aJava

Asian Journal of Animal and Veterinary Advances

Academic
Journals Inc.

www.academicjournals.com

Asian Journal of Animal and Veterinary Advances 9 (4): 229-242, 2014
ISSN 1683-9919 / DOI: 10.3923/ajava.2014.229.242

Bioinformatics Study Involving Characterization of Synonymous Codon Usage Bias in the Duck Enteritis Virus Glycoprotein D (gD) Gene

[1]S. Aravind, [1]N.M. Kamble, [1]S.S. Gaikwad, [1]S.A. Khulape, [1]Sohini Dey, [2]K. Dhama and [1]C. Madhan Mohan

[1]Laboratory of Recombinant DNA, Division of Veterinary Biotechnology,
[2]Avian Disease Section, Division of Pathology, Indian Veterinary Research Institute, Izatnagar, Bareilly, Uttar Pradesh, 243122, India

Corresponding Author: C. Madhan Mohan, Laboratory of Recombinant DNA, Division of Veterinary Biotechnology, Indian Veterinary Research Institute, Izatnagar, Bareilly, Uttar Pradesh, 243122, India Tel: +91 9897749358

ABSTRACT

The aim of the study was to identify the codon usage bias between the newly identified Duck Enteritis Virus (DEV) gD gene (GenBank accession No. KC915041) and the gD like gene of 23 other reference herpesviruses. The codon usage bias analysis of gD gene of DEV may improve our understanding of the evolution and pathogenesis of DEV and provide a basis for understanding the relevant mechanism for biased usage of synonymous codons; and for selecting appropriate heterologous expression systems to improve the expression of target genes. The results showed that codon of gD gene of DEV was having strong bias towards the synonymous codons with A and T at the third codon position. A high level of diversity in codon usage bias existed; and the effective number of codons used in a gene plot revealed that the genetic heterogeneity in gD gene of herpesviruses was constrained by the G+C content. The phylogenetic analysis suggested that DEV was evolutionarily closer to Alphaherpesvirinae, there was no significant deviation in codon usage in different virus strains. There were 17 codons showing distinct usage differences between DEV and *Escherichia coli*, 22 between DEV and *Homo sapiens* but only 15 codons between DEV and yeast. Therefore, the yeast expression system may be more suitable for the expression of DEV genes. The results are encouraging regarding bioinformatics data of an economically important poultry pathogen and could be highly useful for development of new generation vaccines, diagnostics and studying the evolution of the duck enteritis virus.

Key words: Duck enteritis virus, bioinformatics, glycoprotein D (gD), codon usage bias, codon adaptation index, synonymous codon usage

INTRODUCTION

Most amino acids are coded by more than one codon (synonymous codon usage), due to degeneracy of genetic code. These synonymous codons are non-random and specific to a particular species (Bishal *et al*., 2013). Studies revealed that synonymous codon usage vary between genomes, genes and even different parts of a gene (Zhao *et al*., 2008). Rather some codons that are used more frequently than others. Mutational pressure, translational selection and secondary structure of proteins are believed to be the important reasons for codon usage variation among genes in different organisms (Karlin and Mrazek, 1996; Lesnik *et al*., 2000). Amino acid usage was shown to be governed by hydrophobicity, aromaticity, cysteine content and mean molecular weight

(Banerjee *et al.*, 2004; Basak *et al.*, 2004). Codon variation is considered as an indicator of the forces shaping genome evolution. So, understanding the extent and causes of biases in codon usage is essential to the understanding the viral evolution, particularly the interplay between viruses and the immune response (Shackelton *et al.*, 2006). Codon usage variation is represented by two major paradigms. Either mutational bias or selection determines codon usage, or it is determined by mutational bias alone. Compositional constraints and translation selection are the accepted factors for codon usage bias; recent studies suggest that fine tuning translation kinetics selection and escape from antiviral cellular responses also play an inevitable role in codon usage bias (Sugiyama *et al.*, 2005; Aragones *et al.*, 2008; Aragones *et al.*, 2010). 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 (Gouy and Gautier, 1982; Grosjean and Fiers, 1982). Even for functionally homologous gene there are differences in codon usage across species (Karlin and Mrazek, 1996).

Recent studies of virus codon usage has shown that mutation bias may be a more important factor than translational selection in determining codon usage bias of some viruses, such as baculoviruses, plant viruses and vertebrate DNA viruses (Shackelton *et al.*, 2006; Jiang *et al.*, 2008; Jia *et al.*, 2009). Codon usage bias is an important feature, with signs of mutational bias in genes and genomes. The study of codon usage can provide some evidence about the molecular evolution of viruses and their individual genes. Moreover, information about codon usage bias can be relevant for understanding the regulation of viral gene expression. It can be applied in vaccine design where the efficient expression of viral proteins may be required to generate immunity. It can also be used for better understanding the interaction between viruses and their hosts from the perspective of their respective codon usage patterns (Karlin *et al.*, 1990).

The herpesviruses are a large group of double stranded DNA viruses that cause diseases in a wide range of organisms, including humans (Whitley, 1996). Herpesviruses infect a wide range of vertebrates and even some invertebrates such as oysters. Based on details of characteristic morphology, tissue tropism, pathogenicity and more recently on molecular phylogenetic analysis, There are five clades within the herpesvirus family-Alloherpesviridae, Malacoherpesviridae and Herpesviridae. The family Herpesviridae is divided into three subfamilies, Alpha-, Beta- and Gamma-Herpesvirinae based on the phylogenetic analysis (McGeoch *et al.*, 2000; Fu *et al.*, 2008). Duck Viral Enteritis (DVE), is an acute, contagious infection of ducks, geese, swans of all age and species. The disease has been responsible for significant economic losses in domestic and wild water fowl as a result of mortality and decreased egg production (Shawky and Sandhu, 1997). The disease is caused by DEV and is characterized by vascular damage, tissue haemorrhage, eruptions on the digestive mucosa lesions of lymphoid organs and degenerative changes in parenchymatous organs (Shawky *et al.*, 2000). The disease is difficult to monitor and control as the virus establishes an asymptomatic carrier state in water fowl that is detectable only during periods of intermittent virus shedding (Burgess *et al.*, 1979). Upon primary infection, DEV establishes latency in trigeminal ganglia (TG) from which the virus can reactivate leading to further disease outbreaks (Shawky and Schat, 2002). DEV is taxonomically classified as Anatid herpesvirus 1 in the Alphaherpesvirinae subfamily but is not assigned to any genus (Wu *et al.*, 2012). Most of the previous research has focused on the epidemiology and prevention of this disease. However, the molecular biological information about the DEV genome is limited. The data in respect to codon usage bias in the newly identified DEV gD gene is not yet available. Given this background, it has become important to analyze the codon preference used in the DEV UL35 gene, as the gD gene plays an important role in the initial entry of the virus in to the host cell. The present study aimed to analyze the bioinformatic analysis of the synonymous codon usage data of the DEV gD gene and

compare it with the gD-like genes of 23 reference herpesviruses; examination of how other factors may affect codon usage variation in the DEV gD gene and its reference species. A comparison of the codon usage preference of the DEV gD gene with those of *Escherichia coli*, yeast and humans.

MATERIALS AND METHODS

Measure of synonymous codon usage (RSCU): The Relative Synonymous Codon Usage (RSCU) values for the glycoprotein D (gD) gene of DEV were calculated to investigate the characteristics of synonymous codon usage without the confounding influence of amino acid composition of different gene sample (Sharp and Li, 1986). The codons with RSCU values >1.0 have positive codon usage bias (abundant codons), while those with RSCU values <1.0 have negative codon usage bias (less-abundant codons); and when the RSCU values is 1.0, it means that these codons are chosen equally or randomly, indicates lack of bias (Tsai *et al*., 2007). The RSCU is the observed frequency of a codon divided by the frequency expected, if all synonymous codons for that amino acid were used equally. The synonymous codons with RSCU more than 1.6 were thought to be over-represented, while the synonymous codons with RSCU less than 0.6 were regarded as under-represented (Wong *et al*., 2010). The RSCU values are particularly useful in comparing codon usage between genes that differ in size and amino acid composition.

Compositional properties measures: Each general nucleotide composition (U, A, C and G%) and each nucleotide composition in the third site of codon (U3, A3, C3 and G3%) in DEV gD coding sequence was calculated. Also the GC3s (the frequencies of nucleotide G+C at the third codon position) and the GC content of gD gene was calculated to examine the compositional properties. Comparison of gD codon usage in DEV with that in *E. coli*, yeast and humans.

To examine whether different species follow with the same codon usage rule, codon usage bias in the DEV gD gene was determined with the SPSS 13.0 software and compared the gD codon usage bias among DEV, *E. coli*, yeast and *H. sapiens*. The database of the codon usage in *E. coli*, yeast and *H. sapiens* is available at http://www.kazusa.or.jp/codon.

Analysis of codon usage: The 'effective number of codons' (ENC) was used to quantify the codon usage bias of an Open Reading Frame (ORF) (Wright, 1990) like gD gene of DEV. The ENC value ranges from 20-61. The ENC value is inversely proportional to codon bias in a gene. In this study, this measure was used to evaluate the degree of codon usage bias of DEV gD gene.

Phylogenetic analysis of DEV gD gene: The gD ORF sequences of different strains were aligned using Clustal W. The phylogenetic tree was constructed by maximum likelihood with 2000 replicates for boot strap 2000 were kept as out group. Tree was visualized with Figure.

Statistical analysis: In order to detect whether different species follow the same codon usage rule and to select the best suitable expression system, comparison of DEV gD codon usage bias among *E. coli*, yeast and *H. sapiens* was made. Correlation analysis was carried out using Spearman's rank correlation analysis method. Results were statistically analyzed using SPSS software (version 11; SPSS, Inc., Chicago, IL).

RESULTS

Characterization of DEV glycoprotein D (gD) gene: In order to study the extent of codon bias in DEV gD gene, the average values for all triplets and the codon preferences of gD were calculated. The results of these studies are shown in Table 1. From the RSCU values of the codons

Table 1: Synonymous codon usage of DEV gD gene analyzed with CUSPS

Codon	AA[a]	Fraction[b]	Frequency[c] (1/1000)	Number[d]	RSCU[e]
GCA*	A	0.300	20.785	9	1.200
GCC	A	0.233	16.166	7	0.933
GCG	A	0.267	18.476	8	1.067
GCT	A	0.200	13.857	6	0.800
TGC	C	0.000	0.000	0	0.000
TGT*	C	1.000	20.785	9	2.000
GAC*	D	0.500	30.023	13	1.000
GAT*	D	0.500	30.023	13	1.000
GAA*	E	0.765	30.023	13	1.529
GAG	E	0.235	9.238	4	0.471
TTC	F	0.462	13.857	6	0.923
TTT*	F	0.538	16.166	7	1.077
GGA*	G	0.320	18.476	8	1.280
GGC	G	0.240	13.857	6	0.960
GGG	G	0.280	16.166	7	1.120
GGT	G	0.160	9.238	4	0.640
CAC	H	0.125	4.619	2	0.250
CAT*	H	0.875	32.333	14	1.750
ATA*	I	0.400	27.714	12	1.200
ATC	I	0.200	13.857	6	0.600
ATT*	I	0.400	27.714	12	1.200
AAA*	K	0.522	27.714	12	1.043
AAG	K	0.478	25.404	11	0.957
CTA	L	0.083	4.619	2	0.500
CTC	L	0.208	11.547	5	1.250
CTG	L	0.042	2.309	1	0.250
CTT	L	0.167	9.238	4	1.000
TTA	L	0.208	11.547	5	1.250
TTG*	L	0.292	16.166	7	1.750
ATG	M	1.000	25.404	11	
AAC	N	0.455	23.095	10	0.909
AAT*	N	0.545	27.714	12	1.091
CCA	P	0.250	13.857	6	1.000
CCC	P	0.125	6.928	3	0.500
CCG*	P	0.375	20.785	9	1.500
CCT	P	0.250	13.857	6	1.000
CAA*	Q	0.727	18.476	8	1.455
CAG	Q	0.273	6.928	3	0.545
AGA*	R	0.250	11.547	5	1.500
AGG	R	0.100	4.619	2	0.600
CGA	R	0.150	6.928	3	0.900
CGC	R	0.200	9.238	4	1.200
CGG	R	0.050	2.309	1	0.300
CGT*	R	0.250	11.547	5	1.500
AGC	S	0.206	16.166	7	1.235
AGT*	S	0.235	18.476	8	1.412
TCA	S	0.176	13.857	6	1.059

Table 1: Continue

Codon	AA[a]	Fraction[b]	Frequency[c] (1/1000)	Number[d]	RSCU[e]
TCC	S	0.059	4.619	2	0.353
TCG	S	0.118	9.238	4	0.706
TCT	S	0.206	16.166	7	1.235
ACA*	T	0.475	43.880	19	1.900
ACC	T	0.100	9.238	4	0.400
ACG	T	0.175	16.166	7	0.700
ACT	T	0.250	23.095	10	1.000
GTA	V	0.280	16.166	7	1.120
GTC	V	0.200	11.547	5	0.800
GTG	V	0.120	6.928	3	0.480
GTT*	V	0.400	23.095	10	1.600
TGG	W	1.000	30.023	13	
TAC	Y	0.368	16.166	7	0.737
TAT*	Y	0.632	27.714	12	1.263
TAA	*	1.000	2.309	1	
TAG	*	0.000	0.000	0	
TGA	*	0.000	0.000	0	

[a]AA: Amino acid, [b]"Fraction" column shows the proportion of each synonymous codons encoding the same amino acid, [c]"frequency/1000" value represents the No. of codons present per 1000 bases in the input sequence(s), [d]No. represents the No. of occurrence of each sense codons, [e]"RSCU" is nsed to investigate the pattern of relative synonymous codon usage

in the table, it was found that most of the A and T ended codons used for coding amino acid are much higher than the C and G ended codons; except TTG and CCG codons which are G ended. A high level of diversity in codon usage bias existed for coding the Tyr, Thr, Glu, Leu, His, Gln amino acids. Also the GC content varies form 0.41-0.76 with a mean of 0.59 and SD 0.022. The GC content of DEV gD gene is 0.44 (Table 2), indicating that this ORF is AT rich. Therefore nucleotide composition is a major contributing factor shaping the codon usage pattern.

Codon usage analysis of the gD gene of DEV 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 (GC3s%) of 24 herpesviruses species are shown in Table 2. The ENC values of different gD genes of herpes viruses vary from 28.2-60.1, with a mean value of 43.5 and a standard deviation of 12.2. In general speaking, if the ENC value of a gene is 35 or less, that gene is thought to possess strong codon bias (Jiang *et al*., 2008). Since an average of 43.5 in case of gD was observed, codon usage bias in herpes virus gD gene is moderate. There is a high variation in the usage of codon pattern among different herpes viral gD genes (SD = 12.2). Similarly, the GC3 content of each gD gene also confirm the homogeneity of synonymous codon usage among the different herpes viruses, which range from 33.1-97.7 with a mean of 67.9 and SD of 28.12.

It has been reported that the plot of ENC and GC3s content is another effective way to explore codon usage variation among different genes (Wright, 1990). In Fig. 1, the solid line represents the curve if codon usage is only determined by GC3s content. If GC3s 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 (Jiang *et al*., 2008). For gD genes, only a few ENC values were

Table 2: Codon preferences of DEV gD gene analyzed with CUSPS program

Name	L[a]	CAI[b]	Gc[c] (%)	GC3[d] (%)	ENC[e]
DEV (Duck enteritis virus kerala isolate)	1299	0.725	44.0	41.1	57.6
CeHV1 (Cervid herpes virus 1)	1185	0.507	74.6	96.7	29.4
BoHV1 (Bovine herpes virus 1)	1254	0.533	69.7	84.4	38.0
BoHV5 (Bovine herpes virus 5)	1254	0.493	73.4	94.7	30.2
CpHV1 (Caprine herpes virus 1)	1227	0.528	71.4	88.3	37.0
EHV1 (Equine herpes virus 1)	1401	0.694	48.2	49.0	60.1
EHV4 (Equine herpes virus 5)	1209	0.764	42.3	36.0	57.1
FeHV1 (Feline herpes virus 1)	1125	0.773	41.0	33.1	53.9
GaHV1 (Gallid herpes virus 1)	1305	0.639	53.1	58.4	59.0
GaHV2 (Gallid herpes virus 2)	1212	0.758	39.5	34.4	51.9
HHV2 (Human herpes virus 2)	1182	0.577	64.3	82.0	38.9
HHV1 (Human herpes virus 1)	1185	0.577	64.8	83.5	39.3
HHV5 (Human herpes virus 5)	552	0.676	52.4	56.0	59.9
MDV (Mareks disease virus)	1212	0.761	39.4	33.9	52.3
MeHV1 (Meleagrid herpes virus 1)	1116	0.734	42.4	41.7	58.3
PaHV2 (Panine herpes virus 2)	579	0.591	64.6	80.3	41.6
CaHV2 (Camine herpes virus 2)	1038	0.874	27.9	13.3	33.8
SuHV1 (Suid herpes virus 1)	1203	0.613	74.5	96.8	31.0
PRV (Pseudo rabies virus)	1254	0.611	74.5	95.7	31.3
CeHV16 (Cervid herpes virus 16)	1185	0.618	75.6	99.0	28.2
CeHV2 (Cervid herpes virus 2)	1188	0.611	75.7	97.7	28.8
PsHV (Psittacid herpes virus)	1152	0.552	59.3	69.0	52.9

[a]L: Length of identified ORF, [b]CAI: Codon adaptation index, [c]GC: Frequency of the nncleotide G+C of codons, [d]GC3s: Frequency of the nucleotide G+C at the synonymous third position of codons, [e]ENC: Effective No. of codous

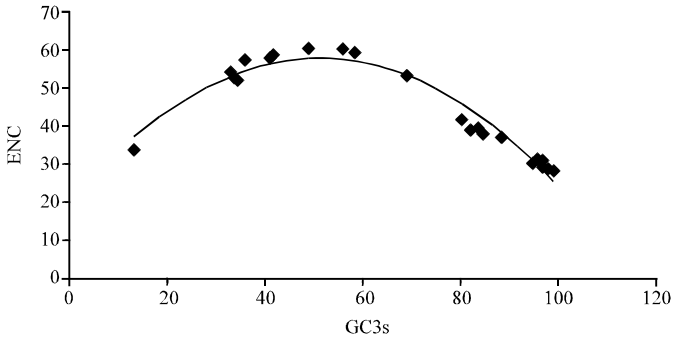

Fig. 1: Effective No. of Codons (ENC) and Guanine (G)+Cytosine (C) frequency at the synonymous third position of codons (GC3s) of the gD gene in the DEV and 23 reference herpes viruses

lined on the curve and others were plotted against both GC3s content and the expected ENC value. Result suggested that there are other factors that contributed to the codon usage pattern in the gD genes besides the genomic composition.

Phylogenetic analysis: Using mega 5 a phylogenetic tree was established from the deduced amino acids encoded by the 1.3 kbp ORF of the gD gene of DEV and the 44 reference herpesviruses (Fig. 2). The distance method was used to draw the phylogenetic tree with a bootstrap frequency

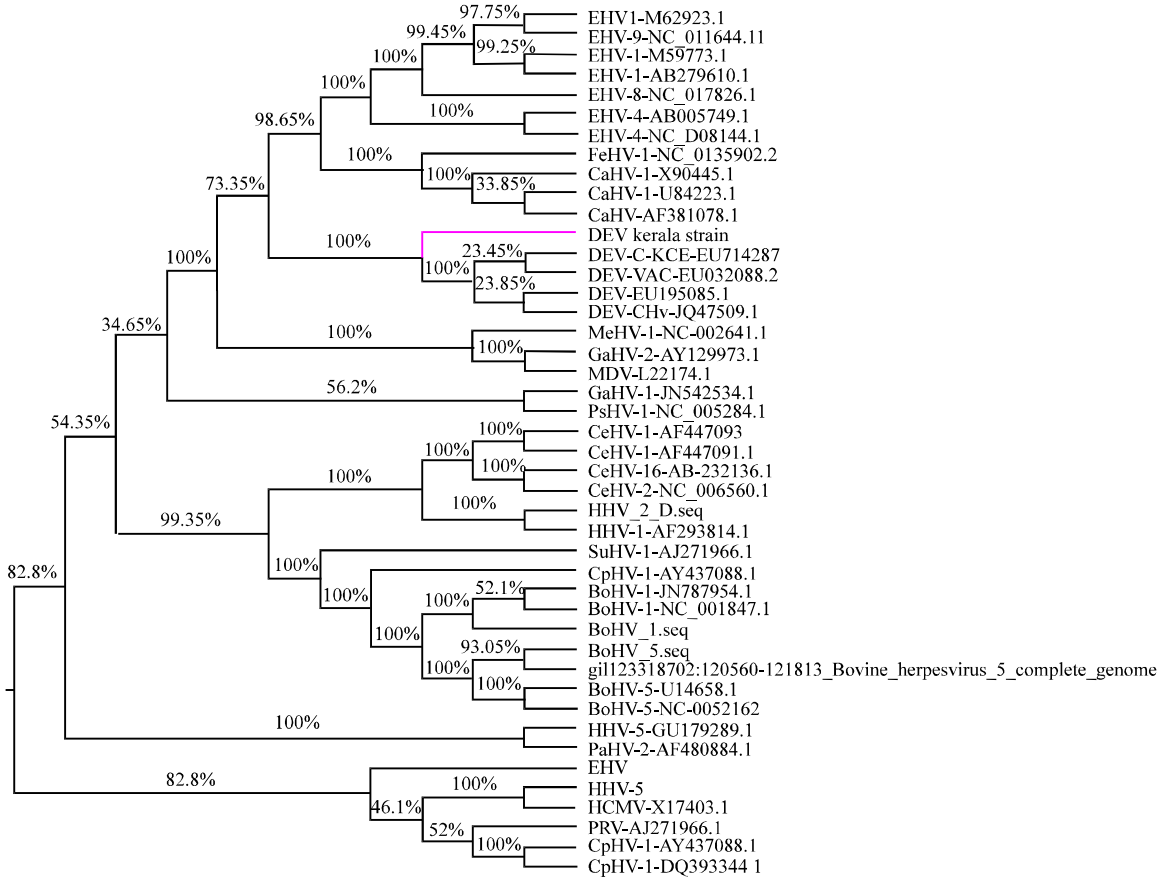

Fig. 2: Evolutionary relationships of the putative DEV gD gene with its 23 reference herpes viruses from different species (Table 1). Phylogenetic tree of this gene was generated by using the MEGA 5.1 β 4 program with clustal W and TREEVIEW software

of 2000. In the above phylogenetic tree, there are 3 main branches and DEV viruses are clustered in a separate clade. In that particular clade Indian DEV sequence forms a separate lineage from other DEV viruses. From Fig. 2 it can be seen that DEV Indian isolate and other DEV strains are different from other herpesviruses as they first cluster together and form a separate branch; then cluster with the avian herpesviruses of GaHV-2 and GaHV-1 and afterwards cluster with other herpesviruses such as FeHV, EHV and CaHV. This result indicates the internal relations of the codon usage pattern between DEV and other herpesviruses, suggesting that the codon usage pattern of DEV has differences with other herpesviruses: the more distant the relationship, the bigger the variation of the codon usage bias and vice versa. Thus, it can be concluded that the codon usage pattern of DEV is fairly close to that of avian herpesviruses and is largely different from other types of herpesviruses and the gD gene's function contributes to the codon usage pattern.

Comparison of codon preference of DEV gD gene with those of *E. coli*, yeast and human: Normally, the codon usage bias in genes remains at a certain level among species. In this study, comparisons in the ratio usage frequency (1/1000) of DEV to *E. coli*, yeast and *H. sapiens* for the

selection of suitable host system for optimal protein production was used. Thus, the codon preferences of DEV, as a representative of herpesviruses, were compared with those of *E. coli*, yeast and *H. sapiens* to see which will be the suitable host for the optimal expression of DEV genes (Table 3). In Fig. 3, there are 17 codons showing a DEV to *E. coli* ratio higher than 2 or lower than

Table 3: Comparison of codon preferences between the DEV gD gene and *E. coli*, yeast and humans

Codon	AA	DEV (1/1000)	Human (1/1000)	Yeast (1/1000)	*E. coli* (1/1000)	DEV/*E .coli*	DEV/Yeast	DEV/Human
GCA	A	20.785	16.1	16.2	23.00	0.90	1.28	1.29
GCC	A	16.166	28.1	12.6	21.60	0.75	1.28	0.56
GCG	A	18.476	7.6	6.2	21.10	0.86	2.98	2.43
GCT	A	13.857	18.4	21.2	18.90	0.73	0.65	0.75
TGC	C	0.000	12.2	4.8	5.50	0.00	0.00	0.00
TGT	C	20.785	10.6	8.1	5.90	3.52	2.57	1.96
GAC	D	30.023	25.1	20.2	17.90	1.68	1.49	1.20
GAT	D	30.023	22.3	37.6	33.70	0.89	0.80	1.35
GAA	E	30.023	30.4	45.6	35.10	0.86	0.67	0.99
GAG	E	9.238	40.7	19.2	19.40	0.48	0.48	0.23
TTC	F	13.857	19.5	18.4	13.90	0.99	0.75	0.71
TTT	F	16.166	17.1	26.1	24.40	0.66	0.62	0.95
GGA	G	18.476	16.5	10.9	13.60	1.36	1.70	1.12
GGC	G	13.857	22.3	9.8	20.60	0.67	1.41	0.62
GGG	G	16.166	16.5	6.0	12.30	1.31	2.69	0.98
GGT	G	9.238	10.6	23.9	23.70	0.39	0.39	0.87
CAC	H	4.619	15.2	7.8	7.30	0.63	0.59	0.30
CAT	H	32.333	11.1	13.6	12.40	2.60	2.38	2.91
ATA	I	27.714	7.5	17.8	13.30	2.08	1.56	3.70
ATC	I	13.857	19.9	17.2	19.40	0.71	0.80	0.69
ATT	I	27.714	15.8	30.1	29.60	0.94	0.92	1.75
AAA	K	27.714	25.1	41.9	37.20	0.75	0.66	1.10
AAG	K	25.404	31.9	30.8	15.30	1.66	0.82	0.80
CTA	L	4.619	7.2	13.4	5.60	0.82	0.34	0.64
CTC	L	11.547	19.2	5.4	9.50	1.22	2.14	0.60
CTG	L	2.309	39.4	10.5	37.40	0.06	0.22	0.06
CTT	L	9.238	13.3	12.3	14.50	0.64	0.75	0.69
TTA	L	11.547	7.8	26.2	17.40	0.66	0.44	1.48
TTG	L	16.166	12.9	27.2	12.90	1.25	0.59	1.25
ATG	M	25.404	21.3	20.9	23.70	1.07	1.22	1.19
AAC	N	23.095	18.6	24.8	20.30	1.14	0.93	1.24
AAT	N	27.714	17.1	35.7	29.30	0.95	0.78	1.62
CCA	P	13.857	17.3	18.3	9.10	1.52	0.76	0.80
CCC	P	6.928	20.3	6.8	6.20	1.11	1.01	0.34
CCG	P	20.785	7.3	5.3	14.50	1.43	3.92	2.84
CCT	P	13.857	17.9	13.5	9.50	1.46	1.02	0.77
CAA	Q	18.476	12.7	27.3	14.40	1.28	0.68	1.45
CAG	Q	6.928	34.9	12.1	26.70	0.26	0.57	0.20
AGA	R	11.547	12.0	21.3	7.10	1.63	0.54	0.96
AGG	R	4.619	11.9	9.2	4.00	1.15	0.50	0.39
CGA	R	6.928	6.2	3.0	4.80	1.44	2.30	1.11
CGC	R	9.238	10.5	2.6	14.00	0.66	3.55	0.88

Table 3: Continue

Codon	AA	DEV (1/1000)	Human (1/1000)	Yeast (1/1000)	*E. coli* (1/1000)	DEV/*E .coli*	DEV/Yeast	DEV/Human
CGG	R	2.309	11.7	1.7	7.90	0.29	1.36	0.20
CGT	R	11.547	4.5	6.4	15.90	0.73	1.80	2.57
AGC	S	16.166	19.9	9.8	14.30	1.13	1.65	0.81
AGT	S	18.476	12.6	14.2	13.20	1.40	1.30	1.47
TCA	S	13.857	12.5	18.7	13.10	1.05	0.74	1.11
TCC	S	4.619	17.7	14.2	9.70	0.48	0.33	0.26
TCG	S	9.238	4.5	8.6	8.20	1.13	1.07	2.05
TCT	S	16.166	15.4	23.5	13.10	1.23	0.69	1.05
ACA	T	43.880	15.1	17.8	15.10	2.91	2.47	2.91
ACC	T	9.238	18.4	12.7	18.90	0.49	0.73	0.50
ACG	T	16.166	6.0	8.0	13.60	1.19	2.02	2.69
ACT	T	23.095	13.3	20.3	13.10	1.76	1.14	1.74
GTA	V	16.166	7.1	11.8	13.10	1.23	1.37	2.28
GTC	V	11.547	14.0	11.8	13.10	0.88	0.98	0.82
GTG	V	6.928	27.6	10.8	19.90	0.35	0.64	0.25
GTT	V	23.095	10.9	22.1	21.60	1.07	1.05	2.12
TGG	W	30.023	12.1	10.4	13.40	2.24	2.89	2.48
TAC	Y	16.166	14.6	14.8	11.70	1.38	1.09	1.10
TAT	Y	27.714	12.0	18.8	21.60	1.28	1.47	2.31
TAA	*		0.5	1.1	2.00	0.00	0.00	0.00
TAG	*		0.4	0.5	0.30	0.00	0.00	0.00
TGA	*		0.9	0.7	0.42	0.00	0.00	0.00

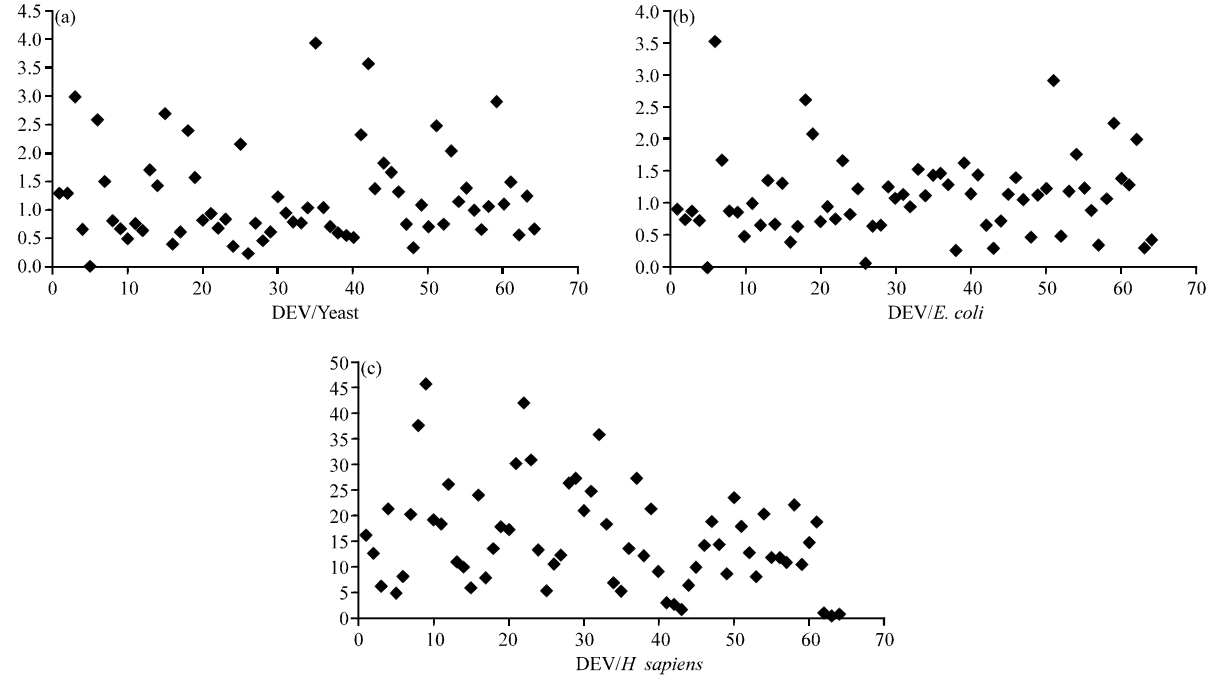

Fig. 3(a-c): Comparisons in the ratio of codon usage frequency (1/1000) of DEV to (a) *E. coli*, (b) Yeast and (c) *H. sapiens*. Ratio higher than 2 or lower than 0.5 indicates that the codon usage preference differs and vice-versa

0.50, with 22 codons between DEV-to *H. sapiens*; but only 15 codons showing a DEV-to-yeast ratio higher than 2 or lower than 0.50, suggesting that codon usage of DEV genes more closely resembles that of yeast genes than that of *E. coli* and *H. sapiens* genes. Thus, to express DEV genes efficiently in *E. coli* and *H. sapiens* codon optimization of the DEV genes may be required. At the same time, it could be speculated that the DEV genes may be more efficiently expressed in the yeast system.

DISCUSSION

As a common evolutionary phenomenon it is well known that synonymous codon usage bias exists in each organism from prokaryotes to eukaryotes. Codon usage bias is found to be related to different biological factors, such as tRNA abundance, strand specific mutational bias, gene expression level, gene length, amino acid composition, protein structure, mRNA structure and GC composition (Sueoka, 1988; Blake *et al*., 2003; Jiang *et al*., 2008). 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. In *E. coli* and yeast, synonymous codon usage patterns are related to the abundance of isoaccepting tRNAs (Ikemura, 1985; Wan *et al*., 2004), where 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 (Hoekema *et al*., 1987; Lesnik *et al*., 2000). In mammals and birds, the diverse patterns of codon usage may arise from the compositional constraints of the genomes (Sharp *et al*., 1986; Karlin and Mrazek, 1996; Jiang *et al*., 2008).

In order to examine synonymous codon usage without the confounding influence of amino acid composition of different gene samples, RSCU with direct assessment codon usage bias of different codons in each gene sample were calculated. Here, it was discovered that some strong bias toward A-ended or T-ended codon at the third codon position were observed and a high level of diversity in codon usage bias is existed for coding the Tyr, Thr, Glu, Leu, His, Gln amino acids. Although the overall RSCU values in a gene could unveil the codon usage pattern of a whole genome, two important codon usage indices, ENC value and GC3s content, have been widely used to explore the codon usage variation among different genes (Romero *et al*., 2003; Gu *et al*., 2004).

In the present study, a comprehensive analysis of codon usage including ENc, CAI value, GC content and the RSCU values of DEV Kerala isolate gD gene was performed by using CodonW 1.4 and EMBOSS CUSP programs and these values were subsequently compared with those of the 23 reference herpesvirus species. The data of synonymous codon usage bias showed certain disparity of each herpesvirus from the different organisms with the result revealing that: (1) The DEV gD gene and its 23 reference herpesviruses adopt relatively similar codon usage patterns, although the DEV gD gene shows a few differences of codon usage bias with its reference herpesvirus species; (2) The DEV gD gene prefers to use the codons with A and T at the 3rd codon position. Furthermore, the biased trend towards A and T coincides in with high A + T content in the DEV gD gene. Since the DEV gD gene is an AT-rich gene, it is reasonable that codons ending in A and/or T are predominant in the gene. In order to show the codon usage variation, ENc-plot was used to analyze the factors affecting codon usage variation among genes. Here, genetic heterogeneity in the DEV and its reference herpesviruses was observed to be constrained by GC content, while the gene length had only minor impact on the codon choice.

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 (Lu *et al*., 2005). 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 gD gene and relational herpesvirus vary from 28.8-0.1, with an average value of 43.47 and the SD value of 12.19. The ENC values of DEV gD gene is 57.6, indicating that the codon usage bias of this gene is weak.

G+C composition and GC3s has been widely reported being 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 gD gene range from 13.3-97.8%, with an average value of 67.9% and SD of 27.475.

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 is calculated from the frequency of the use of all codons in that gene sequence (Novembre, 2002). It is useful for predicting the expression level of a gene, adaptation of viral genes to their hosts and for comparing codon usage in different organisms. The CAI value is much closer to 1, the codon usage is much stronger and the gene expression level is much higher. 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 (Xia, 2007). Thus, the index may also provide an approximate indication of the possible success of heterologous gene expression. Comparative analysis of gD genes in DEV and the reference herpesviruses indicated that synonymous codon usage in these genes is phylogenetically conserved. Data in Table 2 shows that the gD genes in DEV, MeHV-1, GaHV-1 and GaHV-2, whose natural host is avian, have a stronger correlation than the gD genes of herpesviruses with other hosts. In the Table 2, the CAI value of DEV gD gene is 0.725. The CAI values of DEV gD gene and relational herpesvirus vary from 0.493-0.864, with an average value of 0.645 and inferred that gD gene is lowly expressed gene in DEV genome. Moreover, the SD of CAI value is 0.1003, indicating that there is a good variation in codon usage pattern among different herpesvirus gD genes.

From the phylogenetic tree analysis, it can be concluded that DEV is evolutionarily closer to GaHV-2, MeHV-1 than GaHv-1. Simultaneously, the codon usage pattern of DEV (0.725) is closer to MeHV-1 (0.734) and GaHV-1 (0.758) than GaHV-1 (0.639). Similar studies have showed that the codon usage pattern of DEV dUTPase, UL24, gC, UL-27, UL28, UL30, UL35 and UL-21 genes were also similar to herpes viruses Mardi virus; but UL-25, UL26, UL26.5 and UL 29 gene exhibited a close relationship with varicella virus of the Simplex virus. So, it can be inferred from these studies that the codon usage pattern of DEV is similar to that of avian alpha-herpesvirus; however, the conclusive evidence may require further analysis of other DEV genes, such as viral genes for maintenance or replication.

The weak codon bias for the majority of the members of herpesvirus family might be due to the herpesviral infection in host (such as due to Herpes simplex virus-HSV); which may lead to the complete or partial suppression of the cellular protein synthesis and degradation of mRNAs of the host, further reducing competition between the virus and its host (Carbone *et al*., 2003; Matis and

Kudelova, 2001; Smiley, 2004). In the infected cell, herpesviruses use general evasive approaches, such as blocking the induction of programmed cell death and shutting down host protein synthesis (Strand *et al*., 2004). In contrast to most of the herpesviruses, a significantly strong codon bias has been observed in five herpesviruses SuHV-1, CeHV-1,CeHV-2, CeHV-16 and BoHV-5. In these viruses, more than two thirds (>65%) of their total genes had a high codon bias (ENc<35). Noticeably, all the five viruses are also representative of the corresponding rich GC and GC3 content among herpesviruses.

Only alphaherpesviruses (except VZV) encode members of the gD family (Roizman and Pellet, 2001). Glycoprotein D (gD) is essential for alphaherpes viral penetration into cells (Spear and Longnecker, 2003). Herpesvirus glycoproteins C (gC), gB and gD act on the viral entry process in permissive cells and play important roles in pathogenicity by mediating cell-to-cell spread of virus (Spear, 1993). The initial viral attachment on cellular receptors is mediated by gC (Dasika and Letchworth, 1999) and gB, while gD acts on the viral envelope and the plasma membrane fusion, resulting in conformational changes that allow gH and gL interaction (Babiuk *et al*., 1996; Csellner *et al*., 2000), which is essential for viral penetration into host cells and virus cell-to-cell spread.

CONCLUSON

The destination of the analysis on the codon usage bias is to guide the expression of the genes. Among the codon usage bias patterns in *E. coli*, yeast and humans, the codon usage bias pattern in the DEV gD gene is similar to that of yeast. Hence the assumption was made in the present study that the DEV gD gene may express more efficiently in the yeast system. This may serve as a guide for manipulating the expression of the targeted gene. The optimization of the DEV gD gene with host-preferred codons is likely to improve the expression level of the DEV gD gene in a given host. Therefore, the yeast expression system may be better applied to the production of the DEV gD protein. In summary, the present study provides a basic understanding of the evolution and pathogenesis of the DEV gD gene and offer some insights into the mechanisms for codon usage bias, which may assist to the area of herpesvirus research and possibly studies with other viruses. To conclude, the results are promising regarding bioinformatics data for an important avian pathogen and may be useful for development of recombinant protein based new generation vaccines, novel diagnostics and revealing the evolutionary mechanisms of the duck enteritis virus.

REFERENCES

Aragones, L., A. Bosch and R.M. Pinto, 2008. Hepatitis A virus mutant spectra under the selective pressure of monoclonal antibodies: Codon usage constraints limit capsid variability. J. Virol., 82: 1688-1700.

Aragones, L., S. Guix, E. Ribes, A. Bosch and R.M. Pinto, 2010. Fine-tuning translation kinetics selection as the driving force of codon usage bias in the hepatitis A virus capsid. PLoS Pathogens, Vol. 6, No. 3. 10.1371/journal.ppat.1000797

Babiuk, L.A., D.H.S. van Drunen Little-Van and S.K. Tikoo, 1996. Immunology of bovine herpesvirus 1 infection. Vet. Microbiol., 53: 31-42.

Banerjee, T., S. Basak, S.K. Gupta and T.C. Ghosh, 2004. Evolutionary forces in shaping the codon and amino acid usages in *Blochmannia floridanus*. J. Biomol. Struct. Dyn., 22: 13-24.

Basak, S., T. Banerjee, S.K. Gupta and T.C. Ghosh, 2004. Investigation on the causes of codon and amino acid usages variation between thermophilic *Aquifex aeolicus* and mesophilic *Bacillus subtilis*. J. Biomol. Struct. Dyn., 22: 205-214.

Bishal, A.K., R. Mukherjee and C. Chakraborty, 2013. Synonymous codon usage pattern analysis of Hepatitis D virus. Virus Res., 173: 350-353.

Blake, W.J., M. Kaern, C.R. Cantor and J.J. Collins, 2003. Noise in eukaryotic gene expression. Nature, 422: 633-637.

Burgess, E.C., J. Ossa and T.M. Yuill, 1979. Duck plague: A carrier state in waterfowl. Avian Dis., 23: 940-949.

Carbone, A., A. Zinovyev and F. Kepes, 2003. Codon adaptation index as a measure of dominating codon bias. Bioinformatics, 19: 2005-2015.

Csellner, H., C. Walker, J.E. Wellington, L.E. McLure, D.N. Love and J.M. Whalley, 2000. EHV-1 glycoprotein D (EHV-1 gD) is required for virus entry and cell-cell fusion and an EHV-1 gD deletion mutant induces a protective immune response in mice. Arch. Virol., 145: 2371-2385.

Dasika, G.K. and G.J. Letchworth III, 1999. Cellular expression of bovine herpesvirus 1 gD inhibits cell-to-cell spread of two closely related viruses without blocking their primary infection. Virology, 245: 24-36.

Fu, M.H., R.Q. Deng, J.W. Wang and X.Z. Wang, 2008. Whole-genome phylogenetic analysis of herpesviruses. Acta Virol., 52: 31-40.

Gouy, M. and C. Gautier, 1982. Codon usage in bacteria: Correlation with gene expressivity. Nucl. Acids Res., 10: 7055-7074.

Grosjean, H. and W. Fliers, 1982. Preferential codon usage in prokaryotic genes: The optimal codon-anticodon interaction energy and the selective codon usage in efficiently expressed genes. Gene, 18: 209-299.

Gu, W., T. Zhou, J. Ma, X. Sun and Z. Lu, 2004. Analysis of synonymous codon usage in SARS *Coronavirus* and other viruses in the *Nidovirales*. Virus Res., 101: 155-161.

Hoekema, A., R.A. Kastelein, M. Vasser and H.A. de Boer, 1987. 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., 7: 2914-2924.

Ikemura, T., 1985. Codon usage and transfer-RNA content in unicellular and multicellular organisms. Mol. Biol. Evol., 2: 13-34.

Jia, R.Y., A.C. Cheng, M.S. Wang, H.Y. Xin and Y.F. Guo *et al.*, 2009. Analysis of synonymous codon usage in the UL24 gene of duck enteritis virus. Virus Genes, 38: 96-103.

Jiang, Y., F. Deng, H.L. Wang and Z.H. Hu, 2008. An extensive analysis on the global codon usage pattern of baculoviruses. Arch. Virol., 153: 2273-2282.

Karlin, S. and J. Mrazek, 1996. What drives codon choices in human genes? J. Mol. Biol., 262: 459-472.

Karlin, S., B.E. Blaisdell and G.A. Schachtel, 1990. Contrasts in codon usage of latent versus productive genes of Epstein-Barr virus: Data and hypotheses. J. Virol., 64: 4264-4273.

Lesnik, T., J. Solomovici, A. Deana, R. Ehrlich and C. Reiss, 2000. Ribosome traffic in *E. coli* and regulation of gene expression. J. Theor. Biol., 202: 175-185.

Lu, H., W.M. Zhao, Y. Zheng, H. Wang, M. Qi and X.P. Yu, 2005. Analysis of synonymous codon usage bias in *Chlamydia*. Acta Biochim. Biophys. Sin., 371: 1-10.

Matis, J. and M. Kudelova, 2001. Early shutoff of host protein synthesis in cells infected with herpes simplex viruses. Acta Virol., 45: 269-277.

McGeoch, D.J., A. Dolan and A.C. Ralph, 2000. Toward a comprehensive phylogeny for mammalian and avian herpesviruses. J. Virol., 74: 10401-10406.

Novembre, J.A., 2002. Accounting for background nucleotide composition when measuring codon usage bias. Mol. Biol. Evol., 19: 1390-1394.

Roizman, B. and P.E. Pellet, 2001. The Family Herpesviridae: A Brief Introduction. In: Fields Virology, Knipe, D.M., P.M. Howley, D.E. Griffin, R.A. Lamb, M.A. Martin, B. Roizman and S.E. Strauss (Eds.). 4th Edn., Lippincott Williams and Wilkins, Philadelphia, USA., pp: 2381-2397.

Romero, H., A. Zavala, H. Musto and G. Bernardi, 2003. The influence of translational selection on codon usage in fishes from the family Cyprinidae. Gene, 317: 141-147.

Shackelton, L.A., C.R. Parrish and E.C. Holmes, 2006. Evolutionary basis of codon usage and nucleotide composition bias in vertebrate DNA viruses. J. Mol. Evol., 62: 551-563.

Sharp, P.M. and W.H. Li, 1986. Codon usage in regulatory genes in *Escherichia coli* does not reflect selection for rare codons. Nucleic Acids Res., 14: 7737-7749.

Sharp, P.M., T.M.F. Tuohy and K.R. Mosurski, 1986. Codon usage in yeast cluster-analysis clearly differentiates highly and lowly expressed genes. Nucleic Acids Res., 14: 5125-5143.

Shawky, S.A. and K. Schat, 2002. Latency sites and reactivation of duck enteritis virus. Avian Dis., 46: 308-313.

Shawky, S.A. and T.S. Sandhu, 1997. Inactivated vaccine for protection against duck virus enteritis. Avian Dis., 41: 461-468.

Shawky, S.A., T. Sandhu and H.L. Shivaprasad, 2000. Pathogenicity of a low virulence duck virus enteritis isolate with immunosuppressive ability. Avian. Dis., 44: 590-599.

Smiley, R.J., 2004. Herpes simplex virus virion host shutoff protein: Immune evasion mediated by a viral RNase? J. Virol., 78: 1063-1068.

Spear, P.G. and R. Longnecker, 2003. Herpesvirus entry: An update. J. Virol., 77: 10179-10185.

Spear, P.G., 1993. Entry of alphaherpesviruses into cells. Semin. Virol., 4: 167-180.

Strand, S.S., T.K. Vanheyningen and D.A. Leib, 2004. The virion host shutoff protein of herpes simplex virus type 1 has RNA degradation activity in primary neurons. J. Virol., 78: 8400-8403.

Sueoka, N., 1988. Directional mutation pressure and neutral molecular evolution. Proc. Natl. Acad. Sci., 85: 2653-2657.

Sugiyama, T., M. Gursel, F. Takeshita, C. Coban and J. Conover *et al.*, 2005. CpG RNA: Identification of novel single-stranded RNA that stimulates human $CD14^{+}CD11c^{+}$ monocytes. J. Immunol., 174: 2273-2279.

Tsai, C.T., C.H. Lin and C.Y. Chang, 2007. Analysis of codon usage bias and base compositional constraints in iridovirus genomes. Virus Res., 126: 196-206.

Wan, X.F., D. Xu, A. Kleinhofs and J. Zhou, 2004. Quantitative relationship between synonymous codon usage bias and GC composition across unicellular genomes. BMC Evol. Biol., Vol. 4. 10.1186/1471-2148-4-19

Whitley, R.J., 1996. Herpesviruses. In: Baron's Medical Microbiology, Baron, S. (Ed.). 4th Edn., University of Texas. Galveston, USA.

Wong, E.H., D.K. Smith, R. Rabadan, M. Peiris amd L.L. Poon, 2010. Codon usage bias and the evolution of influenza A viruses. Codon usage biases of influenza virus. BMC Evol. Biol., Vol: 10. 10.1186/1471-2148-10-253

Wright, F., 1990. The effective number of codons used in a gene. Gene, 87: 23-29.

Wu, Y., A. Cheng, M. Wang, Q. Yang and D. Zhu *et al.*, 2012. Complete genomic sequence of Chinese virulent duck enteritis virus. J. Virol., 86: 5965-5965.

Xia, X.H., 2007. An improved implementation of codon adaptation index. Evol. Bioinform., 3: 53-58.

Zhao, L., A. Cheng, M. Wang, G. Yuan and M. Cai, 2008. Characterization of codon usage bias in the dUTPase gene of duck enteritis virus. Progr. Nat. Sci., 18: 1069-1076.