Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Available online at www.sciencedirect.com

SciVerse ScienceDirect

AASRI Procedia 1 (2012) 112 – 123

AASRI Procedia

www.elsevier.com/locate/procedia

2012 AASRI Conference on Computational Intelligence and Bioinformatics

Characterization of Codon Usage Bias in UL21gene from Duck Enteritis Virus

Yong Shi[a,b] ,Anchun Cheng[a,b,c,*], Mingshu Wang[a,b,c]

[a]*Institute of Preventive Veterinary Medicine, Sichuan Agricultural University, Wenjiang, Chengdu city, Sichuan, 611130, P.R.China 2.*
[b]*Avian Disease Research Center, College of Veterinary Medicine of Sichuan Agricultural University, 46 Xinkang Road, Ya'an, Sichuan 625014, P.R. China*
[c]*Key Laboratory of Animal Disease and Human Health of Sichuan Province, Sichuan Agricultural University,Wenjiang, Chengdu city, Sichuan, 611130, P.R.China)*

Abstract

In this paper, the comprehensive analysis of codon usage bias of Duck enteritis virus (DEV) UL21 gene was performed by using CAI, CHIPS and CUSP program of EMBOSS. Our study showed that codon usage bias of DEV UL21 had strong bias towards the A-ended or T-ended codons, and GC3s contents of the codon usage bias in DEV UL21 gene were significantly varied compared with those of other 27 reference herpesviruses. The CAI, ENC value of DEV CHv strain UL21 gene is 0.615 and 55.167, respectively, indicating that the codon usage bias of this gene is weak and lowly expressed. The plot of ENC versus GC3S content revealed that DEV UL21 gene is subject to GC compositional constraints. The phylogentic analysis about amino acids codon usage bias of DEV UL21 and the27 reference herpesviruses showed that DEV was evolutionarily closer to herpesviruses Mardivirus. In addition, the codon usage bias of DEV UL21 gene was compared with those of E. coli, yeast and humans. There are 42，45, 39 same codons usage bias between the DEV UL21 to E.coli, Yeast, H.sapiens, respectively, indicaiting that UL21 gene of DEV may be more efficiently expressed in the yeast system

.

Selection and/or peer review under responsibility of American Applied Science Research Institute

Keyworlds: Codon Usage Bias; Duck virus enteritis(DVE); Duck plague(DP); UL21 Gene

* Anchun Cheng; Tel.: +86 835 2885774; fax: +86 835 2885774
E-mail address: chenganchun@vip.163.com.

doi:10.1016/j.aasri.2012.06.020

1. Introduction

Duck virus enteritis(DVE), also known as duck plague(DP), is an serious and contagious disease, and highly lethal in all ages of birds from Anseriformes, such as ducks, geese, and swans. Since the first report of DVE in domestic ducks in 1923, more outbreaks were reported in many countries, and it has become a significant threat to the commercial duck industry in global [1-2]. Duck enteritis virus (DEV) is the causative agent of DVE, and which is one of the members of the subfamily Alphaherpesvirinae, but it has not been grouped into any genus [3].

With the standard genetic codes used in a many ways, all amino acids（except Met and Trp） are coded by more than one codons, and these synonymous codons are not equally used among organisms[4], namely one certain synonymous codons of each amino acid is preferred over other codons, a phenomenon termed codon usage bias. The frequencies of alternative synonymous codons used among all organisms[5]. It has been reported that some factors related to the codon usage in various specieses, such as mutational bias[6-8], translational selection[9-10], the structure of proteins[11–12], tRNA abundance[13-14], and GC compositions[15-17], gene expression level and gene length[18-19]. Study on the codon usage bias can provide some evidences about the molecular evolution of viruses and their individual genes. So far, some detailed comparative analysis of herpesvirus codon usage are reported[20-21], but the study of DEV UL21 gene codon usage bias is still absent.

In this paper codon usage bias in the UL21 gene of DEV was analyzed and also compared with those of 27 other herpesvirus. In addition, the codon usage bias of DEV UL21 gene was compared with those of E. coli, yeast, and humans. All these analyses might provide some reference for future study of the UL21 gene.

2. Materials and methods

2.1. Virus Species and Gene Sequences

The DEV was obtained from Key Laboratory of Animal Disease and Human Health of Sichuan Province. The UL21 gene(GeneBank Accession No. EU195090) of the DEV was isolated and identified by our laboratory. And the 27 nucleotide sequences of the UL21 like genes reference herpesviruses were obtained from the NCBI GenBank nucleotide database .

2.2. Analysis on codon usage in UL21 genes of DEV and 27 reference herpesviruses

In order to analazy the extent of codon usage bias in DEV UL21 gene, some codon usage normalizes are used as followed:

•RSCU: Relative synonymous codon usage. It is the observed frequency of codon usage divided by the sum of codons per amino acid and multiplied by the actual number of codons per amino acid [22].

•ENC: Effective number of codons. It can measure the degree of the codon usage of a gene departs from equal synonymous codon usage [23, 24].

•CAI: Codon Adaptation Index. It is a simple and effective measure of the overall synonymous codon usage bias of genes of different organisms [25, 26].

•GC and GC3s: The frequency of the nucleotide G+C of codons, and frequency of the nucleotide G + C at the synonymous third position of codons (excluding Met, Trp, and termination codons). They can reflect the extent of base composition bias [27-28].

The value of these normalization of UL21 genes of DPV and 27 reference herpesviruses were calculated with the European Molecular Biology Open Software Suite (EMBOSS) CHIPS online service program (http://emboss.bioinformatics.nl/).

2.3. Phylogenetic analysis of the DEV UL21 gene

The phylogenetic tree was constructed by employing the Clustal V in MegAlign program of DNAStar software, according to the amino acid sequences of the UL21 gene in DPV and other 27 herpesviruses.

2.4. Comparison of codon preferences of DEV UL21 gene with those of E. coli, yeast and humans

In order to detect whether different species follow the same codon usage rule, and select the best suitable expression system, we compared the DEV UL21 gene codon usage bias among E. coli, yeast and H. sapiens. The database of the codon usage in E. coli, yeast and humans is obtained from http://www.kazusa.or.jp/codon.

3. Results

3.1. A. Characterization of DEV UL21gene

The overall RSCU values of fifty-nine sense codons (with exception of Trp, Met, and the termination codons) and the codon preferences of the DEV UL21 gene were analyzed through the program of CodonW and CUSP program of EMBOSS, respectively (display in Table 1). Analyzing the RSCU values of the codons from the table, we find that most the A-ended or T-ended codons used for coding amino acid are much higher than the C-ended or G-ended codons. Meanwhile, a high level of diversity in codon usage bias is found among the Ala, Cys, Gly, Leu, Pro, Arg, Ser, Thr and Val amino acids, for they own a 6-fold or 4-fold coding degeneracy.

Table 1 The Analysis of Synonymous Codon Usage of DEV UL21 Gene

Codon	AA[a]	Fraction[b]	Frequency/1000[c]	NO.[d]	RSCU[e]	Codon	AA	Fraction	Frequency/1000	NO.	RSCU
GCA	A[Ala]	0.359	24.911	14	1.436	CCA	P[Pro]	0.5	12.456	7	2
GCC	A	0.205	14.235	8	0.821	CCC	P	0.214	5.338	3	0.857
GCG	A	0.205	14.235	8	0.821	CCG	P	0.214	5.338	3	0.857
GCT	A	0.231	16.014	9	0.923	CCT	P	0.071	1.779	1	0.286
TGC	C	0.429	10.676	6	0.857	CAA	Q[Gln]	0.5	14.235	8	1
TGT	C[Cys]	0.571	14.235	8	1.143	CAG	Q	0.5	14.235	8	1
GAC	D[Asp]	0.263	17.794	10	0.526	AGA	R[Arg]	0.175	12.456	7	1.05
GAT	D	0.737	49.822	28	1.474	AGG	R	0.075	5.338	3	0.45
GAA	E[Glu]	0.75	42.705	24	1.5	CGA	R	0.15	10.676	6	0.9
GAG	E	0.25	14.235	8	0.5	CGC	R	0.375	26.69	15	2.25
TTC	F[Phe]	0.217	8.897	5	0.435	CGG	R	0.05	3.559	2	0.3
TTT	F	0.783	32.028	18	1.565	CGT	R	0.175	12.456	7	1.05
GGA	G[Gly]	0.234	19.573	11	0.936	AGC	S[Ser]	0.083	7.117	4	0.5

GGC	G	0.255	21.352	12	1.021	AGT	S	0.188	16.014	9	1.125
GGG	G	0.234	19.573	11	0.936	TCA	S	0.271	23.132	13	1.625
GGT	G	0.277	23.132	13	1.106	TCC	S	0.083	7.117	4	0.5
CAC	H[His]	0.167	3.559	2	0.333	TCG	S	0.167	14.235	8	1
CAT	H	0.833	17.794	10	1.667	TCT	S	0.208	17.794	10	1.25
ATA	I[Ile]	0.489	39.146	22	1.467	ACA	T[Thr]	0.368	24.911	14	1.474
ATC	I	0.156	12.456	7	0.467	ACC	T	0.158	10.676	6	0.632
ATT	I	0.356	28.47	16	1.067	ACG	T	0.237	16.014	9	0.947
AAA	K[Lys]	0.647	19.573	11	1.294	ACT	T	0.237	16.014	9	0.947
AAG	K	0.353	10.676	6	0.706	GTA	V[Val]	0.424	24.911	14	1.697
CTA	L[Leu]	0.096	8.897	5	0.484	GTC	V	0.182	10.676	6	0.727
CTC	L	0.135	12.456	7	0.677	GTG	V	0.091	5.338	3	0.364
CTG	L	0.096	8.897	5	0.484	GTT	V	0.303	17.794	10	1.212
CTT	L	0.192	17.794	10	0.968	TGG	W[Trp]	1	10.676	6	1
TTA	L	0.288	26.69	15	1.452	TAC	Y[Tyr]	0.25	7.117	4	0.5
TTG	L	0.192	17.794	10	0.968	TAT	Y	0.75	21.352	12	1.5
ATG	M[Met]	1	17.794	10	1	TAA	*	1	1.779	1	1
AAC	N[Asn]	0.333	12.456	7	0.667	TAG	*	0	0	0	0
AAT	N	0.667	24.911	14	1.333	TGA	*	0	0	0	0

a AA represents Amino Acid.

b The “Fraction” column shows the proportion of each synonymous codons encoding the same amino acid.

c The “frequency/1000” value represents the number of codons present per 1000 bases in the input sequence(s).

d Number represents the number of occurrence of each sense codons.

e The “RSCU” is used to investigate the pattern of relative synonymous codon usage.

3.2. Codon usage bias analysis

The values of the ENC and CAI, coding GC and GC3s content analysis of DEV UL21 and other 27 relational herpesviruses UL21 genes are calculated by EMBOSS, and the results are shown in Table 2. We find that codon usage in UL21 gene is highly non-random in these herpesviruses, and the overall base composition of UL21 gene in these viruses are still differs dramatically. The CAI values of different herpesviruses UL21 genes vary from 0.613 to 0.768, with an average value of 0.681 and a standard deviation (SD) of 0.0579; The ENC values vary from 29.499 to 59.205, with an average value of 47.293 and SD of 9.9667; the GC contents vary from 27.66% to 77.70%, with an average value of 55.41% and SD of 13.858; the GC3s contents vary from 13.00% to 96.58%, with an average value of 60.48% and SD of 24.249. Interestingly, there were no differences or few differences in the codon usage bias parameters of the UL21 gene indicated by CAI, ENC, coding GC content among the DEV CHv strain, DEV UL21-like strain and DEV VAC strain, so we presumed that UL21 genes variation existed among different strains in the same species.

Table 2. Summary Analysis of The UL21 Gene In Herpesviruses

Genus	Virus name (Abbreviation)	GeneBank accession NO	L[a][bp]	CAI[b]	ENC[c]	GC[d] [%]	GC3s[e] [%]
Undesigned	Duck enteritis virus CHv strain (DEV CHv)	EU 195090	1686	0.615	55.167	43.89	36.65
	Duck enteritis virus VAC strain (DEV VAC)	NC 013036	1686	0.615	55.066	43.83	36.48
	Duck enteritis virus UL21-like genes	EF 203707	1686	0.615	55.483	43.89	36.48
Simplexvirus	Cercopithecine herpesvirus 1(CeHV-1)	AF 533768	1581	0.75	35.321	71.6	89.56
	Cercopithecine herpesvirus 2(CeHV-2)	NC 006560	1581	0.741	31.302	74.95	95.64
	Bovine herpesvirus 2(BoHV-2)	AF 387490	1569	0.709	43.426	65.2	78.97
	Human herpesvirus 1(HSV-1)	NC 001806	1608	0.735	42.355	66.23	77.8
	Human herpesvirus 2(HSV-2)	NC 001798	1599	0.76	37.294	70.36	87.62
	Papiine herpesvirus 2(PaHV-2)	NC 007653	1587	0.752	31.405	75.36	96.41
	Saimiriine herpesvirus 1(SaHV-1)	HM 625781	1614	0.664	40.935	67.29	78.07
Varicellovirus	Bovine herpesvirus 1(BoHV-1)	NC 001847	1725	0.718	38.859	72.29	85.57
	Bovine herpesvirus 5(BoHV-5)	NC 005261	1812	0.75	29.961	77.7	96.19
	Equine herpesvirus 1(EHV-1)	AY 464052	1593	0.679	53.51	52.86	56.69
	Equine herpesvirus 4(EHV-4)	AF 030027	1590	0.641	55.365	46.67	44.91
	Equid herpesvirus 9(EHV-9)	AP 010838	1593	0.665	55.342	52.92	56.5
	Felid herpesvirus 1（FeHV-1）	FJ 478159	1584	0.65	57.273	44.76	38.26
	Human herpesvirus 3(HSV-3)	DQ 674250	1626	0.632	56.06	46.43	42.25
	Suid herpesvirus 1（SuHV-1）	AY 363172	1578	0.753	29.499	74.4	96.58
	Canine Herpesvirus(CHV)	AY 768815	1569	0.579	37.49	27.66	13
Iltovirus	Psittacid herpesvirus 1 (PsHV-1)	NC 005264	1710	0.678	49.666	60.53	64.39
	Gallid herpesvirus 1(GaHV-1)	NC 006623	1599	0.647	57.711	42.71	41.84
Mardivirus	Gallid herpesvirus 2(GaHV-2)	AF 439271	1641	0.615	54.959	41.07	34.73
	Gallid herpesvirus 3(GaHV-3)	HQ 840738	1599	0.635	59.205	49.78	51.97
	Meleagrid herpesvirus 1 (MeHV-1)	AF 282130	1746	0.613	58.361	45.3	42.96
Roseolovirus	Human herpesvirus 7 (HSV-7)	AF 037218	1293	0.632	46.86	35.73	31.79
Lymphocryptovirus	Human herpesvirus 4(HSV-4)	NC 007605	1215	0.768	48.27	55.97	70.62
Rhadinovirus	Murid herpesvirus 4 (MHV-4)	AF 105037	1935	0.721	55.18	48.11	46.82
Macavirus	Ovine herpesvirus 2 (OvHV-2)	AY 839756	1203	0.735	52.891	53.87	64.59

a L represents the length of identified ORF. b CAI represents the codon Adaptation Index. c ENC represents the effective number of codons. d GC represents the frequency of the nucleotide G + C of codons. e GC3s represents the frequency of the nucleotide G + C at the synonymous third position of codons.

The plot of ENC versus GC3S content is another effective way to investigate codon usage variation among genes. If GC3s a unique determined conditions to shape the codon usage pattern, the values of ENC would lie considerably below the continuous curve. ENC values of the 28 UL21 genes were plotted against their corresponding GC3s contents, as was shown from Fig. 1, we can firm that a large number of points do not follow the theoretical curve, which lie near the solid curve of this distribution, indicating that the codon preference of these genes are subject to GC compositional constraints.

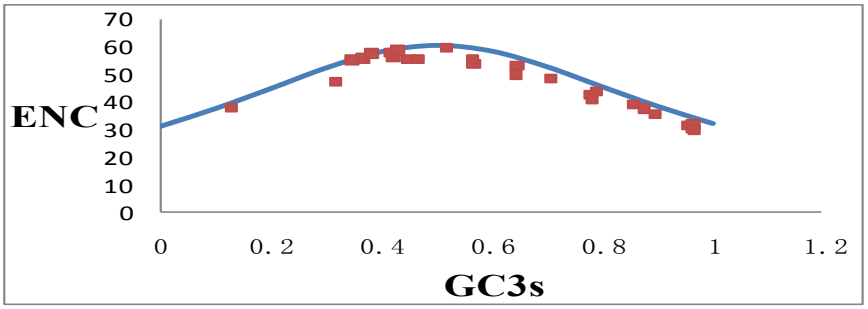

Fig.1. The plot of ENC and GC3s of UL21 gene of DEV and those of 27 reference herpesviruses. The solid line represents the expected curve between GC3S and ENC under random codon usage assumption.

3.3. Phylogenetic analysis

The phylogenetic tree was established in order to analyse the gene evolution among herpesviruses UL21 gene, based on the UL21 amino acid sequence of 28 herpesviruses. As it is showed in fig.2, the DEV, DEV-VAC and the DEV UL21 like gene are different from otherherpesviruses as they first cluster together and form a separate branch, then clustered with HSV-7 and 2 kind genus of virus (Mardivirus and Varicellovirus, respectively) in a monophyletic clade. We conclude that the UL21 gene of DEV CHv strain are quite homology to those of DEV VAC strain and DEV UL21-like gene, the shorter distance between DEV and Mardivirus(GaHV-2, GaHV-3 and MeHV-1) suggests the amino acid sequence are similar between DEV and GaHV-2, GaHV-3, MeHV-1, and the UL21 protein of DEV is closely related to GaHV-2, GaHV-3 and MeHV-1.

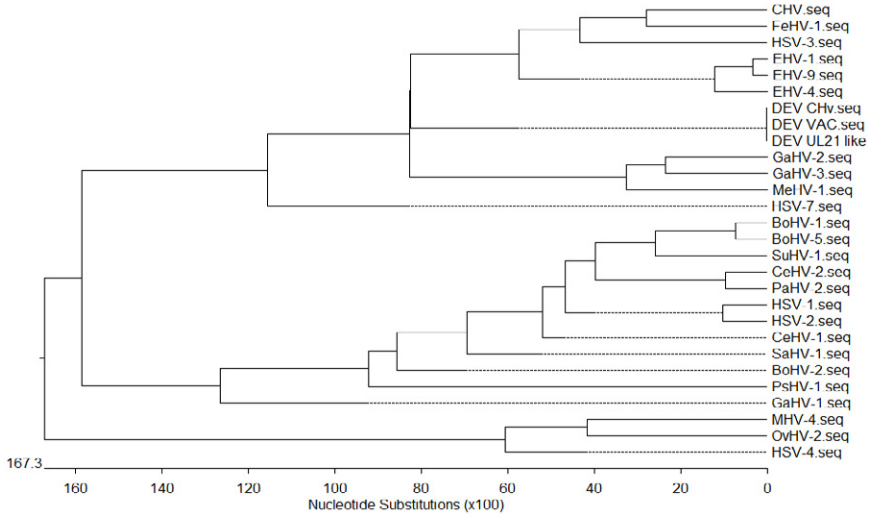

Fig.2 Phylogenetic tree based on the UL21 amino acid sequence of 27 herpesviruses (Table 2), and generated by using the MegAlign program with Clustal V multiple alignments in DNAstar 7.1.

3.4. Comparison of codon preference of DPV UL21 gene with those of E. coli, Yeast and Human

In general, the codon usage bias in genes remains at a certain level among species. In this study, we used the comparisons in the ratio of codon usage frequency (1/1000) of DEV to E.coli, yeast, H.sapiens to select suitable host for optimal expression.In Table 3, there are 42, 45, 39 codons showing the DEV UL21 to E.coli , Yeast, H.sapiens ratio between 0.5 to 2.0 respectively, indicating that codon usage of the DEV UL21gene resembles more closely that of yeast than that of E. coli and human. So we presumed that the DEV CHv strain UL21 gene may be more efficiently expressed in the yeast system.

Table 3. Comparison of Codon Preferences Between the DEV UL21 Gene and E. Coli, Yeast And H. Sapiens

Condon	AA	E.coli (1/1000)	Yeast (1/1000)	Human (1/1000)	UL21 (1/1000)	UL21 /E.coli[a]	UL21 /Yeast[b]	UL21 /Human[c]
GCA	A(Ala)	20.6	16.1	16.1	24.911	1.209	1.547	1.547
GCC	A	25.5	12.5	28.4	14.235	0.558	1.139	0.501
GCG	A	31.7	6.1	7.5	14.235	0.449	2.334	1.898
GCT	A	15.6	21.1	18.6	16.014	1.027	0.759	0.861
TGC	C(Cys)	6.9	4.7	12.2	10.676	1.547	2.271	0.875
TGT	C	5.5	8	10	14.235	2.588	1.779	1.424
GAC	D(Asp)	18.6	20.2	25.6	17.794	0.957	0.881	0.695
GAT	D	32.1	37.8	21.9	49.822	1.552	1.318	2.275
GAA	E(Glu)	38.2	48.5	29	42.705	1.118	0.881	1.473
GAG	E	17.7	19.1	39.9	14.235	0.804	0.745	0.357
TTC	F(Phe)	16.9	18.2	20.6	8.897	0.526	0.489	0.432
TTT	F	23.2	26.1	17.1	32.028	1.381	1.227	1.873
GGA	G(Gly)	9	10.9	16.4	19.573	2.175	1.796	1.193
GGC	G	27.9	9.7	22.5	21.352	0.765	2.201	0.949
GGG	G	11.3	6	16.3	19.573	1.732	3.262	1.201
GGT	G	24.4	24	10.8	23.132	0.948	0.964	2.142
CAC	H(His)	9.8	7.7	15	3.559	0.363	0.462	0.237
CAT	H	13.6	13.7	10.5	17.794	1.308	1.299	1.695
ATA	I(Ile)	5.4	17.8	7.7	39.146	7.249	2.199	5.084
ATC	I	24.2	17	21.6	12.456	0.515	0.733	0.577
ATT	I	29.8	30.4	16.1	28.47	0.955	0.937	1.768
AAA	K(Lys)	33.2	42.2	24.1	19.573	0.59	0.464	0.812
AAG	K	10.7	30.7	32.2	10.676	0.998	0.348	0.332
CTA	L(Leu)	4	13.3	7.8	8.897	2.224	0.669	1.141
CTC	L	11	5.4	19.8	12.456	1.132	2.307	0.629
CTG	L	50.9	10.4	39.8	8.897	0.175	0.855	0.224
CTT	L	11.7	12.1	13	17.794	1.521	1.471	1.369
TTA	L	13.9	26.7	7.5	26.69	1.92	1	3.559
TTG	L	14	27	12.6	17.794	1.271	0.659	1.412
ATG	M(Met)	27	20.9	22.2	17.794	0.659	0.851	0.802

AAC	N(Asn)	21.4	24.9	19.5	12.456	0.582	0.5	0.639
AAT	N	18.6	36.3	16.7	24.911	1.339	0.686	1.492
CCA	P(Pro)	8.5	18.2	16.7	12.456	1.465	0.684	0.746
CCC	P	5.8	6.8	20.1	5.338	0.92	0.785	0.266
CCG	P	21.8	5.3	6.9	5.338	0.245	1.007	0.774
CCT	P	7.3	13.6	17.3	1.779	0.244	0.131	0.103
CAA	Q(Gln)	15	27.5	12	14.235	0.949	0.518	1.186
CAG	Q	29.5	12.1	34.1	14.235	0.483	1.176	0.417
AGA	R(Arg)	2.9	21.3	11.5	12.456	4.295	0.585	1.083
AGG	R	1.9	9.2	11.4	5.338	2.809	0.58	0.468
CGA	R	3.9	3	6.3	10.676	2.737	3.559	1.695
CGC	R	21	2.6	10.7	26.69	1.271	10.265	2.494
CGG	R	6.3	1.7	11.6	3.559	0.565	2.094	0.307
CGT	R	20.3	6.5	4.6	12.456	0.614	1.916	2.708
AGC	S(Ser)	16	9.7	19.3	7.117	0.445	0.734	0.369
AGT	S	9.5	14.2	11.9	16.014	1.686	1.128	1.346
TCA	S	7.8	18.8	12	23.132	2.966	1.23	1.928
TCC	S	8.9	14.2	11.9	7.117	0.8	0.501	0.598
TCG	S	8.7	8.5	4.4	14.235	1.636	1.675	3.235
TCT	S	8.7	23.5	14.7	17.794	2.045	0.757	1.21
ACA	T(Thr)	8 .2	17.8	15.1	24.911	3.038	1.399	1.65
ACC	T	22.8	12.6	19.4	10.676	0.468	0.847	0.55
ACG	T	14.8	7.9	6.1	16.014	1.082	2.027	2.625
ACT	T	9.1	20.3	13	16.014	1.76	0.789	1.232
GTA	V(Val)	11.1	11.8	7.2	24.911	2.244	2.111	3.46
GTC	V	15.1	11.6	14.6	10.676	0.707	0.92	0.731
GTG	V	25.5	10.6	28.4	5.338	0.209	0.504	0.188
GTT	V	18.5	22	11	17.794	0.962	0.809	1.618
TGG	W(Trp)	15.2	10.3	12.7	10.676	0.702	1.037	0.841
TAC	Y(Tyr)	12.1	14.6	15.5	7.117	0.588	0.487	0.459
TAT	Y	16.5	18.9	12.1	21.352	1.294	1.13	1.765
TAA	*	2	1	0.7	1.779	0.89	1.779	2.541
TAG	*	0.3	0.5	0.6	0	0	0	0
TGA	*	1.1	0.7	1.5	0	0	0	0

a,b,c represent the ratio of codon usage frequency in DEV UL21 to that in E.coli, yeast and human, respectively.

4. Discussion

As a common evolutionary phenomenon, it is well known that synonymous codon usage bias exists in each organism from prokaryotes to eukaryotes [10,29]. Many reports have revealed that codon usage bias are associated with a variety of biological factors, and the codon usage has been considered to be the equilibrium between natural selection and mutation pressure. It is seemed that the main factor for the codon usage in different viraceae and pestivirus is different, for instance, mutation bias may be a more important factor than natural selection in determining codon usage bias of some viruses, such as herpesviruses and Picornaviridae[20, 30]. Meanwhile, some other reports also showed that the G+C content is the main factor

that determines the codon usage bias in iridovirus genomes[31, 32]. Investigating of codon usage can help us know more about the molecular evolution. of the gene.

As far, the value of RSCU, ENC, CAI, and GC and GC3s contents are widely used to explore the codon usage variation among different genes. In this study, a comprehensive analysis of these codon usage indices of DEV UL21 gene and other 27 herpesvirus UL21 genes nucleotide sequences were performed through the program of CodonW and CUSP .

Relative synonymous codon usage (RSCU), which is the observed frequency of codon usage divided by the sum of codons per amino acid and multiplied by the actual number of codons per amino acid [22], used to examine the synonymous codon usage without the confounding influence of different amino acid compositions. RSCU values greater than 1.0 indicate that the corresponding codon is more frequently used than expected, the opposite is true for RSCU values less than 1.0. From Table 1, we found there are 8 A-ended and 11 T-ended codons at the third codon position which showed strong usage bias for coding the Ala, Asp, Glu, Pro, Phe, Gly, Lys, Leu, Val, and Thr amino acids. So we presumed that codons ended with A and T is preferentially used in DEV UL21 gene.

Codon adaptation index (CAI), it is a simple and effective measure of the overall synonymous codon usage bias of genes. The CAI value, ranging from 0 to 1.0, which is much closer to 1, the codon usage is much stronger.Customarily highly expressed genes correspond to high CAI values[25-26]. The CAI values of DEV UL21 gene and relational herpesvirus vary from 0.613 to 0.768, with a average value of 0.681, and we infer UL21 gene is lowly expressed gene in DEV genome. Moreover, the SD of CAI value is 0.0579, indicating that there is small variation in codon usage pattern among different herpesvirus UL21 genes.

Effective number of codons(ENC), it is used to quantify how far the codon usage of a gene departs from equal usage of synonymous codons and without dependence on sequence length or specific knowledge of preferred codons [33]. ENC ranges from 20 to 61. In an extremely biased gene where only one codon is used for each amino acid, this value would be 20; if all codons are used equally, it would be 61; and if the value of ENC is greater than 40, the codon usage bias was regarded as low. If the sequences in which ENC values less than 30, they are highly expressed while those more than 55 are poorly expressed genes [34]. The ENC values of DEV UL21 gene and relational herpesvirus vary from 29.499 to 59.205, with an average value of 47.293, and the SD value of 9.9667. The ENC values of DEV UL21 gene is 55.167, indicating that the codon usage bias of this gene is weak.

G + C composition and GC3s has been widely reported to be correlated with synonymous codon usage bias. GC3s is a good indicator of the extent of base composition bias, which represents the frequency of the nucleotide G+C at the synonymous third position of codons(excluding Met, Trp and the stop codons). The GC3S contents of each UL21 gene range from 13.00% to 96.58%, with an average value of 60.48% and SD of 24.249.

The plot of ENC against GC3S can be effectively used to analyse the heterogeneity of codon usage among genes.If G+C compositional constraint influences the codon usage, the GC3S and ENC correlated spots would lie on or just below the expected curve[27]. If a gene is subject to selection for translationally optimal codons, it will lie considerably below the expected curve.Fig.1 shows a large number of points lie near the solid curve of this distribution, so we conclude that the UL21 genes of herpesvirus are subject to G+C compositional constraints.

In order to investigate whether existing UL21 gene synonymous codon usage bias difference among different DEV strain, we compared UL21gene of DEV strain with those of DEV strain VAC and DEV UL21-like strain. Interestingly, all of them are same in the value of CAI, there was no difference or little difference in the codon usage bias parameters of the UL21 gene indicated by ENC, coding GC content. we conclude there is little different codon usage pattern of UL21 gene in different strains of DEV. It is reported that

UL21gene is conserved in alphaherpesvirus, and we presumed that there is no significant deviation in codon usage of conserved gene in different virus strains.

The phylogenetic tree analysis showed that the UL21gene of DEV CHv strain, DEV VAC strain and DEV UL21-like gene are quite homology to each other, and DEV UL21 gene evolutionarily is closely related to GaHV-2, GaHV-3 and MeHV-1, which belong to Mardivirus. Meanwhile, similar gene lenghth and CAI, ENC, GC content value are shared. Similar study have showed that the codon usage pattern of the DPV dUTPase, UL24, gC, UL27, UL28, UL30 and UL35 gene were also similar to herpesviruses Mardivirus, but UL25 gene, UL26 gene, UL26.5 gene, and UL29 gene exhibited a close relationship with the varicellovirus or the simplexvirus [35-39]. More researches are required to define which genus of herpesvirus the DEV belongs to.

Previous studies have revealed that there was a strongly significant correlation between gene expression level and codon usage bias in Escherichia coli, yeast and Humans[40]. The synonymous codon usage patterns are related to the abundance of isoaccepting tRNAs [13,14], highly expressed genes have a strong selective preference for the codons complementary to the most abundant tRNA species, whereas lowly expressed genes display more uniform codon usage patterns largely compatible with the mutational bias in the absence of translational selection [41,42]. Selecting relative abundance of isoaccepting tRNAs between the gene and expression system codon usage bias is more important for gene expression level. Table 3 shows there are 42, 45, 39 codons showing the DEV UL21 to E.coli, Yeast, H.sapiens ratio between 0.5 to 2.0 respectively, and the codon usage bias pattern in the DPV UL21 gene is similar to that of Yeast. We can conclude that the yeast expression system is more effective for expression of the UL21gene.

In summary, we can conclude that the UL21 gene of DEV is lowly pressed and weak codon usage bias. Meanwhile, our work also has provided a basic understanding of the evolution of herpesvirus UL21 gene and predicted effective expression system for DEV UL21 gene. All these analyses might provide some reference of future study of UL21 gene.

Acknowledgment

The research was supported by China 973 program（2011CB111606）,China Agricultural Research System(CARS-43-8), Changjiang Scholars and Innovative Research Team in University(PCSIRT0848) and National science and technology support program for agriculture（2011BAD34B03） An-chun Cheng and Ming-shu Wang are the corresponding authors. Institute of Preventive Veterinary Medicine, Sichuan Agricultural University, Wenjiang, Chengdu city, Sichuan, 611130, P.R.China. Avian Disease Research Center, College of Veterinary Medicine of Sichuan Agricultural University, 46# Xinkang Road, Ya'an, Sichuan 625014, P.R. China.Key Laboratory of Animal Disease and Human Health of Sichuan Province, Sichuan Agricultural University, Wenjiang, Chengdu city, Sichuan, 611130, P.R.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] Thomas NT, Gerry M.D, Saunders Ltd. Waterfowl. Handbook of Avian Medicine (Second Edition), USA: Saunders Ltd,2009.
[2] Sandhu TS, Metwally SA. Diseases of poultry.Blackwell Publishing, 2008
[3] Fauquet CM, Mayo MA, Maniloff J . Virus Taxonomy: Eighth Report of the International Committee on Taxonomy of Viruses, Elsevier Academic Press, California, 2005

[4] Duret L.Evolution of synonymous codon usage in metazoans. Curr Opin Genet 2002:12: 640-9.
[5] Paul MS, Elizabeth B, Russell J. Grocock. Variation in the strength of selected codon usage bias among bacteria. Nucleic Acids Research, 2005, 33: 1141-3
[6] Yang ZH, Nielsen R. Mutation-selection models of codon substitution and their use to estimate selective strengths on codon usage. Molecular Biology and Evolution,2007:25:568-79.
[7] David SL, Dmitri AP, Philipp WM. Faster than Neutral Evolution of Constrained Sequences: The Complex Interplay of Mutational Biases and Weak Selection. Genome Biol : 2011:3: 383-95
[8] Nicolas Rodrigue, Hervé Philippe, Nicolas Lartillotb. Mutation-selection models of coding sequence evolution with site-heterogeneous amino acid fitness profiles. Proc. Natl. Acad. Sci. 2010:107: 4629-34
[9] Emery LR, Sharp PM. Impact of translational selection on codon usage bias in the archaeon Methanococcus marripaludis. Biology Letters, 2010:7: 131-5.
[10] Supek F,Skunca N, Repar J, et al. Translational Selection Is Ubiquitous in Prokaryotes. PLoS Genet 6(6): e1001004. doi:10.1371/journal.pgen.1001004
[11] Plotkin JB, Dushoff J, Desai MM,et al. Codon usage and selection on proteins. Jouranl of Molecular Evolution, Spinger 2006:63:635–53
[12] Gu W, Zhou T, Ma J, et al. Analysis of synonymous codon usage in SARS Coronavirus and other viruses in the Nidovirales. Virus Research,2004:101:155-61
[13] Kanaya S, Yamada, Y Kudo Y, et al. 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, Elsevier Ltd 1999:238:143–55
[14] Noguchi S, Satow Y. Purification of human β2-adrenergic receptor expressed in methylotrophic yeast Pichia pastoris", Journal of Biochemistry 2006:140:799–804
[15] Wan XF, Xu D, Kleinhofs A, et al. Quantitative relationship between synonymous codon usage bias and GC composition across unicellular genomes. BMC Evolutionary Biology 2004:4, doi:10.1186/1471-2148-4-19
[16] Knight RD, Freeland SJ, Landweber LF. A simple model based on mutation and selection explains trends in codon and amino-acid usage and GC composition within and across genomes. Genome Biology2001.:2: research0010.1–research0010.13, 2001.
[17] Wang B, Liu J, Jin L,et al. Complex Mutation and Weak Selection together Determined the Codon Usage Bias in Bryophyte Mitochondrial Genomes. Journal of Integrative Plant Biology 2010:52: 8
[18] Nakamura M, Sugiura M. Translation efficiencies of synonymous codons are not always correlated with codon usage in tobacco chloroplasts. The Plant Journal 2007:4:128–34
[19] Eyre WA. Synonymous codon bias is related to gene length in Escherichia coli: selection for translational accuracy?", Molecular Biology and Evolution 1996:13: 864–72
[20] Fu MF. Codon usage bias in herpesviru. Arch Virol 2010:155:391–96
[21] Sourav RC, Debaprasad M. A detailed comparative analysis on the overall codon usage pattern in herpesviruses. Virus Research 2010:148:31–43
[22] S. L. Fennoy, J. Bailey-Serres. Synonymous codon usage in Zea mays L. Nuclear genes is varied by levels of C and G-ending codons .Nucleic Acids Researc1993:21:5294-300.
[23] Wright F. The effective number of codons' used in a gene. Gene 1990:87:23-9.
[24] Zhang L, Wang M, Liu WQ. Analysis of codon usage and nucleotide composition bias in polioviruses. Virology Journal, 2011: 8:146, doi:10.1186/1743-422X-8-146
[25] Xia XH. An improved implementation of codon adaptation index. Evolutionary Bioinformatics 2007:3: 53-8.
[26] Carbone1 A, Zinovyev A. Codon adaptation index as a measure of dominating codon bias. Bioinformatics 2003:19:2005-15.
[27] Lu H, Zhao WM, Zheng Y.Analysis of synonymous codon usage bias in Chlamydia," Acta Biochim Biophys Sin, Oxford University Press, Oxford, 2005, 37:.1-10.

[28] Fennoy SL, BaileySJ. Synonymous codon usage in Zea mays L. Nuclear genes is varied by levels of C and G-ending codons. Nucleic Acids Research 1993: 21: 5294-300.
[29] Mario dos Reis, Lorenz Wernisch. Estimating Translational Selection in Eukaryotic Genomes. Mol Biol Evol 2009:26: 451-61.
[30] Shackelton LA, Parrish CR, Holmes EC. Evolutionary basis of codon usage and nucleotide composition bias in vertebrate DNA viruses. J Mol Evol 2006:62:551-63.
[31] Zhong J, Li Y, Zhao S, et al. Mutation pressure shapes codon usage in the GC-Rich genome of foot-and-mouth disease virus. Virus Genes 2007:35:767-76.
[32] Tsai CT, Lin CH, Chang CY. Analysis of codon usage bias and base compositional constraints in iridovirus genomes. Virus Res 2007:126:196-206.
[33] Novembre JA. Accounting for background nucleotide composition when measuring codon usage bias", Mol Biol Evo 2002:19: 1390-94
[34] Hassan S, Mahalingam V, Kumar V. Synonymous codon usage analysis of thirty two mycobacteriophage genomes. Advances in Bioinformatics 2009, doi: 10.1155/2009/316936.
[35] Jia RY, Cheng AC, Wang MS,et al. Analysis of synonymous codon usage in the UL24 gene of duck enteritis virus. Virus Genes 2009: 38: 96–103.
[36] Zhao LC, Cheng AC, Wang MS. Characterization of codon usage bias in the dUTPase gene of duck enteritis virus," Prog Nat Sci, 2008:18:1069-76.
[37] Liu SW, Chen SH, Li HX, et al. Molecular characterization of the herpes simplex virus 1 (HSV-1) homologues, UL25 to UL30, in duck enteritis virus (DEV). Gene 2007:401:88-96.
[38] Cai MS , Cheng AC, Wang MS. Characterization of synonymous codon usage bias in the duck plague virus UL35 gene. Intervirology 2009:52: 266-78
[39] Shen FX, Cheng AC, Wang MS, et al. Factors affecting synonymous codon usage bias in the gC gene of DPV CHv strain. African Journal of Microbiology Research 2010:4:343-53,
[40] Chang H, Cheng AC, Wang MS, et al. Characterization of codon usage bias in the newly identified DPV gE gene, IEEE, the 2nd International Conference on BioMedical Engineering and Informatics, 2009:4: 1836-1841.
[41]Lesnik T, Solomovici J, Deana A,et al. Ribosome traffic in E. coli and regulation of gene expression." J Theor Biol 2000;202: 175185.
[42] Sharp PM, Tuohy TM, Mosurski KR. Codon usage in yeast: cluster analysis clearly differentiates highly and lowly expressed genes. Nucleic Acids Res 1986; 14: 5125–43.