Analysis of Synonymous Codon Usage in Dengue Viruses

Jun-Jun Ma, Feng Zhao, Jie Zhang, Jian-Hua Zhou, Li-Na Ma, Yao-Zhong Ding,
Hao-Tai Chen, Yuan-Xing Gu and Yong-Sheng Liu
State Key Laboratory of Veterinary Etiological Biology, Lanzhou Veterinary Research Institute,
Chinese Academy of Agricultural Sciences, 730046 Lanzhou, Gansu, P.R. China

Abstract: Dengue Viruses (DENV) are the most common arboviral pathogens in tropical and subtropical regions of the world. Analysis of codon usage patterns of DENV can reveal much about the molecular evolution of the viruses. The codon usage patterns of 119 samples of DENV were analysed including the Relative Synonymous Codon Usage (RSCU) and Codon Usage Bias (CUB). The overall extent of codon usage bias is low in DENV. The codon usage patterns of the 4 serotypes of DENV are significant different indicating that the evolutionary genetics of them are independent. Compositional constraint is a significant factor that affects codon usage variation. Mutation pressure is the main factor in codon usage variation of these viruses. Furthermore, natural selection may play a role in codon usage pattern of DENV as well. However, more comprehensive analysis is needed for show the deeper characteristic of synonymous codon usage and other responsible factors of codon usage bias in DENV.

Key words: Dengue virus, relative synonymous codon usage, effective number of codons, significant factor, codon usage pattern, codon usage bias, mutation pressure

INTRODUCTION

Dengue Viruses (DENV) are the most common arboviral pathogens in tropical and subtropical regions of the world, putting at risk of infection nearly a third of the global human population (Weaver and Vasilakis, 2009). The viruses cause significant human diseases, ranging from a mild acute febrile illness-Dengue Fever (DF) to Dengue Haemorrhagic Fever (DHF) and Dengue Shock Syndrome (DSS). In recent years, there are an estimated 36 million cases of Dengue Fever (DF), 2.1 million cases of Dengue Haemorrhagic Fever (DHF) and 21,000 deaths annually. Approximately 3.6 billion people in 124 countries are at risk (Pediatric Dengue Vaccine Initiative). There are four serotypes of Dengue Virus (DENV-1, -2, -3 and -4) which belonging to the Flaviviridae family, *Flavivirus* genus (Gibbons, 2010). They are transmitted to humans by *Aedes* sp. mosquitoes, primarily *Ae. aegypti* but also *Ae. albopictus.* The DENV genome is a single-stranded positive sense RNA molecule of approximately 11 kb nucleotides which encodes three structural proteins-Capsid (C), Membrane (M) and Envelope (E) and seven non-structural proteins-NS1, NS2A, NS2B, NS3, NS4A, NS4B and NS5 (Trent *et al.*, 1990).

Since, treatment of diseases caused by DENV is supportive with no specific drugs and no effective licensed tetravalent vaccine, it is urgent to get some information about the genetic evolution or DENV. Analysis of codon usage patterns of DENV can reveal much about the molecular evolution. It has been reported that synonymous codons are not chosen equally both with and between genomes (Lloyd and Sharp, 1992; Dittmar *et al.*, 2006). This uneven codon usage is related to nucleotide usage (Adams and Antoniw, 2004), gene expression (Percudani and Ottonello, 1999; Pepin *et al.*, 2008), protein structure formation (Makhoul and Trifonov, 2002; Parmley and Hurst, 2007; Zhang *et al.*, 2009) and even viral RNA packaging (Gog *et al.*, 2007; Marsh *et al.*, 2008). Biased usage of synonymous codons may be due to various factors. In general, natural selection and/or mutation pressure are thought to be the main factors accounting for codon usage variation in different organisms (Gu *et al.*, 2004; Lesnik *et al.*, 2000; Sharp *et al.*, 1986; Zhong *et al.*, 2007; Zhou *et al.*, 2005), compositional constrains are the dominant factor in determining the codon usage variation in organisms with skewed base composition (Ohama *et al.*, 1990; Andersson *et al.*, 1998; Musto *et al.*, 1998). However in some RNA virus, mutation pressure rather than translation selection play an important role in synonymous codon usage pattern (Gu *et al.*, 2004; Levin and Whittome, 2000; Jenkins and Holmes, 2003).

Corresponding Author: Yong-Sheng Liu, State Key Laboratory of Veterinary Etiological Biology,
Lanzhou Veterinary Research Institute, Chinese Academy of Agricultural Sciences,
730046 Lanzhou, Gansu, P.R. China

Codon usage data of DENV might reveal some envoilutionary information of these viruses. Here, researchers analysed the codon usage patterns of 119 samples of DENV including the Relative Synonymous Codon Usage (RSCU) and Codon Usage Bias (CUB), to get some information about the genetic evolution of DENV.

MATERIALS AND METHODS

Sequence data: The 119 complete RNA sequences of 4 DENV serotypes, including 36 serotype 1, 33 serotype 2,

25 serotype 3 and 25 serotype 4 were downloaded randomly from the National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/ GenBank/). Serial number, length value, the isolated area and the GenBank accession numbers of the viruses were listed in Table 1.

The calculation of the Relative Synonymous Codon Usage (RSCU): To investigate the pattern of synonymous codon usage avoiding influence of amino acid composition among all sequences, the RSCU values of codons in the 4 DENV serotypes ORFs were calculated according to the published formula (Sharp *et al.*, 1986).

Table 1: Information of DENV genomes used in this study

Serotypes	No.	Strain	Location	Accession No.
1	1	HawO3663	USA	DQ672564
	2	HawO3758	USA	DQ672563
	3	HawM2540	USA	DQ672562
	4	HawM3430	USA	DQ672561
	5	HawM2516	French	DQ672560
	6	FP1104	French	DQ672559
	7	FP0908	French	DQ672558
	8	FP0705	French	DQ672557
	9	FP0203	French	DQ672556
	10		Nauru Island	U88537
	11		Nauru Island	U88537
	12	297arg00	Argentina	AF514889
	13	D1.Myanmar.31459/98	Myanmar	AY726555
	14	D1.Myanmar.31987/98	Myanmar	AY726554
	15	D1.Myanmar.49440/02	Myanmar	AY726553
	16	D1.Myanmar.44988/02	Myanmar	AY726552
	17	D1.Myanmar.44168/01	Myanmar	AY726551
	18	D1.Myanmar.38862/01	Myanmar	AY726550
	19	D1.Myanmar.37726/01	Myanmar	AY726549
	20	D1.Myanmar.32514/98	Myanmar	AY722803
	21	D1.Myanmar.23819/96	Myanmar	AY722802
	22	D1.Myanmar.40568/76	Myanmar	AY722801
	23	D1.Myanmar.305/01	Myanmar	AY713476
	24	D1.Myanmar.206/01	Myanmar	AY713475
	25	D1.Myanmar.194/01	Myanmar	AY713474
	26	D1.Myanmar.40553/71	Myanmar	AY713473
	27	D1.Myanmar.059/01	Myanmar	AY708047
	28	293arg00	Argentina	AY206457
	29	259par00	Paraguay	AF514883
	30	295arg00	Argentina	AF514885
	31	301arg00	Argentina	AF514876
	32	280par00	Paraguay	AF514878
	33	98901530 DF DV-1	Indonesia	AB189121
	34	98901518 DHF DV-1	Indonesia	AB189120
	35			NC_001477
	36	CHI3336-02	Chile	EU863650
2	37		Nicaragua	JF357907
	38	DENV-2/NI/BID-V4636/2005	Nicaragua	HQ541793
	39	DENV-2/NI/BID-V520/2005	Nicaragua	EU482752/EU482765
	40	DENV-2/TH/BID-V2311/2001	Thailand	FJ744725
	41	DENV-2/PR/13DN/1994	Puerto Rico	GQ398314
	42	DENV-2/PR/49DN/1994	Puerto Rico	GQ398272
	43	DENV-2/ID/1022DN/1975	Indonesia	GQ398268
	44	DENV-2/SG/07K3608DK1/2008	Singapore	GQ398265
	45	D2/TO/UH04/1974	Tonga	HM582117
	46	D2/PF/UH00/1973	French	HM582110
	47	D2/AS/UH73/1972	American	HM582107
	48	D2/NC/UH37/1971	New Caledonia	HM582102
	49	D2/FJ/UH21/1971	Fiji	HM582099
	50	DENV-2/GU/BID-V2950/2001	Guam	HM488257

Table 1: Continue

Serotypes	No.	Strain	Location	Accession No.
	51	Tonga/74	Tonga	AY744147
	52	DENV-2/BR/BID-V3637/2008	Brizal	HM181971
	53	DENV-2/MX/BID-V3715/2007	Mexico	GU131974
	54	New Guinea C	New Guinea	AF038403
	55	43	China	AF204178
	56	DEN2/H/IMTSSA-MART/98-703	Martinique	AF208496
	57	ThNH45/93		AF169680
	58	16681		U87411
	59	D2/SG/05K3295DK1/2005	Singapore	EU081177
	60	Dar Ar D75505	Senegal	EF457904
	61	TB16i	Indonesia	AY858036
	62	Dar Ar 578	Cote d'Ivoire	EF105380
	63	950-DF-11/12/2001	China	DQ645542
	64	ThD2_0055_99	Thailand	DQ181798
	65	Cuba165/97	Cuba	AY702038
	66	DENV-2/IPC/BID-V4270/2008	Cambodia	GU131930
	67	DENV-2/CO/BID-V3371/2005	Colombia	GQ868555
	68	DENV-2/VE/BIE-V3362/1991	Venezuela	GQ868595
	69	DENV-2/NI/BID-V648/2005	Nicaragua	FJ898435
3	70	DENV-2/KH/BID-V4307/2007	Cambodia	JF295012
	71	DENV-3/NI/BID-V4738/2009	Nicaragua	HQ166030
	72	ND143	India	FJ644564
	73	DENV-3/VE/BID-V2186/2001	Venezuela	FJ744700
	74	Sleman/78	Indonesia	AY648961
	75	Feb-80	China	AF317645
	76	D3/SG/05K4648DK1/2005	Singapore	EU081225
	77	Singapore	Singapore	EU081185
	78	TB16	Indonesia	AY858047
	79	ThD3_0104_93	Thailand	AY676350
	80	DENV-3/BR/BID-V3588/2007	Brazil	GU131867
	81	DENV-3/CO/BID-V3400/2004	Colombia	GQ868575
	82	DENV-3/LK/BID-V2409/1997	Sri Lanka	GQ252674
	83	BDH02-7	Bangladesh	AY496877
	84	PF94/136116	French	AY744685
	85	PhMH-J1-97	Philippines	AY496879
	86	DENV-3/VN/BID-V1817/2007	Viet Nam	FJ432743
	87	DENV-3/US/BID-V1737/1999	USA	FJ390377
	88	DENV-3/US/BID-V1476/2002	USA	EU687196
	89	DENV-3/VE/BID-V913/2001	Venezuela	EU482614
	90	DENV-3/GY/BID-V2980/2002	Guyana	FJ898464
	91	DENV-3/LC/BID-V2979/2001	Saint Lucia	FJ898463
	92	DENV-3/WS/BID-V2973/1995	Samoa	FJ898456
	93	DENV-3/MX/BID-V2989/2007	Mexico	FJ898442
	94	DENV-3/MZ/BID-V2418/1985	Mozambique	FJ882575
4	95	341750	Colombia	GU289913
	96	814669		AF326573
	97	DENV-4/US/BID-V860/1994	USA	FJ226067
	98	DENV-4/US/BID-V2438/1996	USA	GQ199884
	99	DENV-4/US/BID-V2447/1999	USA	FJ882600
	100	DENV-4/VE/BID-V2164/1998	Venezuela	FJ639737
	101	DENV-4/CO/BID-V1600/1997	Colombia	FJ024476
	102	DENV-4/VE/BID-V2173/1999	Venezuela	FJ639745
	103	DENV-4/VE/BID-V2176/2000	Venezuela	FJ850095
	104	DENV-4/VE/BID-V2206/2001	Venezuela	FJ639773
	105	DENV-4/CO/BID-V3408/2001	Colombia	GQ868581
	106	DENV-4/VE/BID-V1158/2007	Venezuela	FJ182016
	107	Singapore 8976/95		AY762085
	108	DENV-4/PH/BID-V3361/1956	Philippines	GQ868594
	109	ThD4_0734_00	Thailand	AY618993
	110	B5	China	AF289029
	111	DENV-4/VE/BID-V2501/2008	Venezuela	FJ882592
	112	DENV-4/CO/BID-V3412/2005	Colombia	GQ868585
	113	DENV-4/CO/BID-V3411/2004	Colombia	GQ868584
	114	DENV-4/US/BID-V1083/1986	USA	EU854295
	115	DENV-4/US/BID-V2431/1995	USA	GQ199880
	116	DENV-4/US/BID-V1094/1998	USA	EU854297
	117	DENV-4/VE/BID-V2177/2000	Venezuela	FJ639748
	118	ThD4_0348_91	Thailand	AY618990
	119	DENV-4/VE/BID-V2173/1999	Venezuela	FJ639745

$$RSCU = \frac{g_{ij}}{\sum\limits_{j} g_{ij}} \cdot n_i \qquad (1)$$

Where, g_{ij} is the observed number of the ith codon for jth amino acid which has n_i type of synonymous codons. The codon has positive codon usage bias when RSCU value is >1.0 when the value is <1.0 the codon has relative negative codon usage bias. When RSCU value is = 1.0, it means that this codon is chosen equally and randomly (Gupta and Ghosh, 2001).

Furthermore, the RSCU values of human cell were cited for analyzing the relationship between codon usage pattern of DENV and human. Codons with RSCU values <0.6 were regarded as under-represented while with RSCU values >1.6 were said to be over-represented.

Codon usage bias analysis: The Effective Number of Codons (ENC) which is the overall estimator of absolute codon usage bias (Comeron and Aguade, 1998) was used to quantify the codon usage bias of the ORF of DENV. The value of ENC is always between 20 (when only one codon is used for each amino acid) and 61 (when all codons are used equally) (Wright, 1990). Here, the ENC was used to measure the degree of codon usage bias of coding regions of DENV.

The plot of ENC against GC_3 (GC content at synonymous third codon positions) which is also an effective way to explore codon usage variation among genes was used in this study (Wright, 1990). GC_3 is a good indicator of the extent of base composition bias which is less remarkable if all codon positions are considered. Furthermore, the ENC value is independent of the base composition at non-synonymous positions since, it takes into account uneven amino acid usage (Jenkins and Holmes, 2003). The ENC-plots will lie on or just below the expected curve, if codon usage bias is only due to biased base composition (i.e., G+C content) (Wright, 1990).

Correlation analysis: Correlation analysis was used to indentify the relationship between codon usage bias and synonymous codon usage pattern (Ewens and Grant, 2005). The relationship between each general nucleotide composition (U, A, C and G%) and each nucleotide composition in the third site (U_3, A_3, C_3 and G_3%) of codon in DENV coding regions was implemented based on the Pearson's rank correlation analysis method. Additionally, linear regression analysis was carried out to analysis correlation between synonymous codon usage bias and nucleotide compositions.

Principal component analysis: Principal Component Analysis (PCA) was used to investigate the major trend in codon usage variation among different strains of DENV. Each ORF was represented as a 59 dimensional vector and each dimension corresponds to the RSCU value of one sense codon excluding Met, Trp and three stop codons (Mardia *et al.*, 1979). All statistical processes were carried out by statistical software SPSS11.5 for windows.

RESULTS

Synonymous codon usage in DENV: The nucleotide contents (A, U, C and G%) and each nucleotide contents in the third site of codon (A_3, U_3, C_3 and G_3%) in coding region were quite analogous in the 4 serotypes of DENV (Table 2). It is obvious that A content was distinctly high. By comparing the values of A_3, U_3, C_3 and G%, researchers found that A_3% was the highest of all.

The overall Relative Synonymous Codon Usage (RSCU) values of 59 sense codons in DENV were listed in Table 3, respectively. The codons ended with A are favored and the global pattern of codon usage is partially similar among the 4 serotypes of DENV examined. Due to compositional limitation, it is expected that A-ended codons should be preferentially used in these genomes.

Six codons (AGA, GGA, CCA, UCA, ACA and GUG) were over-represented in all DENV (RSCU>1.6). Except for GUG, all these codons were A-ended codons. Five codons (GCG, CGU, CCG, UCG and ACG), all of which contain the dinucleotide CpG were under-represented in both DENV and human (RSCU<0.6). Furthermore, four codons (CGA, CGC, CGG and GGU) were under-represented in all the genomes of DENV. Except for GGU, all these codons cotained CpG. However, only CpG dinucleotides were under-represented in DENV which indicating that codon usage was highly influenced by nucleotide composition.

The ENC values of DENV-1 varied from 48.97-50.923 with a mean value of 49.70 and SD of 0.52, in DENV-2 fluctuated from 48.37-49.858 with a mean value of 48.78 and SD of 0.32, among DENV-3 ranged 49.279-49.724 with a mean value of 49.52 and SD of 0.13, in DENV-4 endulated from 50.503-51.339 with a mean value of 50.81 and SD of 0.19 (Table 2). All the ENC values of 4 DENV serotypes are much higher (ENV>40), indicating that codon usage bias in DENV genomes is less biased and keeps at a stable level.

ENC values of each sample were plotted against its corresponding GC_3 (Fig. 1). It is obvious that all of the spots lie below the expected curve and the codon usage indeces are lower. These suggest that the codon usage bias is mainly influenced by mutation pressure.

Table 2: Nucleotide contents and ENC in ORFs of DENV genomes

Serotypes	No.	A%	A_3%	U%	U_3%	C%	C_3%	G%	G_3%	ENC
1	1	32.01	35.46	21.53	18.60	20.71	21.51	25.75	24.43	50.00
	2	32.01	35.22	21.59	18.80	20.71	21.57	25.69	24.40	50.18
	3	31.30	34.73	22.12	19.06	21.37	22.04	25.22	24.16	50.92
	4	32.00	35.22	21.62	18.86	20.70	21.54	25.68	24.37	50.19
	5	32.00	35.22	21.61	18.83	20.71	21.57	25.68	24.37	50.18
	6	32.00	35.22	21.62	18.86	20.70	21.54	25.68	24.37	50.19
	7	32.01	35.25	21.61	18.86	20.71	21.54	25.67	24.34	50.16
	8	31.99	35.19	21.63	18.86	20.70	21.54	25.68	24.40	50.20
	9	32.00	35.22	21.64	18.86	20.68	21.54	25.68	24.37	50.20
	10	31.90	35.01	21.57	18.74	20.71	21.49	25.82	24.76	50.50
	11	31.95	35.10	21.57	18.72	20.72	21.54	25.76	24.64	50.31
	12	32.02	35.22	21.73	18.77	20.30	20.87	25.96	25.14	49.71
	13	32.11	35.57	21.72	19.13	20.49	20.87	25.68	24.43	49.35
	14	32.16	35.84	21.53	18.54	20.61	21.25	25.69	24.37	49.02
	15	32.10	35.69	21.51	18.63	20.67	21.28	25.73	24.40	49.17
	16	31.96	35.28	21.67	18.86	20.53	21.10	25.84	24.76	49.58
	17	32.20	35.90	21.51	18.54	20.66	21.37	25.63	24.20	48.97
	18	32.05	35.57	21.67	18.98	20.52	20.95	25.76	24.49	49.37
	19	32.16	35.81	21.50	18.60	20.71	21.28	25.63	24.31	49.05
	20	31.99	35.10	21.43	18.33	20.74	21.72	25.85	24.85	49.10
	21	31.98	35.16	21.43	18.36	20.72	21.57	25.88	24.90	49.28
	22	32.11	35.54	21.52	18.45	20.63	21.54	25.74	24.46	49.06
	23	32.13	35.81	21.42	18.45	20.70	21.34	25.75	24.40	48.97
	24	32.02	35.34	21.70	19.07	20.51	20.93	25.77	24.67	49.34
	25	31.95	35.22	21.67	18.95	20.55	21.04	25.83	24.79	49.47
	26	32.04	35.28	21.56	18.39	20.61	21.63	25.79	24.70	49.33
	27	32.12	35.68	21.42	18.36	20.73	21.51	25.73	24.45	49.09
	28	32.03	35.22	21.71	18.74	20.32	20.90	25.95	25.14	49.73
	29	31.94	35.04	21.70	18.69	20.38	21.07	25.98	25.20	49.48
	30	31.94	35.04	21.68	18.63	20.39	21.10	25.99	25.23	49.46
	31	31.93	35.04	21.71	18.69	20.36	21.04	26.00	25.23	49.48
	32	32.01	35.22	21.72	18.69	20.32	20.95	25.96	25.14	49.63
	33	32.06	35.54	21.79	19.30	20.46	20.84	25.69	24.31	49.72
	34	31.97	35.13	21.58	18.74	20.69	21.51	25.76	24.61	50.33
	35	31.95	35.10	21.57	18.72	20.72	21.54	25.76	24.64	50.31
	36	31.95	35.19	21.65	19.01	20.71	21.34	25.69	24.46	50.25
2	37	33.39	36.97	21.25	18.87	20.19	20.96	25.17	23.20	48.88
	38	33.43	36.97	21.23	18.93	20.24	21.08	25.10	23.02	48.71
	39	33.43	36.97	21.23	18.93	20.24	21.08	25.10	23.02	48.71
	40	33.27	36.41	20.93	17.87	20.60	22.08	25.20	23.64	48.70
	41	33.06	36.17	21.04	18.16	20.47	21.88	25.42	23.79	48.55
	42	33.20	36.59	21.07	18.19	20.49	21.90	25.25	23.32	48.49
	43	33.04	35.88	20.92	17.63	20.62	22.41	25.42	24.09	48.62
	44	33.08	35.94	21.24	18.54	20.39	21.82	25.29	23.70	49.14
	45	32.97	35.86	21.05	18.02	20.52	22.00	25.46	24.12	48.45
	46	32.98	35.82	21.04	18.01	20.50	21.99	25.48	24.17	48.45
	47	32.98	35.89	21.08	18.02	20.48	21.97	25.45	24.12	48.47
	48	32.95	35.79	21.05	18.01	20.53	22.02	25.46	24.17	48.45
	49	32.94	35.79	21.07	18.01	20.50	21.99	25.49	24.20	48.46
	50	33.08	36.11	21.13	18.25	20.49	21.93	25.30	23.70	49.26
	51	32.94	35.70	21.06	17.98	20.51	22.02	25.49	24.29	48.37
	52	33.19	36.56	21.17	18.84	20.40	21.29	25.25	23.32	49.09
	53	33.41	36.94	21.23	18.81	20.27	21.20	25.09	23.05	48.59
	54	33.06	35.85	20.96	17.66	20.58	22.41	25.40	24.09	48.59
	55	33.05	35.85	20.95	17.60	20.61	22.44	25.39	24.12	48.50
	56	33.31	36.82	21.16	18.78	20.34	21.26	25.19	23.14	49.00
	57	33.24	36.38	20.89	17.69	20.59	22.05	25.28	23.88	48.91
	58	33.17	36.17	21.11	17.95	20.43	22.08	25.29	23.79	48.85
	59	33.14	36.26	21.09	18.25	20.44	21.82	25.33	23.67	49.01
	60	32.71	35.94	20.83	17.33	20.46	21.58	25.99	25.15	48.97
	61	33.14	36.23	21.15	18.34	20.43	21.79	25.28	23.64	49.23
	62	32.60	35.52	20.99	17.81	20.32	21.26	26.09	25.41	49.14
	63	32.88	35.70	21.16	18.54	20.50	21.76	25.46	24.00	49.86
	64	33.28	36.41	20.91	17.90	20.62	22.08	25.19	23.61	48.97
	65	33.28	36.73	21.12	18.66	20.41	21.46	25.19	23.14	48.74
	66	33.32	36.50	20.95	17.95	20.58	22.05	25.15	23.50	48.81
	67	33.32	36.85	20.90	18.04	20.54	21.82	25.24	23.29	48.84
	68	33.24	36.76	21.09	18.43	20.42	21.55	25.26	23.26	48.39
	69	33.42	36.97	21.20	18.93	20.29	21.11	25.09	23.00	48.74
3	70	32.29	34.47	21.40	18.73	20.53	21.65	25.77	25.15	49.52
	71	32.35	34.44	21.42	19.08	20.39	20.91	25.84	25.57	49.62

Table 2: Continue

Serotypes	No.	A%	A₃%	U%	U₃%	C%	C₃%	G%	G₃%	ENC
	72	32.17	33.94	21.39	18.96	20.50	21.20	25.94	25.89	49.62
	73	32.31	34.36	21.61	19.46	20.22	20.58	25.86	25.60	49.60
	74	32.25	34.41	21.44	18.76	20.47	21.47	25.84	25.36	49.32
	75	32.35	34.68	21.34	18.58	20.50	21.44	25.81	25.30	49.65
	76	32.26	34.30	21.35	18.81	20.52	21.35	25.86	25.54	49.45
	77	32.26	34.36	21.36	18.87	20.52	21.29	25.85	25.48	49.44
	78	32.27	34.36	21.35	18.46	20.51	21.59	25.87	25.60	49.34
	79	32.28	34.39	21.43	18.93	20.52	21.47	25.77	25.21	49.51
	80	32.37	34.47	21.57	19.40	20.24	20.61	25.82	25.51	49.65
	81	32.32	34.41	21.53	19.29	20.29	20.73	25.86	25.57	49.61
	82	32.30	34.36	21.50	19.23	20.42	21.06	25.78	25.36	49.28
	83	32.33	34.77	21.30	18.55	20.57	21.65	25.79	25.04	49.39
	84	32.25	34.36	21.42	18.64	20.44	21.44	25.89	25.57	49.70
	85	32.31	34.62	21.09	17.93	20.83	22.35	25.77	25.10	49.40
	86	32.40	34.77	21.40	18.76	20.45	21.35	25.75	25.13	49.32
	87	32.39	34.56	21.57	19.40	20.26	20.64	25.78	25.39	49.45
	88	32.34	34.47	21.58	19.40	20.25	20.64	25.83	25.48	49.58
	89	32.35	34.47	21.59	19.43	20.24	20.61	25.82	25.48	49.55
	90	32.29	34.33	21.63	19.52	20.21	20.55	25.87	25.60	49.60
	91	32.33	34.41	21.63	19.46	20.21	20.58	25.83	25.54	49.52
	92	32.21	34.33	21.42	18.64	20.45	21.44	25.92	25.60	49.63
	93	32.30	34.33	21.46	18.99	20.35	20.91	25.89	25.77	49.72
	94	32.30	34.33	21.54	19.40	20.35	20.79	25.81	25.48	49.48
4	95	31.08	32.67	22.07	19.36	20.56	21.84	26.29	26.12	50.71
	96	31.11	32.79	22.05	19.27	20.58	21.93	26.26	26.00	50.74
	97	31.06	32.73	22.00	19.21	20.62	21.96	26.32	26.09	50.86
	98	31.04	32.67	22.08	19.48	20.55	21.72	26.33	26.12	50.67
	99	31.08	32.82	22.03	19.33	20.59	21.81	26.30	26.03	50.50
	100	31.05	32.70	21.98	19.13	20.68	22.17	26.29	26.00	50.74
	101	31.08	32.82	22.03	19.36	20.57	21.78	26.32	26.03	50.60
	102	31.06	32.70	22.06	19.45	20.58	21.75	26.30	26.09	50.71
	103	31.02	32.53	22.07	19.42	20.56	21.81	26.35	26.24	50.86
	104	31.08	32.67	22.06	19.36	20.57	21.90	26.29	26.06	50.82
	105	31.01	32.41	22.06	19.36	20.65	22.08	26.28	26.15	50.84
	106	31.08	32.70	22.09	19.51	20.56	21.78	26.27	26.00	50.91
	107	30.93	32.33	22.14	19.66	20.61	21.82	26.32	26.19	51.09
	108	30.98	32.35	22.00	19.13	20.69	22.23	26.33	26.30	50.90
	109	31.05	32.64	21.90	19.10	20.78	22.23	26.27	26.03	51.09
	110	30.94	32.35	22.03	19.27	20.71	22.26	26.32	26.12	50.63
	111	31.12	32.73	22.06	19.48	20.58	21.78	26.24	26.00	50.91
	112	31.10	32.64	22.02	19.33	20.63	21.96	26.25	26.06	50.86
	113	31.03	32.62	22.07	19.45	20.58	21.84	26.32	26.09	50.91
	114	31.08	32.82	22.03	19.33	20.59	21.81	26.30	26.03	50.50
	115	31.05	32.64	22.13	19.60	20.53	21.66	26.29	26.09	50.96
	116	31.05	32.67	22.01	19.30	20.62	21.90	26.32	26.12	50.68
	117	31.11	32.88	22.02	19.36	20.62	21.81	26.25	25.94	50.72
	118	30.84	32.02	22.07	19.27	20.72	22.34	26.37	26.36	51.34
	119	31.06	32.70	22.06	19.45	20.58	21.75	26.30	26.09	50.71

The codon usage pattern between DENV and that of the human cells: Researchers found that synonmous codon usage pattern of 4 serotypes of DENV is partially antagonistic to human cells by comparing the patterns of synonymous codon usage of DENV and human cells (Table 3). The optimal codons of 9 amino acids (Ala, Cys, Gln, Glu, Gly, Ile, Lys, Ser and Tyr) in DENV-1 are disfavored codons of the corresponding amino acids in human beings. Particularly, the synonymous codon usage of Cys, Gln, Glu, Tle, Lys and Tyr has evolved to complementary to that of host cells. In DENV-2, optimal codons of Ala, Cys, Gln, Glu, Gly, His, Ile, Lys and Ser are disfavored codons in its host.

Specilly, the synonymous codon usage of Cys, Gln, Glu, His and Lys has evolved to complementary to host cells. Optimal codons of 10 amino acids in DENV-3, involving Ala, Cys, Gln, Glu, Gly, Ile, Leu, Lys, Phe and Ser are disfavored codons of human cells. Among these amino acids, the synonymous codon usage of Cys, Gln, Glu, Ile, Lys and Phe has evolved to complementary to host cells. Additionally, the optimal codons of Glu, Gly, Ile, Leu, Lys, Phe and Ser in DENV-4 are disfavored codons of host cells. Especially, the synonymous codon usage of Glu, Ile, Lys and Phe is complementary to that of human cells. However, the synonymous codon usage of Asn, Asp, His, Phe and Val was in accord with that of human cells. Particularly, the AGA and AGG for Arg are chosen preferentially both in DENV and human and the AGA is the most preferentially used by DENV.
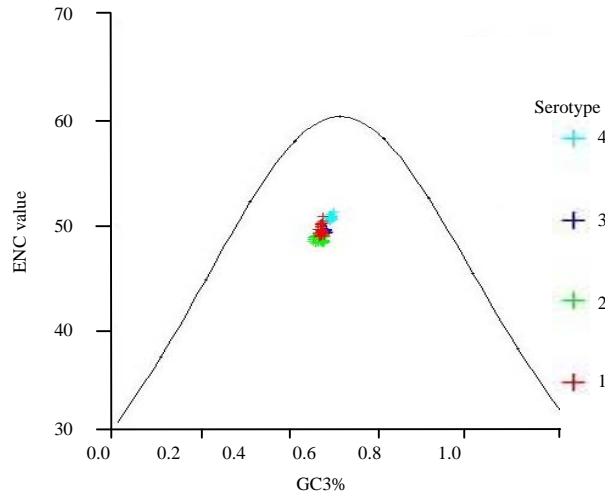
Fig. 1: A plot of Effective Number of Codons (ENC) against the GC content of the third codon position (GC3); The curve indicates the expected codon usage if GC compositional constraints alone account for codon usage bias
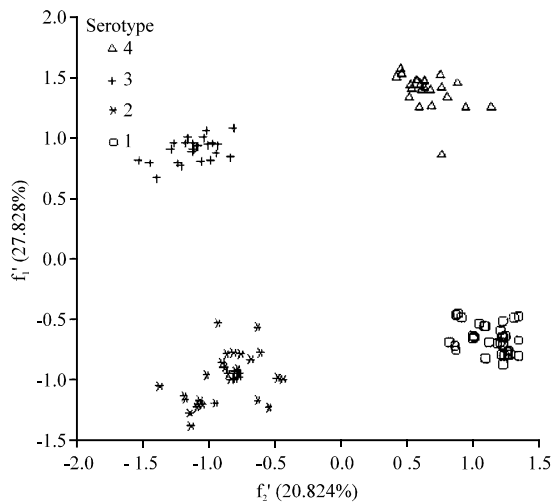

Fig. 2: A plot of values of the first and second axis of each complete coding region in principle component analysis. The first axis (f_1') accounting for 27.828% of the total variation and another major trend in the second axis (f_2') for 20.824% of the total variation

Genetic relationship based on synonymous codon usage in DENV: PCA was implemented for identifing ORFs of the 119 samples selected in this study. PCA detected one major trend in the first axis (f_1') accounting for 27.828% of the total variation and another major trend in the second axis (f_2') for 20.824% of the total variation. The plots of the f_1' and the f_2' of 4 serotypes were shown in Fig. 2. It was

Table 3: Synonymous codon usage and codon usage bias in DENV and human cells

AA[a]	Codon	DENV-1	DENV-2	DENV-3	DENV-4	Human
Ala	GCA	1.374	1.542	1.290	1.161	0.852
	GCC	1.343	1.164	1.114	1.302	1.640
	GCG	0.372	0.324	0.469	0.382	0.418
	GCU	0.909	0.970	1.128	1.157	1.091
Arg	AGA	3.118	3.414	3.322	3.126	1.203
	AGG	1.362	1.185	1.547	1.720	1.260
	CGA	0.523	0.416	0.301	0.450	0.633
	CGC	0.428	0.399	0.348	0.308	1.197
	CGG	0.274	0.192	0.235	0.211	1.196
	CGU	0.299	0.391	0.247	0.184	0.512
Asn	AAC	1.106	1.027	1.150	1.184	1.110
	AAU	0.894	0.973	0.850	0.816	0.890
Asp	GAC	1.140	1.172	1.106	1.116	1.142
	GAU	0.860	0.828	0.894	0.884	0.891
Cys	UGC	0.953	0.988	0.974	1.074	1.142
	UGU	1.047	1.012	1.026	0.926	0.859
Gln	CAA	1.190	1.208	1.294	0.985	0.508
	CAG	0.810	0.792	0.706	1.015	1.493
Glu	GAA	1.226	1.388	1.168	1.252	0.813
	GAG	0.774	0.612	0.832	0.748	1.188
Gly	GGA	2.344	2.248	2.094	2.046	0.977
	GGC	0.528	0.563	0.6489	0.532	1.398
	GGG	0.633	0.728	0.850	0.917	0.984
	GGU	0.495	0.459	0.408	0.506	0.642
His	CAC	1.085	0.954	1.115	1.013	1.191
	CAU	0.915	1.046	0.885	0.987	0.809
Ile	AUA	1.387	1.187	1.307	1.167	0.444
	AUC	0.810	1.035	0.736	0.870	1.522
	AUU	0.804	0.778	0.957	0.964	1.035
Leu	CUA	1.307	1.063	0.911	0.859	0.404
	CUC	0.699	0.950	0.928	1.002	1.215
	CUG	1.455	1.504	1.190	1.290	2.534
	CUU	0.679	0.610	0.824	0.703	0.713
	UUA	0.728	0.716	0.830	0.752	0.388
	UUG	1.133	1.157	1.313	1.394	0.731
Lys	AAA	1.333	1.273	1.161	1.246	0.822
	AAG	0.668	0.727	0.839	0.754	1.178
Phe	UUC	1.044	1.060	0.944	0.776	1.129
	UUU	0.956	0.940	1.056	1.224	0.872
Pro	CCA	2.258	2.324	2.181	1.850	1.066
	CCC	0.744	0.667	0.772	1.077	1.354
	CCG	0.352	0.269	0.177	0.368	0.460
	CCU	0.644	0.740	0.871	0.706	1.121
Ser	AGC	0.806	0.980	0.888	0.720	1.523
	AGU	0.742	0.955	0.722	0.799	0.836
	UCA	2.168	2.067	2.146	2.082	0.836
	UCC	0.971	0.878	0.970	0.878	1.394
	UCG	0.316	0.334	0.457	0.404	0.330
	UCU	0.997	0.789	0.816	1.115	1.107
Thr	ACA	1.767	2.029	2.146	1.823	1.075
	ACC	0.921	0.828	0.747	1.034	1.524
	ACG	0.583	0.500	0.476	0.472	0.464
	ACU	0.730	0.643	0.629	0.673	0.939
Tyr	UAC	0.961	1.223	1.154	1.031	1.161
	UAU	1.039	0.777	0.846	0.969	0.839
Val	GUA	0.575	0.587	0.591	0.619	0.416
	GUC	0.842	0.959	0.903	0.955	0.996
	GUG	1.723	1.635	1.651	1.715	1.897
	GUU	0.862	0.818	0.857	0.712	0.691

[a]AA is the abbreviation of amino acid; [b]RSCU value is a mean value of each codon for a particular amino aicd; [c]The preferentially used codons for each amino acid are described in bold

obvious that the plots of the ORFs belonging to the same serotype aggregated highly while the plots of different

Table 4: Summary of correlation analysis between the A, U, C, G contents and A_3, U_3, C_3, G_3 contents in all selected samples

Serotype	Correlation analysis (%)	A_3%	U_3%	C_3%	G_3%
1	A	r = 0.784**	r = -0.181[NS]	r = -0.337*	r = -0.141[NS]
	U	r = -0.538**	r = 0.557**	r = 0.240[NS]	r = 0.125[NS]
	C	r = -0.126[NS]	r = -0.106[NS]	r = 0.817**	r = -0.668**
	G	r = -0.195[NS]	r = -0.017[NS]	r = -0.553**	r = 0.862**
2	A	r = 0.892**	r = 0.572**	r = -0.368*	r = -0.942**
	U	r = 0.320[NS]	r = 0.787**	r = -0.582**	r = -0.513**
	C	r = -0.545**	r = -0.658**	r = 0.915**	r = 0.412**
	G	r = -0.767**	r = -0.560**	r = 0.188[NS]	r = 0.949**
3	A	r = 0.642**	r = 0.034[NS]	r = -0.260[NS]	r = -0.330*
	U	r = -0.217[NS]	r = 0.533**	r = -0.938**	r = 0.462*
	C	r = 0.161[NS]	r = -0.551**	r = 0.950**	r = -0.482*
	G	r = -0.738**	r = 0.095[NS]	r = -0.039[NS]	r = 0.662**
4	A	r = 0.893**	r = 0.086[NS]	r = -0.449*	r = -0.751**
	U	r = -0.221[NS]	r = 0.334[NS]	r = -0.433*	r = 0.394[NS]
	C	r = -0.463*	r = -0.362[NS]	r = 0.850**	r = 0.225[NS]
	G	r = -0.492*	r = 0.075[NS]	r = 0.301[NS]	r = 0.565**

[a]r value in this table is calculated in each correlation analysis; NS = Non-Significant (p>0.05); *0.01<p<0.05; **p<0.01

Table 5: Analysis of correlation between the first two principle axes and nucleotide contents in samples

Serotypes	Correlation (%)	f_1' (27.828%)	f_2' (20.824%)
1	A_3	r = 0.028[NS]	r = 0.046[NS]
	U_3	r = 0.147[NS]	r = 0.489**
	C_3	r = -0.450**	r = 0.256[NS]
	G_3	r = 0.496**	r = -0.584**
2	A_3	r = -0.756**	r = -0.469**
	U_3	r = -0.417*	r = -0.580**
	C_3	r = 0.196[NS]	r = 0.542**
	G_3	r = 0.794**	r = 0.443**
3	A_3	r = 0.041[NS]	r = 0.028[NS]
	U_3	r = 0.436*	r = 0.370[NS]
	C_3	r = -0.416*	r = -0.365[NS]
	G_3	r = 0.030[NS]	r = 0.053[NS]
4	A_3	r = 0.225[NS]	r = -0.340[NS]
	U_3	r = 0.426*	r = -0.353[NS]
	C_3	r = -0.525**	r = 0.484*
	G_3	r = -0.044[NS]	r = 0.258[NS]

[a]r value in this table is calculated in each correlation analysis; NS = Non-Significant; *0.01<p<0.05; **p<0.01

serotype scattered in different areas. These findings indicated that the codon usage patterns are significant different among the different serotype of DENV samples, implying that the 4 serotypes of DENV are genetically quite distinct.

Compositional properties of DENV genomes: The 4 DENV serotypes A, U, G and C% were compared with A_3, U_3, G_3 and C_3%, respectively (Table 4). A complex correlation was observed in nucleotide compositions. In DENV-1, A_3, C_3 and G_3% have a significant negative correlation with U, G and C%, respectively. In DENV-2, there are strong negative correlations between A_3 and C%, G% between U_3 and C%, G% between C_3 and U%, and between G_3 and A%, U%, respectively. In DENV-3, there are significant negative correlation between A_3 and G%, U_3 and C% and C_3 and U%. In DENV-4, G_3 and A% have a significant negative correlation. All these data suggest that nucleotide constraints play a role in codon usage pattern of DENV. However, significant positive correlation between A and U_3%, C and G_3% in DENV-2, U and G_3% in DENV-3 and no correlation between U and U_3% in DENV-4 might indicate that natural selection plays a role in codon usage pattern of DENV as well. Furthermore, researchers also performed a linear regression analysis on synonymous codon usage bias and nucleotide compositions. The correlations between the first two principle axes (f_1' and f_2') of each DENV genome and each nucleotide content in the third site of codon were not analogue in the four serotypes of DENV (Table 5). In DENV-1, both axis values are closely correlated with base nucleotide G content on the third site of codon, indicating that nucleotide G is the major factor influencing the synonymous codon usage pattern, although A content is the highest in the ORFs. In

DENV-2, f_1' is correlated with most of nucleotide contents in the third site of codon while the f_2' is tightly correlated with all nucleotide contents in the third codon position. Taken together, compositional constraint is a factor responsible for the shaping of synonymous codon usage among these samples. Interestingly, U_3% has stronger correlation with f_2' values than f_1' in DENV-1 and C_3% has stronger correlation with f_2' than f_1' in DENV-2, implying that they have no real correlation with codon usage indices for f_1' being the first principal component. Furthermore in DENV-3 and -4, the f_1' and f_2' values have non-correlation with most of nucleotide contents on the third codon position.

DISCUSSION

The mean values in 4 serotypes of DENV were 49.70, 48.78, 49.52 and 50.81, respectively which suggesting that the synonymous codon usage bias in coding regions of DENV are low. The result is similar to some other RNA virus, such as SARSCoV (mean ENC = 48.99), H5N1 influenza virus (mean ENC = 50.91) and poliovirus (mean ENC = 53.754) (Gu *et al.*, 2004; Zhou *et al.*, 2005; Zhang *et al.*, 2011). The reason for DENV had a low codon usage bias may be that a low bias is advantageous to replicate efficiently in both mosquito and vertebrate cells which are two very different cell types with potentially distinct codon preferences (Jenkins and Holmes, 2003). However, earlier report suggested that insect and mammalian cells may constitute similar adaptive environments for arbovirus (Novella *et al.*, 1999). Consequently, the reasons for the low codon usage bias in DENV are still uncertain.

The ratio of A and A_3 were much higher than other general nucleotide composition and nucleotide

composition in the third position, respectively which explained why most optimal codons were A-ended in DENV. By comparing the A, U, C and G% with A_3, U_3, C_3 and G_3% of 4 DENV serotypes (Table 4), researchers found that composition contraints under mutational pressure played a role in codon usage pattern. Moreover, significant positive correlation between A and U_3, C and G_3% in DENV-2, U and G_3% in DENV-3 and no correlation between U and U_3% in DENV-4 might indicate that natural selection plays a role in codon usage pattern of DENV as well.

Codon usage is known to be influenced by nucleotide composition (Jenkins and Holmes, 2003). Statistically significant differences in nucleotide of codons were found when the less used codons were compared with more used codons or with the overall sequence composition in DENV. It was obvious that nearly all of the under-represented codons contained the dinucleotide CpG. In fact, some previous studies have reported the under-representation of CpG or TpA dinucleotides in many organisms (Kariin and Burge, 1995; Greenbaum *et al.*, 2008). These findings were consistent with previous report that there is a selection pressure to eliminate CG dinucleotides in some single-stranded RNA viruses (Greenbaum *et al.*, 2008). Furthermore, the five codons (GCG, CGU, CCG, UCG and ACG) which contain the dinucleotide CpG were under-represented in both DENV and human cells suggesting that DENV may evolve by mimicking some of the features of their host's synonymous codon usages. Additionally, the synonymous codon usage pattern of DENV is partially antagonistic to host cells. The antagonistic portion of codon usage pattern probably enable DENV proteins to fold properly, although, the translation efficiency of the corresponding amino acids may decreased. While the coincident portion enable the corresponding amino acids to be translated efficiently (Liu *et al.*, 2011).

It had been reported that mutational pressure, rather than natural selection is the most important cause of codon usage bias in human RNA virus, accounting for 71-85% of the observed bias (Jenkins and Holmes, 2003). The significant correlation between f_1' and C_3%, G_3% and A_3%, G_3% in DENV-2 and f_1' and C_3% in DENV-4 suggested that mutation pressure played an important role in codon usage variation. Although, A_3% were significant higher than other nucleotide composition in the third position, f_1' has no correlation with it in DENV-1. Nevertheless, f_1' has significant correlation with G_3% which was much higher than C_3% and U_3%. These also suggested that mutation pressure is the determinant in codon usage variation. Furthermore, the plot of ENC

against GC_3 indicated that the codon usage bias was mainly influenced by mutation pressure in DENV. However, the non-correlation between most of nucleotide contents on the third codon position and the two axis values, especially in DENV-3 suggesting that there might be some additional factors affecting the synonymous codon usage variation.

Synonymous codon usage pattern may reflect the evolutional relationship among the 4 serotypes. That the codon usage pattern of the 4 serotypes of DENV are significant different may indicate that the evolutionary genetics of the viruses are independent of each other (Fig. 2). Some phylogenetic trees have shown that the 4 serotypes of DENV are phylogenetically distinct (Kuno *et al.*, 1998; Holmes and Twiddy, 2003). There are two hypothesis which can explain why DENV exists as four distinct serotypes. At present, most evidence supports the one that DENV became separated into distinct lineages because of geographic or ecological partitioning in different primate populations, so that the 4 serotypes envolved independently (Holmes and Twiddy, 2003). Maintenance of the current levels of antigenic and genetic diversity among the four DENV serotypes may be achieved by immune enhancement, if each virus gains replication efficiency in the primate host due to limited cross-reactive immunity (Ferguson *et al.*, 1999).

The research is the first report of synonymous codon usage analysis on DENV. The codon usage bias in DENV is low and mutation pressure is the main factor that affects codon usage variation in DENV. However, more comprehensive analysis is needed for showing the deeper characteristic of synonymous codon usage and other responsible factors of codon usage bias in DENV.

CONCLUSION

This is the first report of synonymous codon usage analysis on DENV. It may have theoretical value in understanding the evolution of DENV. The codon usage patterns of the 4 serotypes of DENV are significant different indicating that the evolutionary genetics of them are independent. Compositional constraint is a significant factor that affects codon usage variation. Mutation pressure is the main factor that affects codon usage variation in these viruses. Furthermore, natural selection may play a role in codon usage pattern of DENV as well. However, more comprehensive analysis is needed for show the deeper characteristic of synonymous codon usage and other responsible factors of codon usage bias in DENV.

REFERENCES

Adams, M.J. and J.F. Antoniw, 2004. Codon usage bias amongst plant viruses. Arch. Virol., 149: 113-135.

Andersson, S.G.E., A. Zomorodipour, J.O. Andersson, T. Sicheritz-Ponten and U.C.M. Alsmark *et al.*, 1998. The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. Nature, 396: 133-140.

Comeron, J.M. and M. Aguade, 1998. An evaluation of measures of synonymous codon usage bias. J. Mol. Evol., 47: 268-274.

Dittmar, K.A., J.M. Goodenbour and J. Pan, 2006. Tissue-specific differences in human transfer RNA expression. PLoS Genet., Vol. 2, No. 12. 10.1371/journal.pgen.0020221.

Ewens, W.J. and G.R. Grant, 2005. Statistical Methods in Bioinformatics: An Introduction. 2nd Edn., Springer, New York, USA., ISBN-13: 9780387400822, Pages: 597.

Ferguson, N., R. Anderson and S. Gupta, 1999. The effect of antibodydependent enhancement on the transmission dynamics and persistence of multiple-strain pathogens. Proc. Natl. Acad. Sci. USA., 96: 790-794.

Gibbons, R.V., 2010. Dengue conundrums. Int. J. Antimicrob. Agents, 36: S36-S39.

Gog, J.R., E.D.S. Afonso, R.M. Dalton, I. Leclercq and L. Tiley *et al.*, 2007. Codon conservation in the influenza A virus genome defines RNA packaging signals. Nucl. Acids Res., 35: 1897-1907.

Greenbaum, B.D., A.J. Levine, G. Bhanot and R. Rabadan, 2008. Patterns of evolution and host gene mimicry in influenza and other RNA viruses. PLoS Pathog., Vol. 4.

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.

Gupta, S.K. and T.C. Ghosh, 2001. Gene expressivity is the main factor in dictating the codon usage variation among the genes in *Pseudomonas aeruginosa*. Gene, 273: 63-70.

Holmes, E.C. and S.S. Twiddy, 2003. The origin, emergence and evolutionary genetics of dengue virus. Infect. Genet. Evol., 3: 19-28.

Jenkins, G.M. and E.C. Holmes, 2003. The extent of codon usage bias in human RNA viruses and its evolutionary origin. Virus Res., 92: 1-7.

Kariin, S. and C. Burge, 1995. Dinucleotide relative abundance extremes: A genomic signature. Trends Genet., 11: 283-290.

Kuno, G., G.J. Chang, K.R. Tsuchiya, N. Karabatasos and C.B. Cropp, 1998. Phylogeny of the genus *Flavivirus*. J. Virol., 72: 73-83.

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.

Levin, D.B. and B. Whittome, 2000. Codon usage in nucleopolyhedroviruses. J. Gen. Virol., 81: 2313-2325.

Liu, Y.S., J.H. Zhou, H.T. Chen, L.N. Ma, Z. Pejsak, Y.Z. Ding and J. Zhang, 2011. The characteristics of the synonymous codon usage in enterovirus 71 virus and the effects of host on the virus in codon usage pattern. Infect. Genet. Evol., 11: 1168-1173.

Lloyd, A.T. and P.M. Sharp, 1992. Evolution of codon usage patterns: The extent and nature of divergence between Candida albicans and Saccharomyces cerevisiae. Nucleic Acids Res., 20: 5289-5295.

Makhoul, C.H. and E.N. Trifonov, 2002. Distribution of rare triplets along mRNA and their relation to protein folding. J. Biomol. Struct. Dyn., 20: 413-420.

Mardia, K., J. Kent and J. Bibby, 1979. Multivariate Analysis. Academic Press, USA., ISBN-10: 0124712525.

Marsh, G.A., R. Rabadan, A.J. Levine and P. Palese, 2008. Highly conserved regions of influenza A virus polymerase gene segments are critical for efficient viral RNA packaging. J. Virol., 82: 2295-2304.

Musto, H., H. Romero and H. Rodriguez-Maseda, 1998. Heterogeneity in codon usage in the fatworm *Schistosoma mansoni*. J. Mol. Evol., 46: 159-167.

Novella, I.S., C.L. Hershey, C. Escarmis, E. Domingo and J.J. Holland, 1999. Lack of evolutionary stasis during alternating replication of an arbovirus in insect and mammalian cells. J. Mol. Biol., 287: 459-465.

Ohama, T., A. Muto and S. Osawa, 1990. Role of GC-biased mutation pressure on synonymous codon choice in Micrococcus luteus, a bacterium with a high genomic GC-content. Nucleic. Acids Res., 18: 1565-1569.

Parmley, J.L. and L.D. Hurst, 2007. How do synonymous mutations affect fitness? BioEssays, 29: 515-519.

Pepin, K.M., J. Domsic and R. McKenna, 2008. Genomic evolution in a virus under specific selection for host recognition. Infect. Genet. Evol., 8: 825-834.

Percudani, R. and S. Ottonello, 1999. Selection at the wobble position of codons read by the same tRNA in *Saccharomyces cerevisiae*. Mol. Biol. Evol., 16: 1752-1762.

Sharp, P.M., T.M. 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.

Trent, D.W., C.L. Manske, G.E. Fox, M.C. Chu, C.S. Kliks and T.P. Monath, 1990. The molecular epidemiology of dengue viruses: Genetic variation and microevolution. Appl. Virol. Res., 2: 293-315.

Weaver, S.C. and N. Vasilakis, 2009. Molecular evolution of dengue viruses: Contributions of phylogenetics to understanding the history and epidemiology of the preeminent arboviral disease. Infect. Genet. Evol., 9: 523-540.

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

Zhang, G., M. Hubalewska and Z. Ignatova, 2009. Transient ribosomal attenuation coordinates protein synthesis and co-translational folding. Nat. Struct. Mol. Biol., 16: 274-280.

Zhang, J., M. Wang, W.Q. Liu, J.H. Zhou and H.T. Chen *et al.*, 2011. Analysis of codon usage and nucleotidecomposition bias in polioviruses. Virol. J., Vol. 8 10.1186/1743-422X-8-146.

Zhong, J., Y. Li, S. Zhao, S. Liu and Z. Zhang, 2007. Mutation pressure shapes codon usage in the GC-Rich genome of foot-and-mouth disease virus. Virus Genes, 35: 767-776.

Zhou, T., W. Gu, J. Ma, X. Sun and Z. Lu, 2005. Analysis of synonymous codon usage in H5N1 virus and other influenza a viruses. Bio. Syst., 81: 77-86.