Iranian Journal of Biotechnology. 2013 December;11(4): 214-22. DOI: 10.5812 / ijb.13811

Published Online 2013 November 20. **Research Article**

Identification of Synonymous Codon Usage Bias in the *Pseudorabies Virus UL31* Gene

Mingsheng Cai [1, 2], Zhiyao Zhao [1], Junyi Zhu [1], Jianhong Chen [2], Bingyun Wang [2], Zi Li [1], Meili Li [1, *]

[1] Department of Pathogenic Biology and Immunology, Guangzhou Medical University, Guangzhou, China
[2] Department of Veterinary Medicine, Foshan Science and Technology University, Foshan, China

*Corresponding author: Meili Li, Department of Pathogenic Biology and Immunology, Guangzhou Medical University, 195 Dongfeng Xi Road, Yuexiu, Guangzhou 510182, Guangdong, P.R. China. Tel: +1186-2081340200, Fax: +1186-2081340200, E-mail: meili_2011@hotmail.com

Received: July 25, 2013; **Revised:** August 30, 2013; **Accepted:** September 03, 2013

Background: Little knowledge of synonymous codon usage pattern of *pseudorabies virus* (*PRV*) genome, especially the *UL31* gene in the process for its evolution is available.
Objectives: In the present study, the codon usage bias between *PRV UL31* sequence and the *UL31*-like sequences was identified.
Materials and Methods: We used a comprehensive analysis on codon usage pattern in the *PRV UL31* gene and the *UL31*-like genes of 48 reference *herpesviruses* by calculating codon adaptation index, *ENc*, *RSCU* and EMBOSS assays.
Results: Cluster analysis demonstrated that the codon usage bias of *UL31*-like genes of 49 *herpesviruses* had a very close relation to their gene functions. In addition, comparison of codon preferences in the *UL31* gene of *PRV* with those of *E. coli*, yeast and human showed that there were 33 codons showing discrete usage differences between *PRV* and yeast, 24 between *PRV* and *E. coli*, but 22 between *PRV* and human. Although there were slightly fewer differences in codon usages between *PRV* and human, the difference is unlikely to be statistically significant, and experimental studies are necessary to establish the most suitable expression system for *PRV UL31*.
Conclusions: These results may further our comprehending of the evolution, pathogenesis and functional studies of *PRV*.

Keywords: Alpha herpesvirus; Codon Usage Bias; *Pseudorabies* Virus; *UL31* gene

1. Background

Within the ordinary genetic codes exploited in a large amount of varied ways, all amino acids (aa) are coded by two to six synonymous codons, except Met and Trp. However, degenerate codons are not employed at identical frequencies within organism. A phenomenon designated codon usage bias (1-3). Codon usage bias among synonymous codons has been documented for numerous genes in variant species (3-9). It is reported that synonymous codon usage bias may be connected with different biological factors (6, 10-13). Further analysis found that synonymous codon usage pattern changed at distinct sites along a coding sequence (14), balances of strong versus weak base pair bonding (15, 16), maintenance of DNA and RNA secondary structure (17), and translational efficiency and fidelity (6). Aujeszky's disease, which is provoked by the pathogenic factor of *Pseudorabies* virus (*PRV*) (also known as *Suid herpesvirus 1, SuHV-1*), is a regularly lethal disease with a global distribution that influences swine, mainly and other domestic and wild animals incidentally (18-22). PRV belongs to the genus *Varicellovirus*, subfamily *alpha herpesvirinae*, which is a swine *alpha herpesvirus* (20, 23-26)

2. Objectives

PRV UL31 gene, an 816-base pair sequence encodes a putative polypeptide of 271 aa residues named *UL31* protein. Regarding the role of *UL31* gene product played in the herpesvirus life cycle, herpes simplex virus 1 (HSV-1) *UL31* (27-36) and *Epstein-Barr* virus (EBV) *BFLF2* (37-39), the homologue of *PRV UL31*, have been extensively studied; however, the precise function propers of *PRV UL31* gene, as well as its codon usage bias is weakly understood.

3. Materials and Methods

3.1. Virus Species and Gene Sequences

Implication for health policy / practice / research / medical education:
These results may further our comprehending 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.

The nucleotide sequences of *PRV Becker* strain *UL31* gene (GenBank accession no. JF797219) and the *UL31*-like genes of 48 reference *herpesviruses* were gained from the GenBank.

3.2. Molecular Phylogenetic Tree of UL31-Like Proteins of the 49 Reference Herpesviruses

To compare with those of *UL31*-like proteins of the 49 reference herpesviruses, then multiple sequence alignment and phylogenetic analysis (rooted tree) were carried out by employing the DNAStar (version 7.0, DNAStar, Inc.) (40).

3.3. Codon Usage Analysis of the PRV Becker Strain UL31 Gene and other 48 Reference Herpesviruses

For each gene, codon usage was estimated by using CAI, CHIPS and CUPS programs of EMBOSS. *ENc*, GC3s and RSCU were analyzed (41, 43). 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 extent of base composition bias, and stands for the frequency of the nucleotide G + C at the synonymous third position of codons, except for Met, Trp and the stop codons. A heat map to represent the clustering of RSCU values was generated by the CIMMiner software tool (44) with each row representing a specific codon and each column representing a different species. Clustering was accomplished based on Euclidean distance and the average linkage method. Curves were created using a logarithmic distribution curve where y = -18.564 Ln(x) + 36.503, y = 1.8179 Ln (x) + 33.257 and y = 0.4539 Ln (x) - 2.4428 were used for calculating the points for *ENc*-GC3s, *ENc*-Length and GC3s-Length, respectively.

3.4. Statistical Analysis

The correlations between codon usage variations among the *PRV UL31* gene and 48 reference *herpesviruses* and four indicators (CAI, ENc, GC3s and gene length) were estimated by using the SPSS 12.0 software package.

4. Results

4.1. Molecular Phylogenetic tree of the UL31-Like Proteins in PRV Becker Strain and the Reference Herpesviruses

A phylogenetic tree as the basis of the deduced *UL31* and its *UL31*-like proteins in the reference herpesviruses was generated. We can see that the proteins could be preliminary separated into different subfamilies, i.e. *Alpha herpesvirinae*, *Beta herpesvirinae* and *Gamma herpesvirinae* (20, 21). Simultaneously, it is shown that the *UL31* of *PRV Becker* strain clusters with *Bartha*, *Kaplan* and *Ea* strains are initially placed in a monophyletic clade and then clustered with Bovine *herpesvirus 1* (*BoHV-1*) and *BoHV-5* of the genus *Varicellovirus* of subfamily *alpha herpesvirus*, sequently they clustered with other members of the reference species.

4.2. Codon Usage Analysis of the UL31 Gene in PRV Becker Strain and the Reference Herpesviruses

Codon usage in the *PRV UL31* gene and its homologous genes is highly nonrandom. However, there are some diverse patterns in the codon usage bias parameters of the *UL31* gene among the PRV *Becker*, *Kaplan*, *Bartha* and *Ea* strains. It can be seen in Table 1 that the CAI values of distinct *herpesviruses* vary from 0.602 to 0.842, with a mean value of 0.720 and a standard deviation (SD) of 0.062 and their *ENc* values range from 37.345 to 59.619, with a mean value of 45.644 and SD of 9.958. Compared to other species, the *ENc* values of different PRV strains are much lower (*ENc* < 40), the codon usage bias in the *UL31* -like genes of 49 reference species, especially the PRV is therefore, slightly high. If a specific gene is exposed to G + C compositional restriction for shaping the codon usage pattern, it will lie on a continuous curve, representing random codon usage (45). The *ENc* values of each *UL31* -like gene in the 49 reference *herpesviruses* are plotted against their corresponding *GC3s* in Figure 1.

Here, the plot of gene length against *ENc* (Figure 1 B) or against *GC3s* (Figure 1 C) shows the distribution for each gene. It seems that in the *UL31*-like genes of the 49 reference *herpesviruses*, shorter or longer genes both have a similar variance in *ENc* values and *GC3s*. It suggested that gene length may not play a role in shaping the codon usage bias of the 49 reference species. Similar results were also established in *P. aeruginosa;* duck plagued virus and SARS coronavirus (5, 46, 47).

4.3. Variation in the PRV Becker Strain UL31 Gene Codon Usage and aa Composition

While the *CAI*, *ENc* and the related measures present the overall codon bias of *PRV UL31* gene. Table 2 shows the overall codon preference of the *UL31* gene in the *PRV Becker* strain. Moreover, Cys, Asp, Glu, His, Ile, Lys, Asn, Gln and Tyr also have a high level of variety in codon usage bias, even though they only have two-fold or three-fold coding degeneracy. Altogether, although the most and the least frequencies' utilized codons of all the aa are disparate, the analyzed PRV Becker strain, *UL31* gene discloses meaningful preference for one or more than one suppose codon for each aa. However, a similar bias also exists at the first position, indicating a more complicated situation exists in reality.

Table 1. Summary Analysis of the *PRV Becker* Strain *UL31* Gene and the *UL31*-Like Genes of 40 Reference *Herpesviruses* from Different Species

Rank	Virus name	Strain	CAI [a]	ENc [b]	Coding GC [c] (%)	GC3s [d] (%)
1	*SuHV-1*	*Becker*	0.781	29.887	69.49	93.38
2	*SuHV-1*	*Bartha*	0.781	29.887	69.49	93.38
3	*SuHV-1*	*Kaplan*	0.781	29.887	69.49	93.38
4	*SuHV-1*	*Ea*	0.763	30.441	69.49	93.01
5	*EHV-1*	*Ab4*	0.698	55.871	54.03	58.72
6	*EHV-4*	*NS80567*	0.660	56.065	49.34	46.79
7	*EHV-9*	*P19*	0.696	55.147	54.03	59.33
8	*CeHV-9*	*Delta*	0.602	47.173	39.01	27.33
9	*FeHV-1*	*C-27*	0.614	55.089	42.41	35.21
10	*BoHV-1*	*Composite of 5 strains*	0.733	38.540	68.97	81.49
11	*BoHV-5*	*SV507 / 99*	0.787	31.869	73.60	93.42
12	*HHV-3*	*Dumas*	0.636	59.061	46.75	44.09
13	*HHV-1*	*CR38*	0.773	39.023	65.47	84.36
14	*HHV-2*	*HG52*	0.806	34.518	67.97	89.54
15	*CeHV-1*	*E2490*	0.802	31.731	71.08	93.05
16	*CeHV-2*	*B264*	0.807	31.340	72.02	93.77
17	*CeHV-16*	*X313*	0.803	29.793	73.20	96.73
18	*SaHV-1*	*MV 5-4*	0.740	37.991	67.98	85.41
19	*GaHV-2*	*GA*	0.619	52.085	39.20	31.23
20	*GaHV-3*	*HPRS24*	0.658	56.283	47.98	51.80
21	*MeHV-1*	*FC126*	0.646	57.498	41.83	38.89
22	*AnHV-1*	*VAC*	0.644	57.824	47.27	42.44
23	*GaHV-1*	*A489*	0.662	58.452	45.50	44.13
24	*PsHV-1*	*97-0001*	0.720	48.045	60.40	77.46
25	*HGTHV*		0.682	46.908	56.28	64.94
26	*HHV-5*	*U11*	0.748	36.433	60.74	84.08
27	*SaHV-3*	*SqSHV*	0.737	59.619	46.16	56.12
28	*CeHV-5*	*2715*	0.809	40.102	56.22	80.28
29	*MuHV-2*	*Maastricht*	0.778	27.345	65.90	94.72
30	*MuHV-1*	*C*	0.759	35.975	62.28	85.03
31	*HHV-7*	*RK*	0.624	47.309	32.44	28.46
32	*HHV-6*	*U1102*	0.672	50.639	42.39	49.43
33	*CavHV-2*	*21222*	0.714	47.403	53.19	71.43
34	*PCMV*	*489*	0.719	48.387	48.44	70.31
35	*MuHV-4*	*g2.4*	0.739	54.592	50.85	60.75
36	*AtHV-3*	*73*	0.648	46.789	35.11	30.53
37	*SaHV-2*	*11*	0.666	41.686	34.22	30.15
38	*HHV-8*	*GK18*	0.695	51.238	54.79	65.35
39	*CeHV-17*	*17577*	0.686	52.416	53.91	65.10
40	*BoHV-4*		0.710	46.979	40.94	48.99

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

Table 2. The Result of Codon Preferences in *PRV Becker* Strain *UL31* Gene Analyzed with the CUSP Program

Codon	AA	Fraction	Frequency	Number	RSCU	Codon	AA	Fraction	Frequency	Number	RSCU
GCA [a]	A(Ala)	0.029	3.676	1	0.118	**CCA**	P(Pro)	0.000	0.000	0	0.000
GC C[b]	A	0.529	66.176	18	2.118	CCC [b]	P	0.688	40.441	11	2.750
GCG	A	0.441	55.147	15	1.765	CCG	P	0.250	14.706	4	1.000
GCT	A	0.000	0.000	0	0.000	CCT [a]	P	0.062	3.676	1	0.250
TGC [b]	C(Cys)	1.000	22.059	6	2.000	**CAA**	Q(Gln)	0.000	0.000	0	0.000
TGT	C	0.000	0.000	0	0.000	CAG [b]	Q	1.000	22.059	6	2.000
GAC [b]	D(Asp)	0.944	62.500	17	1.889	AGA [a]	R(Arg)	0.069	7.353	2	0.414
GAT [a]	D	0.056	3.676	1	0.111	**AGG**	R	0.000	0.000	0	0.000
GAA [a]	E(Glu)	0.077	3.676	1	0.154	CGA [a]	R	0.069	7.353	2	0.414
GAG [b]	E	0.923	44.118	12	1.846	CGC [b]	R	0.621	66.176	18	3.724
TTC [b]	F(Phe)	0.615	29.412	8	1.231	CGG	R	0.241	25.735	7	1.448
TTT [a]	F	0.385	18.382	5	0.769	**CGT**	R	0.000	0.000	0	0.000
GGA [a]	G(Gly)	0.062	3.676	1	0.250	AGC [a]	S(Ser)	0.357	18.382	5	2.143
GGC [b]	G	0.625	36.765	10	2.500	**AGT**	S	0.000	0.000	0	0.000
GGG	G	0.312	18.382	5	1.250	**TCA**	S	0.000	0.000	0	0.000
GGT	G	0.000	0.000	0	0.000	**TCC**	S	0.000	0.000	0	0.000
CAC [b]	H(His)	1.000	22.059	6	2.000	TCG [b]	S	0.643	33.088	9	3.857
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 [b]	I	0.800	14.706	4	2.400	ACC [a]	T	0.375	11.029	3	1.500
ATT [a]	I	0.200	3.676	1	0.600	ACG [b]	T	0.625	18.382	5	2.500
AAA [a]	K(Lys)	0.125	3.676	1	0.250	**ACT**	T	0.000	0.000	0	0.000
AAG [b]	K	0.875	25.735	7	1.750	**GTA**	V(Val)	0.000	0.000	0	0.000
CTA	L(Leu)	0.000	0.000	0	0.000	GTC [a]	V	0.083	7.353	2	0.333
CTC [b]	L	0.484	55.147	15	2.903	GTG [b]	V	0.917	80.882	22	3.667
CTG	L	0.419	47.794	13	2.516	**GTT**	V	0.000	0.000	0	0.000
CTT	L	0.000	0.000	0	0.000	TGG	W(Trp)	1.000	3.676	1	1.000
TTA	L	0.000	0.000	0	0.000	TAC [b]	Y(Tyr)	0.929	47.794	13	1.857
TTG [a]	L	0.097	11.029	3	0.581	TAT[a]	Y	0.071	3.676	1	0.143
ATG	M(Met)	1.000	22.059	6	1.000	**TAA**	*	0.000	0.000	0	0.000
AA C[b]	N(Asn)	1.000	11.029	3	2.000	**TAG**	*	0.000	0.000	0	0.000
AAT	N	0.000	0.000	0	0.000	TGA [b]	*	1.000	3.676	1	3.000

[a] codons appear during the lower frequency coding of the amino acid.
[b] codons indicate the highest frequency in coding the amino acid. Triplets in bold face indicate the lowest frequency (frequency is zero) in coding the amino acid.

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.

4.4. Phylogenetic Persistence in Codon Usage Bias of the PRV Becker Strain UL31 Gene

To provide a visual representation of the variation in codon bias (48-50), we carried out a cluster analysis (Figure 2) of the codon usage pattern on the basis of the *PRV Becker* strain *UL31* gene and its 48 reference *herpesviruses* in accord with the RSCU values. From the figure, we can see that *PRV Becker*, *Kaplan*, *Bartha*and *Ea* strains appear different from other *herpesviruses*. They, firstly, cluster together and from a segregated branch, then they cluster with *BoHV-1* and *BoHV-5* of the genus *Varicellovirus* of subfamily *alphaherpesvirus* and *Cercopithecine herpesvirus 2* (*CeHV-2*) of the genus *Simplexvirus* of a subfamily *alphaherpesvirus*, subsequently they cluster with other mem-

bers of the reference species. This consequence wholly indicates the internal relations of the codon usage pattern between *PRV* and other *herpesviruses*, particularly the *alphaherpesviruses*, suggesting that the codon usage pattern of *PRV* has distinctions with other members of the reference species, the more distant the genetic relationship, the bigger the expected variation in the codon usage bias, and vice versa. Consequently, we can conclude that the codon usage pattern of *PRV* is fairly close to that of the members of genus *Varicellovirus* of *alpha herpesvirus*.

4.5. Comparison of the UL31 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 extent, across species. Here, the codon usage of *PRV Becker* strain *UL31* gene was compared with those of *E. coli*, yeast, and human to see which would be the most appropriate host for optimal expression. From Table 2 we can see that there are 33 codons showing a *PRV*-to-yeast ratio higher than 2 or lower than 0.50, and 24 codons showing a *PRV*-to-*E. coli* ratio higher than 2 or lower than 0.50, but 22 codons showing a *PRV*-to-human ratio higher than 2 or lower than 0.50, indicating that large diversities in the codon preferences exist for all three hosts. Although there were slightly fewer differences in codon usages between *PRV* and human, the difference is unlikely to be statistically significant, and experimental studies would be necessary to assess the most suitable expression system for this virus.

5. Discussion

In this study, the data of synonymous codon usage bias exhibited certain different distinctions existed for each *herpesvirus* from different species, and the result exposed that: a) *PRV Becker* strain *UL31* gene and its 48 reference *herpesviruses* adopt comparatively similar codon usage patterns; and b) the *PRVBecker* strain *UL31* gene opts to employ the codons with C and G at the third codon position. Furthermore, the biased tendency towards C and G is consistent with the high C + G content in *PRV Becker* strain *UL31* gene. Since the *UL31* gene in the *PRV Becker* strain is a CG-rich gene, it is rational that C and / or G ending codons are prevalent in the gene. In order to show the codon usage variation.

Table 1 shows that the *UL31* genes in *alpha herpesvirus* member of *PRV*, *BoHV-1*, *BoHV-5*, Human *herpesvirus 1* (*HHV-1*), *HHV-2*, *CeHV-1*, *CeHV-2*, *CeHV-16* and *Saimiriine herpesvirus 1* (*SaHV-1*), etc. whose natural host is mammalian, have a stronger correlation than other *UL31* genes of the reference *alpha herpesviruses* with avian host, such as *GaHV-1*, *GaHV-2*, *GaHV-3*, *Meleagrid herpesvirus 1* (*MeHV-1*) and *Anatid herpesvirus 1* (*AnHV-1*). It is critical to clarify the fundamental mechanisms of codon usage pattern to perceive the evolution of the species (51, 52). From the phylogenetic tree (Figure 3) and cluster analysis results (Figure 1) we can see that *PRV* is evolutionarily closer with *BoHV-1* and *BoHV-5* than *GaHV-1* and *PsHV-1*, etc. Simultaneously, its codon usage pattern is also closer with *BoHV-1* and *BoHV-5* than other members of the reference species. Accordingly, we can draw a conclusion that species has a certain effect to the preference of codon usage, but is less substantial than the influence of gene function, and the codon usage bias of *PRV UL31* gene has a very close connection with its gene function.

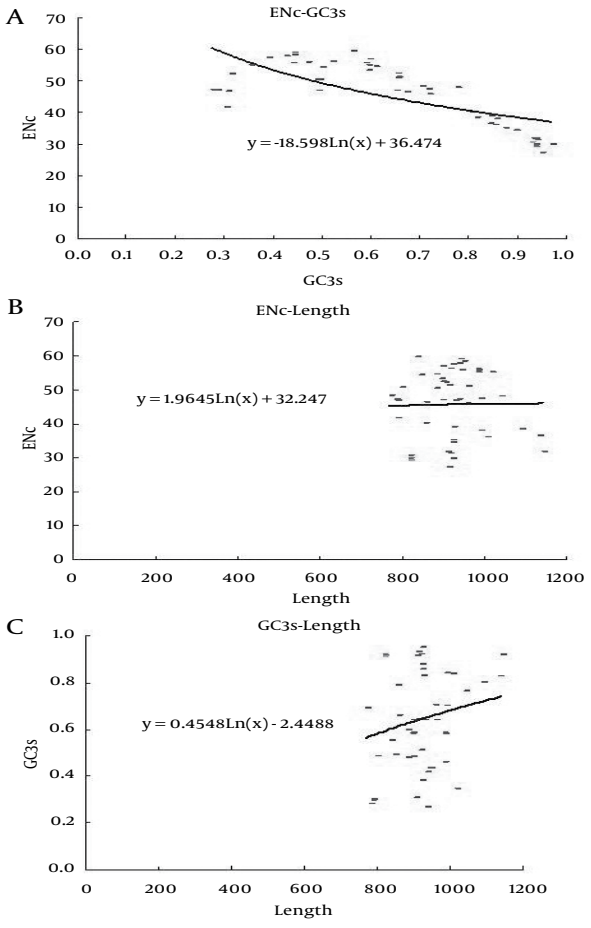

Figure 1. Relationship between ENc, GC3s and gene length of the *PRV Becker* strain *UL31* gene and the *UL31*-like genes of 48 reference herpesviruses. (A) Plot of ENc versus GC3s for the *PRV Becker* strain *UL31* gene and the *UL31*-like genes of 48 reference herpesviruses. ENc denotes the effective 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 *UL31* gene and the *UL31*-like genes of 48 reference *herpesviruses*. (C) Plot of GC3s versus gene length for the PRV Becker strain *UL31* gene and the *UL31*-like genes of 48 reference herpesviruses.

Figure 2. Heat Map of RSCU Values for the 49 Reference Herpesvirus Species (clustered by the RSCU values). Each row represents a various codon. Different species are represented in each column. Cluster is shown to the top based on euclidean distance and average method

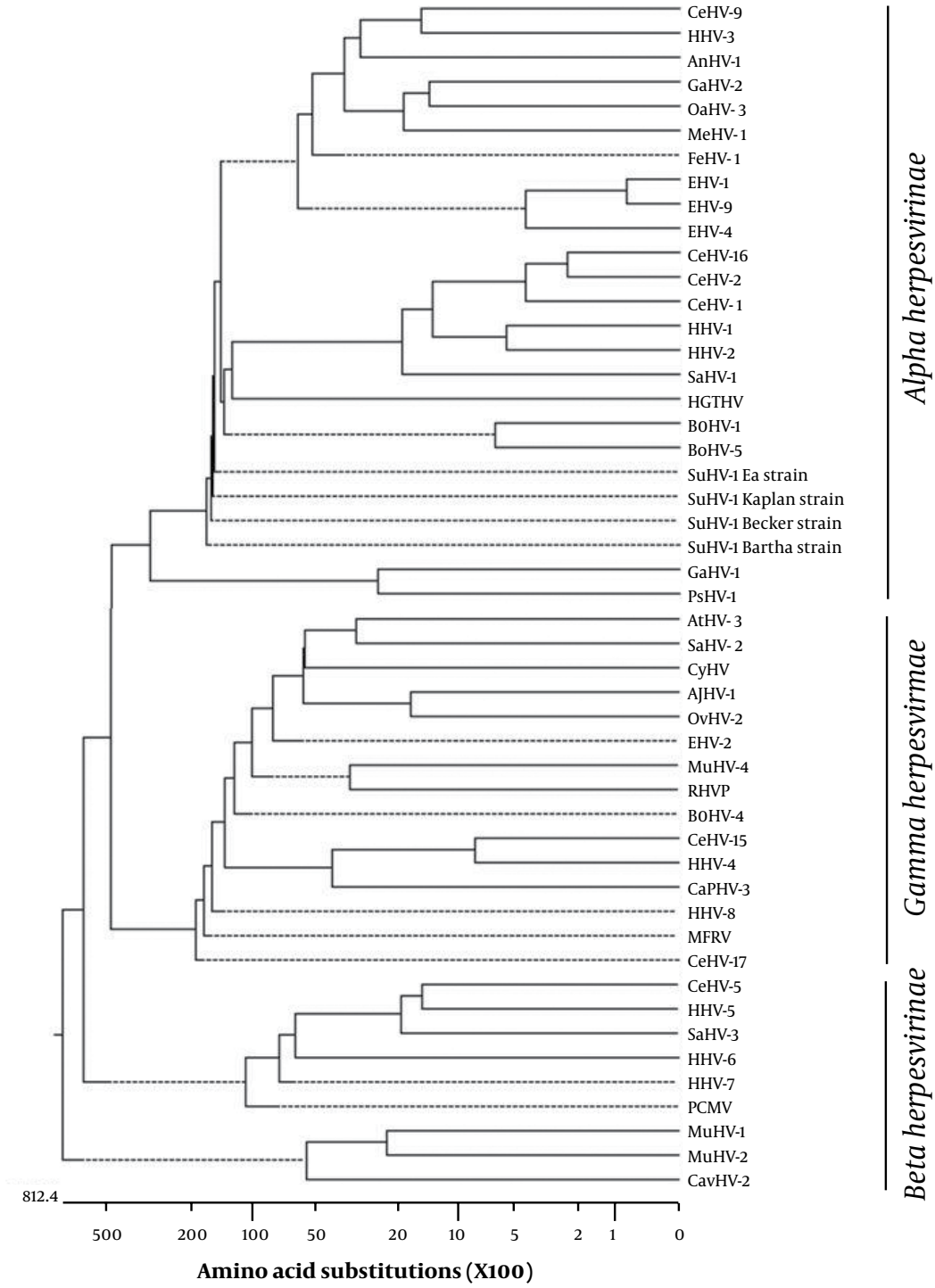

Figure 3. Evolutionary Relationship of the *PRV Becker* Btrain *UL31* Protein with the *UL31*-like Proteins of 48 Reference *Herpesviruses* from Different species. Phylogenetic tree of these proteins was generated by using the MEGALIGN (DNAStar) program with Clustal V multiple alignment software packages and sequence distance indicated by the scale was calculated using the PAM250 matrix in LASERGENE.

Bioinformatic's analysis reveals that *PRV UL31* protein is a member of *PHA03328* superfamily (data not shown), which encodes nuclear egress lamina protein *UL31* and is conserved throughout the *herpesvirus*. Although the biological characteristics of most of the *herpesviral UL31* homologues are ill understood at the present time, a common property is the interaction of *UL31* and *UL34,* and their co-localization at the nuclear or nuclear rim happened at different *herpesvirus* subfamilies, such as *alpha herpesvirus HSV-1* (32) and *HSV-2* (53), *beta herpesvirus* murine *cytomegalovirus* (*MCMV*) (54), and *gammaherpesvirus EBV* (38) and *Kaposi's sarcoma-associated herpesvirus* (*KSHV*) (55). Another interesting feature is their significance for primary envelopment and nuclear egress in all *herpesvirus* subfamily (32, 35, 39, 56). Therefore, because of the crucial roles acted by the counterpart of *PRV UL31* in *HSV*, *MCMV*, *EBV* and *KSHV* in the course of infection, it indicates that *PRV UL31* 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 roles of *UL31* have in the *PRV* life cycle, and the investigation 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 fashions in *E. coli*, yeast and human, no clear definition of the most appropriate host could be made. Although the codon usages between *PRV* and human were slightly better matched compared to the other hosts, they were not significantly different. Nevertheless, in a recent study, we successfully expressed the *PRV UL31* protein in the human embryonic kidney 293T expression system (unpublished data).

Taken together, analysis of codon usage pattern of *PRV UL31* gene and a comparison of codon preference between *PRV UL31* gene and other species can offer a foundation for understanding the relevant mechanism of biased usage of synonymous codons.

Acknowledgements

There is no acknowledgment.

Authors' Contributions

MSC and MLL contributed equally to this study, both of them equally carried out most of the experiments and wrote the manuscript. MSC and MLL have critically revised the manuscript and the experimental design. JYZ, JHC, BYW and ZL helped in experiments. All the authors read and approved the final manuscript.

Funding / Support

This work was supported by grants from the Natural Science Foundation of Guangdong Province (S2013040016596); Science and Technology New Star in Zhu Jiang, Guangzhou City (2013J2200018); National Natural Science Foundation of China (31200120); Medical Scientific Research Foundation of Guangdong Province, China (B2012165); Foundation for Distinguished Young Talents in Higher Education of Guangdong, China; First Batch of Youth Learning Backbone Teacher in Guangzhou Medical University; and Students' extracurricular scientific and technological activities in Guangzhou Medical University (2012A039 and 2012C007).

Financial Disclosure

The authors declare that they have no competing interests.

References

1. Bernardi G, Bernardi G. Compositional constraints and genome evolution. *J Mol Evo.* 1986;**24**(1-2):1-11.
2. 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 Evo.* 1983;**19**(3-4):226-236.
3. Blake RD, Hinds PW. Analysis of the codon bias in *E. coli* sequences. *J Biomol Struct Dyn.* 1984;**2**(3):593-606.
4. Boelaert F, Deluyker H, Maes D, Godfroid J, Raskin A, Varewijck H, et al. Prevalence of herds with young sows seropositive to pseudorabies (Aujeszky's disease) in northern Belgium. *Prev Vete Med.* 1999;**41**(4):239-255.
5. Burland TG. DNASTAR's Lasergene sequence analysis software. *Methods Mol Biol.* 2000;**132**:71-91.
6. Cai MS, Cheng AC, Wang MS, Zhao LC, Zhu DK, Luo QH, et al. Characterization of synonymous codon usage bias in the duck plague virus *UL35* gene. *Intervirology.* 2009;**52**(5):266-78.
7. Chang YE, Roizman B. The product of the *UL31* gene of herpes simplex virus 1 is a nuclear phosphoprotein which partitions with the nuclear matrix. *J Virol.* 1993;**67**(11):6348-56.
8. Comeron JM, Aguade M. An evaluation of measures of synonymous codon usage bias. *J Mol Evol.* 1998;**47**(3):268-74.
9. Dass JF, Sudandiradoss C. Insight into pattern of codon biasness and nucleotide base usage in serotonin receptor gene family from different mammalian species. *Gene.* 2012;**503**(1):92-100.
10. D'Onofrio G, Ghosh TC, Bernardi G. The base composition of the genes is correlated with the secondary structures of the encoded proteins. *Gene.* 2002;**300**(1 – 2):179-187.
11. Duret L. Evolution of synonymous codon usage in metazoans. *Curr Opin Gen & Dev.* 2002;**12**(6):640-649.
12. Fu M. Codon usage bias in herpesvirus. *Arch Virol.* 2010;**155**(3):391-6.
13. Fuchs W, Klupp BG, Granzow H, Osterrieder N, Mettenleiter TC. The interacting *UL31* and *UL34* gene products of pseudorabies virus are involved in egress from the host-cell nucleus and represent components of primary enveloped but not mature virions. *J Virol.* 2002;**76**(1):364-78.
14. Gonnella R, Farina A, Santarelli R, Raffa S, Feederle R, Bei R, et al. Characterization and intracellular localization of the Epstein-Barr virus protein BFLF2: interactions with BFRF1 and with the nuclear lamina. *J Virol.* 2005;**79**(6):3713-27.
15. Gouy M, Gautier C. Codon usage in bacteria: correlation with gene expressivity. *Nucleic Acids Res.* 1982;**10**(22):7055-74.
16. Granato M, Feederle R, Farina A, Gonnella R, Santarelli R, Hub B, et al. Deletion of Epstein-Barr virus BFLF2 leads to impaired viral DNA packaging and primary egress as well as to the production of defective viral particles. *J Virol.* 2008;**82**(8):4042-51.
17. Grantham R, Gautier C, Gouy M, Jacobzone M, Mercier R. Codon catalog usage is a genome strategy modulated for gene expressivity. *Nucleic Acids Res.* 1981;**9**(1):r43-74.
18. Grosjean H, Fiers W. Preferential codon usage in prokaryotic genes: the optimal codon-anticodon interaction energy and

the selective codon usage in efficiently expressed genes. *Gene*. 1982;**18**(3):199-209.
19. Gu W, Zhou T, Ma J, Sun X, Lu Z. Analysis of synonymous codon usage in SARS Coronavirus and other viruses in the Nidovirales. *Virus Res.* 2004;**101**(2):155-61.
20. Gupta SK, Ghosh TC. Gene expressivity is the main factor in dictating the codon usage variation among the genes in Pseudomonas *aeruginosa*. *Gene*. 2001;**273**(1):63-70.
21. Hooper SD, Berg OG. Gradients in nucleotide and codon usage along *Escherichia coli* genes. *Nucleic Acids Res.* 2000;**28**(18):3517-23.
22. Ikemura T. Codon usage and tRNA content in unicellular and multicellular organisms. *Mol Biol Evol.* 1985;**2**(1):13-34.
23. Jiang *Peng*, Sun *Xiao*, Lu *Zuhong*. Analysis of Synonymous Codon Usage in Aeropyrum pernix K1 and Other Crenarchaeota Microorganisms. *J Gene Genom*. 2007;**34**(3):275-284.
24. Kato A, Yamamoto M, Ohno T, Tanaka M, Sata T, Nishiyama Y, et al. Herpes simplex virus 1-encoded protein kinase UL13 phosphorylates viral Us3 protein kinase and regulates nuclear localization of viral envelopment factors *UL34* and *UL31*. *J Virol.* 2006;**80**(3):1476-86.
25. Koppers-Lalic D, Reits EA, Ressing ME, Lipinska AD, Abele R, Koch J, et al. Varicelloviruses avoid T cell recognition by *UL49*.5-mediated inactivation of the transporter associated with antigen processing. *Proc Natl Acad Sci U S A.* 2005;**102**(14):5144-9.
26. Lake CM, Hutt-Fletcher LM. The Epstein-Barr virus BFRF1 and BFLF2 proteins interact and coexpression alters their cellular localization. *Virology.* 2004;**320**(1):99-106.
27. Li M, Wang S, Cai M, Guo H, Zheng C. Characterization of molecular determinants for nucleocytoplasmic shuttling of PRV UL54. *Virology.* 2011;**417**(2):385-93.
28. Li M, Wang S, Cai M, Zheng C. 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.* 2011;**85**(19):10239-51.
29. Li M, Zhao Z, Chen J, Wang B, Li Z, Li J, et al. Characterization of synonymous codon usage bias in the pseudorabies virus US1 gene. *Virol Sin.* 2012;**27**(5):303-15.
30. Liang L, Baines JD. Identification of an essential domain in the *herpes simplex* virus 1 *UL34* protein that is necessary and sufficient to interact with *UL31* protein. *J Virol.* 2005;**79**(6):3797-806.
31. Liu Q, Dou S, Ji Z, Xue Q. Synonymous codon usage and gene function are strongly related in Oryza sativa. *Biosystems*. 2005;**80**(2):123-31.
32. Lobry JR, Gautier C. Hydrophobicity, expressivity and aromaticity are the major trends of amino-acid usage in 999 *Escherichia coli* chromosome-encoded genes. *Nucleic Acids Res.* 1994;**22**(15):3174-80.
33. Lotzerich M, Ruzsics Z, Koszinowski UH. Functional domains of murine cytomegalovirus nuclear egress protein M53 / p38. *J Virol.* 2006;**80**(1):73-84.
34. Maruyama T, Gojobori T, Aota S, Ikemura T. Codon usage tabulated from the GenBank genetic sequence data. *Nucleic Acids Res.* 1986;**14 Suppl**:r151-97.
35. Miller MS, Furlong WE, Pennell L, Geadah M, Hertel L. RASCAL is a new human cytomegalovirus-encoded protein that localizes to the nuclear lamina and in cytoplasmic vesicles at late times postinfection. *J Virol.* 2010;**84**(13):6483-96.
36. Mou F, Wills E, Baines JD. Phosphorylation of the U(L)31 protein of herpes simplex virus 1 by the U(S)3-encoded kinase regulates localization of the nuclear envelopment complex and egress of nucleocapsids. *J Virol.* 2009;**83**(10):5181-91.
37. Müller T, Bätza HJ, Schlüter H, Conraths FJ, Mettenleiter TC. Eradication of Aujeszky's Disease in Germany. *Journal of Veterinary Medicine, Series B.* 2003;**50**(5):207-213.
38. Muller T, Hahn EC, Tottewitz F, Kramer M, Klupp BG, Mettenleiter TC, et al. Pseudorabies virus in wild swine: a global perspective. *Arch Virol.* 2011;**156**(10):1691-705.
39. Najafabadi HS, Goodarzi H, Salavati R. Universal function-specificity of codon usage. *Nucleic Acids Res.* 2009;**37**(21):7014-23.
40. Yang K, Baines JD. Selection of HSV capsids for envelopment involves interaction between capsid surface components p*UL31*, pUL17, and pUL25. *Proc Natl Acad Sci U S A.* 2011;**108**(34):14276-81.
41. Novembre J A. Accounting for Background Nucleotide Composition When Measuring Codon Usage Bias. *Mol Biol Evol.* 2002;**19**(8):1390-1394.
42. Ye GJ, Roizman B. The essential protein encoded by the *UL31* gene of herpes simplex virus 1 depends for its stability on the presence of UL34 protein. *Proc Natl Acad Sci U S A.* 2000;**97**(20):11002-7.
43. Ono E, Amagai K, Yoshino S, Taharaguchi S, Inobe M, Uede. Resistance to pseudorabies virus infection in transformed cell lines expressing a soluble form of porcine herpesvirus entry mediator C. *JGeneral Virol.* 2004;**85**(1):173-178.
44. Pomeranz LE, Reynolds AE, Hengartner CJ. Molecular biology of pseudorabies virus: impact on neurovirology and veterinary medicine. *Microbiol Mol Biol Rev.* 2005;**69**(3):462-500.
45. Reynolds AE, Ryckman BJ, Baines JD, Zhou Y, Liang L, Roller RJ. U(L)31 and U(L)34 proteins of herpes simplex virus type 1 form a complex that accumulates at the nuclear rim and is required for envelopment of nucleocapsids. *J Virol.* 2001;**75**(18):8803-17.
46. Roberts KL, Baines JD. *UL31* of herpes simplex virus 1 is necessary for optimal NF-kappaB activation and expression of viral gene products. *J Virol.* 2011;**85**(10):4947-53.
47. Roychoudhury S, Mukherjee D. A detailed comparative analysis on the overall codon usage pattern in herpesviruses. *Virus Res.* 2010;**148**(1-2):31-43.
48. Santarelli R, Farina A, Granato M, Gonnella R, Raffa S, Leone L, et al. Identification and characterization of the product encoded by ORF69 of Kaposi's sarcoma-associated herpesvirus. *J Virol.* 2008;**82**(9):4562-72.
49. Sharp PM, Averof M, Lloyd AT, Matassi G, Peden JF. DNA sequence evolution: the sounds of silence. *Philos Trans R Soc Lond B Biol Sci.* 1995;**349**(1329):241-7.
50. Simpson-Holley M, Baines J, Roller R, Knipe DM. Herpes simplex virus 1 U(L)31 and U(L)34 gene products promote the late maturation of viral replication compartments to the nuclear periphery. *J Virol.* 2004;**78**(11):5591-600.
51. Wada A, Suyama A. Local stability of DNA and RNA secondary structure and its relation to biological functions. *Prog Biophys Mol Biol.* 1986;**47**(2):113-157.
52. Wain-Hobson S, Nussinov R, Brown RJ, Sussman JL. Preferential codon usage in genes. *Gene*. 1981;**13**(4):355-364.
53. Weinstein JN, Myers TG, O'Connor PM, Friend SH, Fornace AJ, Kohn KW, et al. An Information-Intensive Approach to the Molecular Pharmacology of Cancer. *Science*. 1997;**275**(5298):343-349.
54. 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.* 1996;**57**(5):608-11.
55. Wright F. The 'effective number of codons' used in a gene. *Gene*. 1990;**87**(1):23-29.
56. Yamauchi Y, Shiba C, Goshima F, Nawa A, Murata T, Nishiyama Y. Herpes simplex virus type 2 UL34 protein requires *UL31* protein for its relocation to the internal nuclear membrane in transfected cells. *J Gen Virol.* 2001;**82**(Pt 6):1423-8.