Translation Elongation Rate Measurement of Epstein-Barr Virus Strain GD1

Gholamreza Motalleb[1]

Abstract

Background: Epstein-Barr Virus (EBV) has a great co relationship with human malignancies such as gastric carcinoma. Synonymous codon investigations in viruses could help designing vaccine, to generate immunity.
Codon Adaptation Index (CAI) has measured translation elongation rate, among the highly expressed genes. The aim of this study was: usage of "CAI" to measure translation efficiency to know how fast EBV-GD1 could produce its proteins.
Methods: The complete genomic sequences of human herpes virus 4 strain GD1 have retrieved from http://www.ncbi.nlm.nih.gov/sites/gquery (GenBank accession no. AY961628) to extract all protein-coding genes. The sequences have analyzed with DAMBE software.
Results: The results have shown that CAI values for the EBV-GD1 genes were 0.76356 ± 0.02957. The highest and lowest CAI values were 0.82233 and 0.68321 respectively. The results have shown that highly expressed genes mostly had more codon usage bias than low expressed genes.
Conclusion: The results provide and introduce not only a system, but also the principles in order to understand the pathogenesis and evolution of EBV-GD1, to open a window, in order to make a better product or vaccine to challenge with the virus.
Keywords: Epstein-Barr virus ; Gene expression ; Codon usage bias

Please cite this article as: Motalleb G. Translation Elongation Rate Measurement of Epstein-Barr Virus Strain GD1.Iran J Cancer Prev. 2013; 6(4):214-21.

1. Dept. of Biology, Faculty of Science, University of Zabol, Zabol, Iran

Corresponding Author:
Gholamreza Motalleb, PhD;
Assistant Professor of Molecular Biology
Tel: (+98) 54 22 24 25 03
Email: reza.motaleb@uoz.ac.ir
Received: 12 May. 2013
Accepted: 11 Aug. 2013
Iran J Cancer Prev. 2013; 4:214-21

Introduction

The Codon Adaptation Index (CAI) measures the synonymous codon, using bias for DNA or RNA sequence. CAI has known to be an excellent predictor of gene expression in prokaryotes and unicellular eukaryotes. CAI has evaluated the effect of natural selection in pattern of codon usage, and prediction of gene expression level [1, 2] to find highly expressed genes [3, 4], virus genes adaptation evaluation and their hostages [1], indication of heterologous gene expression [5], comparing organisms for codon usage favorites [1], to find the genes transfection horizontally [6-8] using the genomic codon for bias detection in genomes [9] to study the cell cycle species [10], to optimize DNA vaccines [11], gene therapy [12], vaccine development and recombinant therapeutics [13]. Some have reported the influence of codon usage on the viral cycle among viruses. Adaptation Studies for host codon usage, have indicated viral genes which codify for critical proteins, tend to use the synonymous codons, which mostly represented in the host genome [14], but the synonymous codon usage within and between genomes could not be used equally [15]. Epstein-Barr Virus (EBV) is a ubiquitous double stranded DNA virus, derived of human herpes virus family, which has B-lymphotropism. More than 90% of adults have serologic evidence of infection with this virus. It has acquired during early childhood, but the age of infection is much lower in undeveloped countries with low socioeconomic condition [16]. It has been documented that gastric carcinoma, Burkitt's lymphoma, undifferentiated Nasopharyngeal Carcinoma (NPC), Hodgkin's disease, B and T-cell lymphoma, and B-cell lympho proliferations among the immune compromised patients could cause by EBV [17-20]. EBV infection is ubiquitous. Iran has a high incidence rate of gastric carcinoma with an annual incidence of 26.1 per 100,000 for males and 11.1 for females [21]. In bio pharmacology, researchers have interested to improve translation efficiency that is derived from protein production. Unfortunately, experiments are tedious and the reality is much more complicated. In the current study, DAMBE software (version 5.3.27) has used to assess CAI, to realize how fast EBV-GD1 could produce its proteins. These data might provide and introduce a system and principles in order to understand the pathogenesis and evolution of EBV-GD1.

Materials and Methods

The research study has started in Winter of 2012. All bioinformatics analysis has performed at bioinformatics facility of Faculty of Science in University of Zabol. Sequences of the genome segments of human herpes virus 4 strain GD1 (GenBank accession no. AY961628) have retrieved from http://www.ncbi.nlm.nih.gov/sites/gquery (GenBank accession no. AY961628) to extract all protein coding genes in order to evaluate the effectiveness of CAI from DAMBE [22]. To calculate the CAI for any protein-coding sequence:

(1)

$$CAI = e^{\left(\frac{\sum_{i=1}^{n}[CodFreq_i \ln(w_i)]}{\sum_{i=1}^{n} CodFreq_i}\right)}$$

n is the number of sense codons and the related **wij** value will always be 1 regardless of codon usage bias of the gene. CAI of a coding sequence (CDS) has calculated from 1) the codon frequencies of the CDS and 2) the codon frequencies of a known highly expressed genes set (often referred to as the reference set) which has been used to generate a column of **w** values:

(2)

$$w_{ij} = \frac{f_{ij.\mathrm{ref}}}{\mathrm{Max} f_{i.\mathrm{ref}}}$$

Where $f_{ij.ref}$ is the frequency of codon **j** in synonymous codon family **i**, and **Maxf**$_{i.ref}$ is the maximum codon frequency in synonymous codon family **i**. The codon whose frequency is **Maxf**$_{i.ref}$ has been often referred to as the major codon (whose **w** is 1), and the other codons have referred as minor codons. The major codon has assumed to be the translated optimal codon.

The CAI value of a CDS has calculated as below equation:

(3)

$$CAI = \exp\left(\frac{\sum_{i=1}^{m}\sum_{j=1}^{n_i}[f_{ij}\ln(w_{ij})]}{\sum_{i=1}^{m}\sum_{j=1}^{n_i} f_{ij}}\right)$$

Where **m** is the number of synonymous codon families, n_i is the number of synonymous codons between the codon family **i**, and f_{ij} is the frequency of codon **j** in codon family **i**. The exponent is simply a weighted average of **ln(w)**.

The maximum CAI value is 1 [23]. Relative Synonymous Codon Usage (RSCU) measures codon usage bias for each codon family. It is calculated directly from input sequences. RSCU is a codon-specific index for codon usage, whereas CAI is a gene-specific index for codon usage, which related to gene expression [23]. The general equation for RSCU is:

(4)

$$RSCU_{ij} = \frac{CodFreq_j}{\frac{\left(\sum_{j=1}^{NumCodon_i} CodFreq_{ij}\right)}{NumCodon_i}}$$

i is codon family, **j** is specific codon within the family [23]. For example, **i** for alanine codon family is GCU, GCC, GCA, and GCG, then **j** would be a specific codon such as GCU. RSCU measures codon usage bias for each codon family. RSCU is 1 whencodon usage bias does not exist, but RSCU would be higher than 1 when its codon is either overused or vice versa [22].

Results

Human herpes virus 4 strain GD1 genome segment sequences have used to evaluate the effectiveness of CAI from DAMBE. The results have shown that CAI values for the EBV-GD1 genes were 0.76356 ± 0.02957 (Table 1).

The highest and lowest CAI values were 0.82233 and 0.68321 respectively. The results have shown for alanine codon family (as an example), genes with high-CAI have more codon usage bias with highest RSCU being 2.923 and the lowest being only 0.246. In contrast, for the low-CAI genes, the highest and lowest RSCU is 2.797 and 0.241 (Table 2 and 3). The results have shown that highly expressed genes mostly had more codon usage bias than lowly expressed genes (Figure 1) but ANOVA for RSCU_H and RSCU_L genes , has not significantly shown difference ($P>0.05$).

Table 1. Output of codon adaptation index (CAI) for EBV-GD1 (Mean: 0.76356; STD: 0.02957)

SeqName	SeqLen	CAI	SeqName	SeqLen	CAI
unknown|1736	3954	0.76709	unknown|98764	651	0.78868
unknown|9710	510	0.76776	unknown|C99460	2427	0.78326
unknown|36258	1353	0.70990	unknown|103578	834	0.82115
unknown|46538	1008	0.76740	unknown|106768	1215	0.77914
unknown|47455	1773	0.77478	unknown|C108378	225	0.75569
unknown|49154	528	0.74644	unknown|C111572	996	0.75849
unknown|C49725	9528	0.76936	probable DNA packaging protein|112569	2070	0.78928
unknown|C59248	3717	0.76630	unknown|112569	975	0.76398
unknown|62966	1092	0.76266	unknown|C113494	1008	0.77171
unknown|64136	2478	0.79963	unknown|C114482	1521	0.75317
unknown|66629	906	0.80136	unknown|C115975	675	0.80152
unknown|67628	1212	0.76329	unknown|C117993	702	0.72556
unknown|68847	1071	0.77103	unknown|C118758	1260	0.75635
unknown|C70473	1314	0.77520	unknown|C120031	903	0.78173
unknown|C71899	117	0.76729	unknown|C120952	4143	0.80319
unknown|C71987	2622	0.76897	unknown|125621	1725	0.78192
unknown|74654	654	0.76750	unknown|C128546	2118	0.77372
unknown|C75368	834	0.79476	unknown|C130668	1821	0.75959
unknown|76277	306	0.73176	unknown|132490	744	0.71112
unknown|76655	486	0.71418	unknown|133046	1710	0.76698
unknown|C77160	2568	0.73715	unknown|135557	1815	0.78338
unknown|C77297	444	0.72631	unknown|C137409	744	0.75202
unknown|79820	357	0.68321	unknown|C140486	2772	0.74451
EBNA3B (EBNA4A) latent protein|82903	2814	0.72716	unknown|C149527	708	0.71853

SeqName	SeqLen	CAI	SeqName	SeqLen	CAI
EBNA3C latent protein|85921	3027	0.74198	unknown|C150198	1407	0.74320
unknown|C89046	669	0.73457	unknown|150236	309	0.72173
Z protein|C89811	735	0.73358	unknown|C151616	936	0.78017
unknown|C90996	1815	0.79433	unknown|C153152	3045	0.82233
unknown|92812	930	0.77330	unknown|C156202	2571	0.79823
unknown|93932	1611	0.73806	unknown|C160837	3384	0.80318
unknown|95580	1923	0.70636	unknown|C164308	660	0.81822
unknown|97588	411	0.77062	unknown|164957	663	0.79950
unknown|97983	765	0.76167	unknown|C166757	180	0.72802

Table 2. RSCU genes with low-CAI value (RSCU_L) for EBV-GD1

Codon	AA	ObsFreq	RSCU_L	Codon	AA	ObsFreq	RSCU_L
UAG	*	0	0.000	UGA	*	1	1.000
GCU	A	12	0.361	UAA	*	2	2.000
GCC	A	20	0.602	GCG	A	8	0.241
UGU	C	3	0.750	GCA	A	93	2.797
GAU	D	34	1.172	UGC	C	5	1.250
GAG	E	21	0.792	GAC	D	24	0.828
UUU	F	13	1.444	GAA	E	32	1.208
GGU	G	29	0.410	UUC	F	5	0.556
GGC	G	29	0.410	GGG	G	76	1.074
CAC	H	5	0.455	GGA	G	149	2.106
AUU	I	17	1.821	CAU	H	17	1.545
AUC	I	6	0.643	AUA	I	5	0.536
AAG	K	17	1.478	AAA	K	6	0.522
CUC	L	14	1.167	CUA	L	14	1.167
CUU	L	13	1.083	CUG	L	7	0.583
UUG	L	8	1.000	UUA	L	8	1.000
AAC	N	12	1.143	AUG	M	20	1.000
CCA	P	68	1.744	AAU	N	9	0.857
CCU	P	40	1.026	CCC	P	33	0.846
CAA	Q	28	1.217	CCG	P	15	0.385
AGA	R	19	0.844	CAG	Q	18	0.783
CGA	R	10	1.000	AGG	R	26	1.156
CGG	R	12	1.200	CGC	R	9	0.900
AGC	S	11	0.759	CGU	R	9	0.900
UCA	S	24	2.043	AGU	S	18	1.241
UCG	S	4	0.340	UCC	S	11	0.936
ACC	T	14	1.167	UCU	S	8	0.681
ACG	T	7	0.583	ACA	T	17	1.417
GUU	V	10	0.930	ACU	T	10	0.833
GUC	V	12	1.116	GUG	V	11	1.023
UGG	W	12	1.000	GUA	V	10	0.930
UAU	Y	10	1.429	UAC	Y	4	0.571

ObsFreq: observation frequency; AA: amino acid.RSCU_L: Low relative synonymous codon usage.

Table 3.RSCU genes with high-CAI value (RSCU_H) for EBV-GD1

Codon	AA	ObsFreq	RSCU_ H	Codon	AA	ObsFreq	RSCU_ H
UAG	*	1	1.000	UGA	*	0	0.000
GCU	A	8	0.246	UAA	*	2	2.000
GCC	A	95	2.923	GCG	A	18	0.554
UGU	C	8	0.390	GCA	A	9	0.277
GAU	D	22	0.518	UGC	C	33	1.610
GAG	E	70	1.750	GAC	D	63	1.482
UUU	F	33	0.971	GAA	E	10	0.250
GGU	G	3	0.132	UUC	F	35	1.029
GGC	G	41	1.802	GGG	G	39	1.714
CAC	H	31	1.442	GGA	G	8	0.352
AUU	I	15	0.703	CAU	H	12	0.558
AUC	I	40	1.875	AUA	I	9	0.422
AAG	K	63	1.703	AAA	K	11	0.297
CUC	L	58	1.415	CUA	L	9	0.220
CUU	L	4	0.098	CUG	L	93	2.268
UUG	L	10	1.818	UUA	L	1	0.182
AAC	N	37	1.609	AUG	M	29	1.000
CCA	P	15	0.779	AAU	N	9	0.391
CCU	P	14	0.727	CCC	P	36	1.870
CAA	Q	10	0.417	CCG	P	12	0.623
AGA	R	9	0.529	CAG	Q	38	1.583
CGA	R	4	0.219	AGG	R	25	1.471
CGG	R	29	1.589	CGC	R	34	1.863
AGC	S	27	1.636	CGU	R	6	0.329
UCA	S	7	0.444	AGU	S	6	0.364
UCG	S	16	1.016	UCC	S	31	1.968
ACC	T	35	1.892	UCU	S	9	0.571
ACG	T	25	1.351	ACA	T	12	0.649
GUU	V	4	0.143	ACU	T	2	0.108
GUC	V	37	1.321	GUG	V	66	2.357
UGG	W	15	1.000	GUA	V	5	0.179
UAU	Y	11	0.379	UAC	Y	47	1.621

ObsFreq: observation frequency; AA: amino acid.RSCU_H: High relative synonymous codon usage.

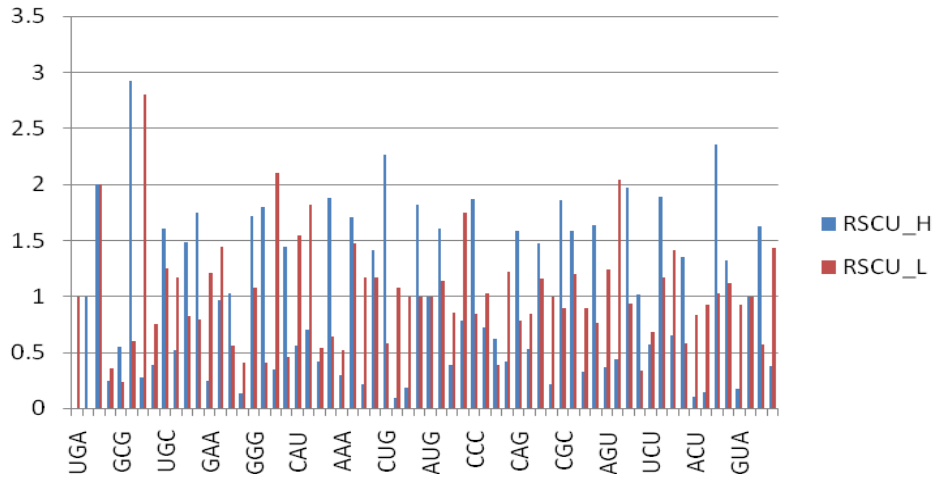

Figure 1. It shows relative synonymous codon usage (RSCU) for high-CAI and low-CAI genes (RSCU_H and RSCU_L, respectively) for 64 codons of EBV-GD1.

Discussion

In molecular biology, one of the fundamental questions is genetic codes. In microorganisms, the unequal usage of synonymous codons, due to both of the mutation and the pressure of usual normal selection, has been accepted as the most common hypothesis which could effect on translation level.

The CAI has used highly expressed genes from a species to evaluate the relative merits of each codon. CAI has also used for gene expression and translation efficiency [23]. The mRNA translation efficiency has depended partially on mRNA coding strategy, and has reflected codon usage bias. Codon usage bias has often determined by codon-specific, as well as the other existing gene-specific.

A representative of codon-specific could be the RSCU or relative synonymous codon usage [24], and a representative of the gene-specific could be the codon adaptation index or CAI. CAI is a measure index of translation elongation rate according to our finding of highly expressed genes [25]. Clarifying in a different better way, highly expressed genes would be under pressure to use abundant, or common, or cheap amino acids. On the other hand, we couldn't produce a big mass of the protein that its amino acids components would be rare or expensive. According to previous data, highly expressed genes which would use codons,have distinguished by the most abundant tRNA, in order to code each amino acid. For this matter, highly biased codon has used in highly expressed genes, especially in organisms with rapidly replication [23-28].

Finding the highly and lowly expressed genes in organisms, we might be able to select them as the main targets in pharmacology, especially in vaccine production. CAI has calculated with a reference set of highly expressed genes. The maximum CAI is 1, and the minimum is 0. In general, the higher that the CAI value would be, caused the mRNA have translatedmuch more efficient. Highly expressed human genes typically have CAI value above 0.7, have given the human reference set of highly expressed genes [23]. The results have shown CAI values for the EBV-GD1 genes were 0.76356 $\pm$ 0.02957. Our result have agreed with Knipe et al. (2001) that EBV is an extremely efficient virus, which has infected a large majority of the adult population, as well as following primary infection, EBV has remained in the infected host as a lifelong asymptomatic infection [26]. Xia (2007) has determined that the viruses which have caused acute diseases, as well as being pathogen, need to translate their mRNAs efficiently [27].Figure 1 plots the RSCU for the high-CAI genes (RSCU_H) and low-CAI genes (RSCU_L) of the 64 codons. It has shown that high-CAI genes (representing highly expressed genes) have RSCU values deviating much more from 1 than the low-CAI genes (representing lowly expressed genes) relatively. The results have shown that highly expressed genes mostly had more codon usage bias than lowly expressed genes (Figure 1) but ANOVA has not shown a significant difference ($P>0.05$). This might be related to EBV, that has two different form of existence: latent and productive. The EBV genes that have been expressed during latency, has show codon usage highly different from the genes that would be expressed during lytic growth [29]. For example, what could we say about the tRNA carrying alanine? From the results, GCC is the most frequently used codon, but we might predict that $tRNA^{Ala/AGG}$ might be the most abundant. How could we test this prediction? Unfortunately this is extremely difficult experiment and all these data could be used in order to highlight the genes with high rate of expressions, related to its importance in EBV-GD1, then for this important reason might introduce a basis to understand the pathogenesis of EBV-GD1 to open a window to produce a better product or vaccine, in order to challenge with the virus.

Conclusion

The results might provide and introduce a system and its principles, in order to understand the pathogenesis then evolution of EBV-GD1 and opening a window to make a better product or vaccine to challenge with the virus. Based on the results, we could find which genes or sequences would be highly expressed, or under strong natural selection to maximize translation efficiency and accuracy in order to optimize their codon usage. To say in a different way, selection should be weak for lowly expressed genes that codon usage might largely depend on mutation bias [27].

Acknowledgment

I wish to thank my wife (Niloufar Nabi) for her support and encouragement throughout my study.

Conflict of Interest

The author has no conflict of interest in this article.

Authors' Contribution

The subject selection, study design,data entry and analysis,literature review and writing-up the article structure made and wrote by GholamrezaMotalleb.

References

1. Sharp PM, Li WH. The codon Adaptation Index- a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res. 1987;15:1281-95.

2. Goetz RM, Fuglsang A.Correlation of codon bias measures with mRNA levels: analysis of transcriptome data from Escherichia coli. BiochemBiophys Res Commun. 2005;327(1):4-7.

3. Wu G, Culley DE, Zhang W.Predicted highly expressed genes in the genomes of Streptomyces coelicolor and Streptomyces avermitilis and the implications for their metabolism. Microbiology. 2005;151(7):2175-87.

4. Wu G, Nie L, Zhang W.Predicted highly expressed genes in Nocardiafarcinica and the implication for its primary metabolism and nocardial virulence. Antonie Van Leeuwenhoek. 2006;89(1):135-46.

5. Puigbo P, Guzman E, Romeu A, Garcia-vallve S. Optimizer: A web server for optimizing the codon usage of DNA sequences. Nucleic Acids Res. 2007;35:W126-W131.

6. Lawrence JG, Ochman H. Molecular archaelogy of the Escherichia coli genome. ProcNatlAcadSci USA. 1998;95(16):9413-17.

7. Garcia-Vallve S, Palau J, Romeu A. Horizontal gene transfer in glycosyl hydrolases inferred from codon usage in Escherichia coli and Bacillus subtilis. MolBiolEvol. 1999;16(9):1125-34.

8. Garcia-Vallve S, Guzman E, Montero MA, Romeu A. GT-DB: a database of putative horizontally transferred genes in prokaryotic complete genomes. Nucleic Acids Res. 2003;31:187-9.

9. Carbone A, Zinovyev A, Kepes F. Codon adaptation index as a measure of dominating codon bias. Bioinformatics.2003; 19(16):2005-15.

10. Willenbrock H, Friis C, Juncker AS, Ussery DW.An environmental signature for 323 microbial genomes based on codon adaptation indices. Genome Biol. 2006; 7(12):R114.

11. Ruiz LM, Armengol G, Habeych E, Orduz S. A theoretical analysis of codon adaptation index of the Boophilusmicroplus bm86 gene directed to the optimization of a DNA vaccine. J Theor Biol.2006;239:445-9.

12. Kim YH, Chung JK, Lee DS, Youn H. Development of a codon-optimized sodium/iodide symporter (NIS) for a sensitive imaging reporter gene and an efficient therapeutic gene. J Nucl Med. 2013;54(Supplement 2):65.

13. Bauer AP, Leikam D, Krinner S, Notka F, Ludwig C, Gernot La¨ ngst, et al. The impact of intragenic CpG content on gene expression.Nucleic Acids Research. 2010;38(12):3891-908.

14. Tello M, Saavedra JM, Spencer E. Analysis of the use of codon pairs in the HE gene of the ISA virus shows a correlation between bias in HPR codon-pair use and mortality rates caused by the virus. Virol J. 2013;10:180.

15. Qin H, Mingshu W, Anchun C, Dekang Z, Xiaoyue C, Renyong JIA, Qihui et al. Analysis of Synonymous Codon Usage in the newly identified DPV UL43 Gene. IJEME. 2011;1(5):31-45.

16.Knipe DM., Howley PM. Roizman, B., Knipe, DM. Herpes simplex viruses and their replication. In (eds.), Fields Virology Lippincott Williams & Wilkins; Philadelphia. 2001; p. 2399–2459.

17. Magrath IT. African Burkitt's lymphoma.History, biology, clinical features, and treatment.Am J PediatrHematolOncol. 1991;13:222-46.

18. Harabuchi Y, Imai S, Wakashima J, Hirao M, Kataura A, Osato T, et al. Nasal T-cell lymphoma causally associated with Epstein-Barr virus: clinicopathologic, phenotypic, and genotypic studies. Cancer. 1996;77:2137-49.

19. Glaser SL, Lin RJ, Stewart SL, Ambinder RF, Jarrett RF, Brousset P,Et al. Epstein-Barr virus-associatedHodgkin's disease: epidemiologic characteristics in interna-tional data. Int J Cancer. 1997; 70: 375-82.

20. Hamilton-Dutoit SJ, Raphael M, Audouin J, Diebold J, Lisse I, PedersenC, et al. In situ demonstration of Epstein-Barr virus small RNAs (EBER 1) in acquired immunodeficiency syndrome-related lymphomas: correlation with tumor morphology and primary site. Blood. 1993;82:619-24.

21. Sadjadi A, Nouraie M, Mohagheghi MA, Mousavi-Jarrahi A, Malekeza-deh R, Parkin DM. Cancer occurrence in Iran in 2002, an inter-national perspective. Asian Pac J Cancer Prev. 2005;6:359-63.

22. Xia X. Data analysis in molecular biology and evolution. Kluwer Academic Publishers; Boston: 2001. Xia X, Xie Z. DAMBE: Software package for data analysis in molecular biology and evolution. Journal of Heredity.2001;92:371-3.

23. Xia X. An Improved Implementation of Codon Adaptation Index.EvolBioinform Online.2007; 3: 53-8.

24. Sharp PM, Tuohy TM, Mosurski KR. Codon usage in yeast: cluster analysis clearly differentiates highly and lowly expressed genes. Nucleic Acids Res. 1986;14(13):5125-43.

25. 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.

26.Knipe DM., Howley PM., Kieff E., Rickinson AB. Epstein–Barr virus and its replication. In (eds.), Fields Virology Philadelphia: Lippincott Williams & Wilkins. 2001; p. 2511–2574.

27.Xia X. Bioinformatics and the cell: Modern computational approaches in genomics, proteomics and transcriptomics. New York: Springer US; 2007.

28. Zhao L, Cheng A, Wang M, Yuan G, Cai M. Characterization of codon usage bias in the dUTPase gene of duck enteritis virus. Progress in natural science. 2008; 18:1069-76.

29. Karlin S, BlaisdellBE, 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.