Available online at www.sciencedirect.com
ScienceDirect
Progress in Natural Science 18 (2008) 1069–1076
Progress in Natural Science
www.elsevier.com/locate/pnsc

Characterization of codon usage bias in the dUTPase gene of duck enteritis virus

Lichan Zhao [a], Anchun Cheng [a,b,*], Mingshu Wang [a,b,*], Guiping Yuan [c], Mingsheng Cai [a]

[a] *Avian Disease Research Center, College of Veterinary Medicine of Sichuan Agricultural University, Xinkang Road 46, Ya'an, Sichuan 625014, China*
[b] *Key Laboratory of Animal Disease and Human Health of Sichuan Province, Ya'an 625014, China*
[c] *Analytical & Testing Center, Sichuan University, Chengdu 610064, China*

Received 24 December 2007; received in revised form 20 February 2008; accepted 17 March 2008

Abstract

A comparative analysis of the codon usage bias in the newly discovered dUTPase gene (Assigned Accession No.: DQ486149) of the duck enteritis virus (DEV) and the dUTPase gene of 32 reference herpesviruses was performed. The results indicated that the DEV dUTPase gene encodes a protein of 477 amino acids, which includes five conserved motifs with a 3–1–2–4–5 arrangement. The codon adaptation index (CAI), effective number of codons (ENC), and GC_{3S} values indicated synonymous codon usage bias in the dUTPase gene of herpesviruses, and this synonymous bias was correlated with host evolution. The codon usage patterns of the DEV dUTPase gene were phylogenetically conserved and similar to that of the dUTPase genes of the avian alphaherpesvirus. Although codon usage in each microorganism was different, there were no strain-specific differences among them. Sixty-one codons in the predicted polypeptide, with a strong bias towards A and T at the third codon position, were used. Comparison of the codon usage in the dUTPase gene of different organisms revealed that there were 19 codons showing distinct codon usage differences between the DEV and *Escherichia coli* dUTPase genes; 16 between the DEV and yeast dUTPase genes; and 15 between the DEV and human dUTPase genes. Analysis of variance (ANOVA) showed significant differences between the DEV and yeast dUTPase genes ($r = 0.536$, $P < 0.01$). The extent of codon usage bias in the DEV dUTPase gene was highly correlated with the gene expression level, therefore the results may provide useful information for gene classification and functional studies.

Keywords: DEV; dUTPase; Codon usage bias; EMBOSS

1. Introduction

DNA sequence analyses have demonstrated that synonymous codons are used differently by living organisms, and each type of genome has a specific coding strategy [1–3]. However, degenerate codons are not present at equal frequencies in the genes, giving rise to the phenomenon termed "codon usage bias" [4,5]. Studies on synonymous codons and amino acid usages in living organisms revealed that they vary between genomes, between genes, and even between the different parts of a gene. While factors such as mutational pressure [6–8], translational selection [9–12], and secondary structure of proteins [13–17] influence codon usage in various organisms, amino acid usage was shown to be governed by hydrophobicity, aromaticity, cysteine content, and mean molecular weight [18–20]. In general, highly expressed genes have a strong preference for a subset of codons, while lowly expressed genes have a more uniform pattern of codon usage [21,22]. The biases in

* Corresponding authors. Tel./fax: +86 835 2885774.
E-mail addresses: chenganchun@vip.163.com (A. Cheng), mshwang@163.com (M. Wang).

1002-0071/$ - see front matter © 2008 National Natural Science Foundation of China and Chinese Academy of Sciences. Published by Elsevier Limited and Science in China Press. All rights reserved.
doi:10.1016/j.pnsc.2008.03.009

synonymous codon usage and amino acid usage have been studied in only a few bacteriophage genomes. In *Escherichia coli* and yeast, synonymous codon usage patterns are related to the abundance of isoaccepting tRNAs [10,23]. The interpretation of codon usage patterns in these species is complicated by cell-specific, tissue-specific, and developmentally regulated expression of genes [24,25]. Moreover, even for functionally homologous genes, remarkable differences in codon usage exists across species [26–28]. Amino acid compositions of the proteins encoded by the pseudorabies virus (PRV) genes have been reported and indicated that the codon usage bias in the genes could be related to the different functions of the encoded proteins [29]. Recent analyses of herpesvirus codon composition and codon usage are primarily focused on the PrV [29], herpes simplex virus type 1 (HSV-1) [30], and Epstein–Barr virus (EBV) [31]. However, data regarding codon usage bias in the duck enteritis virus (DEV) genome is not yet available.

DEV is a member of the *Herpesviridae* family, which can induce viral enteritis (VE) in ducks; it has caused significant economic losses in domestic and wild water fowl [32]. Most of the previous research work has focused on the epidemiology and prevention of this disease; however, limited molecular biology data is available regarding the DEV genome. Recently, a DEV genomic library was constructed successfully for the first time [33] and the dUTPase gene was discovered in our laboratory. In this study, synonymous codon usage in the DEV dUTPase gene was analyzed and compared with the codon usage in the dUTPase gene of 32 species of herpesvirus. Moreover, the codon usage bias in the DEV dUTPase gene was compared with that in the dUTPase gene of *E. coli*, yeast, and humans. Codon usage data of the DEV dUTPase gene and the comparison results might provide some insights into the features of the DEV genome. This study may also provide insights into the expression and possible function of DEV dUTPase.

2. Materials and methods

2.1. Virus species and gene sequences

The DEV CHv strain, which is a high-virulence field strain of DEV, was obtained from Key Laboratory of Animal Disease and Human Health of Sichuan Province. The dUTPase gene (GenBank Accession No.: DQ486149) of the DEV CHv strain was first discovered by constructing a DEV genomic library in our laboratory. The nucleotide sequences of the dUTPase gene of 32 reference herpesviruses were obtained from the NCBI GenBank nucleotide database (Table 1).

2.2. Molecular characterization and phylogenetic analysis of the DEV dUTPase gene

SHOWORF, an EMBOSS nucleotide translation program, was used to read and transfer the nucleotide sequence data to a computer file. For characterization of DEV dUTPase DNA and protein sequences, the GenBank sequence databases were scanned using the BLASTN and BLASTP programs, respectively. Multiple sequence alignment and phylogenetic analysis were performed for the dUTPase gene of 33 herpesviruses (Table 1) with CLUSTAL-X and TREEVIEW software [34].

2.3. Codon usage analysis of the dUTPase gene in DEV and the 32 reference herpesviruses

Generally, the effective number of codons (ENC) of a gene is used to quantify its codon usage bias, which is essentially independent of gene length. The ENC value [35] of the dUTPase gene of each reference herpesvirus was computed with the EMBOSS CHIPS online service program [5,36]. Another simple and effective method of examining synonymous codon usage bias is the codon adaptation index (CAI) value, which was calculated with the EMBOSS CAI program [37]. The peculiarity in codon usage frequency and the G + C content of the gene sequences were also calculated with the EMBOSS CUSP program [38].

2.4. Analysis of the phylogenetic persistence in codon usage bias in the DEV dUTPase gene

Codon usage bias in the DEV dUTPase gene was determined with the SPSS 11.0 software, and we performed one-way analysis of variance (ANOVA) analysis to compare the dUTPase codon usage bias between DEV, *E. coli*, yeast, and humans. The database of the codon usage in *E. coli*, yeast, and humans is available at http://www.kazusa.or.jp/condon.

3. Results

3.1. Characterization of the DEV dUTPase gene

The deduced amino acid sequence encoded by the 1344 bp open reading frame (ORF) of the dUTPase gene is shown in Fig. 1. Five conserved motifs (motif 1, PKRLEDAGYDI; motif 2, GRSS; motif 3, GVVDAGYRG; motif 4, GDRVAQ; motif 5, RREGGFGS) boxed in Fig. 1 in the DEV dUTPase were rearranged in the order 3–1–2–4–5. It is now termed the Class 2 dUTPases with amino acid residues around twice as long to the Class 1 (motifs ordered as 1–2–3–4–5, sequences typically of around 150 residues) [39]. A phylogenetic tree based on the amino acid sequences of the dUTPase in the 33 herpesviruses is shown in Fig. 2; the general branching pattern coincided with other previously published phylogenetic analyses [40,41]. As shown in Fig. 2, the dUTPase gene within the same herpesvirus subfamily (*alphaherpesvirinae*, *betaherpesvirinae*, *gammaherpesvirinae*) or in the same microorganism is clustered together.

Table 1
Nucleotide sequences of the dUTPase gene of DEV and 32 reference herpesviruses

Species	Virus name (abbreviation)	Natural host	GenBank Accession No.	Length (bp)
Alphaherpesvirinae	Duck enteritis virus CHv strain (DEV CHv)	Duck	DQ486149	1344
	Duck enteritis virus clone-03 strain (DEV clone-03)	Duck	EF492886	1344
	Meleagrid herpesvirus 1 (MeHV-1)	Meleagrid	AF291866	1314
	Gallid herpesvirus 2 (GaHV-2)	Avian	EF523390	1311
	Marek's disease virus serotype 2 (MDV2)	Avian	AB012572	1188
	Equid herpesvirus 1 (EHV-1)	Equid	AY665713	1698
	Equid herpesvirus 4 (EHV-4)	Equid	AF030027	1697
	Bovine herpesvirus 5 (BoHV-5)	Bovine	NC 005261	966
	Bovine herpesvirus 1 (BoHV-1)	Bovine	Z54206	978
	Suid herpesvirus 1 (SuHV-1)	Swine	U38547	1095
	Pseudorabies virus (PRV)	Swine	U38548	807
	Cercopithecine herpesvirus 1 (CeHV-1)	Cercopithecine	NC 004812	1110
	Cercopithecine herpesvirus 2 (CeHV-2)	Cercopithecine	NC 006560	1104
	Psittacid herpesvirus 1 (PsHV-1)	Psittacid	NC 005264	1245
	Gallid herpesvirus 1(GaHV-1)	Avian	Y14300	1017
	Human herpesvirus 1 (HHV-1)	Human	NC 001806	1116
	Human herpesvirus 2 (HHV-2)	Human	NC 001798	1110
	Cercopithecine herpesvirus 16 (CeHV-16)	Cercopithecine	NC 007653	1104
Betaherpesvirinae	Human herpesvirus 6 (HHV-6)	Human	X92436	447
	Human herpesvirus 7 (HHV-7)	Human	AF037218	1140
	Human herpesvirus 5 (HHV-5)	Human	BK000394	1167
	Cercopithecine herpesvirus 8 (CeHV-8)	Cercopithecine	NC 006150	1032
	Murid herpesvirus 1 (MuHV-1)	Murine	NC 004065	1206
	Murid herpesvirus 2 (MuHV-2)	Murine	NC 02512	1044
	Pongine herpesvirus 4 (PoHV-4)	Chimpanzee	AF480884	1161
Gammaherpesvirinae	Human herpesvirus 4 (HHV-4)	Human	AJ507799	837
	Epstein–Barr virus (EBV)	Human	L07923	825
	Human herpesvirus 8 type P (HHV-8)	Human	NC 009333	888
	Bovine herpesvirus 4 (BoHV-4)	Bovine	NC 002665	849
	Equid herpesvirus 2 (EHV-2)	Equid	NC 001650	870
	Saimiriine herpesvirus 2 (SaHV-2)	Squirrel	NC 001350	864
	Alcelaphine herpesvirus 1 (AlHV-1)	Alcelaphine	NC 002531	897
	Ovine herpesvirus 2 (OvHV-2)	Ovine	AY839756	882

3.2. Codon usage analysis of the dUTPase gene in DEV and the reference herpesviruses

The results obtained by EMBOSS analysis of the CAI, ENC, and coding GC and GC_{3S} content of 33 herpesviruses species are shown in Table 2. Codon usage in the dUTPase gene is highly nonrandom in all the herpesviruses; the overall base composition of the dUTPase genes in these species also differs dramatically. However, interestingly, there was no difference in the codon usage bias parameters of the dUTPase gene indicated by CAI, ENC, coding GC content, and GC_{3S} values, in both the DEV CHv strain and DEV clone-03 strain, which is identical to the suid herpesvirus 1 and pseudorabies virus. Thus, we presumed that there is no significant deviation in codon usage in different virus strains. The CAI value of different herpesviruses varied from 0.62 to 0.83, with a mean value of 0.70 and a standard deviation (SD) of 0.05; their ENC values ranged from 31.77 to 60.32, with a mean value of 48.51 and standard deviation (SD) of 9.19. Since the approximate 60.32% ENC value of the dUTPase gene in the DEV CHv strain was the highest among the reference herpesviruses, its codon usage bias is less. The GC_{3S} varied from 22.92% to 93.75%, with a mean of 62.03% and SD of 18.75%.

It has been reported that a plot of ENC against GC_{3S} can be effectively used to explore the heterogeneity of codon usage among genes [35]. If the codon usage pattern of genes influences parameters other than the GC content, comparing the actual distribution of genes with the expected distribution could be indicative. In other words, if GC_{3S} is the only determinant factor shaping the codon usage pattern, the values of ENC would fall on a continuous curve, which represents random codon usage [42]. Fig. 3 shows the distribution plot of the ENC and GC_{3S} values for the dUTPase gene in the reference herpesviruses. The points in the plot were fairly spread out and the bulk of genes did not appear to follow the theoretical curve, which suggests that factors other than gene composition contribute to the codon usage pattern in the reference herpesviruses.

3.3. Variation in DEV dUTPase codon usage and amino acid composition

While the CAI, ENC, and the related measures indicate the overall DEV dUTPase codon bias, it is also

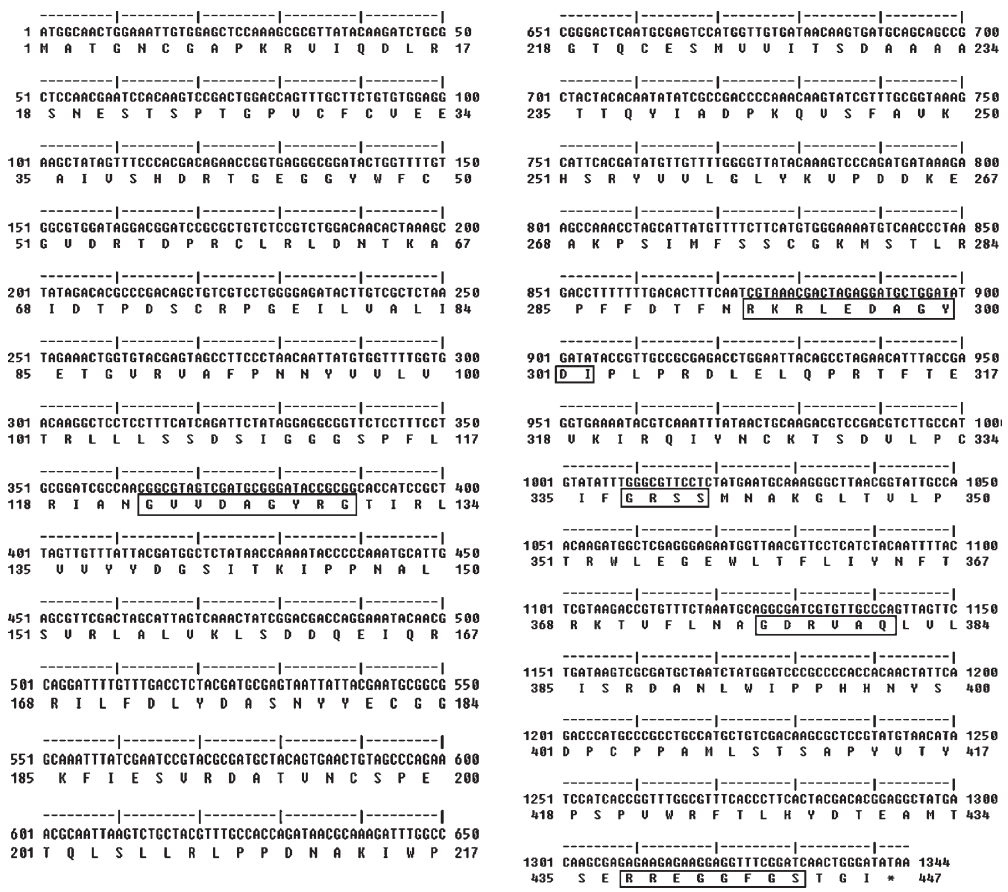

Fig. 1. Nucleotide sequence of the 1344 bp segment of the DEV dUTPase gene and the derived amino acid sequence. Five highly conserved motifs are shown in the boxes.

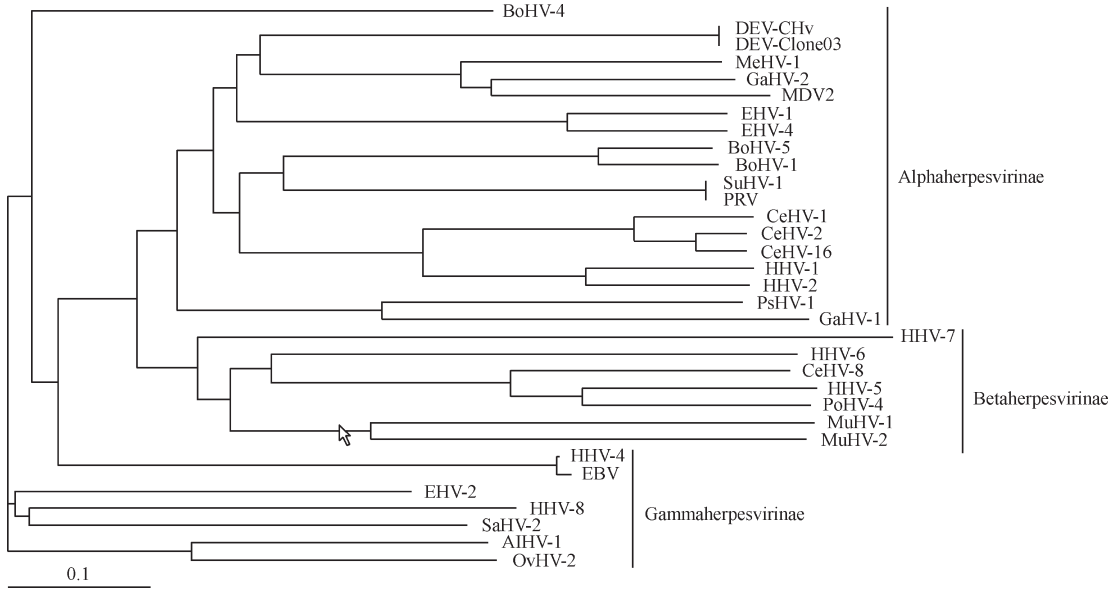

Fig. 2. Phylogenetic tree based on the dUTPase amino acid sequences in 33 herpesviruses (Table 1); the tree was constructed with CLUSTAL-X and TREEVIEW software.

important to closely investigate the pattern of codon bias. Table 3 shows the codon preferences of DEV dUTPase gene. Sixty-one codons (excluding the stop codons) in the polypeptide, with a strong bias towards the codons with A and T at the third codon position, were used. A high level of diversity in codon usage bias exists for coding the Ala, Val, Phe, Gly, Ile, Lys, and Gln amino acids.

Table 2
Summary of dUTPase gene analysis in different herpesvirus species

Rank	Virus name	CAI[a]	ENC[b]	Coding GC (%)	GC_{3S} (%)
1	Duck enteritis virus CHv strain	0.65	60.32	47.47	43.97
2	Duck enteritis virus clone-03 strain	0.65	60.32	47.47	43.97
3	Meleagrid herpesvirus 1	0.65	58.30	46.88	44.98
4	Gallid herpesvirus 2	0.62	56.78	44.16	37.30
5	Marek's disease virus serotype 2	0.63	57.44	50.67	53.28
6	Equid herpesvirus 1	0.70	54.23	56.18	63.78
7	Equid herpesvirus 4	0.67	55.69	49.50	48.67
8	Bovine herpesvirus 5	0.72	48.44	72.15	74.84
9	Bovine herpesvirus 1	0.71	49.97	68.81	70.86
10	Suid herpesvirus 1	0.75	31.77	73.36	91.45
11	Pseudorabies virus	0.75	31.77	73.36	91.45
12	Cercopithecine herpesvirus 1	0.74	33.69	75.14	92.16
13	Cercopithecine herpesvirus 2	0.75	32.06	75.91	92.93
14	Psittacid herpesvirus 1	0.69	51.75	58.47	60.00
15	Gallid herpesvirus 1	0.73	56.20	44.44	51.62
16	Human herpesvirus 1	0.69	47.40	66.67	75.54
17	Human herpesvirus 2	0.72	42.76	68.74	81.35
18	Cercopithecine herpesvirus 16	0.74	32.10	76.54	93.75
19	Human herpesvirus 6	0.69	58.95	42.73	38.93
20	Human herpesvirus 7	0.71	46.25	33.42	37.89
21	Human herpesvirus 5	0.69	53.69	55.78	56.04
22	Cercopithecine herpesvirus 8	0.67	48.99	48.45	50.87
23	Murid herpesvirus 1	0.67	47.57	61.94	59.45
24	Murid herpesvirus 2	0.67	47.01	66.86	60.34
25	Pongine herpesvirus 4	0.68	46.45	64.69	58.40
26	Human herpesvirus 4	0.70	58.60	61.17	59.86
27	Epstein–Barr virus	0.70	56.68	61.45	60.73
28	Human herpesvirus 8 type P	0.73	55.04	53.83	61.15
29	Bovine herpesvirus 4	0.70	53.72	44.05	42.05
30	Equid herpesvirus 2	0.83	36.81	56.90	84.14
31	Saimiriine herpesvirus 2	0.63	39.46	34.72	22.92
32	Alcelaphine herpesvirus 1	0.73	50.01	53.29	67.22
33	Ovine herpesvirus 2	0.79	40.73	56.35	75.17

[a] EMBOSS codon adaptation index.
[b] Effective number of codons.

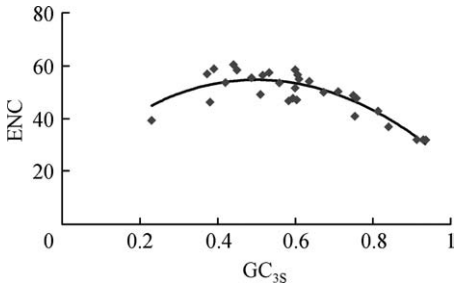

Fig. 3. The plot of effective number of codons (ENC) and guanine (G) + cytosine (C) frequency at the synonymous third position of codons (GC_{3S}) of the dUTPase gene in the DEV CHv strain and 32 reference herpesviruses.

3.4. Comparison of dUTPase codon usage in DEV with that in E. coli, yeast, and humans

Generally, the codon usage bias in genes remains at a certain level across species. Data revealed that there are 19 codons showing distinct usage differences between the DEV and *E. coli* dUTPase genes; 16, between the DEV and yeast dUTPase genes; and 15, between the DEV and human dUTPase genes. ANOVA revealed that the codon usage frequencies of the dUTPase gene of DEV, *E. coli,* yeast, and humans showed great variation (refer to Table 4 and Fig. 4). Significantly high differences were observed between the DEV and yeast dUTPases ($r = 0.536$, $P < 0.01$). Additionally, based on Fig. 4, we conclude that the codon usage pattern of the DEV and human dUTPase genes is similar and that the eukaryotic expression system may be more suitable for dUTPase gene expression.

4. Discussion

Among microorganisms, the most commonly accepted hypothesis for the unequal usage of synonymous codons states that it is the result of mutational biases and natural selection acting at the level of translation. Several measures of the degree of codon usage bias in a certain gene have been developed. The CAI value, which uses a reference set of highly expressed genes from a species to assess the relative merits of each codon, and a value for a gene sequence are calculated from the frequency of the use of

Table 3
Codon preferences in DEV dUTPase gene analyzed with the CUSP program

Codon	Amino acid	Fract[a]	Frequency[b]/1000	No.	Codon	Amino acid	Fract	Frequency/1000	No.
GCA	A	0.286	17.857	8	CCA	P	0.355	24.554	11
GCC	A	0.250	15.625	7	CCC	P	0.161	11.161	5
GCG	A	0.107	6.696	3	CCG	P	0.258	17.857	8
GCT	A	0.357	22.321	10	CCT	P	0.226	15.625	7
TGC	C	0.385	11.161	5	CAA	Q	0.700	15.625	7
TGT	C	0.615	17.857	8	CAG	Q	0.300	6.969	3
GAC	D	0.448	29.018	13	AGA	R	0.188	13.393	6
GAT	D	0.552	35.714	16	AGG	R	0.094	6.696	3
GAA	E	0.550	24.554	11	CGA	R	0.156	11.161	5
GAG	E	0.450	20.089	9	CGC	R	0.250	17.857	8
TTC	F	0.389	15.625	7	CGG	R	0.031	2.232	1
TTT	F	0.611	24.554	11	CGT	R	0.281	20.089	9
GGA	G	0.333	22.321	10	AGC	S	0.176	13.393	6
GGC	G	0.333	22.321	10	AGT	S	0.147	11.161	5
GGG	G	0.200	13.393	6	TCA	S	0.235	17.857	8
GGT	G	0.133	8.929	4	TCC	S	0.206	15.625	7
CAC	H	0.800	8.929	4	TCG	S	0.088	6.696	3
CAT	H	0.200	2.232	1	TCT	S	0.147	11.161	5
ATA	I	0.600	33.482	15	ACA	T	0.303	22.321	10
ATC	I	0.240	13.393	6	ACC	T	0.212	15.625	7
ATT	I	0.160	8.929	4	ACG	T	0.212	15.625	7
AAA	K	0.647	24.554	11	ACT	T	0.273	20.089	9
AAG	K	0.353	13.393	6	GTA	V	0.235	17.857	8
CTA	L	0.200	17.857	8	GTC	V	0.147	11.161	5
CTC	L	0.150	13.393	6	GTG	V	0.235	17.857	8
CTG	L	0.175	15.625	7	GTT	V	0.382	29.018	13
CTT	L	0.075	6.696	3	TGG	W	1.00	13.393	6
TTA	L	0.200	17.857	8	TAC	Y	0.444	17.857	8
TTG	L	0.200	17.857	8	TAT	Y	0.556	22.321	10
ATG	M	1.00	15.625	7	TAA	*	1.000	2.232	1
AAC	N	0.471	17.857	8	TAG	*	0.000	0.000	0
AAT	N	0.529	20.089	9	TGA	*	0.000	0.000	0

*Refers to stop codon.
[a] The "Fract" column shows the proportion of usage of a given codon in its redundant set (i.e., the set of codons that code for the same amino acid).
[b] The "Frequency" column lists the number of codons present per 1000 bases in the input sequence(s).

Table 4
Codon usage frequencies ANOVA analysis of the DEV, *Escherichia coli,* yeast, and human dUTPase genes

r	DEV	*Escherichia coli*	Yeast	Human
DEV	1	0.312*	0.536**	0.300*
Escherichia coli		1	0.418**	0.604**
Yeast			1	0.384**
Human				1

* Significant correlation at the 0.05 level (2-tailed).
** Significant correlation at the 0.01 level (2-tailed).

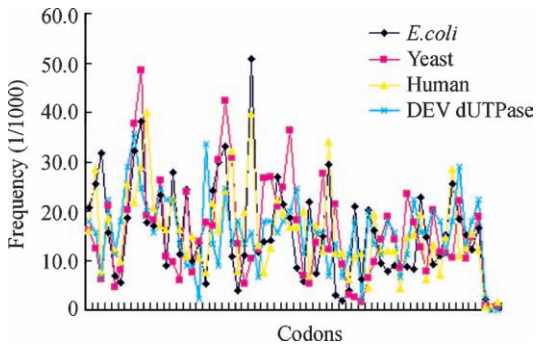

Fig. 4. Comparison between the dUTPase codon preferences in DEV, *Escherichia coli*, yeast, and humans.

all codons in that gene sequence [37]. The index assesses the extent to which selection has been effective in molding the pattern of codon usage. In this respect, it is useful for predicting the expression level of a gene, for assessing the adaptation of viral genes to their hosts, and for comparing codon usage in different organisms. It has been reported that proteins encoded by genes with high CAI values are rich in amino acids carried by the most abundant major tRNA; this implies that the forces shaping codon usage can also influence protein sequences [43]. Thus, the index may also provide an approximate indication of the possible success of heterologous gene expression. In our study, we used the EMBOSS CHIPS, EMBOSS CAI, and EMBOSS CUSP programs to deduce the ENC and CAI values and the GC content, respectively, of the dUTPase gene from its nucleotide sequence in the DEV CHv strain; subsequently these values were compared with those of the 32 reference herpesvirus species. The data of synonymous codon usage bias showed certain disparity of each herpesvirus from different organism. Multiple sequence alignment of the amino acid sequences demonstrated that the DEV dUTPase is a Class 2 dUTPase with the 3–1–2–4–5 motif arrangement, which are found only in *alphaherpesvirinae*

and *gammaherpesvirinae* [41,44–46]. Here, the phylogenetic tree analysis based on the dUTPase gene in the reference herpesviruses revealed that the dUTPase of the DEV CHv strain and some fowl herpesviruses such as DEV clone-03 strain, GaHV-2, GaHV-3, MDV2, and MeHV-1 were clustered within a monophyletic clade and grouped with some alphaherpesviruses. Interestingly, the synonymous codon usage pattern of the dUTPase gene in the DEV CHv strain is similar to that in other alphaherpesviruses. Thus, we concluded that the synonymous codon usage bias is correlated with the molecular characterization of the gene, and the DEV may be a part of the alphaherpesvirus subfamily.

The most plausible and well-documented selection-based explanation for codon usage bias is the selection for efficient translation related to the relative abundance of isoaccepting tRNAs [47,48]. Here we show the base composition of the DEV dUTPase gene, and its pattern of codon bias contributing to the existing variation in codon usage bias within and between other species. It is obvious that greater the skew in base composition, greater the bias in codon usage. Moreover, the correlation analysis with the 61 dUTPase codons of DEV, *E. coli*, yeast, and humans revealed a very wide range in codon frequencies and their proportion in a redundant set of codons. Among the codon usage bias patterns in *E. coli*, yeast, and humans, the codon usage bias pattern in the DEV dUTPase gene is similar to that in the human dUTPase gene (Fig. 1). Thus, we can assume that the eukaryotic expression system is suitable for heterologous expression of the DEV dUTPase gene.

Comparative analysis of dUTPase gene in DEV and the reference herpesviruses indicated that synonymous codon usage in the gene was phylogenetically conserved. Data in Table 2 show that the dUTPase genes in DEV, MeHV-1, GaHV-2, and MDV2, whose natural host is avian, have a stronger correlation than the dUTPase genes of herpesviruses with mammalian hosts, such as EHV-1, BoHV-5, BoHV-1, and SuHV-1. This indicates that the dUTPase genes of herpesviruses belonging to the same host have similar sequence length, and their CAI value is almost the same. Although the dUTPase from each herpesvirus was different in their codon usages, no strain-specific codon usage in the dUTPase gene was observed among the reference species. The codon usage pattern among different species is a complex phenomenon since it is influenced by many factors. It is important to elucidate the underlying mechanisms of codon usage pattern in order to understand the evolution of the species. Thus, we confirmed the finding that the G + C content and gene length are relative to the codon usage bias, shorter genes tend to have a higher bias than longer genes. As discussed above, some explanations of codon usage bias may be affected by the length of a gene and the gene's evolutionary history. Furthermore, it is clear that dUTPase is a ubiquitous and important enzyme that hydrolyzes dUTP to dUMP. It has been reported that many viruses encode virus-specific dUTPases that play an essential role in maintaining the integrity of the viral DNA, both by reducing the dUTP levels and by providing the substrate for thymidylate synthase (TS) [33,39,49]. We also predict that the DEV dUTPase may greatly influence the multiplication of DEV, however, further studies are required to confirm this hypothesis. Research on synonymous codon usage can prove helpful in genetic engineering to increase the output of target proteins. It is also a useful tool for gene classification and gene function prediction. Thus, analysis of codon usage bias in the newly discovered DEV dUTPase gene may be of great importance for gene characterization and for assessing the possible role of dUTPase in viral pathogenesis.

Acknowledgements

The research was supported by the National Natural Science Foundation of China (3047129/30771598), New Century Excellent Talents Program in University (NCET-04-0906/NCET-06-0818), Scientific and Technological Innovation Major Project Funds in University (706050), Sichuan Province Outstanding Youths Fund (03ZQ026-029), Sichuan Province Basic Research Program (04JY021-100/04JY029-006-1/05JY029-109/05JY029-003/07JY029-016/07JY029-016) and Program for Key Disciplines Construction of Sichuan Province (SZD0418).

References

[1] Jarrett N, Payne S, Turner P, et al. Someone to talk to' and pain control': what people expect from a specialist palliative care team. Palliat Med 1999;13(2):139–44.
[2] Ikemura T. Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes. J Mol Biol 1981;146(1):1–21.
[3] Ikemura T. Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the *E. coli* translational system. J Mol Biol 1981;151(3):389–409.
[4] Ellis J, Morrison DA, Kalinna B. Comparison of the patterns of codon usage and bias between *Brugia*, *Echinococcus*, *Onchocerca* and *Schistosoma* species. Parasitol Res 1995;81(5):388–93.
[5] Novembre JA. Accounting for background nucleotide composition when measuring codon usage bias. Mol Biol Evol 2002;19(8):1390–4.
[6] Levin DB, Whittome B. Codon usage in nucleopolyhedroviruses. J Gen Virol 2000;81(Pt. 9):2313–25.
[7] Jenkins GM, Pagel M, Gould EA, et al. Evolution of base composition and codon usage bias in the genus Flavivirus. J Mol Evol 2001;52(4):383–90.
[8] Jenkins GM, Holmes EC. The extent of codon usage bias in human RNA viruses and its evolutionary origin. Virus Res 2003;92(1):1–7.
[9] Grantham R, Gautier C, Gouy M, et al. Codon catalog usage is a genome strategy modulated for gene expressivity. Nucleic Acids Res 1981;9(1):43–74.
[10] Ikemura T. Codon usage and tRNA content in unicellular and multicellular organisms. Mol Biol Evol 1985;2(1):13–34.
[11] Sharp PM, Cowe E. Synonymous codon usage in *Saccharomyces cerevisiae*. Yeast 1991;7(7):657–78.
[12] Ghosh TC, Gupta SK, Majumdar S. Studies on codon usage in *Entamoeba histolytica*. Int J Parasitol 2000;30(6):715–22.

[13] Oresic M, Shalloway D. Specific correlations between relative synonymous codon usage and protein secondary structure. J Mol Biol 1998;281(1):31–48.
[14] Das S, Paul S, Dutta C. Synonymous codon usage in adenoviruses: influence of mutation, selection and protein hydropathy. Virus Res 2006;117(2):227–36.
[15] Gu W, Zhou T, Ma J, et al. The relationship between synonymous codon usage and protein structure in *Escherichia coli* and *Homo sapiens*. Biosystems 2004;73(2):89–97.
[16] Gupta SK, Majumdar S, Bhattacharya TK, et al. Studies on the relationships between the synonymous codon usage and protein secondary structural units. Biochem Biophys Res Commun 2000;269(3):692–6.
[17] Mukhopadhyay P, Basak S, Ghosh TC. Synonymous codon usage in different protein secondary structural classes of human genes: implication for increased non-randomness of GC3 rich genes towards protein stability. J Biosci 2007;32(5):947–63.
[18] Basak S, Banerjee T, Gupta SK, et al. Investigation on the causes of codon and amino acid usages variation between thermophilic *Aquifex aeolicus* and mesophilic *Bacillus subtilis*. J Biomol Struct Dyn 2004;22(2):205–14.
[19] Garat B, Musto H. Trends of amino acid usage in the proteins from the unicellular parasite *Giardia lamblia*. Biochem Biophys Res Commun 2000;279(3):996–1000.
[20] Banerjee T, Basak S, Gupta SK, et al. Evolutionary forces in shaping the codon and amino acid usages in *Blochmannia floridanus*. J Biomol Struct Dyn 2004;22(1):13–23.
[21] Gouy M, Gautier C. Codon usage in bacteria: correlation with gene expressivity. Nucleic Acids Res 1982;10(22):7055–74.
[22] 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.
[23] Hoekema A, Kastelein RA, Vasser M, et al. Codon replacement in the PGK1 gene of *Saccharomyces cerevisiae*: experimental approach to study the role of biased codon usage in gene expression. Mol Cell Biol 1987;7(8):2914–24.
[24] Fennoy SL, Bailey-Serres J. Synonymous codon usage in Zea mays L. nuclear genes is varied by levels of C and G-ending codons. Nucleic Acids Res 1993;21(23):5294–300.
[25] Chiapello H, Lisacek F, Caboche M, et al. Codon usage and gene function are related in sequences of *Arabidopsis thaliana*. Gene 1998;209(1–2):1–38.
[26] Sharp PM, Li WH. The codon adaptation index – a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res 1987;15(3):1281–95.
[27] Sharp PM, Li WH. The rate of synonymous substitution in enterobacterial genes is inversely related to codon usage bias. Mol Biol Evol 1987;4(3):222–30.
[28] Karlin S, Ladunga I, Blaisdell BE. Heterogeneity of genomes: measures and values. Proc Natl Acad Sci USA 1994;91(26):12837–41.
[29] Ma XR, Xiao SB, Fang LR, et al. Bias of base composition and codon usage in pseudorabies virus genes. Yi Chuan Xue Bao 2005;32(6):616–24.
[30] Hall JD, Gibbs JS, Coen DM, et al. Structural organization and unusual codon usage in the DNA polymerase gene from herpes simplex virus type 1. DNA 1986;5(4):281–8.
[31] Karlin S, Blaisdell BE, Schachtel GA. Contrasts in codon usage of latent versus productive genes of Epstein–Barr virus: data and hypotheses. J Virol 1990;64(9):4264–73.
[32] Sandhu TS, Leibovitz L. Duck plague. In: Diseases of poultry. 11th ed. Ames: Iowa State Press, 2003. p. 354–63.
[33] Cheng A, Wang MS, Wen M, et al. Construction of duck enteritis virus gene libraries and discovery, cloning and identification of viral nucleocapsid protein gene. High Technol Lett 2006;16(9):948–53 (in Chinese).
[34] Jeanmougin F, Thompson JD, Gouy M, et al. Multiple sequence alignment with Clustal X. Trends Biochem Sci 1998;23(10):403–5.
[35] Wright F. The effective number of codons used in a gene. Gene 1990;87(1):23–9.
[36] Comeron JM, Aguade M. An evaluation of measures of synonymous codon usage bias. J Mol Evol 1998;47(3):268–74.
[37] Sakai H, Washio T, Saito R, et al. Correlation between sequence conservation of the 5′ untranslated region and codon usage bias in *Mus musculus* genes. Gene 2001;276(1–2):101–5.
[38] Lu H, Zhao WM, Zheng Y, et al. Analysis of synonymous codon usage bias in Chlamydia. Acta Biochim Biophys Sin (Shanghai) 2005;37(1):1–10.
[39] McGeehan JE, Depledge NW, McGeoch DJ. Evolution of the dUTPase gene of mammalian and avian herpesviruses. Curr Protein Pept Sci 2001;2(4):325–33.
[40] Oliveros M, Garcia-Escudero R, Alejo A, et al. African swine fever virus dUTPase is a highly specific enzyme required for efficient replication in swine macrophages. J Virol 1999;73(11): 8934–43.
[41] Prangishvili D, Klenk HP, Jakobs G, et al. Biochemical and phylogenetic characterization of the dUTPase from the archaeal virus SIRV. J Biol Chem 1998;273(11):6024–9.
[42] Jiang P, Sun X, Lu Z. Analysis of synonymous codon usage in *Aeropyrum pernix K* 1 and other Crenarchaeota microorganisms. J Genet Genomics 2007;34(3):275–84.
[43] 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.
[44] Bjornberg O, Bergman AC, Rosengren AM, et al. dUTPase from herpes simplex virus type 1: purification from infected green monkey kidney (Vero) cells and from an overproducing *Escherichia coli* strain. Protein Expr Purif 1993;4(2):149–59.
[45] Jons A, Mettenleiter TC. Identification and characterization of pseudorabies virus dUTPase. J Virol 1996;70(2):1242–5.
[46] Liang X, Tang M, Manns B, et al. Identification and deletion mutagenesis of the bovine herpesvirus 1 dUTPase gene and a gene homologous to herpes simplex virus UL49.5. Virology 1993;195(1): 42–50.
[47] Bachtrog D. Protein evolution and codon usage bias on the neo-sex chromosomes of *Drosophila miranda*. Genetics 2003;165(3): 1221–32.
[48] Powell JR, Moriyama EN. Evolution of codon usage bias in *Drosophila*. Proc Natl Acad Sci USA 1997;94(15):7784–90.
[49] Baldo AM, McClure MA. Evolution and horizontal transfer of dUTPase-encoding genes in viruses and their hosts. J Virol 1999;73(9):7710–21.