HHV-1 RL2), its product is a ring-finger protein (functions in modulating cell state and gene expression), which contains RING-finger (Really Interesting New Gene) domain, a specialized type of Zn-finger of 40 to 60 residues that binds two atoms of zinc, defined by the 'cross-brace' motif C-X2-C-X(9-39)-C-X(1-3)-H-X(2-3)-(N/C/H)-X2-C-X(4-48) C-X2-C.	Homo sapiens (Human)	DQ479955	1404
18	Leporid herpesvirus 4 (LHV-4)		LHV4012612	RL2 gene, its product is ICP0, a ubiquitin E3 ligase (functions in proteasome-dependent degradation of several cellular proteins, disrupting ND10, gene regulation and latency), which contains RING-finger (Really Interesting New Gene) domain, a specialized type of Zn-finger of 40 to 60 residues that binds two atoms of zinc, defined by the 'cross-brace' motif C-X2-C-X(9-39)-C-X(1-3)-H-X(2-3)-(N/C/H)-X2-C-X(4-48) C-X2-C.	Bunny (Rabbit)	JQ596859	1710
19	Saimiriine herpesvirus 1 (SaHV-1) Marmoset herpesvirus (MarHV)		MV 5-4		Saimiri (Squirrel monkeys)	NC_014567	2190
20	Human herpesvirus 1 (HHV-1) Herpes simplex virus 1 (HSV-1)		OD4		Homo sapiens (Human)	JN420342	2328
21	Human herpesvirus 2 (HHV-2) Herpes simplex virus 2 (HSV-2)		HG52			NC_001798	2475
22	Cercopithecine herpesvirus 1 (CeHV-1) Macacine herpesvirus 1 (McHV-1) Monkey B virus	Simplexvirus	E2490		Macaca mulatta (Monkey)	NC_004812	2076
23	Cercopithecine herpesvirus 2 (CeHV-2) Simian agent 8 (SA8)		B264		Cercopithecus aethiops (Monkey)	NC_006560	2130
24	Cercopithecine herpesvirus 16 (CeHV-16) Papiine herpesvirus 2 (PaHV-1) Herpesvirus papio 2 (HVP-2)		X313		Papio cynocephalus (Baboons)	NC_007653	2142
25	Gallid herpesvirus 2 (GaHV-2) Marek's disease virus type 1 (MDV-1)	Mardivirus	Md11	RLORF1 gene, its product is ICP0	Gallus domesticus (Chicken)	AY510475	597

representing a different species. Clustering was performed based on Euclidean distance and the average linkage method. The codon usage pattern across different genes was also analyzed by the ENc-plot, which is a plot of ENc versus GC3s and length or GC3s versus length. Curves were generated using a logarithmic distribution curve where $y = -15.038\text{Ln}(x) + 36.854$, $y = -2.939\text{Ln}(x) + 64.148$ and $y = 0.1259\text{Ln}(x) - 0.1995$, were used for calculating the points for ENc-GC3s, ENc-Length and GC3s-Length, respectively.

Comparison of codon preferences of PRV Becker strain *EP0* gene with those of *E. coli*, yeast and human

To test whether distinct species follow a similar codon usage rule, we compared the codon preferences among the PRV *EP0* gene with those of *E. coli*, yeast, and human. 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 program in the EMBOSS software suite[35].

Statistical analysis

The correlations between codon usage variations among the PRV *EP0* gene and 24 reference alpha herpesviruses and four indicators (CAI, ENc, GC3s and gene length) were estimated using the SPSS 12.0 software package.

Results and Discussion

Molecular phylogenetic tree of the EP0-like proteins in PRV Becker strain and the reference alpha herpesviruses

A phylogenetic tree on the basis of the deduced EP0 and its EP0-like proteins in the reference alpha herpesviruses (Table 1) is shown in Fig. 1. From Fig. 1, we can see that the proteins could be preliminary separated into different genera, *i.e. Simplexvirus*, *Varicellovirus*, and *Mardivirus*, consistent with other previously published phylogenetic analyses[17,18], and the EP0-like proteins within the same genus are clustered together. Simultaneously, it is shown that the EP0 of PRV Becker, Bartha, Kaplan, Indiana-Funkhauser, Fa, PRV-FZ and Ea strains were different from other species. Firstly, they clustered together and formed a monophyletic clade, and then clustered with Gallid herpesvirus 2 (GaHV-2) of genus *Mardivirus*, and subsequently clustered with other members of the

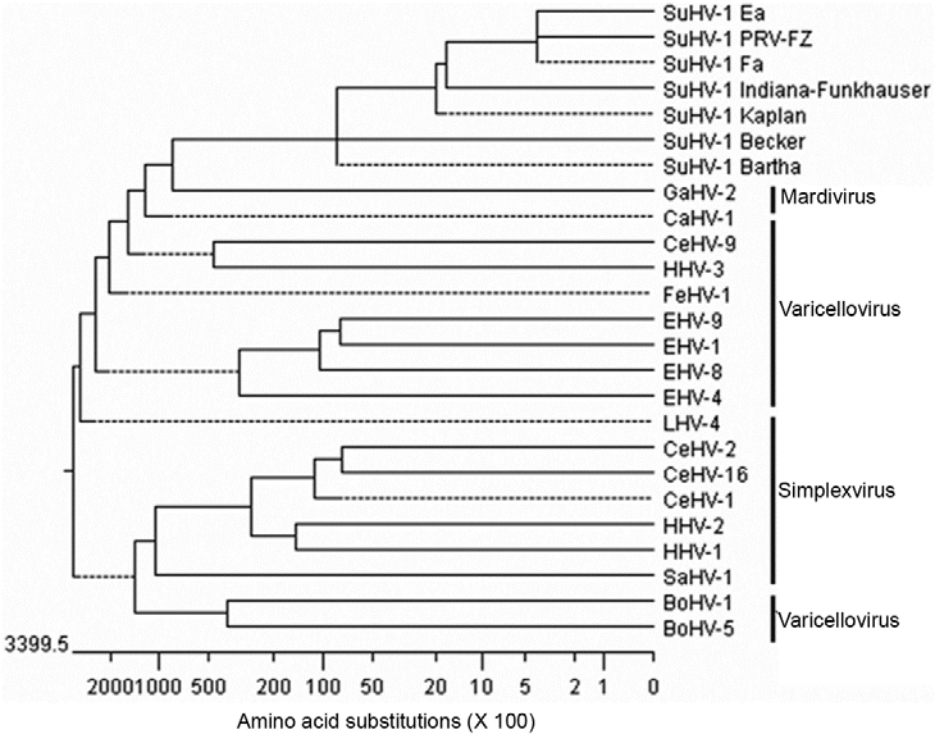

Fig. 1 — Evolutionary relationship of the PRV Becker strain EP0 protein with the EP0-like proteins of 24 reference alpha herpesviruses from different species (Table 1). [Phylogenetic tree of these proteins was generated by using the MEGALIGN (DNAStar) program with Jotun Hein multiple alignment software package and sequence distance indicated by the scale was calculated using the Structural matrix in LASERGENE]

genera *Varicellovirus* and Simplexvirus of alphaherpesvirus. Therefore, we can conclude from the phylogenetic tree and the high aa sequence homology that the PRV EP0 protein has a closer evolutionary relationship with the members of the genus *Mardivirus* than *Varicellovirus* and *Simplexvirus*, but certain differences nevertheless exist.

Codon usage analysis of the *EP0* gene in PRV Becker strain and the reference alpha herpesviruses

The results obtained by using CAI, CHIPS and CUPS programs of EMBOSS analysis of 25 alpha herpesviruses species are shown in Table 2. Codon usage in the PRV *EP0* gene and its homologous genes is extremely non-random, and the overall base composition of the *EP0* gene and its homologous genes in these species also shows similar variation. However, there are some distinct patterns in the codon usage bias parameters of the *EP0* gene among the PRV Becker, Bartha, Fa, PRV-FZ, Ea and Indiana-Funkhauser strains. It can be seen in Table 2 that the CAI values of different alpha herpesviruses vary from 0.648 to 0.783, with a mean value of 0.730 and a standard deviation (SD) of 0.041 and their ENc values range from 30.907 to 58.597, with a mean value of 42.602 and SD of 9.462. Compared to other species, the ENc values of different PRV strains are much lower (ENc<40), the codon usage bias in the *EP0*-like genes of 25 reference species especially the PRV is accordingly slightly high. Moreover, there is a same slight variation in codon usage pattern among different reference species *EP0*-like genes (SD =9.462). Similarly, the $GC3_S$ content of each *EP0*-like gene also confirms the homogeneity of synonymous codon usage among the different alpha herpesviruses, which vary from 18.26 to 92.54%, with a mean of 72.31% and a SD of 20.85%.

Table 2— Summary analysis of the PRV Becker strain *EP0* gene and the *EP0*-like genes of 24 reference alphaherpes viruses from different species

Rank	Virus name	Strain	CAI[a]	ENc[b]	Coding GC[c](%)	GC3s[d](%)
1	SuHV-1	Becker	0.767	35.715	69.38	85.87
2		Bartha	0.768	36.157	69.38	85.60
3		Kaplan	0.769	36.741	69.42	85.40
4		Fa	0.761	37.261	69.43	85.61
5		PRV-FZ	0.761	37.206	69.51	85.85
6		Ea	0.762	36.808	69.59	85.85
7		Indiana-Funkhauser	0.768	37.660	69.26	85.16
8	EHV-1	T-529	0.735	51.738	58.46	58.83
9	EHV-4	NS80567	0.704	55.795	50.71	44.69
10	EHV-8	wh	0.716	56.792	57.05	55.64
11	EHV-9	P19	0.745	53.051	58.68	59.74
12	FeHV-1	C-27	0.676	57.406	46.96	40.28
13	BoHV-1	Composite of 5 strains	0.705	40.220	76.56	84.79
14	BoHV-5	SV507/99	0.705	40.325	76.65	82.39
15	CaHV-1		0.648	39.613	31.94	18.26
16	HHV-3	11	0.673	58.597	49.50	45.30
17	HHV-1	OD4	0.746	34.761	71.69	84.54
18	HHV-2	HG52	0.730	34.443	77.94	91.15
19	CeHV-1	E2490	0.791	31.189	79.67	91.91
20	CeHV-2	B264	0.783	30.907	79.01	92.54
21	CeHV-9	Delta	0.652	55.577	48.61	41.27
22	CeHV-16	X313	0.775	32.185	79.04	91.04
23	LHV-4	LHV40126	0.705	35.590	81.05	84.74
24	SaHV-1	MV 5-4	0.716	53.286	62.56	63.42
25	GaHV-2	Md11	0.697	46.039	77.39	77.89

[a]codon adaptation index, [b] effective number of codons, [c] G+C content in the *EP0*-like gene, [d] G+C content at the third positions of codons. All these indices were calculated by using CAI, CHIPS and CUPS programs of EMBOSS.

A plot of ENc against GC3s is an effective way of examining the heterogeneity of codon usage among a set of homologous genes[33]. If a specific gene is subject to the G+C compositional constraint for shaping the codon usage pattern, it will lie on a continuous curve, representing random codon usage[36]. Conversely, if a gene is subject to selection for translationally optimal codons, it will lie considerably below the expected curve. The ENc values of each *EP0*-like gene in the 25 reference alpha herpesviruses are plotted against their corresponding GC3s in Fig. 2A. From Fig. 2A, we can see 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 described for *Drosophila melanogaster*, *E. coli*, *S. cerevisiae*, *Pseudomonas aeruginosa* and *Yersinia pestis*[37-39]. Here, the plot of gene length against ENc (Fig. 2B) or GC3s (Fig. 2C) shows the distribution for each gene. It appears that in the *EP0*-like genes of the 25 reference alpha herpesviruses, shorter or longer genes both have a similar variance in ENc values, and GC3s. We have

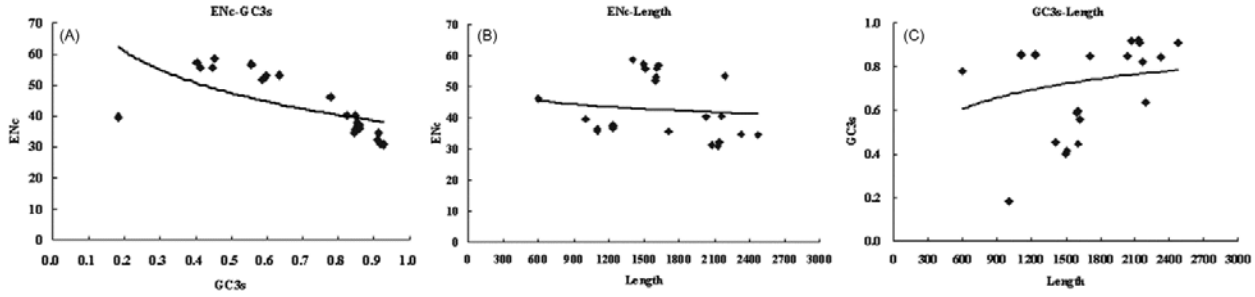

Fig. 2 — Relationship between ENc, GC3s and gene length of the PRV Becker strain *EP0* gene and the *EP0*-like genes of 24 reference alpha herpesviruses. (A) Plot of ENc versus GC3s for the PRV Becker strain *EP0* gene and the *EP0*-like genes of 24 reference alpha herpesviruses. ENc denotes the effective number of codons of each gene, and GC3s denotes the G+C content at the third synonymous codon position of each gene. The solid curve shows the expected position of genes whose codon usage is only determined by the variation in GC3s; (B) Plot of ENc versus gene length for the PRV Becker strain *EP0* gene and the *EP0*-like genes of 24 reference alpha herpesviruses; and (C) Plot of GC3s versus gene length for the PRV Becker strain *EP0* gene and the *EP0*-like genes of 24 reference alpha herpesviruses.

analyzed the relationship between ENc value and gene length, and the relationship between GC3s and gene length of the 25 reference species. However, none of the correlations were statistically significant, suggesting that gene length may not play a role in shaping the codon usage bias of the 25 alpha herpesviruses. Similar results were also found in *P. aeruginosa*[39], duck plague virus[5] and SARS coronavirus[40].

Variation in the PRV Becker strain *EP0* gene codon usage and aa composition

While the CAI, ENc and the related measures indicate the overall codon bias of PRV *EP0* gene, it is also important to examine more closely the pattern of codon bias. Table 3 shows the overall codon preference of the *EP0* gene in the PRV Becker strain. From the RSCU values we can see that the aas, 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 6-fold and 4-fold coding degeneracy. Moreover, Cys, Asp, Glu, His, Ile, Lys, Asn, Gln and Tyr also have a high level of diversity in codon usage bias, even though they only have 2 or 3-fold coding degeneracy. Altogether, although the most and the least frequencies used codons of all the aa are different, the analyzed PRV Becker strain *EP0* gene shows significant preference for one or more than one postulate codon for each aa. Meanwhile, the RSCU values of these codons whose third positions are C or G are much more than those codons whose third positions are A or T, suggesting that the PRV Becker strain *EP0* gene shows a high bias of codon usage toward the codons with C and/or G ending rather than T and/or A ending for all degenerate codons. However, a similar bias also exists at the first position, indicating a more complex situation exists in reality.

Phylogenetic persistence in codon usage bias of the PRV Becker strain *EP0* gene

To provide a visual representation of the variation in codon bias[41-43], we performed a cluster analysis (Fig. 3) of the codon usage pattern based on the PRV Becker strain *EP0* gene and its 24 reference alpha herpesviruses according to the RSCU values (Table 4). From the figure we can see that PRV Becker, Bartha, Fa, PRV-FZ, Ea and Indiana-Funkhauser strains appear distinct from other alpha herpesviruses, they firstly cluster together and form a separate branch, then they cluster with GaHV-2 of genus *Mardivirus* and Leporid herpesvirus 4 (LHV-4), Cercopithecine herpesvirus 1 (CeHV-1), CeHV-2, CeHV-16, Human herpesvirus 2 (HHV-2) of genus Simplexvirus, subsequently they cluster with other members of genera of *Simplexvirus* and *Varicellovirus*, such as HHV-1, Saimiriine herpesvirus 1 (SaHV-1), Bovine herpesvirus 1 (BoHV-1), BoHV-5, CeHV-9 and Equid herpesvirus 1 (EHV-1), etc. This result fully indicates the internal relations of the codon usage pattern between PRV and other alpha herpesviruses, suggesting that the codon usage pattern of PRV has differences with other alpha herpesviruses, the more distant the genetic relationship, the bigger the expected variation in the codon usage bias. Accordingly, we can conclude that the codon usage pattern of PRV is fairly close to that of the members of genus *Mardivirus* of alphaherpesvirus and is most different with other genera of alphaherpesvirus.

Table 3 — The result of codon preferences in PRV Becker strain *EP0* gene analyzed with the CUSP program

Codon	AA	Fraction	Frequency	Number	RSCU	Codon	AA	Fraction	Frequency	Number	RSCU
GCA	A(Ala)	0.050	5.435	2	0.200	CCA	P(Pro)	0.038	2.717	1	0.154
GCC	A	0.550	59.783	22	2.200	CCC	P	0.538	38.043	14	2.154
GCG	A	0.300	32.609	12	1.200	CCG	P	0.385	27.174	10	1.538
GCT	A	0.100	10.870	4	0.400	CCT	P	0.038	2.717	1	0.154
TGC	C(Cys)	0.875	19.022	7	1.750	**CAA**	Q(Gln)	0.000	0.000	0	0.000
TGT	C	0.125	2.717	1	0.250	CAG	Q	1.000	40.761	15	2.000
GAC	D(Asp)	0.833	54.348	20	1.667	AGA	R(Arg)	0.061	5.435	2	0.364
GAT	D	0.167	10.870	4	0.333	AGG	R	0.152	13.587	5	0.909
GAA	E(Glu)	0.133	10.870	4	0.267	CGA	R	0.091	8.152	3	0.545
GAG	E	0.867	70.652	26	1.733	CGC	R	0.455	40.761	15	2.727
TTC	F(Phe)	0.571	10.870	4	1.143	CGG	R	0.242	21.739	8	1.455
TTT	F	0.429	8.152	3	0.857	**CGT**	R	0.000	0.000	0	0.000
GGA	G(Gly)	0.100	8.152	3	0.400	AGC	S(Ser)	0.146	19.022	7	0.875
GGC	G	0.267	21.739	8	1.067	**AGT**	S	0.000	0.000	0	0.000
GGG	G	0.533	43.478	16	2.133	TCA	S	0.021	2.717	1	0.125
GGT	G	0.100	8.152	3	0.400	TCC	S	0.333	43.478	16	2.000
CAC	H(His)	0.900	24.457	9	1.800	TCG	S	0.292	38.043	14	1.750
CAT	H	0.100	2.717	1	0.200	TCT	S	0.208	27.174	10	1.250
ATA	I(Ile)	0.077	2.717	1	0.231	**ACA**	T(Thr)	0.000	0.000	0	0.000
ATC	I	0.846	29.891	11	2.538	ACC	T	0.522	32.609	12	2.087
ATT	I	0.077	2.717	1	0.231	ACG	T	0.435	27.174	10	1.739
AAA	K(Lys)	0.000	0.000	0	0.000	ACT	T	0.043	2.717	1	0.174
AAG	K	1.000	5.435	2	2.000	GTA	V(Val)	0.045	2.717	1	0.182
CTA	L(Leu)	0.000	0.000	0	0.000	GTC	V	0.500	29.891	11	2.000
CTC	L	0.286	10.870	4	1.714	GTG	V	0.455	27.174	10	1.818
CTG	L	0.643	24.457	9	3.857	**GTT**	V	0.000	0.000	0	0.000
CTT	L	0.071	2.717	1	0.429	TGG	W(Trp)	1.000	10.870	4	1.000
TTA	L	0.000	0.000	0	0.000	**TAC**	Y(Tyr)	0.000	0.000	0	0.000
TTG	L	0.000	0.000	0	0.000	TAT	Y	1.000	8.152	3	2.000
ATG	M(Met)	1.000	24.457	9	1.000	**TAA**	*	0.000	0.000	0	0.000
AAC	N(Asn)	1.000	16.304	6	2.000	**TAG**	*	0.000	0.000	0	0.000
AAT	N	0.000	0.000	0	0.000	TGA	*	1.000	2.717	1	3.000

Fract refers to the proportion of all synonymous codons encoding the same amino acid. The frequency of each codon that appears in the coding sequence of the individual gene is 1/1000. Shaded codons indicate the highest frequency in coding the amino acid. Rimmed 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.

Comparison of the *EP0* gene codon usage in PRV Becker strain with those of *E. coli*, yeast and human

Generally, the codon usage bias in a gene remains conserved to a certain degree across species. Here, the codon usage of PRV Becker strain *EP0* gene was compared with those of *E. coli*, yeast and human to see which would be the most suitable host for optimal expression. From Table 5, we can see that there are 40 codons showing a PRV-to-yeast ratio higher than 2 or lower than 0.50 and 27 codons showing a PRV-to-*E. coli* ratio higher than 2 or lower than 0.50, but only 22 codons showing a PRV-to-human ratio higher

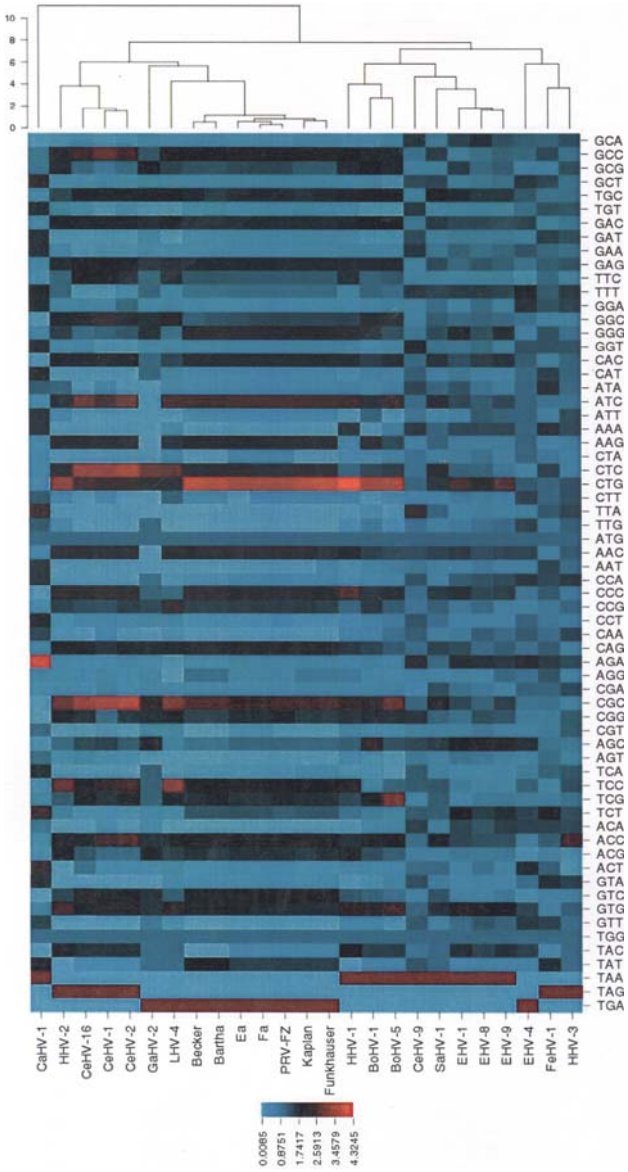

Fig. 3 — Heat map of RSCU values for the 25 reference alphaherpesvirus species (clustered by the RSCU values, Table 4). [Each row represents a various codon. Different species are represented in each column (identifiers as in Table 1). Cluster is shown to the top based on euclidean distance and average method]

than 2 or lower than 0.50, indicating that large differences in the codon preferences exist for all three hosts. Since PRV has the lowest distinct usage differences with a human than yeast and *E. coli*, suggesting that the codon usage of PRV Becker strain *EP0* gene more closely resembles to that of human than that of yeast and *E. coli*. Therefore, to express the PRV Becker strain *EP0* gene efficiently in yeast or *E. coli* system, codon optimization of the PRV Becker strain *EP0* gene may be required. Simultaneously, we can speculate that the PRV Becker strain *EP0* gene may be more efficiently expressed in the human expression system such as human embryonic kidney 293T (HEK293T) cells expression system.

In our study, a comprehensive analysis of codon usage including ENc, CAI value, GC content and the RSCU values of PRV Becker strain *EP0* gene was carried out using analytical techniques implemented in the CAI, CHIPS and CUPS programs of EMBOSS. Subsequently, these values were compared with those of the 24 reference alphaherpesvirus species. The data of synonymous codon usage bias demonstrated certain distinct differences existed for each herpesvirus from different species, and the result revealed that: (i) PRV Becker strain *EP0* gene and its 24 reference alpha herpesviruses take relatively similar codon usage patterns, although PRV Becker strain *EP0* gene shows a few disparities of codon usage bias with its reference alpha herpesvirus species; and (ii) The PRV Becker strain *EP0* gene prefers to use the codons with C and G at the third codon position. Furthermore, the biased inclination towards C and G is consistent with the high C + G content in PRV Becker strain *EP0* gene. Since the *EP0* gene in the PRV Becker strain is a CG-rich gene, it is reasonable that C and G -ending codons are predominant in the gene. In order to show the codon usage variation, we also used the ENc-plot to analyze the factors influencing codon usage variation among genes. Here, genetic heterogeneity in the PRV and its reference alpha herpesviruses is restricted by the GC content, but not gene length.

Comparative analysis of *EP0* genes in PRV and the reference herpesviruses indicated that synonymous codon usage in these genes is relatively phylogenetically conserved. Although the codon usage pattern among different species is a complicated phenomenon, it is vital to elucidate the underlying mechanisms of codon usage pattern so as to understand the evolution of the species[44,45]. From the phylogenetic tree (Fig. 1) and cluster analysis results (Fig. 3) we can see that PRV is evolutionarily closer with GaHV-2 of genus *Mardivirus* than other members of alpha herpesvirus. Simultaneously, its codon usage pattern is also closer with GaHV-2 and LHV-4 than other members of alpha herpesvirus. Therefore, we can draw a conclusion that species has a certain influence on the preference of codon usage, but is less substantial than the influence of gene function, and the codon usage bias of PRV *EP0* gene has a very close relation with its gene function.

Table 4 — RSCU values of the *EP0* gene of PRV Becker strain and the *EP0*-like genes of 24 reference alphaherpesviruses from different species

AA	Codon	Becker	Bartha	Kaplan	Fa	PRV-FZ	Ea	Funkhauser	EHV-1	EHV-4	EHV-8	EHV-9	CeHV-9	FeHV-1	BoHV-1	BoHV-5	CaHV-1	HHV-3	LHV-4	SaHV-1	HHV-1	HHV-2	CeHV-1	CeHV-2	CeHV-16	GaHV-2
A	GCA	0.200	0.195	0.182	0.093	0.093	0.091	0.186	1.302	1.059	1.545	1.156	1.486	1.000	0.139	0.513	0.000	1.200	0.167	0.693	0.158	0.085	0.029	0.122	0.186	0.387
A	GCC	2.200	2.146	2.182	2.140	2.140	2.182	2.233	0.651	0.941	0.818	0.978	0.800	1.286	1.722	1.641	0.800	0.800	2.167	1.387	1.941	1.957	2.964	2.718	2.667	1.161
A	GCG	1.200	1.171	1.091	1.302	1.302	1.273	1.116	1.395	0.471	0.909	1.067	0.457	0.429	1.861	1.718	0.800	1.200	1.333	0.640	1.663	1.816	0.777	0.947	0.992	1.935
A	GCT	0.400	0.488	0.545	0.465	0.465	0.455	0.465	0.651	1.529	0.727	0.800	1.257	1.286	0.278	0.128	2.400	0.800	0.333	1.280	0.238	0.142	0.230	0.214	0.155	0.516
C	TGC	1.750	1.500	1.556	1.778	1.778	1.800	1.778	1.600	1.125	1.467	1.467	0.545	0.800	2.000	2.000	0.250	0.857	2.000	1.818	1.846	1.333	2.000	1.857	1.818	2.000
C	TGT	0.250	0.500	0.444	0.222	0.222	0.200	0.222	0.400	0.875	0.533	0.533	1.455	1.200	0.000	0.000	1.750	1.143	0.000	0.182	0.154	0.667	0.000	0.143	0.182	0.000
D	GAC	1.667	1.667	1.667	1.667	1.667	1.667	1.667	1.282	1.353	1.294	1.268	0.762	0.596	1.941	1.889	0.316	0.765	2.000	1.455	1.795	1.950	1.958	2.000	2.000	1.600
D	GAT	0.333	0.333	0.333	0.333	0.333	0.333	0.333	0.718	0.647	0.706	0.732	1.238	1.404	0.059	0.111	1.684	1.235	0.000	0.545	0.205	0.050	0.042	0.000	0.000	0.400
E	GAA	0.267	0.267	0.258	0.267	0.267	0.267	0.258	1.048	1.171	0.857	0.952	1.500	0.929	0.340	0.302	1.622	0.842	0.211	0.750	0.400	0.313	0.091	0.087	0.000	0.286
E	GAG	1.733	1.733	1.742	1.733	1.733	1.733	1.742	0.952	0.829	1.143	1.048	0.500	1.071	1.660	1.698	0.378	1.158	1.789	1.250	1.600	1.688	1.909	1.913	2.000	1.714
F	TTC	1.143	1.143	1.111	1.111	1.111	1.111	1.111	0.500	0.125	0.533	0.667	0.429	0.909	1.000	1.111	0.167	0.462	1.600	0.615	1.556	1.000	2.000	1.750	2.000	1.000
F	TTT	0.857	0.857	0.889	0.889	0.889	0.889	0.889	1.500	1.875	1.467	1.333	1.571	1.091	1.000	0.889	1.833	1.538	0.400	1.385	0.444	1.000	0.000	0.250	0.000	1.000
G	GGA	0.400	0.267	0.242	0.242	0.250	0.250	0.242	0.387	1.600	0.875	0.645	0.625	1.217	0.395	0.650	1.778	1.412	0.603	0.857	0.225	0.408	0.519	0.985	0.610	0.444
G	GGC	1.067	1.200	1.333	1.333	1.375	1.250	1.333	0.903	0.800	1.000	0.903	0.625	0.348	1.827	1.850	0.000	0.588	2.192	1.071	1.438	1.837	2.370	1.785	2.034	1.630
G	GGG	2.133	2.133	2.061	2.061	2.000	2.125	2.061	1.935	0.400	1.250	1.806	0.750	1.217	1.630	1.250	0.444	0.824	0.877	1.286	2.157	1.510	1.037	1.169	0.949	1.630
G	GGT	0.400	0.400	0.364	0.364	0.375	0.375	0.364	0.774	1.200	0.875	0.645	2.000	1.217	0.148	0.250	1.778	1.176	0.329	0.786	0.180	0.245	0.074	0.062	0.407	0.296
H	CAC	1.800	1.800	1.800	1.800	1.800	1.800	1.800	1.250	1.077	1.500	1.714	0.933	0.571	1.500	1.750	0.000	1.000	1.375	1.636	1.867	2.000	2.000	2.000	2.000	1.000
H	CAT	0.200	0.200	0.200	0.200	0.200	0.200	0.200	0.750	0.923	0.500	0.286	1.067	1.429	0.500	0.250	2.000	1.000	0.625	0.364	0.133	0.000	0.000	0.000	0.000	1.000
I	ATA	0.231	0.231	0.214	0.200	0.200	0.200	0.214	1.167	1.000	0.789	0.667	1.375	1.737	0.200	0.500	1.111	0.913	0.333	0.955	0.462	1.000	0.333	0.000	0.000	0.000
I	ATC	2.538	2.538	2.571	2.600	2.600	2.600	2.571	1.333	1.000	1.105	1.333	0.875	0.789	1.800	2.500	0.222	0.783	2.667	1.909	2.538	1.750	2.667	3.000	3.000	0.000
I	ATT	0.231	0.231	0.214	0.200	0.200	0.200	0.214	0.500	1.000	1.105	1.000	0.750	0.474	1.000	0.000	1.667	1.304	0.000	0.136	0.000	0.250	0.000	0.000	0.000	0.000
K	AAA	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.333	1.000	1.333	1.333	1.571	1.750	0.000	0.667	1.714	1.333	0.333	1.000	1.684	0.000	0.000	0.000	0.000	0.000
K	AAG	2.000	2.000	2.000	2.000	2.000	2.000	2.000	0.667	1.000	0.667	0.667	0.429	0.250	2.000	1.333	0.286	0.667	1.667	1.000	0.316	2.000	2.000	2.000	2.000	0.000
L	CTA	0.000	0.000	0.000	0.353	0.353	0.353	0.000	0.686	0.973	0.811	0.545	0.811	0.714	0.150	0.558	0.462	0.529	0.353	0.558	0.167	0.000	0.000	0.000	0.000	0.000
L	CTC	1.714	1.846	1.875	1.765	1.765	2.118	1.765	1.029	0.324	0.973	0.727	0.811	1.429	1.350	1.535	0.462	0.529	2.824	1.953	1.000	2.000	3.429	3.632	3.450	3.000
L	CTG	3.857	3.692	3.750	3.529	3.529	3.529	3.882	2.571	1.135	2.108	2.727	0.973	1.000	3.600	3.628	0.231	1.235	1.765	1.674	4.333	3.333	2.400	2.368	2.550	2.000
L	CTT	0.429	0.462	0.375	0.353	0.353	0.000	0.353	0.686	1.459	0.649	0.545	0.324	1.000	0.000	0.140	1.154	1.235	1.059	0.279	0.333	0.333	0.171	0.000	0.000	0.000
L	TTA	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.343	0.649	0.649	0.545	2.270	1.143	0.150	0.000	2.538	1.412	0.000	0.977	0.000	0.000	0.000	0.000	0.000	0.000
L	TTG	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.686	1.459	0.811	0.909	0.811	0.714	0.750	0.140	1.154	1.059	0.000	0.558	0.167	0.333	0.000	0.000	0.000	1.000
M	ATG	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
N	AAC	2.000	2.000	2.000	2.000	2.000	2.000	1.714	1.556	1.400	1.273	1.412	1.263	0.621	1.500	1.333	0.381	1.111	2.000	1.440	1.889	2.000	2.000	2.000	2.000	0.000
N	AAT	0.000	0.000	0.000	0.000	0.000	0.000	0.286	0.444	0.600	0.727	0.588	0.737	1.379	0.500	0.667	1.619	0.889	0.000	0.560	0.111	0.000	0.000	0.000	0.000	0.000
P	CCA	0.154	0.154	0.250	0.364	0.364	0.364	0.375	1.091	1.617	1.214	1.309	1.191	1.375	0.229	0.405	1.714	1.600	0.107	0.547	0.130	0.037	0.000	0.037	0.000	0.923
P	CCC	2.154	2.154	2.000	1.939	1.818	1.939	2.000	1.745	0.766	1.357	1.527	0.851	1.125	1.600	1.772	0.000	0.700	0.964	1.558	2.696	2.312	2.272	2.128	2.146	1.077
P	CCG	1.538	1.538	1.625	1.576	1.576	1.576	1.500	0.727	0.596	0.929	0.582	0.766	0.750	1.600	1.114	0.286	1.000	2.321	1.053	1.000	1.505	1.280	1.541	1.301	1.385
P	CCT	0.154	0.154	0.125	0.121	0.242	0.121	0.125	0.436	1.021	0.500	0.582	1.191	0.750	0.571	0.709	2.000	0.700	0.607	0.842	0.174	0.147	0.448	0.294	0.553	0.615
Q	CAA	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.700	1.154	1.100	0.900	1.100	0.889	0.571	0.857	1.400	1.429	0.133	0.533	0.400	0.100	0.345	0.045	0.000	0.000
Q	CAG	2.000	2.000	2.000	2.000	2.000	2.000	2.000	1.300	0.846	0.900	1.100	0.900	1.111	1.429	1.143	0.600	0.571	1.867	1.467	1.600	1.900	1.655	1.955	2.000	2.000
R	AGA	0.364	0.364	0.615	0.632	0.615	0.632	0.615	1.789	2.118	1.684	1.579	2.200	1.543	0.676	0.395	4.000	1.313	0.076	0.931	0.375	0.343	0.190	0.369	0.277	0.308
R	AGG	0.909	0.909	0.769	0.632	0.615	0.632	0.769	0.632	0.824	0.526	0.947	0.400	1.200	0.592	0.316	0.667	0.563	0.000	0.310	0.750	0.514	0.476	0.185	0.369	0.615
R	CGA	0.545	0.545	0.462	0.316	0.308	0.316	0.462	0.842	1.059	0.737	0.421	0.200	1.029	0.592	0.474	0.333	1.313	0.304	1.241	0.375	0.257	0.667	0.185	0.185	0.462
R	CGC	2.727	2.727	2.615	2.526	2.615	2.526	2.615	1.368	0.824	1.053	1.474	1.000	0.857	2.197	2.921	0.000	0.563	3.190	1.966	2.438	3.000	3.619	3.692	3.508	2.308
R	CGG	1.455	1.455	1.538	1.737	1.846	1.737	1.538	0.842	0.706	1.263	1.158	1.600	0.686	1.775	1.737	0.000	1.125	1.823	1.034	1.688	1.886	0.952	1.569	1.662	1.692
R	CGT	0.000	0.000	0.000	0.158	0.000	0.158	0.000	0.526	0.471	0.737	0.421	0.600	0.686	0.169	0.158	1.000	1.125	0.608	0.517	0.375	0.000	0.095	0.000	0.000	0.615
S	AGC	0.875	0.857	0.982	1.019	1.019	1.038	1.000	2.000	1.647	2.250	2.250	1.406	0.808	2.400	1.277	0.158	0.750	0.692	1.548	1.091	1.000	1.500	1.143	1.350	2.182
S	AGT	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.824	0.471	0.750	0.625	0.750	0.923	0.120	0.000	0.947	0.643	0.000	0.452	0.312	0.176	0.167	0.143	0.600	0.000

S	TCA	0.125	0.122	0.218	0.113	0.113	0.115	0.222	0.235	0.706	0.250	0.250	1.125	1.038	0.120	0.000	1.737	1.179	0.000	0.645	0.078	0.000	0.000	0.000	0.000	1.091
S	TCC	2.000	1.959	1.855	1.925	1.925	1.846	1.889	0.706	1.059	0.500	0.500	1.125	0.577	0.600	1.021	0.158	1.286	3.231	1.161	2.260	2.941	2.500	2.571	1.950	1.091
S	TCG	1.750	1.837	1.745	1.698	1.811	1.731	1.667	0.353	0.706	0.750	0.750	0.469	0.692	2.040	3.064	0.474	0.857	1.846	1.097	1.403	1.176	1.667	2.000	1.950	1.091
S	TCT	1.250	1.224	1.200	1.245	1.132	1.269	1.222	1.882	1.412	1.500	1.625	1.125	1.962	0.720	0.638	2.526	1.286	0.231	1.097	0.857	0.706	0.167	0.143	0.150	0.545
T	ACA	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.375	1.486	1.158	1.500	1.488	1.438	0.400	0.485	0.870	1.333	0.000	0.655	0.082	0.000	0.143	0.000	0.000	0.267
T	ACC	2.087	2.286	2.308	2.069	2.069	2.069	2.143	1.375	0.686	1.263	1.250	0.930	1.125	2.000	2.061	0.522	2.578	1.846	1.818	1.714	2.047	2.571	2.732	1.846	1.600
T	ACG	1.739	1.524	1.538	1.655	1.655	1.655	1.714	0.750	0.114	0.737	0.750	0.651	0.375	1.333	1.091	0.174	0.089	1.538	0.582	1.959	1.953	1.000	0.976	1.231	1.867
T	ACT	0.174	0.190	0.154	0.276	0.276	0.276	0.143	0.500	1.714	0.842	0.500	0.930	1.063	0.267	0.364	2.435	0.000	0.615	0.945	0.245	0.000	0.286	0.293	0.923	0.267
V	GTA	0.182	0.182	0.167	0.174	0.174	0.167	0.333	0.258	0.903	0.400	0.235	1.257	1.636	0.133	0.432	1.818	1.455	0.320	0.649	0.000	0.000	0.000	0.000	0.000	0.000
V	GTC	2.000	2.000	2.167	2.087	2.087	2.000	2.000	0.516	0.129	0.400	0.706	0.571	0.364	1.067	0.865	0.000	0.545	1.120	1.189	1.209	1.391	1.800	1.857	2.162	1.600
V	GTG	1.818	1.818	1.667	1.739	1.739	1.833	1.667	2.323	1.419	2.000	2.000	1.029	0.545	2.400	2.595	0.727	1.091	2.400	1.622	2.419	2.609	2.200	2.143	1.838	1.600
V	GTT	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.903	1.548	1.200	1.059	1.143	1.455	0.400	0.108	1.455	0.909	0.160	0.541	0.372	0.000	0.000	0.000	0.000	0.800
W	TGG	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
Y	TAC	0.000	0.000	0.500	0.500	0.500	0.500	0.500	1.667	1.500	1.429	1.667	0.800	0.222	1.500	1.500	0.000	1.000	1.000	1.000	2.000	2.000	1.714	1.714	1.714	1.000
Y	TAT	2.000	2.000	1.500	1.500	1.500	1.500	1.500	0.333	0.500	0.571	0.333	1.200	1.778	0.500	0.500	2.000	1.000	1.000	1.000	0.000	0.000	0.286	0.286	0.286	1.000
*	TAA	0.000	0.000	0.000	0.000	0.000	0.000	0.000	3.000	0.000	3.000	3.000	3.000	0.000	3.000	3.000	3.000	0.000	0.000	3.000	3.000	0.000	0.000	0.000	0.000	0.000
*	TAG	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	3.000	0.000	0.000	0.000	3.000	0.000	0.000	0.000	3.000	3.000	3.000	3.000	0.000
*	TGA	3.000	3.000	3.000	3.000	3.000	3.000	3.000	0.000	3.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	3.000	0.000	0.000	0.000	0.000	0.000	0.000	3.000

Table 5 — Comparison of codon preferences between PRV Becker strain *EP0* gene and *E. coli*, yeast and human (*contd.*)

Condon	Amino acid	*E. coli* (1/1000)	Yeast (1/1000)	Human (1/1000)	SuHV-1 (1/1000)	SuHV-1 /*E. coli*	SuHV-1 /Yeast	SuHV-1 /Human
GCA	A(Ala)	20.6	16.1	16.1	5.4	**0.3**	**0.3**	**0.3**
GCC	A	25.5	12.5	28.4	59.8	**2.3**	**4.8**	**2.1**
GCG	A	31.7	6.1	7.5	32.6	1	**5.3**	**4.3**
GCT	A	15.6	21.1	18.6	10.9	0.7	0.5	0.6
TGC	C(Cys)	6.9	4.7	12.2	19	**2.8**	**4**	1.6
TGT	C	5.5	8	10	2.7	0.5	**0.3**	**0.3**
GAC	D(Asp)	18.6	20.2	25.6	54.3	**2.9**	**2.7**	**2.1**
GAT	D	32.1	37.8	21.9	10.9	**0.3**	**0.3**	0.5
GAA	E(Glu)	38.2	48.5	29	10.9	**0.3**	**0.2**	**0.4**
GAG	E	17.7	19.1	39.9	70.7	**4**	**3.7**	1.8
TTC	F(Phe)	16.9	18.2	20.6	10.9	0.6	0.6	0.5
TTT	F	23.2	26.1	17.1	8.2	**0.4**	**0.3**	0.5
GGA	G(Gly)	9	10.9	16.4	8.2	0.9	0.8	0.5
GGC	G	27.9	9.7	22.5	21.7	0.8	**2.2**	1
GGG	G	11.3	6	16.3	43.5	**3.8**	**7.3**	**2.7**
GGT	G	24.4	24	10.8	8.2	**0.3**	**0.3**	0.8
CAC	H(His)	9.8	7.7	15	24.5	**2.5**	**3.2**	1.6
CAT	H	13.6	13.7	10.5	2.7	**0.2**	**0.2**	**0.3**
ATA	I(Ile)	5.4	17.8	7.7	2.7	0.5	**0.2**	**0.4**
ATC	I	24.2	17	21.6	29.9	1.2	1.8	1.4
ATT	I	29.8	30.4	16.1	2.7	**0.1**	**0.1**	**0.2**
AAA	K(Lys)	33.2	42.2	24.1	0	**0**	**0**	**0**
AAG	K	10.7	30.7	32.2	5.4	0.5	**0.2**	**0.2**

(*contd.*)

Table 5 — Comparison of codon preferences between PRV Becker strain *EP0* gene and *E. coli*, yeast and human

Condon	Amino acid	*E. coli* (1/1000)	Yeast (1/1000)	Human (1/1000)	SuHV-1 (1/1000)	SuHV-1 /*E. coli*	SuHV-1 /Yeast	SuHV-1 /Human
CTA	L(Leu)	4	13.3	7.8	0	**0**	**0**	**0**
CTC	L	11	5.4	19.8	10.9	1	2	0.6
CTG	L	50.9	10.4	39.8	24.5	0.5	**2.4**	0.6
CTT	L	11.7	12.1	13	2.7	**0.2**	**0.2**	**0.2**
TTA	L	13.9	26.7	7.5	0	**0**	**0**	**0**
TTG	L	14	27	12.6	0	**0**	**0**	**0**
ATG	M(Met)	27	20.9	22.2	24.5	0.9	1.2	1.1
AAC	N(Asn)	21.4	24.9	19.5	16.3	0.8	0.7	0.8
AAT	N	18.6	36.3	16.7	0	**0**	**0**	**0**
CCA	P(Pro)	8.5	18.2	16.7	2.7	**0.3**	**0.1**	**0.2**
CCC	P	5.8	6.8	20.1	38	**6.6**	**5.6**	1.9
CCG	P	21.8	5.3	6.9	27.2	1.2	**5.1**	**3.9**
CCT	P	7.3	13.6	17.3	2.7	**0.4**	**0.2**	**0.2**
CAA	Q(Gln)	15	27.5	12	0	**0**	**0**	**0**
CAG	Q	29.5	12.1	34.1	40.8	1.4	**3.4**	1.2
AGA	R(Arg)	2.9	21.3	11.5	5.4	1.9	**0.3**	0.5
AGG	R	1.9	9.2	11.4	13.6	**7.2**	1.5	1.2
CGA	R	3.9	3	6.3	8.2	**2.1**	**2.7**	1.3
CGC	R	21	2.6	10.7	40.8	1.9	**15.7**	**3.8**
CGG	R	6.3	1.7	11.6	21.7	**3.4**	**12.8**	1.9
CGT	R	20.3	6.5	4.6	0	**0**	**0**	**0**
AGC	S(Ser)	16	9.7	19.3	19	1.2	2	1
AGT	S	9.5	14.2	11.9	0	**0**	**0**	**0**
TCA	S	7.8	18.8	12	2.7	**0.3**	**0.1**	**0.2**
TCC	S	8.9	14.2	11.9	43.5	**4.9**	**3.1**	**3.7**
TCG	S	8.7	8.5	4.4	38	**4.4**	**4.5**	**8.6**
TCT	S	8.7	23.5	14.7	27.2	**3.1**	1.2	1.9
ACA	T(Thr)	8.2	17.8	15.1	0	**0**	**0**	**0**
ACC	T	22.8	12.6	19.4	32.6	1.4	**2.6**	1.7
ACG	T	14.8	7.9	6.1	27.2	1.8	**3.4**	**4.5**
ACT	T	9.1	20.3	13	2.7	**0.3**	**0.1**	**0.2**
GTA	V(Val)	11.1	11.8	7.2	2.7	**0.2**	**0.2**	**0.4**
GTC	V	15.1	11.6	14.6	29.9	2	**2.6**	2
GTG	V	25.5	10.6	28.4	27.2	1.1	**2.6**	1
GTT	V	18.5	22	11	0	**0**	**0**	**0**
TGG	W(Trp)	15.2	10.3	12.7	10.9	0.7	1.1	0.9
TAC	Y(Tyr)	12.1	14.6	15.5	0	**0**	**0**	**0**
TAT	Y	16.5	18.9	12.1	8.2	0.5	**0.4**	0.7
TAA	*	2	1	0.7	0	**0**	**0**	**0**
TAG	*	0.3	0.5	0.6	0	**0**	**0**	**0**
TGA	*	1.1	0.7	1.5	2.7	**2.5**	**3.9**	1.8

Note: SuHV-1/*E. coli*, SuHV-1/yeast and SuHV-1/human indicate the ratio of codon usage frequency in SuHV-1 to that in *E. coli*, yeast and human, respectively. A ratio higher than 2 or lower than 0.5 (except 0) underlined and marked with bold indicates that the codon preference differs greatly, and vice versa.

Studies on the conserved RING finger domain have been well documented. However, functions of *EP0* gene product (EP0) involved in the PRV life cycle, is less well understood thus far. The RING-finger region of HSV-1 ICP0 is reported to be essential for its regulation of gene expression, stimulation of lytic infection, enhancement of reactivation from quiescence, disruption of ND10 structures, induction of proteasome-dependent degradation of cellular proteins and interaction with cyclin D3[25]. The RING-finger regions of the EHV-1 ORF63, HSV-1 ICP0 and VZV ORF61 proteins are also reported to bind zinc stably[28], and the transactivation ability of the BICP0 protein was found to be zinc dependent[27]. Therefore, because of the important roles played by the counterpart of EP0 in HSV-1, VZV, EHV-1, and BoHV-1 in the course of infection, it means that EP0 may also play a similar role in the process of infection according to their phylogenetic conservation. However, it is not yet known what real biological functions of EP0 have in the PRV life cycle and the examination of these aspects must, therefore, await further clarification of its functions in viral replication and the interactions between PRV and host.

Among the codon usage bias patterns in *E. coli*, yeast and human, the codon usage bias pattern in the PRV *EP0* gene is more similar to that of human (Table 5). Thus, a supposition made in the present study that PRV *EP0* gene may express more efficiently in a HEK293T cells expression system. This may serve as a guide for manipulating the expression of the targeted gene. PRV *EP0* gene optimizing with host-preferred codons probably help to improve the expression level of the PRV *EP0* gene in a given host. Therefore, HEK293T cells expression system may be better applied to the production of PRV EP0. Our hypothesis was verified in our recent study, in which the PRV EP0 protein was successfully expressed in the HEK293T expression system (unpublished data).

Taken together, analysis of codon usage pattern of PRV *EP0* gene and a comparison of codon preference between PRV *EP0* gene and other species can provide a foundation for understanding the pertinent mechanism of biased usage of synonymous codons and for selecting an appropriate host expression system to improve the expression of PRV *EP0*. It also may provide some insights into the properties of the PRV 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.

Acknowledgements

This work was supported by grants from the National Natural Science Foundation of China (31400150 and 31200120); the Natural Science Foundation of Guangdong Province (2015A030313473); the Training Program for Outstanding Young Teachers in Universities of Guangdong Province (YQ2015132); the Medical Scientific Research Foundation of Guangdong Province, China (A2017055 and B2012165); the Science and Technology Plan Projects of Guangzhou City, China (201607010088); the Scientific Research Projects in Colleges and Universities of Guangzhou (1201430024, 1201610025 and 1201610024); Innovation and entrepreneurship training program in colleges and universities in Guangzhou (2017224104); High-Level Universities Academic Backbone Development Program of Guangzhou Medical University; the Thousand Hundred Ten Projects of Guangzhou Medical University, Guangdong; Guangdong Undergraduate Training Programs of Science & Technology Innovation (Guangdong Climb Plan: pdjh2017b0419,); National, Provincial and School Level Training Programs of Innovation and Entrepreneurship for Undergraduates in Guangzhou Medical University (201610570004, 2016A002, 2016A003, 2016A027, 2016A066, XS201609 and XS201619).

References

1 Duret L, Evolution of synonymous codon usage in metazoans. *Curr Opin Genet Dev*, 12 (2002) 640.

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*, 349 (1995) 241.

3 Grantham R, Gautier C, Gouy M, Jacobzone M & Mercier R, Codon catalog usage is a genome strategy modulated for gene expressivity. *Nucleic Acids Res*, 9 (1981) r43.

4 Blake RD & Hinds PW, Analysis of the codon bias in E. coli sequences. *J Biomol Struct Dyn*, 2 (1984) 593.

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

6 Ikemura T, Codon usage and tRNA content in unicellular and multicellular organisms. *Mol Biol Evol*, 2 (1985) 13.

7 Maruyama T, Gojobori T, Aota S & Ikemura T, Codon usage tabulated from the GenBank genetic sequence data. *Nucleic Acids Res*, 14 Suppl (1986) r151.

8 Wain-Hobson S, Nussinov R, Brown RJ & Sussman JL, Preferential codon usage in genes. *Gene*, 13 (1981) 355.

9 Li ML, Zhao ZY, Chen JH, Wang BY, Li Z, Li J & Cai MS, Characterization of synonymous codon usage bias in the pseudorabies virus US1 gene. *Virol Sin*, 27 (2012) 303.

10 Cai MS, Zhao ZY, Zhu JY, Chen JH, Wang BY, Li Z & Li ML, Identification of Synonymous Codon Usage Bias in the Pseudorabies Virus UL31 Gene. *Iran J Biotechnol*, 11 (2013) 214.
11 Hooper SD & Berg OG, Gradients in nucleotide and codon usage along Escherichia coli genes. *Nucleic Acids Res*, 28 (2000) 3517.
12 Blaisdell 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*, 19 (1983) 226.
13 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*, 18 (1982) 199.
14 Wada A & Suyama A, Local stability of DNA and RNA secondary structure and its relation to biological functions. *Prog Biophys Mol Biol*, 47 (1986) 113.
15 Li ML, Wang S, Cai MS, Guo H & Zheng CF, Characterization of molecular determinants for nucleocytoplasmic shuttling of PRV UL54. *Virol*, 417 (2011) 385.
16 Li ML, Wang S, Cai MS & Zheng CF, Identification of nuclear and nucleolar localization signals of pseudorabies virus (PRV) early protein UL54 reveals that its nuclear targeting is required for efficient production of PRV. *J Virol*, 85 (2011) 10239.
17 Muller T, Hahn EC, Tottewitz F, Kramer M, Klupp BG, Mettenleiter TC & Freuling C, Pseudorabies virus in wild swine: a global perspective. *Arch Virol*, 156 (2011) 1691.
18 Ono E, Amagai K, Yoshino S, Taharaguchi S, Inobe M & Uede T, Resistance to pseudorabies virus infection in transformed cell lines expressing a soluble form of porcine herpesvirus entry mediator C. *J Gen Virol*, 85 (2004) 173.
19 Pomeranz LE, Reynolds AE & Hengartner CJ, Molecular biology of pseudorabies virus: impact on neurovirology and veterinary medicine. *Microbiol Mol Biol Rev*, 69 (2005) 462.
20 Cai MS, Jiang S, Zeng ZC, Li XW, Mo CC, Yang YJ, Chen CK, Xie PP, Bian Y, Wang JL, Huang JL, Chen DX, Peng T & Li ML, Probing the nuclear import signal and nuclear transport molecular determinants of PRV ICP22. *Cell Biosci*, 6 (2016) 3.
21 Muller T, Batza HJ, Schluter H, Conraths FJ & Mettenleiter TC, Eradication of Aujeszky's disease in Germany. *J Vet Med B Infect Dis Vet Public Health*, 50 (2003) 207.
22 White AK, Ciacci-Zanella J, Galeota J, Ele S & Osorio FA, Comparison of the abilities of serologic tests to detect pseudorabies-infected pigs during the latent phase of infection. *Am J Vet Res*, 57 (1996) 608.
23 Koppers-Lalic D, Reits EA, Ressing ME, Lipinska AD, Abele R, Koch J, Marcondes Rezende M, Admiraal P, van Leeuwen D, Bienkowska-Szewczyk K, Mettenleiter TC, Rijsewijk FA, Tampe R, Neefjes J & Wiertz EJ, Varicelloviruses avoid T cell recognition by UL49.5-mediated inactivation of the transporter associated with antigen processing. *Proc Natl Acad Sci U S A*, 102 (2005) 5144.
24 Boelaert F, Deluyker H, Maes D, Godfroid J, Raskin A, Varewijck H, Pensaert M, Nauwynck H, Castryck F, Miry C, Robijns JM, Hoet B, Segers E, Van Vlaenderen I, Robert A & Koenen F, Prevalence of herds with young sows seropositive to pseudorabies (Aujeszky's disease) in northern Belgium. *Prev Vet Med*, 41 (1999) 239.
25 Hagglund R & Roizman B, Role of ICP0 in the strategy of conquest of the host cell by herpes simplex virus 1. *J Virol*, 78 (2004) 2169.
26 Moriuchi H, Moriuchi M, Straus SE & Cohen JI, Varicella-zoster virus (VZV) open reading frame 61 protein transactivates VZV gene promoters and enhances the infectivity of VZV DNA. *J Virol*, 67 (1993) 4290.
27 Fraefel C, Zeng J, Choffat Y, Engels M, Schwyzer M & Ackermann M, Identification and zinc dependence of the bovine herpesvirus 1 transactivator protein BICP0. *J Virol*, 68 (1994) 3154.
28 Everett RD, Barlow P, Milner A, Luisi B, Orr A, Hope G & Lyon D, A novel arrangement of zinc-binding residues and secondary structure in the C3HC4 motif of an alpha herpes virus protein family. *J Mol Biol*, 234 (1993) 1038.
29 Burland TG, DNASTAR's Lasergene sequence analysis software. *Methods Mol Biol*, 132 (2000) 71.
30 Sharp PM & Li WH, The codon Adaptation Index--a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res*, 15 (1987) 1281.
31 Novembre JA, Accounting for background nucleotide composition when measuring codon usage bias. *Mol Biol Evol*, 19 (2002) 1390.
32 Comeron JM & Aguade M, An evaluation of measures of synonymous codon usage bias. *J Mol Evol*, 47 (1998) 268.
33 Wright F, The 'effective number of codons' used in a gene. *Gene*, 87 (1990) 23.
34 Weinstein JN, Myers TG, O'Connor PM, Friend SH, Fornace AJ, Jr., Kohn KW, Fojo T, Bates SE, Rubinstein LV, Anderson NL, Buolamwini JK, van Osdol 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. *Sci*, 275 (1997) 343.
35 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), 37 (2005) 1.
36 Jiang P, Sun X & Lu Z, Analysis of synonymous codon usage in Aeropyrum pernix K1 and other Crenarchaeota microorganisms. *J Genet Genomics*, 34 (2007) 275.
37 Moriyama EN & Powell JR, Gene length and codon usage bias in Drosophila melanogaster, Saccharomyces cerevisiae and Escherichia coli. *Nucleic Acids Res*, 26 (1998) 3188
38 Hou ZC & Yang N, Factors affecting codon usage in Yersinia pestis. *Sheng Wu Hua Xue Yu Sheng Wu Wu Li Xue Bao* (Shanghai), 35 (2003) 580.
39 Gupta SK & Ghosh TC, Gene expressivity is the main factor in dictating the codon usage variation among the genes in Pseudomonas aeruginosa. *Gene*, 273 (2001) 63.
40 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*, 101 (2004) 155.
41 Liu Q, Dou S, Ji Z & Xue Q, Synonymous codon usage and gene function are strongly related in *Oryza sativa*. *Biosystems*, 80 (2005) 123.
42 Najafabadi HS, Goodarzi H & Salavati R, Universal function-specificity of codon usage. *Nucleic Acids Res*, 37 (2009) 7014.
43 Dass JF & Sudandiradoss C, Insight into pattern of codon biasness and nucleotide base usage in serotonin receptor gene family from different mammalian species. *Gene*, 503 (2012) 92.
44 Fu M, Codon usage bias in herpesvirus. *Arch Virol*, 155 (2010) 391.
45 Roychoudhury S & Mukherjee D, A detailed comparative analysis on the overall codon usage pattern in herpesviruses. *Virus Res*, 148 (2010) 31.