VIROLOGICA SINICA, October 2012, 27 (5):303-315
DOI 10.1007/s12250-012-3270-9

Research Article

Characterization of Synonymous Codon Usage Bias in the Pseudorabies Virus *US1* Gene*

Meili Li[1], Zhiyao Zhao[1], Jianhong Chen[2], Bingyun Wang[2], Zi Li[1], Jian Li[1] and Mingsheng Cai[2]✉

1. Department of Pathogenic Biology and Immunology, Guangzhou Medical University, Guangzhou 510182, China;
2. Department of Veterinary Medicine, Foshan Science and Technology University, Foshan 528231, China

In the present study, we examined the codon usage bias between pseudorabies virus (PRV) *US1* gene and the *US1*-like genes of 20 reference alphaherpesviruses. Comparative analysis showed noticeable disparities of the synonymous codon usage bias in the 21 alphaherpesviruses, indicated by codon adaptation index, effective number of codons (ENc) and GC3s value. The codon usage pattern of PRV *US1* gene was phylogenetically conserved and similar to that of the *US1*-like genes of the genus Varicellovirus 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 *US1* gene with its reference alphaherpesviruses demonstrated that the codon usage bias of *US1*-like genes of 21 alphaherpesviruses had a very close relation with their gene functions. ENc-plot revealed that the genetic heterogeneity in PRV *US1* gene and the 20 reference alphaherpesviruses was constrained by G+C content, as well as the gene length. In addition, comparison of codon preferences in the *US1* gene of PRV with those of *E. coli*, yeast and human revealed that there were 50 codons showing distinct usage differences between PRV and yeast, 49 between PRV and human, but 48 between PRV and *E. coli*. Although there were slightly fewer differences in codon usages between *E.coli* and PRV, the difference is unlikely to be statistically significant, and experimental studies are necessary to establish the most suitable expression system for PRV *US1*. In conclusion, these results may improve our understanding of the evolution, pathogenesis and functional studies of PRV, as well as contributing to the area of herpesvirus research or even studies with other viruses.

Pseudorabies virus; *US1* gene; Alphaherpesvirus; Codon usage bias

Within the standard genetic codes utilized in a great deal of 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 organism, a phenomenon called codon usage bias[17,21,50]. Codon usage bias among synonymous codons has been described for many genes in various species[6,10,20,21,26,28,39,53]. Researches of the synonymous codon usage can uncover knowledge concerning the molecular evolution of individual gene. It is reported that synonymous codon usage bias may related with variant biological factors, such as GC compositions, gene length, mutation frequency and patterns, gene expression level, tRNA abundance, gene translation initiation signal and protein structure[4,14,19,27,37]. Further analysis discovered that synonymous codon usage pattern varied at different sites along a coding sequence[24], balances of strong versus weak base pair bonding[5,22], maintenance of DNA and RNA secondary structure[52], and translational efficiency and fidelity[26].

Aujeszky's disease, caused by 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 and wild animals incidentally[34,35,43,46,48]. Most of the previous research works have focused on the epidemiology and prevention of this disease[7,32,42,43,55]. However, the exact molecular biology characteristics about the PRV genome is still not well understood thus far. PRV *US1* gene, a 1050-base pair sequence encodes a putative polypeptide of 349 aa residues designated PICP22. The functions of *US1* gene products, such as herpes simplex virus 1 (HSV-1) ICP22[3,8,16,47] and varicella-zoster virus (VZV) ORF63[2, 11, 12, 41] that are the homologs of PICP22, in the herpesvirus life cycle have been extensively

Received: 2012-07-10, Accepted: 2012-09-12
* Foundation items: This work was supported by grants from the Scientific Research Foundation for the Ph.D., Guangzhou Medical University (2011C20); National Natural Science Foundation of China (31200120); Medical Scientific Research Foundation of Guangdong Province, China (B2012165), and the Guangzhou city-level key disciplines and specialties of Immunology (B127007).
✉ Corresponding author.
Phone/Fax:+86-757-85505103,Email:mingshengcai@hotmail.com

studied; however, the exact functional characteristics of PRV *US1* gene, as well as its codon usage bias is poorly understood. Given this background, it becomes crucial to analyze the codon preference used in PRV *US1* gene. In this study, we analyzed the synonymous codon usage data of PRV *US1* gene and compared it with the *US1*-like genes of 20 reference alphaherpesviruses. Then, we investigated how other factors may impact codon usage variation in the PRV *US1* gene and its reference species. Moreover, we compared the codon usage preference of PRV *US1* 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 *US1* gene (GenBank accession no. JF797219) and the *US1*-like genes of 20 reference alphaherpesviruses were obtained from the GenBank (Bethesda, Maryland, USA; http://www.ncbi.nlm.nih.gov/).

Molecular phylogenetic tree of ICP22-like proteins of the 21 reference alphaherpesviruses

To compare with those of ICP22-like proteins of the 21 reference alphaherpesviruses, for which nucleotide sequences are available in GenBank (listed in Table 1), the nucleotide sequences of PRV *US1* gene and its reference alphaherpesviruses were translated into aa sequence, then multiple sequence alignment and phylogenetic analysis (rooted tree) were performed by employing the Clustal V in MegAlign program of DNAStar (version 7.0, DNAStar, Inc.)[9].

Codon usage analysis of the PRV Becker strain *US1* gene and other 20 reference alphaherpesviruses

For each gene, codon usage was assessed by using the program CodonW 1.4 (http://codonw.sourceforge.net/). 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[51]. 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 equal usage of synonymous codons without reliance on sequence length or a given knowledge of preferred codons, although it is affected by base composition[13,45,56]. 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. A 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[51]. A heat map to represent the clustering of RSCU values was constructed with the CIMMiner software tool (http://discover.nci.nih.gov/cimminer)[54] with each column representing a specific codon and each row representing a different species (in the order as in Table 1). 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

$y = -34.757\ln(x) + 31.407$,

$y = -24.909\ln(x) + 214.24$ and

$y = 0.4553\ln(x) - 2.3871$,

were used for calculating the points for ENc-GC3s, ENc-Length and GC3s-Length, respectively.

Comparison of codon preferences of PRV Becker strain *US1* 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 *US1* 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 (The European Molecular Biology Open Software Suite, http://bioinfo.pbi.nrc.ca:8090/EMBOSS/)[38].

Statistical analysis

The correlations between codon usage variations among the PRV *US1* gene and 20 reference alphaherpesviruses and four indicators (CAI, ENc, GC3s and gene length) were estimated by using the SPSS 12.0 software package.

RESULTS

Molecular phylogenetic tree of the ICP22-like proteins in PRV Becker strain and the reference alphaherpesviruses

A phylogenetic tree on the basis of the deduced PICP22 and its ICP22-like proteins in the reference alphaherpesviruses (Table 1) is shown in Fig. 1. From Fig. 1 we can see that the general branching pattern is consistent with other previously published phylogenetic analyses[43, 46] and the ICP22-like proteins within the same genus or in the same microorganism are clustered together. Simultaneously, it is shown that the PICP22 of PRV Becker strain clusters with Bartha strain and Kaplan strain are initially placed in a monophyletic clade and then clustered with other members of the genus Varicellovirus of alphaherpesvirus, such as bovine herpesvirus 1 (BoHV-1),

Table 1. Nucleotide sequences of the PRV Becker strain *US1* gene and the *US1*-like genes of 20 reference alphaherpes viruses from different species

Rank	Genus	Virus name (Abbreviation)	Description	Natural host	GeneBank accession no.	Sequence length, bp
1	Varicellovirus	Suid herpesvirus 1(SuHV-1) Pseudorabies virus (PRV)	US1 gene, its product is ICP22, a herpesvirus immediate early regulatory protein named Herpes_IE68, which is required for gene regulation, cell cycle regulation and expression of a subset of late genes.	Sus scrofa (Pig)	JF797219	1050
2			US1 gene, its product is ICP22, a herpesvirus immediate early regulatory protein named Herpes_IE68, which is required for gene regulation, cell cycle regulation and expression of a subset of late genes.		JF797217	1140
3			US1 gene, its product is ICP22, a herpesvirus immediate early regulatory protein named Herpes_IE68, which is required for gene regulation, cell cycle regulation and expression of a subset of late genes.		JF797218	1095
4		Equid herpesvirus 1(EHV-1) Equine abortion virus (EAV)	ORF65 gene, its product is an immediate early regulatory protein ICP22, which is an in vitro host range factor (Herpes_IE68) that is required for gene regulation, cell cycle regulation and expression of a subset of late genes.	Equus caballus (Horse)	Z67986	837
5		Equine herpesvirus 4 (EHV-4) Equine rhinopneumonitis virus (ERV)	ORF65 gene, its product is an immediate early regulatory protein ICP22, which is an in vitro host range factor (Herpes_IE68) that is required for gene regulation, cell cycle regulation and expression of a subset of late genes.	Equus caballus (Horse)	NC_001844	855
6		Equid herpesvirus 9 (EHV-9) Gazelle herpesvirus 1 (GHV-1)	ORF65 gene, its product is an immediate early regulatory protein ICP22, which is an in vitro host range factor (Herpes_IE68) that is required for gene regulation, cell cycle regulation and expression of a subset of late genes.	Equus caballus (Horse)	NC_011644	876
7		Cercopithecine herpesvirus 9 (CeHV-9) Simian varicella virus (SVV)	ORF63 gene, its product is an immediate early regulatory protein ICP22, which is an in vitro host range factor (Herpes_IE68) that is required for gene regulation, cell cycle regulation and expression of a subset of late genes.	Erythrocebus patas (Monkey)	NC_002686	786
8		Felid herpesvirus 1 (FeHV-1)	US1 gene, its product is an immediate early regulatory protein ICP22, named Herpes_IE68.	Felidae (Cat)	NC_013590	1005
9		Bovine herpesvirus 1 (BoHV-1) Infectious bovine rhinotracheitis virus (IBRV)	BICP22 gene, its product is an immediate early regulatory protein ICP22, which is an in vitro host range factor (Herpes_IE68) that is required for gene regulation, cell cycle regulation and expression of a subset of late genes.	Bos taurus (Cattle)	CAA54262	903
10		Bovine herpesvirus 5 (BoHV-5) Bovine encephalitis herpesvirus (BEHV)	BICP22 gene, its product is an immediate early regulatory protein ICP22, which is an in vitro host range factor (Herpes_IE68) that is required for gene regulation, cell cycle regulation and expression of a subset of late genes.	Bos taurus (Cattle)	NC_005261	945
11		Human herpesvirus 3 (HHV-3) Varicella-zoster virus (VZV)	ORF63 gene, its product is an immediate early regulatory protein ICP22, which is an in vitro host range factor (Herpes_IE68) that is required for gene regulation, cell cycle regulation and expression of a subset of late genes.	Homo sapiens (Human)	NC_001348	837
12	Simplexvirus	Human herpesvirus 1 (HHV-1) Herpes simplex virus 1 (HSV-1)	US1 gene, its product is ICP22, a herpesvirus immediate early regulatory protein named Herpes_IE68.	Homo sapiens (Human)	X14112	1263

Table 1. Continue

Rank	Genus	Virus name (Abbreviation)	Description	Natural host	GeneBank accession no.	Sequence length, bp
13		Human herpesvirus 2 (HHV-2) Herpes simplex virus 2 (HSV-2)	US1 (ICP22) gene, US1 gene, its product is ICP22, a herpesvirus immediate early regulatory protein named Herpes_IE68.	Homo sapiens (Human)	Z86099	1242
14		Cercopithecine herpesvirus 1 (CeHV-1) Macacine herpesvirus 1 (McHV-1) Monkey B virus	US1 gene, its product is ICP22, a herpesvirus immediate early regulatory protein named Herpes_IE68, which is required for gene regulation, cell cycle regulation and expression of a subset of late genes.	Macaca mulatta (Monkey)	AB074432	1122
15		Cercopithecine herpesvirus 2 (CeHV-2) Simian agent 8 (SA8)	US1 gene, its product is ICP22, a herpesvirus immediate early regulatory protein named Herpes_IE68, which is required for gene regulation, cell cycle regulation and expression of a subset of late genes.	Cercopithecus aethiops (Monkey)	NC_006560	1296
16		Cercopithecine herpesvirus 16 (CeHV-16) Papiine herpesvirus 2 (PaHV-1)	US1 gene, its product is ICP22, a herpesvirus immediate early regulatory protein named Herpes_IE68, which is required for optimal ICP0 expression	Papio cynocephalus (Baboons)	DQ149153	1278
17	Iltovirus	Gallid herpesvirus 1 (GaHV-1) Infectious laryngotrach eitis virus (ILTV)	US1 (ICP22) gene, US1 gene, its product is ICP22, a herpesvirus immediate early regulatory protein named Herpes_IE68.	White Leghorn (Chicken)	AB016432	522
18	Mardivirus	Gallid herpesvirus 2 (GaHV-2)/ Marek's disease virus type 1 (MDV-1)	US1 (ICP22) gene, US1 gene, its product is ICP22, a herpesvirus immediate early regulatory protein named Herpes_IE68.	Gallus domesticus (Chicken)	AF147806	540
19		Gallid herpesvirus 3 GaHV-3) Marek's disease virus type 2 (MDV-2)	US1 (ICP22) gene, US1 gene, its product is ICP22, a herpesvirus immediate early regulatory protein named Herpes_IE68.	Gallus gallus (Chicken)	HQ840738	522
20		Meleagrid herpesvirus 1 (MeHV-1)	US1 (ICP22) gene, US1 gene, its product is ICP22, a herpesvirus immediate early regulatory protein named Herpes_IE68.	Meleagris gallopavo (turkey)	AF282130	498
21		Anatid herpesvirus 1 (AnHV-1) Duck enteritis virus (DEV)	US1 gene, its product is an immediate early regulatory protein ICP22 named Herpes_IE68.	Anatid species (Duck)	NC_013036	990

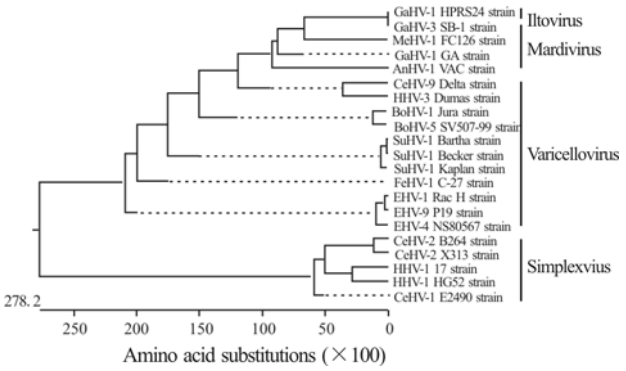

Fig. 1. Evolutionary relationship of the PRV Becker strain ICP22 protein with the ICP22-like proteins of 20 reference alphaherpesviruses from different species (Table 1). Phylogenetic tree of these proteins was generated by using the MEGALIGN (DNAStar) program with Clustal V multiple alignment software package and sequence distance indicated by the scale was calculated using the PAM250 matrix in LASERGENE.

BoHV-5, felid herpesvirus 1 (FeHV-1), equid herpesvirus 1 (EHV-1), EHV-4, EHV-9, human herpesvirus 3 (HHV-3, VZV) and cercopithecine herpesvirus 9 (CeHV-9). Therefore, we can conclude from the phylogenetic tree and the high aa sequence homology that the PRV PICP22 protein has a close evolutionary relationship with the members of the genus Varicellovirus of alphaherpesvirus, but certain differences nevertheless exist.

Codon usage analysis of the *US1* gene in PRV Becker strain and the reference alphaherpesviruses

The results obtained by CodonW analysis of 21 alphaherpesviruses species are shown in Table 2. Codon usage in the PRV *US1* gene and its homologous genes is extremely nonrandom, and the overall base composition of the *US1* 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 *US1* gene among the PRV Becker, Kaplan and Bartha strains. It can be seen in Table 2 that the CAI values of different alphaherpesviruses vary from 0.182 to 0.493, with a mean value of 0.387 and a standard deviation (SD) of 0.084 and their ENc values range from 28.4 to 61.0, with a mean value of 44.2 and SD of 12.1. Since most ENc values of the 21 alphaherpesviruses are lower than the average (ENc<40), the codon usage bias in the *US1*-like genes of the 21 alphaherpesviruses is accordingly slightly higher. Similarly, the $GC3_S$ content of each *US1*-like gene also confirm the homogeneity of synonymous codon usage among the different alphaherpesviruses, which vary from 34.44% to 95.68%, with a mean of 71.68% and a SD of 19.88%.

A plot of ENc against GC3s is an effective way of examining the heterogeneity of codon usage among a set of homologous genes[56]. If a specific gene is subject to G+C compositional constraint for shaping the codon usage pattern, it will lie on a continuous curve, representing random codon usage[29]. 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 *US1*-like gene in the 21 reference alphaherpesviruses 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*, *Saccharomyces cerevisiae*, *Pseudomonas aeruginosa* and *Yersinia pestis*[23,25,40]. Here, the plot of gene length against ENc (Fig. 2B) or against GC3s (Fig. 2C) shows the distribution for each gene. It appears that in the *US1*-like genes of the 21 reference alphaherpesviruses, longer genes have a much wider variance in ENc values and GC3s, suggesting that gene length may also play a role in shaping the codon usage bias of the 21 alphaherpesviruses.

Variation in the PRV Becker strain *US1* gene codon usage and aa composition

While the CAI, ENc and the related measures indicate the overall codon bias of PRV *US1* gene, it is also important to more closely examine the pattern of codon bias. Table 3 shows

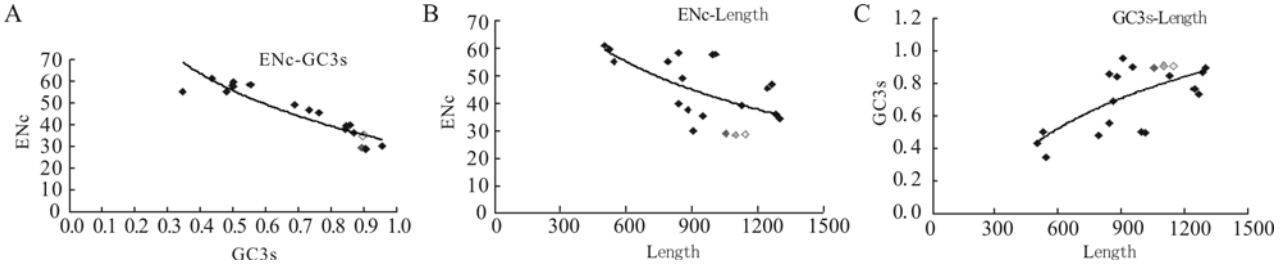

Fig. 2. Relationship between ENc, GC3s and gene length of the PRV Becker strain *US1* gene and the *US1*-like genes of 20 reference alphaherpesviruses. A: Plot of ENc versus GC3s for the PRV Becker strain *US1* gene and the *US1*-like genes of 20 reference alphaherpesviruses. ENc denotes the effective 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 *US1* gene and the *US1*-like genes of 20 reference alphaherpesviruses. C: Plot of GC3s versus gene length (bp) for the PRV Becker strain *US1* gene and the *US1*-like genes of 20 reference alphaherpesviruses. Red point represents the PRV Becker strain, yellow point represents the PRV Bartha strain and green point represents the PRV Kaplan strain.

the overall codon preference of the *US1* gene in the PRV Becker strain. From the RSCU values 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 six-fold and four-fold coding degeneracy. Moreover, Cys, Asp, His, Lys, Asn, Gln and Tyr also have a high level of diversity in codon usage bias, even though they only have two-fold or three-fold coding degeneracy. Altogether, although the most and the least frequencies used codons of all the aa are different, the analyzed PRV Becker strain *US1* gene shows significant preference for one or more than one postulate codon for each aa. 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 *US1* gene

To provide a visual representation of the variation in codon bias[15,36,44], we performed a cluster analysis of the codon usage pattern based on the PRV Becker strain *US1* gene and its 20 reference alphaherpesviruses according to the RSCU values (Table 4 and Fig. 3). From the figure we can see that PRV Becker, Kaplan and Bartha strains appear distinct from other alphaherpesviruses. They firstly cluster together and form a separate branch, then cluster with the members of genus Varicellovirus of alphaherpesvirus, such as BoHV-1, BoHV-5, EHV-1, EHV-4 and EHV-9, and subsequently cluster with other genera of alphaherpesvirus. This result fully indicates the internal relations of the codon usage pattern between PRV and other alphaherpesviruses, suggesting that the codon usage pattern of PRV has differences with other alphaherpesviruses, 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 Varicellovirus of alphaherpesvirus and is most different

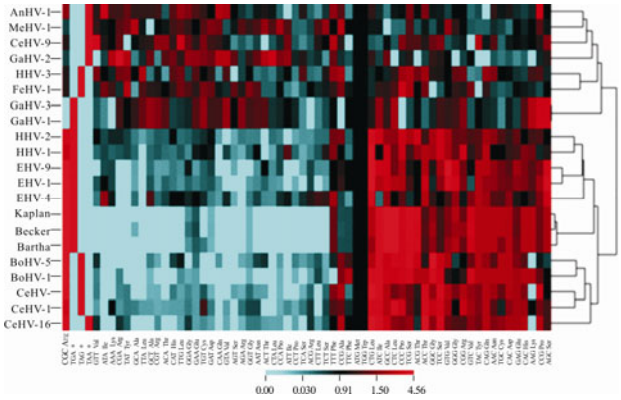

Fig. 3. Heat map of RSCU values for the 21 reference alphaherpesvirus species (clustered by the RSCU values, Table 4). See main text for details.

Table 2. Summary analysis of the PRV Becker strain *US1* gene and the *US1*-like genes of 20 reference alphaherpesviruses from different species

Rank	Virus name	Strain	CAI[a]	ENc[b]	Coding GC[c](%)	GC3s[d](%)
1	SuHV-1	Becker	0.460	29.016	73.24	89.43
2	SuHV-1	Bartha	0.493	28.755	72.81	90.53
3	SuHV-1	Kaplan	0.483	28.434	73.24	90.68
4	EHV-1	Rac H	0.464	39.739	69.06	85.66
5	EHV-4	NS80567	0.359	48.973	62.81	68.77
6	EHV-9	P19	0.455	37.448	68.15	84.25
7	CeHV-9	Delta	0.309	55.196	50.25	48.09
8	FeHV-1	C-27	0.313	57.692	54.03	49.85
9	BoHV-1	Jura	0.402	30.052	79.07	95.68
10	BoHV-5	SV507/99	0.389	35.436	75.13	89.84
11	HHV-3	Dumas	0.366	58.325	57.35	55.20
12	HHV-1	17	0.406	46.709	65.48	73.16
13	HHV-2	HG52	0.421	45.347	67.95	76.33
14	CeHV-1	E2490	0.417	39.344	69.25	84.49
15	CeHV-2	B264	0.485	34.456	73.30	89.58
16	CeHV-16	X313	0.444	36.017	72.38	86.85
17	GaHV-1	HPRS24	0.279	59.442	48.85	50.00
18	GaHV-2	GA	0.182	55.079	44.44	34.44
19	GaHV-3	SB-1	0.279	59.442	48.85	50.00
20	MeHV-1	FC126	0.321	61.000	44.58	43.37
21	AnHV-1	VAC	0.308	57.477	53.64	50.00

[a]: codon adaptation index, [b]: effective number of codons, [c]: G+C content in the *US1*-like gene, [d]: G+C content at the third positions of codons. All these indices were calculated by CodonW 1.4.

Table 3. The result of codon preference analysis in PRV Becker strain *US1* gene analyzed with the CUSP program

Codon	AA	Fraction	Frequency	Number	RSCU	Codon	AA	Fraction	Frequency	Number	RSCU
GCA	A(Ala)	0.036	2.857	1	0.143	CCA	P(Pro)	0.000	0.000	0	0.000
GCC	A	0.821	65.714	23	3.286	CCC	P	0.632	68.571	24	2.526
GCG	A	0.143	11.429	4	0.571	CCG	P	0.368	40.000	14	1.474
GCT	A	0.000	0.000	0	0.000	CCT	P	0.000	0.000	0	0.000
TGC	C(Cys)	1.000	11.429	4	2.000	CAA	Q(Gln)	0.000	0.000	0	0.000
TGT	C	0.000	0.000	0	0.000	CAG	Q	1.000	11.429	4	2.000
GAC	D(Asp)	0.945	148.571	52	1.891	AGA	R(Arg)	0.000	0.000	0	0.000
GAT	D	0.055	8.571	3	0.109	AGG	R	0.000	0.000	0	0.000
GAA	E(Glu)	0.317	57.143	20	0.635	CGA	R	0.000	0.000	0	0.000
GAG	E	0.683	122.857	43	1.365	CGC	R	0.720	51.429	18	4.320
TTC	F(Phe)	0.300	8.571	3	0.600	CGG	R	0.280	20.000	7	1.680
TTT	F	0.700	20.000	7	1.400	CGT	R	0.000	0.000	0	0.000
GGA	G(Gly)	0.103	11.429	4	0.410	AGC	S(Ser)	0.300	17.143	6	1.800
GGC	G	0.436	48.571	17	1.744	AGT	S	0.000	0.000	0	0.000
GGG	G	0.436	48.571	17	1.744	TCA	S	0.000	0.000	0	0.000
GGT	G	0.026	2.857	1	0.103	TCC	S	0.200	11.429	4	1.200
CAC	H(His)	1.000	8.571	3	2.000	TCG	S	0.500	28.571	10	3.000
CAT	H	0.000	0.000	0	0.000	TCT	S	0.000	0.000	0	0.000
ATA	I(Ile)	0.000	0.000	0	0.000	ACA	T(Thr)	0.000	0.000	0	0.000
ATC	I	1.000	5.714	2	3.000	ACC	T	0.333	11.429	4	1.333
ATT	I	0.000	0.000	0	0.000	ACG	T	0.667	22.857	8	2.667
AAA	K(Lys)	0.000	0.000	0	0.000	ACT	T	0.000	0.000	0	0.000
AAG	K	1.000	5.714	2	2.000	GTA	V(Val)	0.000	0.000	0	0.000
CTA	L(Leu)	0.000	0.000	0	0.000	GTC	V	0.611	31.429	11	2.444
CTC	L	0.667	22.857	8	4.000	GTG	V	0.389	20.000	7	1.556
CTG	L	0.333	11.429	4	2.000	GTT	V	0.000	0.000	0	0.000
CTT	L	0.000	0.000	0	0.000	TGG	W(Trp)	1.000	11.429	4	1.000
TTA	L	0.000	0.000	0	0.000	TAC	Y(Tyr)	1.000	17.143	6	2.000
TTG	L	0.000	0.000	0	0.000	TAT	Y	0.000	0.000	0	0.000
ATG	M(Met)	1.000	5.714	2	1.000	TAA	*	0.000	0.000	0	0.000
AAC	N(Asn)	1.000	5.714	2	2.000	TAG	*	0.000	0.000	0	0.000
AAT	N	0.000	0.000	0	0.000	TGA	*	1.000	2.857	1	3.000

Note: Fraction 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. Codons with a box appear to have lower frequency coding that amino acid. Triplets in bold face indicate the lowest frequency (frequency is zero) for coding that amino acid.

with other genera of alphaherpesvirus.

Comparison of the *US1* 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 *US1* 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 50 codons showing a PRV-to-yeast ratio higher than 2 or lower than 0.50 and 49 codons showing a PRV-to-human ratio higher than 2 or lower than 0.50, but 48 codons showing a PRV-to-*E. coli* ratio higher than 2 or lower than 0.50, indicating that large differences in the codon preferences exist for all three hosts. Although there were slightly fewer differences in codon usages between *E.coli* and PRV, the difference is unlikely to be statistically significant, and experimental studies would be necessary to establish the most suitable expression system for this virus.

DISCUSSION

In our study, a comprehensive analysis of codon usage including ENc, CAI value, GC content and the RSCU values of PRV Becker strain *US1* gene was carried out by using analytical techniques implemented in the CodonW 1.4 and EMBOSS CUSP programs. Subsequently these values were compared with those of the 20 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: a. PRV Becker strain *US1* gene and its 20 reference alphaherpesviruses take relatively

Table 4. RSCU values of the *US1* genes of PRV Becker strain and 20 reference alphaherpesviruses from different species

AA	Codon	SuHV-1 Becker strain	SuHV-1 Bartha strain	SuHV-1 Kaplan strain	EIIV-1	EIIV-4	EIIV-9	CeIIV-9	FeIIV-1	BoIIV-1	BoIIV-5	IIIIV-3	IIIIV-1	IIIIV-2	CeIIV-1	CeIIV-2	CeIIV-16	GaIIV-1	GaIIV-2	GaIIV-3	MeIIV-1	AnIIV-1
A	**GCA**	0.143	0.000	0.000	0.190	0.421	0.211	0.889	0.222	0.000	0.000	1.000	0.432	0.381	0.375	0.000	0.000	1.143	0.364	1.143	2.000	1.143
A	**GCC**	3.286	3.407	3.407	2.095	1.684	2.316	0.889	1.111	2.311	2.095	0.800	2.162	1.714	2.000	2.000	2.111	1.143	1.455	1.143	0.000	0.381
A	**GCG**	0.571	0.593	0.593	1.143	0.632	1.053	1.111	1.556	1.600	1.524	1.800	0.973	1.524	1.375	1.895	1.889	0.000	0.727	0.000	2.000	1.333
A	**GCT**	0.000	0.000	0.000	0.571	1.263	0.421	1.111	1.111	0.089	0.381	0.400	0.432	0.381	0.250	0.105	0.000	1.714	1.455	1.714	0.000	1.143
C	**TGC**	2.000	1.600	2.000	1.500	1.250	1.500	0.444	0.462	2.000	2.000	1.000	1.000	1.273	0.800	1.429	1.667	0.333	0.400	0.333	0.667	0.857
C	**TGT**	0.000	0.400	0.000	0.500	0.750	0.500	1.556	1.538	0.000	0.000	1.000	1.000	0.727	1.200	0.571	0.333	1.667	1.600	1.667	1.333	1.143
D	**GAC**	1.891	1.937	1.900	1.444	1.625	1.333	0.833	0.581	2.000	2.000	0.889	1.143	1.133	1.611	1.895	1.795	0.833	0.545	0.833	0.400	1.040
D	**GAT**	0.109	0.063	0.100	0.556	0.375	0.667	1.167	1.419	0.000	0.000	1.111	0.857	0.867	0.389	0.105	0.205	1.167	1.455	1.167	1.600	0.960
E	**GAA**	0.635	0.532	0.551	0.370	0.957	0.250	1.111	1.111	0.235	0.429	1.280	0.828	0.645	0.242	0.229	0.216	1.333	1.667	1.333	1.000	1.250
E	**GAG**	1.365	1.468	1.449	1.630	1.043	1.750	0.889	0.889	1.765	1.571	0.720	1.172	1.355	1.758	1.771	1.784	0.667	0.333	0.667	1.000	0.750
F	**TTC**	0.600	0.600	0.600	1.000	0.250	1.000	0.000	0.857	2.000	1.200	0.286	0.400	0.500	0.750	0.400	1.000	1.000	0.571	1.000	1.000	0.333
F	**TTT**	1.400	1.400	1.400	1.000	1.750	1.000	2.000	1.143	0.000	0.800	1.714	1.600	1.500	1.250	1.600	1.000	1.000	1.429	1.000	1.000	1.667
G	**GGA**	0.410	0.293	0.293	0.500	0.762	0.842	1.455	1.286	0.125	0.000	0.615	0.320	0.457	0.387	0.356	0.683	0.800	1.714	0.800	1.143	1.571
G	**GGC**	1.744	1.659	1.659	0.750	1.143	1.053	1.455	0.000	2.875	3.067	1.077	1.760	1.829	2.065	1.689	1.659	0.800	1.143	0.800	0.571	0.857
G	**GGG**	1.744	1.951	1.951	2.500	1.714	2.105	0.727	1.143	0.875	0.667	1.385	1.440	1.371	0.903	1.511	1.268	1.200	0.000	1.200	1.143	1.000
G	**GGT**	0.103	0.098	0.098	0.250	0.381	0.000	0.364	1.571	0.125	0.267	0.923	0.480	0.343	0.645	0.444	0.390	1.200	1.143	1.200	1.143	0.571
H	**CAC**	2.000	2.000	2.000	1.600	1.000	1.600	0.857	0.857	2.000	1.333	0.667	1.600	1.000	1.750	1.600	2.000	1.333	0.500	1.333	0.800	0.667
H	**CAT**	0.000	0.000	0.000	0.400	1.000	0.400	1.143	1.143	0.000	0.667	1.333	0.400	1.000	0.250	0.400	0.000	0.667	1.500	0.667	1.200	1.333
I	**ATA**	0.000	0.000	0.000	1.000	1.714	0.600	1.071	1.500	0.000	0.000	1.000	0.667	0.600	0.000	0.000	0.000	0.333	0.667	0.333	0.857	2.250
I	**ATC**	3.000	3.000	3.000	2.000	0.429	2.400	0.429	1.125	3.000	3.000	1.000	1.000	2.400	2.667	3.000	3.000	1.667	0.667	1.667	0.857	0.750
I	**ATT**	0.000	0.000	0.000	0.000	0.857	0.000	1.500	0.375	0.000	0.000	1.000	1.333	0.000	0.333	0.000	0.000	1.000	1.667	1.000	1.286	0.000
K	**AAA**	0.000	0.000	0.000	0.667	0.857	0.667	1.000	1.200	0.000	0.000	0.800	0.500	0.800	1.000	0.000	0.000	0.000	2.000	0.000	1.000	1.200
K	**AAG**	2.000	2.000	2.000	1.333	1.143	1.333	1.000	0.800	2.000	0.000	1.200	1.500	1.200	1.000	0.000	0.000	2.000	0.000	2.000	1.000	0.800
L	**CTA**	0.000	0.000	0.000	0.273	0.300	0.300	0.545	0.500	0.000	0.000	0.375	0.188	0.414	0.000	0.000	0.207	0.706	2.250	0.706	1.286	0.857
L	**CTC**	4.000	4.286	4.000	1.909	0.900	1.500	0.545	1.250	2.211	2.571	1.125	1.500	1.655	2.880	1.500	1.655	1.765	0.000	1.765	0.429	0.857
L	**CTG**	2.000	1.714	2.000	3.545	2.700	3.600	0.545	0.750	2.842	2.000	0.750	3.188	2.897	2.640	4.286	3.724	1.059	1.125	1.059	0.857	1.143
L	**CTT**	0.000	0.000	0.000	0.000	1.200	0.300	1.091	0.750	0.632	0.571	0.375	0.563	0.414	0.000	0.000	0.207	0.353	0.750	0.353	1.286	0.286
L	**TTA**	0.000	0.000	0.000	0.000	0.000	0.000	1.091	1.250	0.000	0.000	1.875	0.000	0.207	0.240	0.000	0.000	1.412	0.750	1.412	0.857	1.143
L	**TTG**	0.000	0.000	0.000	0.273	0.900	0.300	2.182	1.500	0.316	0.857	1.500	0.563	0.414	0.240	0.214	0.207	0.706	1.125	0.706	1.286	1.714
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
N	**AAC**	2.000	2.000	2.000	1.667	1.333	1.714	1.333	0.667	2.000	1.333	1.600	1.000	1.636	1.333	1.750	1.600	0.727	0.571	0.727	0.750	1.143
N	**AAT**	0.000	0.000	0.000	0.333	0.667	0.286	0.667	1.333	0.000	0.667	0.400	1.000	0.364	0.667	0.250	0.400	1.273	1.429	1.273	1.250	0.857
P	**CCA**	0.000	0.000	0.000	0.250	0.727	0.606	0.800	0.889	0.105	0.242	0.696	0.649	0.372	0.133	0.098	0.211	0.800	1.600	0.800	1.778	0.444
P	**CCC**	2.526	2.526	2.526	1.875	1.333	1.818	1.600	2.370	2.421	2.545	1.739	2.162	2.233	2.267	1.659	1.579	0.000	0.800	0.000	0.444	0.889

Table 4. Continue

AA	Codon	SuHV-1 Becker strain	SuHV-1 Bartha strain	SuHV-1 Kaplan strain	EHV-1	EHV-4	EHV-9	CeHV-9	FeHV-1	BoHV-1	BoHV-5	HHV-3	HHV-1	HHV-2	CeHV-1	CeHV-2	CeHV-16	GaHV-1	GaHV-2	GaHV-3	MeHV-1	AnHV-1
P	**CCG**	1.474	1.474	1.474	1.750	1.576	1.576	1.067	0.593	1.368	0.848	1.217	0.757	1.209	1.467	2.244	2.000	2.400	0.800	2.400	1.333	2.222
P	**CCT**	0.000	0.000	0.000	0.125	0.364	0.000	0.533	0.148	0.105	0.364	0.348	0.432	0.186	0.133	0.000	0.211	0.800	0.800	0.800	0.444	0.444
Q	**CAA**	0.000	0.000	0.000	0.286	0.250	0.000	1.143	0.857	0.000	0.286	0.286	0.000	0.250	0.000	0.000	0.000	1.200	2.000	1.200	1.500	2.000
Q	**CAG**	2.000	2.000	2.000	1.714	1.750	2.000	0.857	1.143	2.000	1.714	1.714	2.000	1.750	2.000	2.000	2.000	0.800	0.000	0.800	0.500	0.000
R	**AGA**	0.000	0.000	0.000	0.000	0.000	0.000	0.545	1.091	0.000	0.182	0.643	0.113	0.391	0.409	0.113	0.103	1.600	0.667	1.600	1.333	0.692
R	**AGG**	0.000	0.000	0.000	0.667	1.071	0.643	0.545	0.909	0.188	0.364	0.429	0.340	0.261	0.682	0.453	0.724	0.000	0.667	0.000	0.667	0.462
R	**CGA**	0.000	0.000	0.000	0.000	0.000	0.000	1.364	0.727	0.375	0.000	1.500	0.566	0.783	0.273	0.340	0.724	1.200	3.000	1.200	1.333	1.154
R	**CGC**	4.320	4.560	4.560	3.556	3.643	3.643	1.091	1.091	4.313	4.364	1.071	2.038	2.609	2.455	3.057	2.897	0.000	0.667	0.000	0.667	1.615
R	**CGG**	1.680	1.440	1.440	1.556	0.857	1.500	0.818	1.091	1.125	0.909	1.500	2.377	1.435	2.045	1.925	1.241	1.600	0.667	1.600	1.000	0.923
R	**CGT**	0.000	0.000	0.000	0.222	0.429	0.214	1.636	1.091	0.000	0.182	0.857	0.566	0.522	0.136	0.113	0.310	1.600	0.333	1.600	1.000	1.154
S	**AGC**	1.800	1.800	1.800	3.000	2.286	2.634	1.579	0.955	0.900	0.655	0.783	1.067	1.000	1.149	1.000	1.158	2.400	1.091	2.400	0.857	0.955
S	**AGT**	0.000	0.000	0.000	0.000	0.143	0.146	0.316	1.364	0.000	0.000	0.000	0.267	0.143	0.000	0.222	0.000	0.900	1.364	0.900	0.857	0.955
S	**TCA**	0.000	0.000	0.000	0.000	0.714	0.146	0.316	0.545	0.000	0.218	1.304	0.133	0.286	0.383	0.111	0.000	0.300	1.091	0.300	0.429	1.091
S	**TCC**	1.200	1.200	1.200	1.650	1.000	1.756	1.579	0.545	2.400	2.073	1.043	2.267	2.286	2.426	2.778	2.632	1.200	1.091	1.200	0.857	0.818
S	**TCG**	3.000	3.000	3.000	1.200	0.714	1.024	1.263	1.773	2.550	2.945	1.826	1.200	1.857	1.404	1.444	1.684	0.600	1.091	0.600	1.286	1.500
S	**TCT**	0.000	0.000	0.000	0.150	1.143	0.293	0.947	0.818	0.150	0.109	1.043	1.067	0.429	0.638	0.444	0.526	0.600	0.273	0.600	1.714	0.682
T	**ACA**	0.000	0.000	0.000	0.471	0.667	0.235	1.636	1.000	0.000	0.000	0.250	0.414	0.500	0.211	0.000	0.000	1.333	1.333	1.333	0.857	0.522
T	**ACC**	1.333	1.231	1.231	2.118	2.000	1.882	1.091	1.500	2.000	1.091	1.750	1.931	1.167	2.737	2.667	3.059	0.444	0.667	0.444	0.857	1.565
T	**ACG**	2.667	2.769	2.769	0.941	1.000	0.941	1.273	1.000	2.000	2.182	1.500	1.655	2.333	1.053	1.111	0.706	0.889	0.000	0.889	0.571	0.870
T	**ACT**	0.000	0.000	0.000	0.471	0.333	0.941	0.000	0.500	0.000	0.727	0.500	0.000	0.000	0.000	0.222	0.235	1.333	2.000	1.333	1.714	1.043
V	**GTA**	0.000	0.000	0.000	0.000	0.800	0.000	0.857	0.750	0.000	0.000	1.231	0.667	0.000	0.250	0.000	0.182	1.714	2.000	1.714	0.571	1.778
V	**GTC**	2.444	2.316	2.400	0.571	1.600	0.889	0.000	1.500	2.545	1.455	0.308	1.111	1.231	1.750	1.714	1.636	0.571	1.000	0.571	0.571	0.444
V	**GTG**	1.556	1.684	1.600	2.857	1.200	3.111	0.571	1.000	1.091	1.818	0.923	1.778	1.846	1.750	1.905	1.455	1.143	1.000	1.143	1.714	0.667
V	**GTT**	0.000	0.000	0.000	0.571	0.400	0.000	2.571	0.750	0.364	0.727	1.538	0.444	0.923	0.250	0.381	0.727	0.571	0.000	0.571	1.143	1.111
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
Y	**TAC**	2.000	2.000	2.000	2.000	2.000	2.000	2.000	0.000	2.000	2.000	0.400	1.333	1.333	1.600	1.600	1.500	1.000	0.000	1.000	0.500	0.600
Y	**TAT**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	2.000	0.000	0.000	1.600	0.667	0.667	0.400	0.400	0.500	1.000	2.000	1.000	1.500	1.400
*****	**TAA**	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	0.000	0.000	3.000	0.000	3.000	0.000	3.000	3.000
*****	**TAG**	0.000	0.000	0.000	0.000	0.000	0.000	0.000	3.000	3.000	3.000	3.000	0.000	0.000	3.000	3.000	0.000	0.000	0.000	0.000	0.000	0.000
*****	**TGA**	3.000	3.000	3.000	3.000	3.000	3.000	0.000	0.000	0.000	0.000	0.000	3.000	3.000	0.000	0.000	0.000	3.000	0.000	3.000	0.000	0.000

Table 5. Comparison of codon preferences between PRV Becker strain *US1* 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
GCA	A(Ala)	20.6	16.1	16.1	2.9	**0.1**	**0.2**	**0.2**
GCC	A	25.5	12.5	28.4	65.7	**2.6**	**5.3**	**2.3**
GCG	A	31.7	6.1	7.5	11.4	**0.4**	1.9	1.5
GCT	A	15.6	21.1	18.6	0.0	**0.0**	**0.0**	**0.0**
TGC	C(Cys)	6.9	4.7	12.2	11.4	1.7	**2.4**	0.9
TGT	C	5.5	8.0	10.0	0.0	**0.0**	**0.0**	**0.0**
GAC	D(Asp)	18.6	20.2	25.6	148.6	**8.0**	**7.4**	**5.8**
GAT	D	32.1	37.8	21.9	8.6	**0.3**	**0.2**	**0.4**
GAA	E(Glu)	38.2	48.5	29.0	57.1	1.5	1.2	2.0
GAG	E	17.7	19.1	39.9	122.9	**6.9**	**6.4**	**3.1**
TTC	F(Phe)	16.9	18.2	20.6	8.6	0.5	0.5	**0.4**
TTT	F	23.2	26.1	17.1	20.0	0.9	0.8	1.2
GGA	G(Gly)	9.0	10.9	16.4	11.4	1.3	1.0	0.7
GGC	G	27.9	9.7	22.5	48.6	1.7	**5.0**	**2.2**
GGG	G	11.3	6.0	16.3	48.6	**4.3**	**8.1**	**3.0**
GGT	G	24.4	24.0	10.8	2.9	**0.1**	**0.1**	**0.3**
CAC	H(His)	9.8	7.7	15.0	8.6	0.9	1.1	0.6
CAT	H	13.6	13.7	10.5	0.0	**0.0**	**0.0**	**0.0**
ATA	I(Ile)	5.4	17.8	7.7	0.0	**0.0**	**0.0**	**0.0**
ATC	I	24.2	17.0	21.6	5.7	**0.2**	**0.3**	**0.3**
ATT	I	29.8	30.4	16.1	0.0	**0.0**	**0.0**	**0.0**
AAA	K(Lys)	33.2	42.2	24.1	0.0	**0.0**	**0.0**	**0.0**
AAG	K	10.7	30.7	32.2	5.7	0.5	**0.2**	**0.2**
CTA	L(Leu)	4.0	13.3	7.8	0.0	**0.0**	**0.0**	**0.0**
CTC	L	11.0	5.4	19.8	22.9	**2.1**	**4.2**	1.2
CTG	L	50.9	10.4	39.8	11.4	**0.2**	1.1	**0.3**
CTT	L	11.7	12.1	13.0	0.0	**0.0**	**0.0**	**0.0**
TTA	L	13.9	26.7	7.5	0.0	**0.0**	**0.0**	**0.0**
TTG	L	14.0	27.0	12.6	0.0	**0.0**	**0.0**	**0.0**
ATG	M(Met)	27.0	20.9	22.2	5.7	**0.2**	**0.3**	**0.3**
AAC	N(Asn)	21.4	24.9	19.5	5.7	**0.3**	**0.2**	**0.3**
AAT	N	18.6	36.3	16.7	0.0	**0.0**	**0.0**	**0.0**
CCA	P(Pro)	8.5	18.2	16.7	0.0	**0.0**	**0.0**	**0.0**
CCC	P	5.8	6.8	20.1	68.6	**11.8**	**10.1**	**3.4**
CCG	P	21.8	5.3	6.9	40.0	1.8	**7.5**	**5.8**
CCT	P	7.3	13.6	17.3	0.0	**0.0**	**0.0**	**0.0**
CAA	Q(Gln)	15.0	27.5	12.0	0.0	**0.0**	**0.0**	**0.0**
CAG	Q	29.5	12.1	34.1	11.4	**0.4**	0.9	**0.3**
AGA	R(Arg)	2.9	21.3	11.5	0.0	**0.0**	**0.0**	**0.0**
AGG	R	1.9	9.2	11.4	0.0	**0.0**	**0.0**	**0.0**
CGA	R	3.9	3.0	6.3	0.0	**0.0**	**0.0**	**0.0**
CGC	R	21.0	2.6	10.7	51.4	**2.4**	**19.8**	**4.8**
CGG	R	6.3	1.7	11.6	20.0	**3.2**	**11.8**	1.7
CGT	R	20.3	6.5	4.6	0.0	**0.0**	**0.0**	**0.0**
AGC	S(Ser)	16.0	9.7	19.3	17.1	1.1	1.8	0.9
AGT	S	9.5	14.2	11.9	0.0	**0.0**	**0.0**	**0.0**
TCA	S	7.8	18.8	12.0	0.0	**0.0**	**0.0**	**0.0**
TCC	S	8.9	14.2	11.9	11.4	1.3	0.8	1.0
TCG	S	8.7	8.5	4.4	28.6	**3.3**	**3.4**	**6.5**
TCT	S	8.7	23.5	14.7	0.0	**0.0**	**0.0**	**0.0**
ACA	T(Thr)	8.2	17.8	15.1	0.0	**0.0**	**0.0**	**0.0**
ACC	T	22.8	12.6	19.4	11.4	0.5	0.9	0.6
ACG	T	14.8	7.9	6.1	22.9	1.5	**2.9**	**3.7**
ACT	T	9.1	20.3	13.0	0.0	**0.0**	**0.0**	**0.0**
GTA	V(Val)	11.1	11.8	7.2	0.0	**0.0**	**0.0**	**0.0**
GTC	V	15.1	11.6	14.6	31.4	**2.1**	**2.7**	**2.2**
GTG	V	25.5	10.6	28.4	20.0	0.8	1.9	0.7
GTT	V	18.5	22.0	11.0	0.0	**0.0**	**0.0**	**0.0**
TGG	W(Trp)	15.2	10.3	12.7	11.4	0.8	1.1	0.9
TAC	Y(Tyr)	12.1	14.6	15.5	17.1	1.4	1.2	1.1
TAT	Y	16.5	18.9	12.1	0.0	**0.0**	**0.0**	**0.0**
TAA	*	2.0	1.0	0.7	0.0	**0.0**	**0.0**	**0.0**
TAG	*	0.3	0.5	0.6	0.0	**0.0**	**0.0**	**0.0**
TGA	*	1.1	0.7	1.5	2.9	**2.6**	**4.1**	1.9

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 underlined and marked with bold indicates that the codon preference differs greatly, and vice versa[43].

similar codon usage patterns, although PRV Becker strain *US1* gene shows a few disparities of codon usage bias with its reference alphaherpesvirus species; b. the PRV Becker strain *US1* 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 *US1* gene. Since the *US1* gene in the PRV Becker strain is a CG-rich gene, it is reasonable that C and/or 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 alphaherpesviruses is observed to be restricted by the GC content and gene length.

Comparative analysis of *US1* genes in PRV and the reference herpesviruses indicated that synonymous codon usage in these genes are phylogenetically conserved. Table 2 shows that the *US1* genes in PRV, BoHV-1, BoHV-5, EHV-1, EHV-4 and HV-9, whose natural host is mammalian, have a stronger correlation than other *US1* genes of the reference alphaherpesviruses with avian host or human host, such as Anatid herpesvirus 1 (AnHV-1), Gallid herpesvirus 2 (GaHV-2), CeHV-2 and HHV-1.This indicates that the *US1* genes of alphaherpesviruses belonging to the same host may have similar sequence length and CAI value. 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[18, 49]. From the phylogenetic tree (Fig. 1) and cluster analysis results (Fig. 3) we can see that PRV is evolutionarily closer with BoHV-1 and BoHV-5 than FeHV-1, EHV-1, EHV-4 and EHV-9. Simultaneously, its codon usage pattern is also closer with BoHV-1and BoHV-5 than EHV-1, EHV-4 and EHV-9. Therefore, we can draw a conclusion that species has a certain influence to the preference of codon usage, but is less substantial than the influence of gene function, and the codon usage bias of PRV *US1* gene has a very close relation with its gene function.

Pertaining to the functions of *US1* gene product (ICP22) in the alphaherpesvirus life cycle, studies on the HSV-1 ICP22 and VZV ORF63, the homologue of PICP22, have been well documented and show that the *US1* gene, which is acting as a real immediate-early (IE) gene encoding for an IE protein, can modulate viral and cellular gene expression[1,8,30,31,33]. Besides, as an essential protein for HSV-1 replication, ICP22 also plays some other roles during infection, such as inducing the formation of discrete nuclear foci containing cellular chaperone proteins known as VICE domains[3] and ensuring proper virion morphology[47]. Moreover, VZV ORF63 is critical for efficient establishment of latency[2]. Therefore, because of the important roles played by HSV-1 ICP22 and VZV ORF63 in the course of infection, it means that PICP22 may also play a similar role to that of HSV-1 and VZV in the process of infection according to their phylogenetic conservation. However, it is not yet known what real biological functions of PICP22 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, no clear determination of the most suitable host could be made. Nevertheless, determination of an appropriate host remains a priority as the PRV *US1* gene optimized with host-preferential codons will probably improve the expression level of the PRV *US1* gene in a given host. Although the codon usages between PRV and *E. coli* were slightly better matched compared to the other hosts, they were not significantly different. Nevertheless, in a recent study, we successfully expressed the PICP22 protein in the *E. coli* expression system (unpublished data).

Taken together, analysis of codon usage pattern of PRV *US1* gene and a comparison of codon preference between PRV *US1* 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 *US1*. 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.

References

1. **Advani S J, Weichselbaum R R, Roizman B.** 2003. Herpes simplex virus 1 activates cdc2 to recruit topoisomerase II alpha for post-DNA synthesis expression of late genes. **Proc Natl Acad Sci U S A**, 100:4825–4830.
2. **Ambagala A P, Cohen J I.** 2007. Varicella-Zoster virus IE63, a major viral latency protein, is required to inhibit the alpha interferon-induced antiviral response. **J Virol**, 81:7844–7851.
3. **Bastian T W, Livingston C M, Weller S K, *et al.*** 2010. Herpes simplex virus type 1 immediate-early protein ICP22 is required for VICE domain formation during productive viral infection. **J Virol**, 84:2384–2394.
4. **Bernardi G.** 1986. Compositional constraints and genome evolution. **J Mol Evol**, 24:1–11.
5. **Blaisdell B E.** 1983. 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**, 19:226–236.
6. **Blake R D, Hinds P W.** 1984. Analysis of the codon bias in *E. coli* sequences. **J Biomol Struct Dyn**, 2:593–606.
7. **Boelaert F, Deluyker H, Maes D, *et al.*** 1999. Prevalence of herds with young sows seropositive to pseudorabies (Aujeszky's disease) in northern Belgium. **Prev Vet Med**, 41:239–255.
8. **Bowman J J, Orlando J S, Davido D J, *et al.*** 2009. Transient expression of herpes simplex virus type 1 ICP22 represses viral promoter activity and complements the replication of an ICP22

null virus. **J Virol**, 83:8733–8743.

9. **Burland T G.** 2000. DNASTAR's Lasergene sequence analysis software. **Methods Mol Biol**, 132:71–91.
10. **Cai M S, Cheng A C, Wang M S, *et al.*** 2009. Characterization of synonymous codon usage bias in the duck plague virus UL35 gene. **Intervirology**, 52:266–278.
11. **Cohen J I, Cox E, Pesnicak L, *et al.*** 2004. The varicella-zoster virus open reading frame 63 latency-associated protein is critical for establishment of latency. **J Virol**, 78:11833–11840.
12. **Cohen J I, Krogmann T, Bontems S, *et al.*** 2005. Regions of the varicella-zoster virus open reading frame 63 latency-associated protein important for replication in vitro are also critical for efficient establishment of latency. **J Virol**, 79:5069–5077.
13. **Comeron J M, Aguade M.** 1998. An evaluation of measures of synonymous codon usage bias. **J Mol Evol**, 47:268–274.
14. **D'Onofrio G, Ghosh T C, Bernardi G.** 2002. The base composition of the genes is correlated with the secondary structures of the encoded proteins. **Gene**, 300:179–187.
15. **Dass J F, Sudandiradoss C.** 2012. Insight into pattern of codon biasness and nucleotide base usage in serotonin receptor gene family from different mammalian species. **Gene**, 503:92–100.
16. **Durand L O, Roizman B.** 2008. Role of cdk9 in the optimization of expression of the genes regulated by ICP22 of herpes simplex virus 1. **J Virol**, 82:10591–10599.
17. **Duret L.** 2002. Evolution of synonymous codon usage in metazoans. **Curr Opin Genet Dev**, 12:640–649.
18. **Fu M.** 2010. Codon usage bias in herpesvirus. **Arch Virol**, 155:391–396.
19. **Gouy M, Gautier C.** 1982. Codon usage in bacteria: correlation with gene expressivity. **Nucleic Acids Res**, 10:7055–7074.
20. **Grantham R, Gautier C, Gouy M, *et al.*** 1981. Codon catalog usage is a genome strategy modulated for gene expressivity. **Nucleic Acids Res**, 9:r43–r74.
21. **Grantham R, Gautier C, Gouy M, *et al.*** 1980. Codon catalog usage and the genome hypothesis. **Nucleic Acids Res**, 8:r49–r62.
22. **Grosjean H, Fiers W.** 1982. Preferential codon usage in prokaryotic genes: the optimal codon-anticodon interaction energy and the selective codon usage in efficiently expressed genes. **Gene**, 18:199–209.
23. **Gupta S K, Ghosh T C.** 2001. Gene expressivity is the main factor in dictating the codon usage variation among the genes in Pseudomonas aeruginosa. **Gene**, 273:63–70.
24. **Hooper S D, Berg O G.** 2000. Gradients in nucleotide and codon usage along Escherichia coli genes. **Nucleic Acids Res**, 28:3517–3523.
25. **Hou Z C, Yang N.** 2003. Factors affecting codon usage in Yersinia pestis. **Acta Bioch Bioph Sin**, 35:580–586.
26. **Ikemura T.** 1985. Codon usage and tRNA content in unicellular and multicellular organisms. **Mol Biol Evol**, 2:13–34.
27. **Ikemura T.** 1981. 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**, 151:389–409.
28. **Jia R, Cheng A, Wang M, *et al.*** 2009. Analysis of synonymous codon usage in the UL24 gene of duck enteritis virus. **Virus Genes**, 38:96–103.
29. **Jiang P, Sun X, Lu Z.** 2007. Analysis of synonymous codon usage in Aeropyrum pernix K1 and other Crenarchaeota microorganisms. **J Genet Genomics**, 34:275–284.
30. **Jones J O, Arvin A M.** 2005. Viral and cellular gene transcription in fibroblasts infected with small plaque mutants of varicella-zoster virus. **Antiviral Res**, 68:56–65.
31. **Kalamvoki M, Roizman B.** 2011. The histone acetyltransferase CLOCK is an essential component of the herpes simplex virus 1 transcriptome that includes TFIID, ICP4, ICP27, and ICP22. **J Virol**, 85:9472–9477.
32. **Koppers-Lalic D, Reits E A, Ressing M E, *et al.*** 2005. Varicelloviruses avoid T cell recognition by UL49.5-mediated inactivation of the transporter associated with antigen processing. **Proc Natl Acad Sci U S A**, 102:5144–5149.
33. **Kost R G, Kupinsky H, Straus S E.** 1995. Varicella-zoster virus gene 63: transcript mapping and regulatory activity. **Virology**, 209:218–224.
34. **Li M L, Wang S, Cai M S, *et al.*** 2011. Characterization of molecular determinants for nucleocytoplasmic shuttling of PRV UL54. **Virology**, 417:385–393.
35. **Li M L, Wang S, Cai M S, *et al.*** 2011. 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**, 85:10239–10251.
36. **Liu Q, Dou S, Ji Z, *et al.*** 2005. Synonymous codon usage and gene function are strongly related in Oryza sativa. **Biosystems**, 80:123–131.
37. **Lobry J R, Gautier C.** 1994. Hydrophobicity, expressivity and aromaticity are the major trends of amino-acid usage in 999 Escherichia coli chromosome-encoded genes. **Nucleic Acids Res**, 22:3174–3180.
38. **Lu H, Zhao W M, Zheng Y, *et al.*** 2005. Analysis of synonymous codon usage bias in Chlamydia. **Acta Biochim Biophys Sin**, 37:1–10.
39. **Maruyama T, Gojobori T, Aota S, *et al.*** 1986. Codon usage tabulated from the GenBank genetic sequence data. **Nucleic Acids Res**, 14 Suppl:r151–197.
40. **Moriyama E N, Powell J R.** 1998. Gene length and codon usage bias in Drosophila melanogaster, Saccharomyces cerevisiae and Escherichia coli. **Nucleic Acids Res**, 26:3188–3193.
41. **Mueller N H, Walters M S, Marcus R A, *et al.*** 2010. Identification of phosphorylated residues on varicella-zoster virus immediate-early protein ORF63. **J Gen Virol**, 91:1133–1137.
42. **Muller T, Batza H J, Schluter H, *et al.*** 2003. Eradication of Aujeszky's disease in Germany. **J Vet Med B Infect Dis Vet Public Health**, 50:207–213.
43. **Muller T, Hahn E C, Tottewitz F, *et al.*** 2011. Pseudorabies virus in wild swine: a global perspective. **Arch Virol**, 156:1691–1705.
44. **Najafabadi H S, Goodarzi H, Salavati R.** 2009. Universal

function-specificity of codon usage. **Nucleic Acids Res**, 37:7014–7023.

45. **Novembre J A.** 2002. Accounting for background nucleotide composition when measuring codon usage bias. **Mol Biol Evol**, 19:1390–1394.
46. **Ono E, Amagai K, Yoshino S, *et al.*** 2004. Resistance to pseudorabies virus infection in transformed cell lines expressing a soluble form of porcine herpesvirus entry mediator C. **J Gen Virol**, 85:173–178.
47. **Orlando J S, Balliet J W, Kushnir A S, *et al.*** 2006. ICP22 is required for wild-type composition and infectivity of herpes simplex virus type 1 virions. **J Virol**, 80:9381–9390.
48. **Pomeranz L E, Reynolds A E, Hengartner C J.** 2005. Molecular biology of pseudorabies virus: impact on neurovirology and veterinary medicine. **Microbiol Mol Biol Rev**, 69:462–500.
49. **Roychoudhury S, Mukherjee D.** 2010. A detailed comparative analysis on the overall codon usage pattern in herpesviruses. **Virus Res**, 148:31–43.
50. **Sharp P M, Averof M, Lloyd A T, *et al.*** 1995. DNA sequence evolution: the sounds of silence. **Philos Trans R Soc Lond B Biol Sci**, 349:241–247.
51. **Sharp P M, Li W H.** 1987. The codon Adaptation Index--a measure of directional synonymous codon usage bias, and its potential applications. **Nucleic Acids Res**, 15:1281–1295.
52. **Wada A, Suyama A.** 1986. Local stability of DNA and RNA secondary structure and its relation to biological functions. **Prog Biophys Mol Biol**, 47:113–157.
53. **Wain-Hobson S, Nussinov R, Brown R J, *et al.*** 1981. Preferential codon usage in genes. **Gene**, 13:355–364.
54. **Weinstein J N, Myers T G, O'Connor P M, *et al.*** 1997. An information-intensive approach to the molecular pharmacology of cancer. **Science**, 275:343–349.
55. **White A K, Ciacci-Zanella J, Galeota J, *et al.*** 1996. Comparison of the abilities of serologic tests to detect pseudorabies-infected pigs during the latent phase of infection. **Am J Vet Res**, 57:608–611.
56. **Wright F.** 1990. The 'effective number of codons' used in a gene. **Gene**, 87:23–29.