Characterization of Synonymous Codon Usage in the Newly Identified Duck Plague Virus UL16 Gene

Qin He[1], Mingshu Wang[1,2], Anchun Cheng[1,2,3], Dekang Zhu[1,2], Xiaoyue Chen[1,2,3], Renyong Jia[2], Qihui Luo[2], Yi Zhou[2], and Zhengli Chen[2]

[1] Avian Disease Research Center College of Veterinary Medicine of Sichuan Agricultural University, Ya'an, Sichuan, 625014, China
[2] Key Laboratory of Animal Diseases and Human Health of Sichuan Province, Sichuan Agricultural University, Ya'an, Sichuan, 625014, China
[3] Epizootic Diseases Institute of Sichuan Agricultural University, Ya'an, Sichuan, 625014, China
chenganchun@vip.163.com

Abstract. A comparative analysis of the codon usage bias in the newly identified UL16 gene(GenBank accession no.EU195095) of DPV and the UL16 gene of 22 reference herpesviruses was performed. In this study, the synonymous codon usage bias of UL16 gene in the 23 herpesviruses have been analyzed and the results showed obvious differences by the CAI, RSCU, ENC and GC_{3s}. The results revealed that the synonymous codons with A and T at the third codon positon have widely usage in the codon of UL16 gene of DPV. The ENC-GC_{3s} plot revealed that the genetic heterogeneity in UL16 gene of herpesviruses was constrained by G+C content at the third codon position. The phylogenetic analysis suggested that DPV was evolutionarily closer to herpesviruses which further clustered into Alphaherpesvirinae. Furthermore the ORF of DPV UL16 gene has sequential rare codons. There were 21 codons showing distinct usage differences between DPV with Escherichia coli, 19 codons showing distinct usage differences between DPV with yeast, and 20 between DPV and Human. Therefore the Escherichia coli, Yeast and Human expression system were suitable for the expression of DPV UL16 gene if some codons could be optimized.

Keywords: Duck plague virus, UL16 gene, codon usage bias.

1 Introduction

Codon Usage Bias was defined as deviation from equal usage of synonymous codons[1]. Within the standard genetic codes used in a number of different ways, all amino acids except Met and Trp are coded by 2–6 synonymous codons, but the synonymous codon usage are not used equally both within and between genomes[2]. Previous research have been showed that codon usage bias may be very complicated and associated with various biological factors, such as gene expression level[3], gene length[4], gene translation initiation signal[5], protein amino acid composition[6], protein structure[7], tRNA abundance[8,9], mutation frequency and patterns[6,10], GC composition[11,12], and environmental factors[13]. For a number of different

E. Zhu and S. Sambath (Eds.): Information Tech. and Agricultural Eng., AISC 134, pp. 747–761.
springerlink.com

organisms, it was suggested that codon usage is best explained by selection for tRNA abundance, gene expression levels, and translational optimization[14]. Recently, it was also suggested that codon usage is related to gene function[15,16] and the the evolutionary history of an organism in metazoan genomes[17]. Codon usage bias can reveal information about the molecular evolution of individual genes and provide data to train genome-specific gene recognition algorithms which recognize protein coding regions in uncharacterized genomic DNA. Codon usage bias is also widely studied in particular organisms to achieve high expression of heterologous proteins in vitro and to vaccine design where the efficient expression of viral proteins may be required to generate immunity[18,19]. Recently, analyses of the patterns of codon usage bias of herpesviruses are primarily focused on the pseudorabies virus (PRV)[20], herpes simplex virus type 1 (HSV-1)[21], Epstein-Barr virus [22], but the codon usage bias in DPV genome was known little.

Duck plague, which is caused by duck plague virus(DPV), a member of the Herpesviridae family, is an acute, lethal and contagious disease that occurs worldwide among domestic and wild ducks, geese, swans, and other water fowl, with migratory waterfowl contributing to spread between continents[23,24]. Now most of the previous research work has focused on the epidemiology and prevention of this disease. However, the molecular biology information about the DPV genome is limited. Recently, the UL16 gene was isolated and identified from DPV CHv strain in our laboratory[25]. The UL16 gene of herpes simplex virus encodes tegument proteins, which are conserved throughout the herpesvirus family[26]. Little is known about the molecular informations and function of DPV UL16 protein at present. In this study, we first analyzed the synonymous codon usage in the UL16 gene of DPV and compared with those of 22 other species of herpesviruses. Moreover, the codon usage bias in the DPV genes was compared with those of *Escherichia coli*, Yeast and Human. In addtion, we also investigated the rare codons of UL16 gene. All these datas might provide some insights into the features of the DPV genome, the possible function of DPV UL16 gene as well as the suitable expression system in in vitro.

2 Materials and Methods

2.1 Virus Species and Gene Sequences

The DPV CHv strain, a high-virulence strain of DPV, was obtained from Key Laboratory of Animal Disease and Human Health of Sichuan Province. The UL16 gene of the DPV CHv strain was isolated and identified by our laboratory. The nucleotide sequences of the UL16 gene of 22 reference herpesviruses were obtained from the NCBI GenBank nucleotide database (table 2).

2.2 Analysis on Codon Usage in UL16 Gene of DPV and 22 Reference Herpesviruses

For each gene, codon usage was estimated by using CAI, CHIPS and CUSP program of EMBOSS. The RSCU values of UL16 gene was analyzed with the CodonW. Generally, The‘Effective Number of Codons’(ENC) was often used to quantify the codon usage bias of an ORF in an individual gene. The values range from 20 to 61.

In an extremely biased gene where only one codon is used for each amino acid, this value would be 20; in an unbiased gene, it would be 61[27]. The codon adaptation index (CAI) value was regarded as a reference set of highly expressed genes from a species to assess the relative merits of each codon. Higher CAI value expected stronger codon usage bias and higher expression level, whereas the reverse was true for lower CAI value. The relative synonymous codon usage (RSCU) value was used to examine the codon usage variation among the genes without the confounding influence of amino acid composition. It is defined as the ratio of the observed frequency of codons to the expected frequency if all the synonymous codons for those amino acids are used equally[28]. GC_{3s} is a good indicator of the extent of base composition bias, which represents the frequency of the nucleotide G + C at the synonymous 3rd codon position, excluding Met, Trp and the stop codons. The codon usage pattern across genes was examined by the ENC-plot, which is a plot of ENC versus GC_{3s}.

2.3 Molecular Characterization and Phylogenetic Analysis of the DPV UL16 Gene

The nucleotide sequences of the DPV UL16 gene and 22 reference herpesviruses were translated into amino acid sequences by using DNASTAR software. After this, multiple sequence alignment and phylogenetic analysis were performed for the UL16 genes of 23 herpesviruses with CLUSTAL-X and TREEVIEW software.

2.4 Analysis the Rare Codons of DPV UL16 Gene

The proteins in heterologous hosts are often difficult to express or at very low levels. They might contain codons that are rarely used in the desired host. Log to http://nihserver.mbi.ucla.edu/RACC/ to analyze the rare codons of the DPV UL16 gene.

2.5 Comparison of Codon Preferences of DPV UL16 Gene with those of E. coli, Yeast and Human

To examine whether different species follow with the same codon usage rule, Codon usage bias in the DPV UL16 gene was determined with the SPSS 13.0 software, and we compare the UL16 codon usage bias among DPV, *E. coli*, yeast and Human (create a codon usage table). The database of the codon usage in E. coli, yeast and Human is available at http://www.kazusa.or.jp/codon.

3 Results

3.1 Variation in DPV UL16 Codon Usage and Amino Acid Composition

While RSCU and the related measures indicate the overall DPV UL16 codon bias, it is also important to closely investigate the pattern of codon bias. Table 1 shows the codon preferences of DPV UL16 gene. Sixty-one codons (excepting Met and the termination codons) in the polypeptide, with twenty-six synonymous codons strong

bias toward A-ended and T-ended at the third codon position, were used. A high level of diversity in codon usage bias existed for coding the Ala, Gly, Leu, Pro, Arg, Ser, Thr and Val amino acids because they have a 6-fold and 4-fold coding degeneracy.

3.2 Codon Usage Analysis of the UL16 Genes of DPV and Reference Herpesviruses

The results obtained by CodonW and EMBOSS analysis of the ENC, CAI, coding G + C content (GC%) and the G + C contents at the third codon position content (GC_{3S}%) of 23 herpesviruses species are shown in table 2 . Codon usage in the UL16 gene is highly nonrandom in all the herpesviruses, and the overall base composition of the UL16 genes in these species also differs dramatically. From the table 2, the ENC values of different UL16 genes vary from 28.504 to 60.115, with a mean value of 50.519 and standard deviation (S.D.) of 8.206. The GC_{3S} contents of each UL16 gene range from 27.65 to 97.97% with a mean of 56.826% and S.D. of 18.152. The CAI values of different UL16 genes vary from 0.6 to 0.764, with a mean value of 0.681 and standard deviation (S.D.) of 0.045.

Table 1. Synonymous Codon Usage of DEV UL16 Gene Analyzed with Cusp Program

Rank	AA	Codon	Fraction[a]	Frequency[b]	Number[c]	RSCU[d]
1	A(Ala)	GCA	0.297	30.303	11	1.19
2		GCC	0.189	19.284	7	0.76
3		GCG	0.243	24.793	9	0.97
4		GCT	0.270	27.548	10	1.08
5	C(Cys)	TGC	0.385	13.774	5	0.77
6		TGT	0.615	22.039	8	1.23
7	D(Asp)	GAC	0.444	22.039	8	0.89
8		GAT	0.556	27.548	10	1.11
9	E(Glu)	GAA	0.882	41.322	15	1.76
10		GAG	0.118	5.510	2	0.24
11	F(Phe)	TTC	0.444	11.019	4	0.89
12		TTT	0.556	13.774	5	1.11
13	G(Gly)	GGA	0.438	19.284	7	1.75
14		GGC	0.125	5.510	2	0.50
15		GGG	0.188	8.264	3	0.75
16		GGT	0.250	11.019	4	1.00
17	H(His)	CAC	0.444	11.019	4	0.89
18		CAT	0.556	13.774	5	1.11

Table 1. (*continued*)

19	I(Ile)	ATA	0.423	30.303	11	1.27
20		ATC	0.231	16.529	6	0.69
21		ATT	0.346	24.793	9	1.04
22	K(Lys)	AAA	0.778	19.284	7	1.56
23		AAG	0.222	5.510	2	0.44
24	L(Leu)	CTA	0.179	19.284	7	1.08
25		CTC	0.179	19.284	7	1.08
26		CTG	0.077	8.264	3	0.46
27		CTT	0.154	16.529	6	0.92
28		TTA	0.256	27.548	10	1.54
29		TTG	0.154	16.529	6	0.92
30	M(Met)	ATG	1.000	30.303	11	1.00
31	N(Asn)	AAC	0.385	13.774	5	0.77
32		AAT	0.615	22.039	8	1.23
33	P(Pro)	CCA	0.308	22.039	8	1.23
34		CCC	0.077	5.510	2	0.31
35		CCG	0.308	22.039	8	1.23
36		CCT	0.308	22.039	8	1.23
37	Q(Gln)	CAA	0.600	8.264	3	1.20
38		CAG	0.400	5.510	2	0.80
39	R(Arg)	AGA	0.214	16.529	6	1.29
40		AGG	0.107	8.264	3	0.64
41		CGA	0.179	13.774	5	1.07
42		CGC	0.107	8.264	3	0.64
43		CGG	0.071	5.510	2	0.43
44		CGT	0.321	24.793	9	1.93
45	S(Ser)	AGC	0.031	2.755	1	0.19
46		AGT	0.125	11.019	4	0.75
47		TCA	0.312	27.548	10	1.88
48		TCC	0.062	5.510	2	0.38
49		TCG	0.156	13.774	5	0.94

Table 1. (*continued*)

50		TCT	0.312	27.548	10	1.88
51	T(Thr)	ACA	0.368	19.284	7	1.47
52		ACC	0.158	8.264	3	0.63
53		ACG	0.316	16.529	6	1.26
54		ACT	0.158	8.264	3	0.63
56	V(Val)	GTA	0.364	22.039	8	1.45
57		GTC	0.136	8.264	3	0.55
58		GTG	0.273	16.529	6	1.09
59		GTT	0.227	13.774	5	0.91
60	W(Trp)	TGG	1.000	19.284	7	1.00
61	Y(Tyr)	TAC	0.333	5.510	2	0.67
62		TAT	0.667	11.019	4	1.33
63	*	TAA	1.000	2.755	1	3.00

The strong bias towards the codons with A and T at the third codon position and the preferentially used codons for each amino acid are displayed in red.

a. The ‘‘Fract’’ shows the proportion of all synonymous codons encoding the same amino acid
b. The ‘‘Frequency’’ lists the number of codons present per 1000 bases in the input sequence(s)
c. The ‘‘Number’’ lists the number of codons
d. The ‘‘RSCU’’ shows the proportion of relative synonymous codon usage

In general speaking, the gene is thought to possess strong codon bias if the ENC value is less than 35[29]. Analyzing the ENC values of all the UL16 genes, the results showed the majority of them do not have a strong codon bias. The plot of ENC and GC_{3S} content is another effective way to explore codon usage variation among different genes[29]. In fig.1, the solid line represents the curve if codon usage is only determined by GC_{3S} content[30]. In principle, proprietary proportion of points lay near to the solid line on this distribution. It suggested that mutational bias was the main factor determining the codon usage variation among these UL16 genes. But the ENC values were mostly dispersed below the curve. Hence, other than mutational bias, there might be few additional factors driving the codon usage variation among these UL16 genes such as natural selection.

Table 2. Summary Analysis of UL16 Gene In Different Herpesvirus Species

	Virus name	GenBank	L(bp)[a]	CAI[b]	ENC[c]	GC(%)	GC_{3s}(%)[d]
Alphaherpsvirinae	Duck plague virus(DPV)	EU195095	1089	0.600	58.928	46.74	38.29
	Meleagrid herpesvirus 1 (MeHV-1)	NC_002641	1056	0.700	53.775	46.40	52.27
	Bovine herpesvirus 1 (BoHV-1)	NC_001847	1020	0.695	32.900	75.49	90.29
	Bovine herpesvirus 5(BoHV-5)	NC_005261	1032	0.732	28.504	77.91	97.97
	Equid herpesvirus 1(EHV-1)	NC_001491	1113	0.711	50.845	57.41	65.77
	Equid herpesvirus 4(EHV-4)	NC_001844	1110	0.662	54.901	50.45	50.54
	Gallid herpesvirus 2(GaHV-2)	NC_002229	1083	0.712	55.371	43.49	48.20
	Gallid herpesvirus 3(GaHV-3)	NC_002577	1056	0.675	57.994	51.04	51.14
	Human herpesvirus 1 (HHV-1)	NC_001806	1200	0.701	47.551	68.08	80.25
	Human herpesvirus 2 (HHV-2)	NC_001798	1150	0.686	56.366	71.45	59.01
	Human herpesvirus 3(HHV-3)	NC_001348	1092	0.624	56.500	47.44	42.31
	Suid herpesvirus 1(SuHV-1)	NC_006151	982	0.663	42.623	78.39	55.66
	Psittacid herpesvirus 1(PsHV-1)	NC_005264	1068	0.719	48.946	59.83	70.79
	Ceropithecine herpesvirus 2 (CerHV-2)	NC_006560	1089	0.701	42.382	76.95	72.45
	Ceropithecine herpesvirus 9(CerHV-9)	NC_002686	1083	0.605	50.657	41.83	33.52
Betaherpsvirinae	Murid herpesvirus 1(MuHV-1)	NC_004065	1080	0.764	40.985	61.02	79.72
	Human herpesvirus 7(HHV-7)	NC_001716	1020	0.627	48.404	33.14	27.65
	Human herpesvirus 6 (HHV-6)	NC_001664	1080	0.608	55.424	39.07	39.72
	Human herpesvirus 5 (HHV-5)	NC_006273	1080	0.643	60.115	57.31	46.94
Gammaherpsvirinae	Human herpesvirus 4 (HHV-4)	NC_007605	1080 1074	0.714	56.761	53.80	53.33
	Human herpesvirus 8(HHV-8)	NC_009333	1005	0.716	54.946	53.83	61.19
	Murid herpesvirus 4 (MuHV-4)	NC_001826	984	0.693	55.547	43.60	41.77
	Alcelaphine herpesvirus 1 (AlHV-1)	NC_002531	1008	0.723	51.518	47.32	48.21

a Represents the length of identified ORF
b Effective number of codons
c Codon Adaptation Index
d G + C frequency at the synonymous third position of codons

3.3 Characterization of the DPV UL16 Gene

Usinng CLUSTAL-X and TREEVIEW software, A phylogenetic tree was established from the deduced amino acids encoded by the 1089 bp ORF of the UL16 gene of DPV and the 22 reference herpesviruses (Fig. 2). It shows that there are mainly three branches for the 23 herpesvirus. The DPV, MaHV-2, GaHV-3, and MeHV-1 are clustered in a distinct subbranch in Alphaherpesvirinae. The amino acid sequence of DPV UL16 is higher similarity and has a closer evolution with MaHV-2, GaHV-3, and MeHV-1.

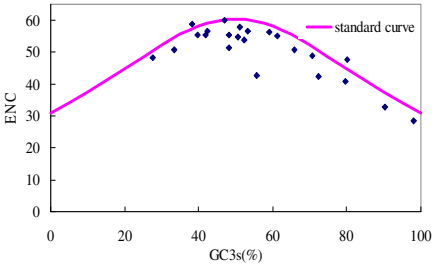

Fig. 1. The plot of ENC and guanine (G) + cytosine (C) frequency at the synonymous third position of codons (GC_{3S}) of the UL16 gene in the DEV CHv strain and those of 23 reference herpesviruses. The curve indicates the expected codon usage if GC compositional constraints alone accout for codon usage bias.

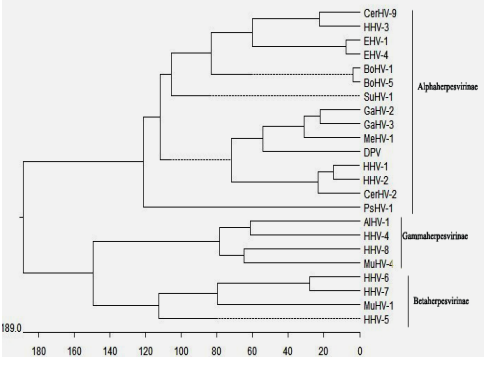

Fig. 2. Phylogenetic tree based on the UL16 amino acid sequences in 24 herpesviruses (Table 2), and constructed with CLUSTAL-X and TREEVIEW software

Rare condons analysis showed that there are 34 rare condons (9.366%) in the ORF of the DPV UL16 gene by using codon usage database on line (http://nihserver.mbi.ucla.edu/RACC/) (shown in Fig.3). The result revealed that there have sequential two rare codons in the ORF of the DPV UL16 gene.atg gct cgc agt act att aca cgt CGA tta tca tca tgt acg gaa ctc gac gat gga gaa ctg aat tcg cca ATA tta ttt tta aat gac ccg tct ctc ggt agt gtt cac CTA gct cgg gca ttg aac aca caa gtg tgt tca tgg cgc ctt att AGA tct gat tct cgt atc aag atc atg atc gca att aca gca ctc ggg gac cgt ctt tgt gct ttc gcg cct cca CTA gaa gat CGA gaa AGG gcg gca atg gtg gaa ATA ATA ttg tac tta acg cgt cct aaa gcg tta gct ctc cca tct gga act ttc cat gcc gtg ttt att gtc aac cgt tca tca atg tat gct gca ATA gca gct ATA cat atc gaa gca CTA aac caa tct gga acc ctg ttc tca tta ttg ttt tcc tca gta gaa acg acc ccg ccg cct ccg gaa gtt cct gac ccg tcg aca gaa att atg ccg cag gcc cct gct tca atc CTA aat CTA gaa gac cat acg gaa aat ATA acg ccg cca AGG gat cct cat aac tgt AGA atg gta tct gtt ggg gcg tgg tgg tct ttt CCC aaa CGA AGG ctc tac tat tta CGA atg gat aca cca ctt tta gct ATA tgc ccg gcg gga tgg aaa gca AGA acg ctt gga gac gtt CTA gcg AGA ctc gta gac cat aca cca ggt tgc gag acc tgc att agc ggc cac gat cac gtg gat tcg tat aat gcc ATA tgg aag cct ggc gaa gtc gca gag gca tgt

tta tgt aaa gga cca tgc ctg tgg ctc aaa tca aaa cag cgg gat atg ATA gta gaa ggg gat gtg agt atg tgt cgc gtt ttg ttc atg gac gct gta gat act ATA CGA ctt gta tct aac cgt aat cca cgt att tct gca aat ttg gcc gaa gta att tcc gcc ttt ggt tca gcc AGA caa gta cct gtc aat gcg gcc gga tgg cac ttg gtg gcg tta tcg gaa att gct agt tcg atc atg ATA tct ggt tgc gcg cgt ctt AGA cgt CTA tgt tat CCC aaa aca taa

Fig. 3. Rare condons analysis of the DPV UL16 gene **Red** = rare Arg codons **AGG, AGA, CGA** ; **Green** = rare Leu codon **CTA** ;**Blue** = rare Ile codon **ATA** ;**Orange** = rare Pro codon **CCC**

3.4 Comparison of Codon Usage between DPV and E. coli , Yeast and Human

The DPV UL16 gene was compared with those of *E. coli*, Yeast and Human to see which will be the suitable host for the optimal expression of DPV genes. From table 3, there are 21 codons distinct usage differences between DPV UL16 gene and E.coli (a DPV-to-E. coli ratio higher than 2 or lower than 0.50), 19 between DPV UL16 gene and yeast (a DPV -to-Yeast ratio higher than 2 or lower than 0.50), 20 between DPV UL16 gene and Human (a DPV -to- Human ratio higher than 2 or lower than 0.50) . Codons usage analysis datas (Fig.4) shows variation between DPV UL16 gene, E.coli, Yeast and Human. All these might suggest that expressing DPV genes more efficiently in Yeast systems.

Table 3. Comparison of Codon Preferences Between the DPV UL16 Gene and E. Coli, Yeast And H. Sapiens

Condon	Amino acid	E.coli (1/1000)	Yeast (1/1000)	Human (1/1000)	UL16 (1/1000)	UL16 /E.coli	UL16 /Yeast	UL16 /Human
GCA	A(Ala)	20.6	16.1	16.1	30.303	1.471	1.882	1.882
GCC	A	25.5	12.5	28.4	19.284	0.756	1.542	0.679
GCG	A	31.7	6.1	7.5	24.793	0.782	4.064	3.306
GCT	A	15.6	21.1	18.6	27.548	1.766	1.306	1.481
TGC	C(Cys)	6.9	4.7	12.2	13.774	1.996	2.931	1.129
TGT	C	5.5	8	10	22.039	4.007	2.755	2.204
GAC	D(Asp)	18.6	20.2	25.6	22.039	1.185	1.091	0.861
GAT	D	32.1	37.8	21.9	27.548	0.858	0.729	1.258
GAA	E(Glu)	38.2	48.5	29	41.322	1.082	0.852	1.425
GAG	E	17.7	19.1	39.9	5.51	0.311	0.288	0.138
TTC	F(Phe)	16.9	18.2	20.6	11.019	0.652	0.605	0.535
TTT	F	23.2	26.1	17.1	13.774	0.594	0.528	0.805
GGA	G(Gly)	9	10.9	16.4	19.284	2.143	1.769	1.176

Table 3. (*continued*)

GGC	G	27.9	9.7	22.5	5.51	0.197	0.568	0.245
GGG	G	11.3	6	16.3	8.264	0.731	1.377	0.507
GGT	G	24.4	24	10.8	11.019	0.452	0.459	1.020
CAC	H(His)	9.8	7.7	15	11.019	1.124	1.431	0.735
CAT	H	13.6	13.7	10.5	13.774	1.013	1.005	1.312
ATA	I(Ile)	5.4	17.8	7.7	30.303	5.612	1.702	3.935
ATC	I	24.2	17	21.6	16.529	0.683	0.972	0.765
ATT	I	29.8	30.4	16.1	24.793	0.832	0.816	1.540
AAA	K(Lys)	33.2	42.2	24.1	19.284	0.581	0.457	0.800
AAG	K	10.7	30.7	32.2	5.51	0.515	0.179	0.171
CTA	L(Leu)	4	13.3	7.8	19.284	4.821	1.450	2.472
CTC	L	11	5.4	19.8	19.284	1.753	3.571	0.974
CTG	L	50.9	10.4	39.8	8.264	0.162	0.795	0.208
CTT	L	11.7	12.1	13	16.529	1.413	1.366	1.271
TTA	L	13.9	26.7	7.5	27.548	1.982	1.032	3.673
TTG	L	14	27	12.6	16.529	1.181	0.612	1.312
ATG	M(Met)	27	20.9	22.2	30.303	1.122	1.450	1.365
AAC	N(Asn)	21.4	24.9	19.5	13.774	0.644	0.553	0.706
AAT	N	18.6	36.3	16.7	22.039	1.185	0.607	1.320
CCA	P(Pro)	8.5	18.2	16.7	22.039	2.593	1.211	1.320
CCC	P	5.8	6.8	20.1	5.51	0.95	0.810	0.274
CCG	P	21.8	5.3	6.9	22.039	1.011	4.158	3.194
CCT	P	7.3	13.6	17.3	22.039	3.019	1.621	1.274
CAA	Q(Gln)	15	27.5	12	8.264	0.551	0.301	0.689
CAG	Q	29.5	12.1	34.1	5.51	0.187	0.455	0.162
AGA	R(Arg)	2.9	21.3	11.5	16.529	5.700	0.776	1.437
AGG	R	1.9	9.2	11.4	8.264	4.349	0.898	0.725
CGA	R	3.9	3	6.3	13.774	3.532	4.591	2.186
CGC	R	21	2.6	10.7	8.264	0.394	3.178	0.772
CGG	R	6.3	1.7	11.6	5.51	0.875	3.241	0.475
CGT	R	20.3	6.5	4.6	24.793	1.221	3.814	5.390

Table 3. (*continued*)

AGC	S(Ser)	16	9.7	19.3	2.755	0.172	0.284	0.143
AGT	S	9.5	14.2	11.9	11.019	1.160	0.776	0.926
TCA	S	7.8	18.8	12	27.548	3.532	1.465	2.296
TCC	S	8.9	14.2	11.9	5.51	0.619	0.388	0.463
TCG	S	8.7	8.5	4.4	13.774	1.583	1.620	3.130
TCT	S	8.7	23.5	14.7	27.548	3.166	1.172	1.874
ACA	T(Thr)	8.2	17.8	15.1	19.284	2.352	1.083	1.277
ACC	T	22.8	12.6	19.4	8.264	0.362	0.656	0.426
ACG	T	14.8	7.9	6.1	16.529	1.117	2.092	2.710
ACT	T	9.1	20.3	13	8.264	0.908	0.407	0.636
GTA	V(Val)	11.1	11.8	7.2	22.039	1.985	1.868	3.061
GTC	V	15.1	11.6	14.6	8.264	0.547	0.712	0.566
GTG	V	25.5	10.6	28.4	16.529	0.648	1.559	0.582
GTT	V	18.5	22	11	13.774	0.745	0.626	1.252
TGG	W(Trp)	15.2	10.3	12.7	19.284	1.269	1.872	1.518
TAC	Y(Tyr)	12.1	14.6	15.5	5.51	0.455	0.377	0.355
TAT	Y	16.5	18.9	12.1	11.019	0.668	0.583	0.911
TAA	*	2	1	0.7	2.755	1.378	2.755	3.936
TAG	*	0.3	0.5	0.6	0	0	0	0
TGA	*	1.1	0.7	1.5	0	0	0	0

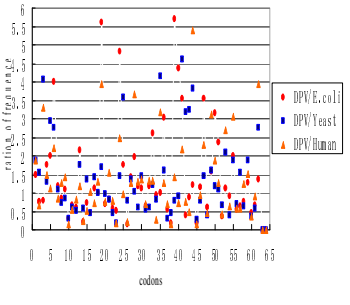

Fig. 4. The comparisons in the ratio of codon usage frequency (1/1000) of DPV to *E. coli*, yeast and H.sapiens. The ratio higher than 2 or lower 0.5 indicates that the codon usage preference differs, and vice versa.

4 Discussions

The degeneracy of the genetic code implies that multiple triplets codes for the same amino acid . The frequencies with which different codons are used vary significantly between organisms and between proteins within the same organism[31]. In this paper, a comprehensive analysis of codon usage including ENC, CAI value, GC content and the RSCU values of the DPV UL16 gene was performed by using EMBOSS programs and CodonW, and these values were subsequently compared with those of the 22 reference herpesvirus species. These datas of synonymous codon usage bias showed certain disparity of each herpesvirus from the different organisms with the result revealing that: (1) the DPV UL16 gene and its 22 reference herpesviruses adopt relatively similar codon usage patterns, although the DPV UL16 gene shows a few differences of codon usage bias with its reference herpesvirus species; and (2) the DPV UL16 gene prefers to use the codons with A and T at the 3rd codon position. From the table 1, twenty-six synonymous codons strong bias toward A- and T-ended at the third codon position were used. The DPV UL16 gene is an AT-rich gene. At the same time, random mutations increase the population of AT rather than CG as a result, undergo spontaneous deaminations[32]. So it is reasonable that codons ending in A and/or T are predominant in the gene. The ENC value and GC_{3S} content, two important codon usage indices, have been widely used to explore the codon usage variation among different genes[33-35]. Values of ENC range From 20(when only one codon is used per amino acid) to 61(when all synonyms are used with equal frequency). The ENC values of herpesvirus UL16 genes are dramatically different (from 28.504 to 60.115 shown in table 2) but mostly the codon usage bias is lower stronger. The ENC value of DPV UL16 gene is high (ENC>50), so that the codon usage bias is low stronger. If G+C compositional constraint influences the codon usage, then the GC_{3S} and ENC correlated spots would lie on the expected curve[30]. If a gene is subject to selection for translationally optimal codons, it will lie considerably below the expected curve[36]. In Fig.1, a large number of points do not follow the theoretical curve suggesting that other factors other than gene composition contribute to the codon usage pattern in the reference herpesviruses, which maybe mutational bias and natural selection, such as translational selection, the tRNA abundance, leading to the codon usage variation among genes in different organisms.

Comparative analysis of UL16 genes in DPV and the reference herpesviruses indicated that synonymous codon usage in these genes is phylogenetically conserved. Datas in table 2 show that the UL16 genes in DPV, MeHV-1, GaHV-2 and GaHV-3, whose natural host is avian, have a stronger correlation than the UL16 genes of herpesviruses with other hosts. In the table 3, the CAI value of DPV UL16 gene is 0.600, which is a little slight higher. The CAI value is much closer to 1, the codon usage is a little stronger and the gene expressing level is much higher. We can infer UL16 gene is highly expressed gene in DPV genome. Simultaneously, the phylogenetic tree analysis based on the UL16 gene products revealed that UL16 protein of the DPV CHv strain and some avian herpesviruses such as MeHV-1, GaHV-2 and GaHV-3, were clustered within a monophyletic clade and grouped within Alphaherpesvirinae. We speculate that the codon usage bias of DPV UL16 gene has a very close relation with its gene function and gene type. DPV UL16 gene is similar to the UL16 gene function of the herpesviruses. Studies on the HHV-1 have

well documented that the tegument UL16 protein is not required for viral replication in cell culture and its function may be in viral DNA packaging, virion assembly, budding, and egress by providing an interaction with the membrane-bound UL11 protein and the UL21 protein[26,37-39]. In the absence of murine Gammaherpesvirus 68 (MuHV-4) ORF33 (UL16 homologue), immature virions were restrained in a state interacting with actin and glycoproteins gB and failed release of infectious virions[40]. Furthermore, bioinformatics analysis indicated that the DPV UL16 protein has higher similarity with these homologues, and may also play a similar role to that of HHV-1 in the viral DNA packaging, virion assembly, budding, and egress. However, further studies are required to confirm this hypothesis.

The most plausible selection-based explanation for codon usage bias is the selection for efficient translation related to the relative abundance of isoaccepting tRNAs[33,34]. In order to show the codon usage variation among genes from different organisms, the codon usage bias of DPV was compared with that of E. coli, Yeast, and Human. There are 21 codons distinct usage differences between DPV UL16 gene and E.coli, 19 between DPV UL16 gene and Yeast, 20 between DPV UL16 gene and Human. Thus, we can assume that the E.coli, Yeast and Human expression system were suitable for heterologous expression of the DPV gene. In addition, we analyzed the rare condons of the DPV UL16 gene. There were 34 rare codons and sequential two rare codons in UL16 gene ORF, which may influence the expression of the UL16 gene in vitro. So if we choose the prokaryotic expression system, we should choose the host bacteria Rosseta, which should impove the expression of the exogenous genes. All of these may be of great importance for gene characterization, for gene classification and for assessing the possible role of UL16 protein in viral pathogenesis.

Acknowledgments. The research were supported by grants from Changjiang Scholars and Innovative Research Team in University (PCSIRT0848), the earmarked fund for Modern Agro-industry Technology Research System (nycytx-45-12), the Cultivation Fund of the Key Scientific and Technical Innovation Project, the Ministry of Education of China (No. 706050), the Cultivation Fund of the Key Scientific and Technical Innovation Project, Department of Education of Sichuan Province (No. 07ZZ028). The first two authors contribute equally to this work and should be considered as first author. An-chun Cheng and Ming-shu Wang are the corresponding authors. Key Laboratory of Animal Disease and Human Health of Sichuan Province & Avian Disease Research Center, College of Veterinary Medicine of Sichuan Agricultural University, 46# Xinkang Road, Yucheng district, Yaan 625014, Sichuan province of China. Tel.: +86 835 2885774; fax: +86 835 2885774. E-mail address: chenganchun@vip.163.com (A. Cheng); mshwang@163.com (M. Wang).

References

1. Jan, B.: Does codon bias have an evolutionary origin? Theor. Biol. Med. Model 5(16), 1–16 (2008)
2. Fuglsang, A.: Accounting for background nucleotide composition when measuring codon usage bias: brilliant idea, difficult in practice. Mol. Biol. Evol. 23(7), 1345–1347 (2006)

3. Nakamura, M., Sugiura, M.: Translation efficiencies of synonymous codons are not always correlated with codon usage in tobacco chloroplasts. Plant J. 49(1), 128–134 (2007)
4. Marais, G., Duret, L.: Synonymous codon usage, accuracy of translation, and gene length in Caenorhabditis elegans. J. Mol. Evol. 52(3), 275–280 (2001)
5. Ma, J., Campbell, A., Karlin, S.: Correlations between shine-dalgarno sequences and gene features such as predicted expression levels and operon structures. J. Bacteriol. 184(20), 5733–5745 (2002)
6. Noboru, S.: Two aspects of DNA base composition: G+C content and translation-coupled deviation from intra-strand rule of A = T and G = C. J. Mol. Evol. 49(1), 49–62 (1999)
7. Plotkin, J.B., Dushoff, J., Desai, M.-M., Fraser, H.-B.: Codon usage and selection on proteins. J. Mol. Evol. 63(5), 635–653 (2006)
8. Kanaya, S., Yamada, Y., Kudo, Y., Ikemura, T.: Studies of codon usage and tRNA genes of 18 unicellular organisms and quantification of Bacillus subtilis tRNAs: gene expression level and speciesspecific diversity of codon usage based on multivariate analysis. Gene 238(1), 143–155 (1999)
9. Noguchi, S., Satow, Y.: Purification of human β2-adrenergic receptor expressed in methylotrophic yeast Pichia pastoris. J. Biochem. 140(6), 799–804 (2006)
10. Sau, K., Gupta, S.-K., Sau, S., Ghosh, T.-C.: Synonymous codon usage bias in 16 Staphylococcus aureus phages: implication in phage therapy. Virus Res. 113(2), 123–231 (2005)
11. Sueoka, N., Kawanishi, Y.: DNA G+C content of the third codon position and codon usage biases of human genes. Gene 261(1), 53–62 (2000)
12. Wan, X.-F., Xu, D., Kleinhofs, A., Zhou, J.: Quantitative relationship between synonymous codon usage bias and GC composition across unicellular genomes. BMC Evol. Biol. 4(9), 19–29 (2004)
13. Lynn, D.-J., Singer, G.-A., Hicker, D.-A.: Synonymous codon usage is subject to selection in thermophilic bacteria. Nucleic Acids Res. 30(19), 4272–4277 (2002)
14. Duret, L.: tRNA gene number and codon usage in the C.elegans genome are co-adapted for optimal translation ofhighly expressed genes. Trends Genet. 16(7), 287–289 (2000)
15. Cai, M.-S., Cheng, A.-C., Wang, M.-S., Zhao, L.-C., Zhu, D.-K., Luo, Q.-H., et al.: Characterization of synonymous codon usage bias in the duck plague virus UL35 gene. Intervirology 52(5), 266–278 (2009)
16. Zhao, L.-C., Cheng, A.-C., Wang, M.-S., Yuan, G.-P., Cai, M.-S.: Characterization of codon usage bias in the dUTPase gene of duck enteritis virus. Prog. Nat. Sci. 18(9), 1069–1076 (2008)
17. Prat, Y., Fromer, M., Linial, N., Linial, M.: Codon usage is associated with the evolutionary age of genes in metazoan genomes. BMC Evol. Biol. 9(1), 285–299 (2009)
18. Hassard, S., Ward, G.: Efficient creation of sequencing libraries from blunt-ended restriction enzyme fragments. Biotechniques 18(3), 396–398 (1995)
19. Jenkins, G.-M., Holmes, E.-C.: The extent of codon usage bias in human RNA viruses and its evolutionary origin. Virus Res. 92(1), 1–7 (2003)
20. Ma, X.-R., Xiao, S.-B., Fang, L.-R., Chen, H.-C.: Bias of base composition and codon usage in pseudorabies virus genes. In: Yi Chuan Xue Bao, vol. 32(6), pp. 616–624 (June 2005)
21. Hall, J.-D., Gibbs, J.-S., Coen, D.-M., Mount, D.-W.: Structural organization and unusual codon usage in the DNA polymerase gene from herpes simplex virus type 1. DNA 5(4), 281–288 (1986)

22. Karlin, S., Blaisdell, B.-E., Schachtel, G.-A.: Contrasts in codon usage of latent versus productive genes of Epstein-Barr virus: data and hypotheses. J. Virol. 64(9), 4264–4273 (1990)
23. Davison, S., Converse, K.-A., Hamir, A.-N., Eckroade, R.-J.: Duck viral enteritis in domestic Muscovy ducks in Pennsylvania. Avian Dis. 37(4), 1142–1146 (1993)
24. Saif, Y.-M., Barnes, H.-J., Glisson, J.-R., Fadly, A.-M., McDougald, L.-R., Swayne, D.-E.: Diseases of Poultry, 11th edn., pp. 354–363. Iowa State University Press, Ames (2003)
25. Cheng, A.-C., Wang, M.-S., Wen, M.: Construction of duck enteritis virus gene libraries and discovery, cloning and identification of viral nucleocapsid protein gene. High Technolagy Letters, Institute of Scientific and Technical Information of China 16, 948–953 (2006)
26. Meckes Jr., D.-G., Wills, J.-W.: Dynamic interactions of the UL16 tegument protein with the capsid of herpes simplex virus. J. Virol., American Society For Microbiology, United States 81(23), 13028–13036 (2007)
27. Comeron, J.-M., Aguade, M.: An evaluation of measures of synonymous codon usage bias. J. Mol. Evol. 47(3), 268–274 (1998)
28. Liu, Q., Dou, S., Ji, Z., Xue, Q.: Synonymous codon usage and gene function are strongly related in Oryza sativa. Biosystems 80(2), 123–131 (2005)
29. Wright, F.: The 'effective number of codons' used in a gene. Gene 87(1), 23–29 (1990)
30. Jiang, P., Sun, X., Lu, Z.: Analysis of Synonymous codon usage in Aeropyrum pernix K1 and other Crenarchaeota Microorganisma. J. Genet. Genomics 34(3), 275–284 (2007)
31. Akashi, H.: Gene expression and molecular evolution. Curr. Opin. Genet. Dev. 11(1), 660–666 (2001)
32. Lindahl, T.: Instability and decay of the primary structure of DNA. Nature 362(6422), 709–715 (1993)
33. Sakai, H., Washio, T., Saito, R., Shinagawa, A., Itoh, M., Shibata, K., et al.: Correlation between sequence conservation of the 5' untranslated region and codon usage bias in Mus musculus genes. Gene 276(1-2), 101–105 (2001)
34. 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(2), 155–161 (2004)
35. Zhou, T., Gu, W., Ma, J., Sun, X., Lu, Z.: Analysis of synonymous codon usage in H5N1 virus and other influenza A viruses. Biosystems 81(1), 77–86 (2005)
36. Zhou, T., Sun, X., Lu, Z.: Synonymous codon usage in environmental chlamydia UWE25 reflects an evolutional divergence from pathogenic chlamydiae. Gene 368(1), 117–125 (2006)
37. Baines, J.-D., Roizman, B.: The open reading frames UL3, UL4, UL10, and UL16 are dispensable for the replication of herpes simplex virus 1 in cell culture. J. Virol. 65(2), 938–944 (1991)
38. Oshima, S., Daikoku, T., Shibata, S., Yamada, H., Goshima, F., Nishiyama, Y.: Characterization of the UL16 gene product of herpes simplex virus type 2. Arch. Virol. 143(5), 863–880 (1998)
39. Nalwanga, D., Rempel, S., Roizman, B., Baines, J.D.: The UL16 gene product of herpes simplex virus 1 is a virion protein that colocalizes with intranuclear capsid proteins. Virol. 226(2), 236–242 (1996)
40. Guo, H., Wang, L., Peng, L., Zhou, Z.-H., Deng, H.: Open Reading Frame 33 of a Gammaherpesvirus Encodes a Tegument Protein Essential for Virion Morphogenesis and Egress. Virol. 83(20), 10582–10595 (2009)