

Full Length Research Paper

Analysis of factors affecting codon usage bias in human papillomavirus

Takaaki Kamatani* and Tatsuo Shiota

Department of Oral and Maxillofacial Surgery, Showa University School of Dentistry, Kitasenzoku 2-1-1, Ota City, Tokyo, 145-8515, Japan.

Received 8 April, 2017; Accepted 29 December, 2017

Indices of codon usage pattern of human papillomavirus (HPV) were analyzed to understand the key determinants of synonymous codon usage in the HPV genome. The complete sequences of 39 HPV genomes were downloaded from the website of the National Center for Biotechnology Information. The relative synonymous codon usage values, effective number of codons, GC content, percentage of GCs at the third position of synonymous codons (GC3s), codon adaptation index, hydrophobicity, aromaticity of conceptually translated gene products were calculated using the Codon W 1.4.2 program. HPV preferentially used codons ending with A/U. By comparing relative synonymous codon usage of the HPV genome and human genome, the codon usage of HPV was almost entirely different from that of humans. Statistical significant of the separation between codons ending with A/U and G/C on the first axis was shown by the principal component analysis. The greater number of the effective number of codon values against the value of GC3s was below the expected values. The correlation between effective number of codon values and both aromaticity and hydrophobicity showed significant high negative correlation. These results showed that composition constraint is likely the key element for codon usage in the HPV genome.

Key words: Principal component analysis, composition constraints, papilloma virus, codon usage.

INTRODUCTION

Analysis of codon usage of virus genomes enhances the understanding of virus evolution and virus-host interaction (Zhong et al., 2012). There are 64 different codons for 20 amino acids and 3 stop codons in nature. These different set of codons for the same amino acids are termed as synonymous codons, which, however, are not used at random (Karin and Mrázek, 1996). The frequency of occurrence of synonymous codons is

different for every gene and each organism (Grantham et al., 1980). A phenomenon called synonymous codon usage or codon usage bias.

Synonymous codon usage is related to DNA replication and transcription, open reading frame length, gene structure, protein secondary structure, mutation pressure, translational selection, natural selection, aromaticity and hydrophobicity of the corresponding polyprotein, and

*Corresponding author. E-mail: kamatanitakaaki@gmail.com.

environmental conditions (Zhao et al., 2003; Bishal et al., 2013; Zhang et al., 2013). Natural selection and/or mutation pressure for efficiency and accuracy are the fundamental forces that influence synonymous codon usage (Jenkins and Holmes, 2003; Hu et al., 2014).

Significantly different synonymous codon usage exists between virus genomes and that of their host species, which reflects different codon usage bias (Zhou et al., 1999; Chen, 2013; Cristina et al., 2016; Xu et al., 2017). Evolution of viruses involves changes in virus nucleotide composition, which ultimately creates variations in the virus genome (Sablok et al., 2011; Zhang et al., 2011). Considering reliant on host's machinery for transcription, replication, protein synthesis and transmission of virus genomes, the interplay of codon usage among viruses and their hosts is expected to affect the overall viral survival (Shackelton and Holmes, 2004). Therefore, to measure all the codon usage in virus genome can improve recognizing of the regulation of viral genes expression (Butt et al., 2014).

The human papillomavirus (HPV) is a non-enveloped, epitheliotropic, double-stranded DNA virus with a genome of approximately 8000 bp, which infects stratified squamous epithelial cells. Gene expression of HPV in squamous epithelial cells is linked to the differentiation and function of the epithelial cells (Zheng and Baker, 2006; Zhao and Chen, 2011). The complete nucleotide sequence of the HPV genome was determined in 1985 (Seedorf et al., 1985). Although, HPV causes a certain cancers such as oral cancer (Lee et al., 2010), uterine cervical cancer (Galloway and McDougall, 1989), and anal cancer (Daling et al., 2004), its pathogenicity is unclear.

Although there are several studies on the HPV genome (Zhou et al., 1999; Zhao et al., 2003; Zhao and Chen, 2011), there is dearth of information on synonymous codon usage of HPV and factors that influence it. To obtain better and integrated understanding of synonymous codon usage of HPV, codon usage patterns of 39 HPV genomes were analyzed. This study will provide new insights into codon usage of HPV genome.

MATERIALS AND METHODS

Sequence collection

The complete sequences of the 39 HPV genomes downloaded from the National Center for Biotechnology Information (NCBI) website were used in this study (Table 1). The indices of codon preferences in the HPV genome were analyzed by the CodonW 1.4.2 program (Peden, 2005).

Relative synonymous codon usage

Relative synonymous codon usage (RSCU) values for all the codons of the 39 HPV genomes (excluding the codons for Met and Trp, each of which has only one codon triplet), were calculated to examine the feature of synonymous codon usage without

Table 1. Accession number and genome size of the human papillomavirus.

Type	Accession number	Genome size
0	NC_019023	7212
4	NC_001457	7353
5	NC_001531	7746
7	NC_001595	8027
9	NC_001596	7434
10	NC_001576	7919
16	NC_001526	7905
26	NC_001583	7855
32	NC_001586	7961
34	NC_001587	7723
41	NC_001354	7614
48	NC_001690	7100
49	NC_001591	7560
50	NC_001691	7184
53	NC_001593	7856
60	NC_001693	7313
63	NC_001458	7348
88	NC_010329	7326
90	NC_004104	8033
92	NC_004500	7461
96	NC_005134	7438
101	NC_008189	7259
103	NC_008188	7263
108	NC_012213	7149
109	NC_012485	7346
112	NC_012486	7227
116	NC_013035	7184
121	NC_014185	7342
126	NC_016157	7326
128	NC_014952	7259
129	NC_014953	7219
131	NC_014954	7182
132	NC_014955	7125
134	NC_014956	7309
135	NC_017993	7293
136	NC_017994	7319
137	NC_017995	7236
140	NC_017996	7341
144	NC_017997	7271

confounding force of amino acid composition of the different gene samples (Sharp and Li, 1986). The formula is as follows:

$$RSCU = (G_{ij} / \sum_j^{ni} G_{ij}) / ni$$

Where, G_{ij} is the observed number of the i th codon for the j th amino acid, which has ni types of synonymous codons. The codons with a RSCU value higher than 1.0 have a positive codon usage bias, while codons with a RSCU value lower than 1.0 have a relatively negative codon usage bias. Additionally, comparative analysis of

the RSCU values between HPV and humans (downloaded from <http://www.kazusa.or.jp/codon/>) was performed.

The principle component analysis

To investigate the major trends in codon usage variation of the 39 HPV genomes, the principle component analysis was performed. By creating a series of orthogonal axes, the major trends present in the dataset were analyzed in the multidimensional space. The codons were plotted on the first two axes due to these two axes and showed the highest fraction of data variance (Gu et al., 2004).

The index of GC3s

Excluding those encoding Met or Trp and the termination codons, the index of GC3s was calculated as the fraction of GC content at the third position of synonymous codon (Epstein et al., 2000).

The effective number of codon values

The effective number of codon (ENC) values is the best estimator of absolute synonymous codon usage bias (Comeron and Aguade, 1998) that was analyzed for the quantification of the codon usage bias of each open reading frame (Wright, 1990). The predicted ENC values were calculated as:

$$ENC = 2 + s + 29 / (s^2 + (1 - s^2))$$

Where, s denotes the value of percentage of GC at the third position of the synonymous codons (GC3s).

Indices of codon preference in the HPV genome

Indices for measuring chemical properties of amino acids

Hydrophobicity (GRAVY) and aromaticity (AROMO) of conceptually translated gene products may be factors that influence codon usage bias patterns (Peden, 1999).

The hydrophobicity index (Peden, 1999) is calculated as:

$$GRAVY = 1/N \sum_{i=1}^N ki$$

Where, N is the number of amino acids and ki is the hydrophobic index of the i th amino acid.

The aromaticity index (Peden, 1999) is calculated as:

$$AROMO = 1/N \sum_{i=1}^N vi$$

Where, vi is either 1 (for aromatic amino acids Phe, Tyr, and Trp) or 0 (for a non-aromatic amino acid), and N is the number of amino acids.

Codon adaptation index

The CAI is a measurement of the relative adaptiveness of codon usage of a gene with the codon usage of highly expressed genes. For each genome sequence G and some set of coding sequences S in G , codon bias is measured with respect to its synonymous codon usage. Given an amino-acid j , its synonymous codons might

have different frequencies in S ; if x_{ij} is the number of times that the codon i for the amino-acid j occurs in S , then one associates to i a weight w_{ij} relative to its sibling of maximal frequency y_j in S .

$$w_{ij} = x_{ij} / y_j$$

A codon with maximal frequency in S is called preferred among its sibling codons. To each gene g in G , Sharp and Li associated a value in $[0, 1]$, called CAI defined as:

$$CAI(g) = (\prod_{k=1}^L w_k)^{1/L}$$

Where, L is the number of codons in the gene, and w_k is the weight of the k th codon in the gene sequence. Genes with CAI value close to 1 are made by highly frequent codons (Sharp and Li, 1987).

Statistical analysis

Correlation analysis was calculated using Spearman's rank correlation method of the R package (R Development Core Team, 2011).

RESULTS

Relative synonymous codon usage of the HPV genome

Overall relative synonymous codon usage of the HPV genome

The overall RSCU values of all the codons in the 39 HPV genomes are summarized in Table 2. A significantly non-random usage of degenerate codons encoding 18 amino acids was found. The amino acids Arg, Leu, and Ser had six-type codon degeneracy. For Arg, AGA, CGA, and CGU codons (RSCU of 1.13-1.37) were more frequently used than other codons (RSCU of 0.63-0.93). Similarly, CUA, UUA, CUU, and UUA codons (RSCU of 1.22-1.30) for Leu and AGU, UCA, and UCU codons (RSCU of 1.21-1.73) for Ser were the preferred codons respectively. The amino acids Ala, Gly, Pro, Thr, and Val had four-type codon degeneracy. xyU or xyA codons (RSCU of 1.28-1.76) were more frequently used than xyG or xyC codons (RSCU of 0.27-0.98). The amino acids Asp, Asn, Cys, Phe, His, and Tyr had two-type codon degeneracy (xyU and xyC). xyU codons (RSCU of 1.42-1.71) were more frequently used than xyC codons (RSCU of 0.29-0.68). The amino acids Gln, Glu, and Lys also showed two-type codon degeneracy. xyA codons (RSCU of 1.16-1.40) were more frequently used than that of xyG codons (RSCU of 0.68-0.84).

Relationship between the codon usage patterns of HPV and the host

Comparison of the genomes of HPV and humans revealed that the codon usage pattern of the virus was

Table 2. Relative synonymous codon usage values of the human papillomavirus and humans with the exception of those encoding Met or Trp. The preferentially used codons for each amino acid are displayed in bold.

Amino acid	Codon	RSCU	
		HPV	HUMAN
Ala	GCU	1.47	1.08
	GCG	0.30	0.44
	GCC	0.58	1.60
	GCA	1.66	0.92
Cys	UGU	1.44	0.92
	UGC	0.56	1.08
Asp	GAU	1.32	0.52
	GAC	0.68	1.08
Glu	GAG	0.68	1.16
	GAA	1.32	0.84
Phe	UUU	1.71	0.92
	UUC	0.29	1.08
Gly	GGU	1.28	0.64
	GGG	0.71	1.00
	GGC	0.71	1.36
	GGA	1.30	1.00
His	CAC	0.58	1.16
	CAU	1.42	0.84
Ile	AUU	1.45	1.08
	AUA	1.28	0.51
	AUC	0.27	1.40
Lys	AAA	1.40	0.86
	AAG	0.60	1.14
Leu	CUA	1.22	0.42
	CUC	0.33	1.20
	CUG	1.15	2.40
	CUU	1.30	0.78
	UUA	1.29	0.48
	UUG	0.71	0.78
Asn	AAC	0.54	1.06
	AAU	1.46	0.94
Pro	CCA	1.42	1.12
	CCC	0.55	1.28
	CCU	1.73	1.16
	CCG	0.30	0.44
Gln	CAA	1.16	0.54
	CAG	0.84	1.46
Arg	AGA	1.37	1.26
	AGG	0.63	1.26
	CGA	1.30	0.66
	CGC	0.93	1.08
	CGG	0.64	1.20
	CGU	1.13	0.48
Ser	AGC	0.68	1.44
	AGU	1.32	0.90
	UCA	1.21	0.90
	UCC	0.75	1.32
	UCG	0.31	0.30

Table 2. Contd.

Thr	UCU	1.73	1.14
	ACC	0.65	1.44
	ACA	1.76	1.12
	ACG	0.27	0.44
Val	ACU	1.33	1.00
	GUU	1.36	0.72
	GUG	0.98	1.84
	GUC	0.32	0.96
Tyr	GUA	1.34	0.48
	UAC	0.45	1.12
	UAU	1.55	0.88

different from that of the host (Figure 1). There were only few similar synonymous codon usage patterns between HPV and humans; these similarities were found in Ala (GGU), Pro (CCA and CCU), Arg (AGA), and Ser (UCU) (Table 2).

Principle component analysis

Principal component analysis was performed for all the genes in the 39 HPV genomes. The analysis detected one major trend in the first axis and another major trend in the second axis. The plots of codons ending with A/U (Figure 2a) and G/C (Figure 2b) were scattered in different ways. Most of the codons ending with A/U were clustered around the origin (0, 0) while codons ending with G/C appeared on both sides of the first axis. The separation of these codons on the first axis is statistically significant by analysis of variance ($p < 0.05$). These results suggest that certain factors might influence codon usage, which results in the observed difference between the characteristics of the codon plots ending with A/U and G/C.

Relationship of ENC values with GC3s

The ENC values for the 39 HPV genomes varied from 25.93 to 61.0 with a mean of 46.98 and a standard deviation of 6.11. The ENC-GC3s plot showed that most of the ENC values were just below the expected curve (Figure 3). Only 2.2% of total genes had high codon bias (ENC < 35). About 28% (76 genes) of the total genes had high ENC values (ranging over 50), indicating that these genes had random codon usage in HPV.

Level of gene expression and codon bias

The level of gene expression of HPV was measured through codon adaptation index values, which varied from 0.108 to 0.268 with the mean of 0.180 and standard

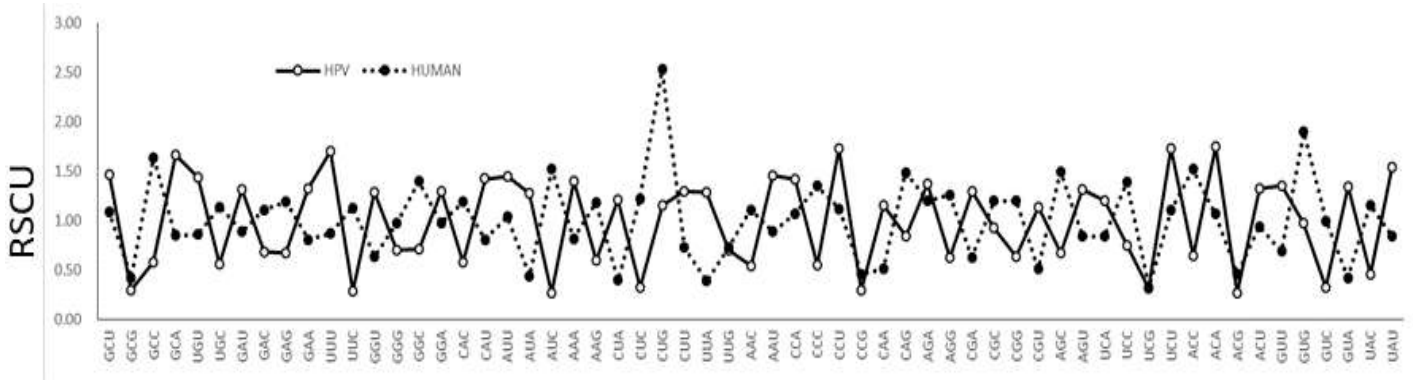


Figure 1. Comparison of the codon usage patterns between the HPV genome and the human genome.

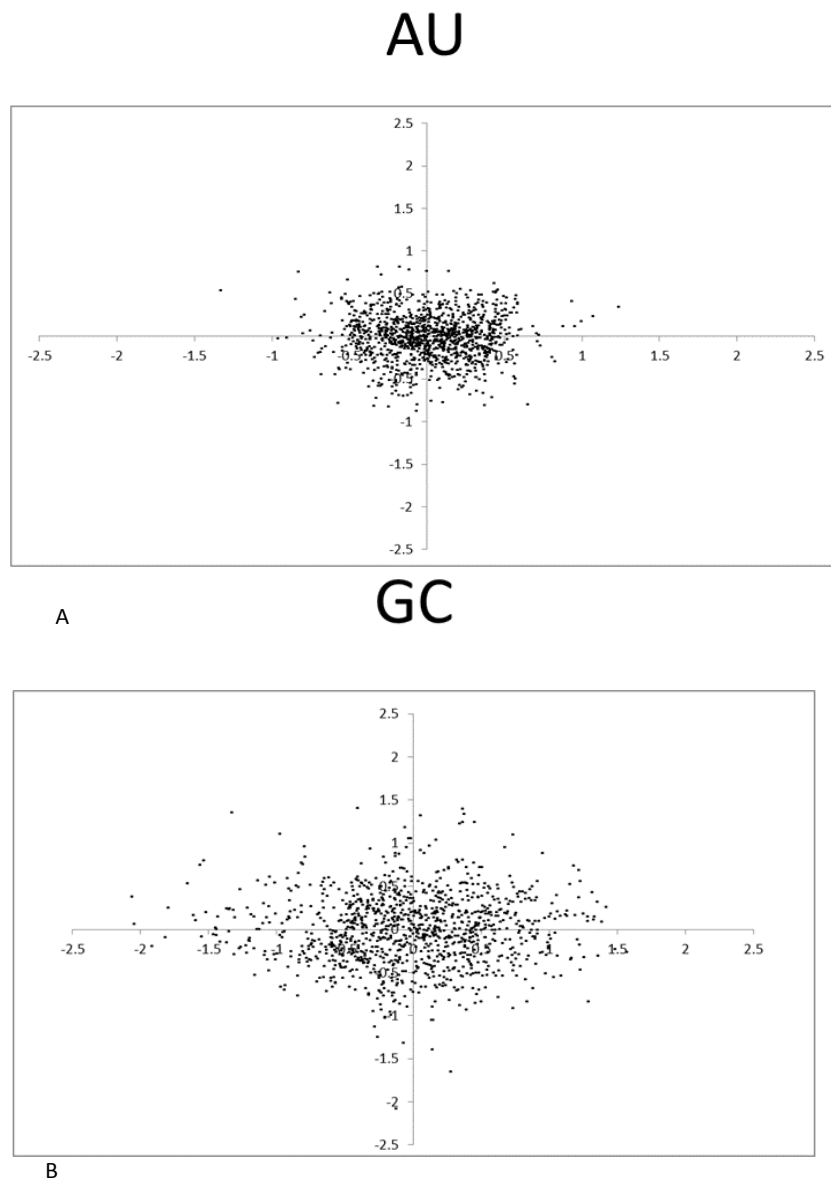


Figure 2. Distributions of codons ending with A/U (a) and with G/C (b) are shown along the first and second axes of the principle component analysis.

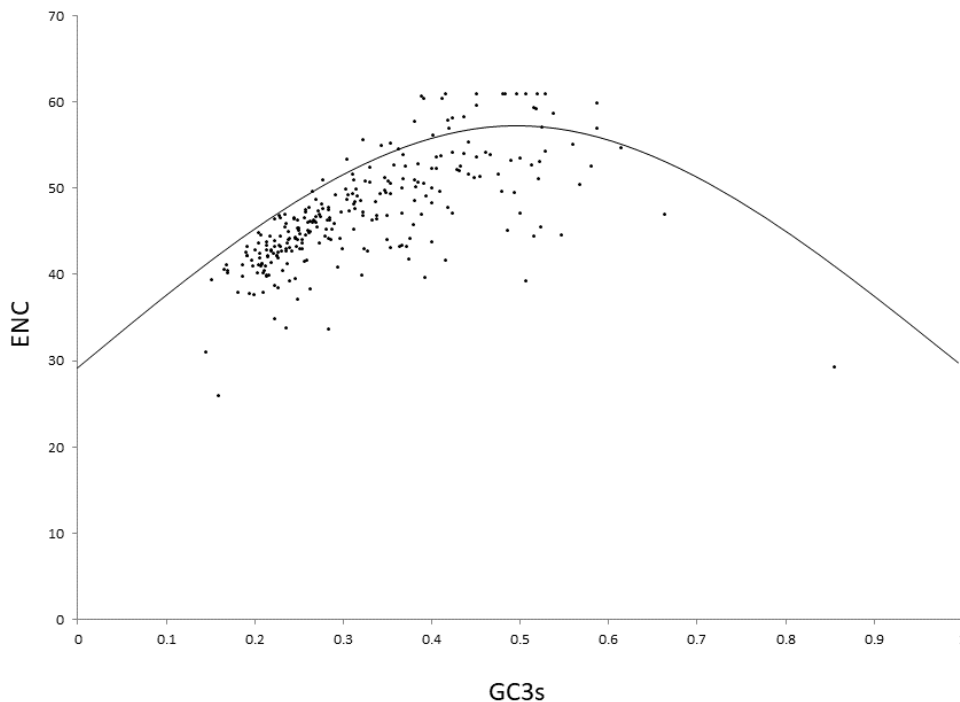


Figure 3. The relationship between the effective number of codons (ENC) and the GC content of the third codon position (GC3s) of the human papilloma virus. The solid curve represents the expected curve between ENC and GC3s under random codon usage. Black dots indicate all the genes.

deviation of 0.0256. A significant negative correlation was observed between ENC and CAI (Figure 4a) (Spearman, $r = -0.37991$, $p < 0.001$).

Correlation between ENC and both AROMO and GRAVY

We also investigated whether other factors could explain the codon usage bias seen in the HPV genome. Significant high negative correlation was observed in between ENC and both aromaticity (Figure 4b) (Spearman, $r = -0.71464$, $p < 0.001$) and GRAVY (Figure 4c) (Spearman, $r = -0.4391$, $p < 0.001$).

DISCUSSION

The term codon usage bias shows the unequal usage of synonymous codons for encoding amino acids which may differ significantly between genomes, genes, and within a single gene. That is the reason that codon usage bias has received much attention and various research about codon usage bias have been reported (Mazumder et al., 2014).

RSCU results showed that all preferred codons ended in A/U, which accounted for the majority of the nucleotide

composition in the HPV genome (Nasrullah et al., 2015). Principle component analysis showed that codons ended in A/U and G/C were statistically different. These results imply that nucleotide composition is a key factor in determining the preferred codon usage in HPV.

Investigation of synonymous codon usage have not only presented insight into the molecular evolution of genes, but also identified potential modulations of gene expression as a result of codon selection that influence efficiency (Heitzer et al., 2007; RoyChoudhury and Mukherjee, 2010). In this study, we observed that the RSCU of the HPV genome showed a complementary trend when compared to the RSCU of the human genome. This might be beneficial for virus survival and persistence by eliminating competition with the host translation machinery (Zhong et al., 2012). Moreover, this pattern might also be induced by a process of selective evolution of the virus. It has been proposed that differential synonymous codon usage of a virus and its host strongly influences both viral replication and gene expression (Zhao et al., 2003).

The ENC value, which is one of the best overall estimators of absolute synonymous codon usage bias, provides an intuitively meaningful measure of the extent of codon preference in a gene (Wright, 1990; Ma et al., 2014). Compared to the ENC values of other DNA viruses, ENC values of the HPV suggest that the low

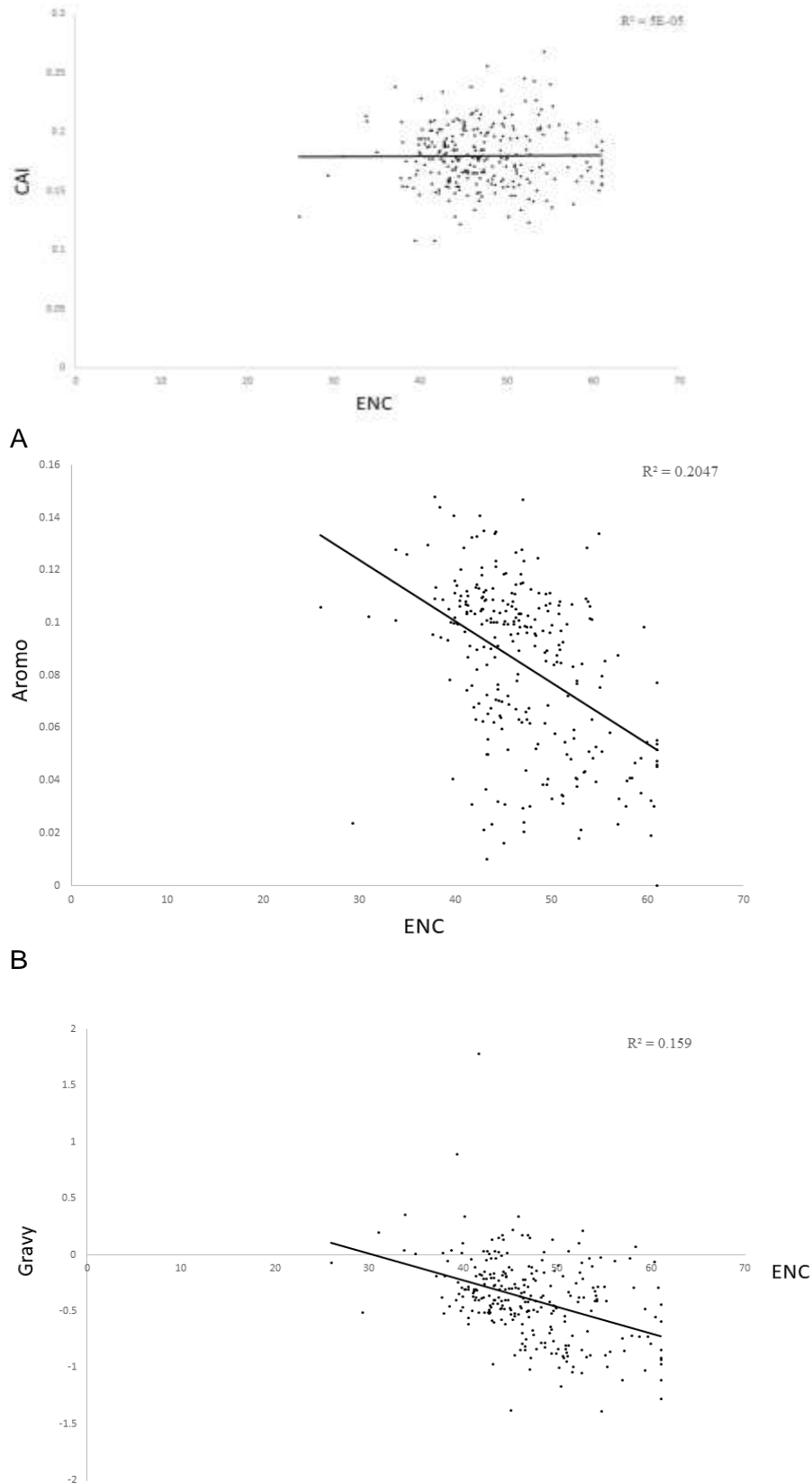


Figure 4. Correlation analysis plots showing (a) the relationship between ENC and CAI, (b) the relationship between ENC and AROMO, and (c) the relationship between ENC and GRAVY.

codon bias may result from an increase in its replication efficiency in order to adapt to the replication system of the host (Liu et al., 2012).

The ENC-GC3s plot is used to investigate patterns of synonymous codon usage visually (Wright, 1990; Comerón and Aguade, 1998; Gupta and Ghosh, 2001). The ENC-GC3s plot has been generally used to resolve whether codon usages of given genes are influenced by mutation alone (corresponding points would lie around the expected curve) or also by other components such as selection (corresponding points would depart away from, markedly below the expected curve) (Chen et al., 2014). The data points follow a curvilinear trend if the synonymous codon usage is only determined by the GC content on the third codon position (Gu et al., 2004). If the synonymous codon usage depends on compositional constraints, the data points occur on or just below the expected curve; however, if the synonymous codon usage is subject to natural selection, the points should be considerably below the expected curve (Wright, 1990; Ma et al., 2014). In this study, all the data points were immediately below the expected curve, suggesting that synonymous codon usage in these 39 HPV genomes was basically influenced by compositional constraints. The results from correlation analysis between percentage of ENC and GRAVY or AROMO indicate that these factors have ineffective on the synonymous codon usage of HPV.

The CAI is used to characterize translationally optimal codons that are used as a choice in highly expressed genes (Xia, 2007; RoyChoudhury and Mukherjee, 2010; Mazumder et al., 2014). This is expressed as a ratio whose value ranges from 0 to 1, where a higher value is likely to indicate stronger codon usage bias and a potential higher expression level. This information is useful for identifying highly expressed genes in any organism (Sharp and Li, 1987). In this study, the CAI value indicates that most of the HPV genes are not highly expressive in nature. Furthermore, significant negative correlation was detected between ENC and CAI. From these results, levels of gene expression have ineffective on the synonymous codon usage of HPV.

In summary, we hypothesize that the HPV codon usage may influence its pathogenic mechanism by striking a balance with the codon usage of the host and ensuring competition-free survival. It would also be useful for understanding the cell-host interaction and evolution of the HPV.

CONFLICT OF INTERESTS

The authors have not declared any conflict of interests.

ACKNOWLEDGEMENTS

Authors thank the anonymous reviewers for their

suggestions and help in data presentation.

REFERENCES

- Bishal AK, Mukherjee R, Chakraborty C (2013). Synonymous codon usage pattern analysis of Hepatitis D virus. *Virus Res.* 173(2):350-353.
- Butt AM, Nasrullah I, Tong Y (2014). Genome-wide analysis of codon usage and influencing factors in chikungunya viruses. *PLoS One* 9(3):e90905.
- Chen H, Sun S, Norenburg JL, Sundberg P (2014). Mutation and selection cause codon usage and bias in mitochondrial genomes of ribbon worms (Nemertea). *PLoS One* 9 (1):e85631.
- Chen Y (2013). A comparison of synonymous codon usage bias patterns in DNA and RNA virus genomes: quantifying the relative importance of mutational pressure and natural selection. *Biomed. Res. Int.* 406342.
- Comerón JM, Aguade M (1998). An evaluation of measures of synonymous codon usage bias. *J. Mol. Evol.* 47(3):268-274.
- Cristina J, Fajardo A, Soñora M, Moratorio G, Musto H (2016). A detailed comparative analysis of codon usage bias in Zika virus. *Virus Res.* 223:147-152.
- Daling JR, Madeleine MM, Johnson LG, Schwartz SM, Shera KA, Wurscher MA, Carter JJ, Porter PL, Galloway DA, McDougall JK (2004). Human papillomavirus, smoking, and sexual practices in the etiology of anal cancer. *Cancer* 101(2):270-280.
- Epstein RJ, Lin K, Tan TW (2000). A functional significance for codon third bases. *Gene* 245(2):291-298.
- Galloway DA, McDougall JK (1989). Human Papillomaviruses and Carcinomas. *Adv. Virus Res.* 37:125-171.
- Grantham R, Gautier C, Gouy M (1980). Codon frequencies in 119 individual genes confirm consistent choices of degenerate bases according to genome type. *Nucleic Acids Res.* 8(9):1893-1912.
- Gu W, Zhou T, Ma J, Sun X, Lu Z (2004). Analysis of synonymous codon usage in SARS Coronavirus and other viruses in the Nidovirales. *Virus Res.* 101(2):155-161.
- Gupta SK, Ghosh TC (2001). Gene expressivity is the main factor in dictating the codon usage variation among the genes in *Pseudomonas aeruginosa*. *Gene* 273(1):63-70.
- Heitzer M, Eckert A, Fuhrmann M, Griesbeck C (2007). Influence of Codon Bias on the Expression of Foreign Genes in Microalgae. *Adv. Exp. Med. Biol.* 616:46-53.
- Hu C, Chen J, Ye L, Chen R, Zhang L, Xue X (2014). Codon usage bias in human cytomegalovirus and its biological implication. *Gene* 545(1):5-14.
- Jenkins GM, Holmes EC (2003). The extent of codon usage bias in human RNA viruses and its evolutionary origin. *Virus Res.* 92(1):1-7.
- Karlin S, Mrázek J (1996). What drives codon choices in human genes? *J. Mol. Biol.* 262(4):459-472.
- Lee SY, Cho NH, Choi EC, Baek SJ, Kim WS, Shin DH, Kim SH (2010). Relevance of human papilloma virus (HPV) infection to carcinogenesis of oral tongue cancer. *Int. J. Oral Maxillofac. Surg.* 39(7):678-683.
- Liu X, Zhang Y, Fang Y, Wang Y (2012). Patterns and influencing factor of synonymous codon usage in porcine circovirus. *Virology* 439(1):68.
- Ma MR, Hui L, Wang ML, Tang Y, Chang YW, Jia QH, Wang XH, Yan W, Ha XQ, Ling H (2014). Overall codon usage pattern of enterovirus 71. *Genet. Mol. Res.* 13(1):336-343.
- Mazumder TH, Chakraborty S, Paul P (2014). A cross talk between codon usage bias in human oncogenes. *Bioinformatics* 10(5):973-2063.
- Nasrullah I, Butt AM, Tahir S, Idrees M, Tong Y (2015). Genomic analysis of codon usage shows influence of mutation pressure, natural selection, and host features on Marburg virus evolution. *BMC Evol. Biol.* 15:174.
- Peden FJ (1999). Analysis of codon usage [WWW Document]. PhD Thesis, Univ. Nottingham, UK. URL http://codonw.sourceforge.net/JohnPedenThesisPressOpt_water.pdf
- Peden FJ (2005). CodonW [WWW Document]. 2005. <http://codonw.sourceforge.net/>
- R Development Core Team (2011). R: A language and environment for

- statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.r-project.org>
- RoyChoudhury S, Mukherjee D (2010). A detailed comparative analysis on the overall codon usage pattern in herpesviruses. *Virus Res.* 148(1-2):31-43.
- Sablok G, Nayak KC, Vazquez F, Tatarinova TV (2011). Synonymous codon usage, GC (3), and evolutionary patterns across plastomes of three pooid model species: emerging grass genome models for monocots. *Mol. Biotechnol.* 49(2):116-128.
- Seedorf K, Krämmer G, Dürst M, Suhai S, Röwekamp WG (1985). Human papillomavirus type 16 DNA sequence. *Virology* 145(1):181-185.
- Shackelton LA, Holmes EC (2004). The evolution of large DNA viruses: combining genomic information of viruses and their hosts. *Trends Microbiol.* 12(10):458-465.
- Sharp PM, Li WH (1986). Codon usage in regulatory genes in *Escherichia coli* does not reflect selection for "rare" codons. *Nucleic Acids Res.* 14(19):7737-7749.
- Sharp PM, Li WH (1987). The codon adaptation index--a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res.* 15(3):1281-1295.
- Wright F (1990). The 'effective number of codons' used in a gene. *Gene* 87(1):23-29.
- Xia X (2007). An improved implementation of codon adaptation index. *Evol. Bioinform. Online* 3:53-58.
- Xu X, Fei D, Han H, Liu H, Zhang J, Zhou Y, Xu C, Wang H, Cao H, Zhang H (2017). Comparative characterization analysis of synonymous codon usage bias in classical swine fever virus. *Microb. Pathog.* 107:368-371.
- Zhang Y, Liu Y, Liu W, Zhou J, Chen H, Wang Y, Ma L, Ding Y, Zhang J (2011). Analysis of synonymous codon usage in hepatitis A virus. *Virol. J.* 8:174.
- Zhang Z, Dai W, Wang Y, Lu C, Fan H (2013). Analysis of synonymous codon usage patterns in torque teno sus virus 1 (TTSuV1). *Arch. Virol.* 158(1):145-154.
- Zhao KN, Chen J (2011). Codon usage roles in human papillomavirus. *Rev. Med. Virol.* 21(6):397-411.
- Zhao KN, Liu WJ, Frazer IH (2003). Codon usage bias and A+T content variation in human papillomavirus genomes. *Virus Res.* 98(2):95-104.
- Zheng ZM, Baker CC (2006). Papillomavirus genome structure, expression, and post-transcriptional regulation. *Front. Biosci.* 11:2286-2302.
- Zhong Q, Xu W, Wu Y, Xu H (2012). Patterns of synonymous codon usage on human metapneumovirus and its influencing factors. *J. Biomed. Biotechnol.* 460837.
- Zhou J, Liu WJ, Peng SW, Sun XY, Frazer I (1999). Papillomavirus capsid protein expression level depends on the match between codon usage and tRNA availability. *J. Virol.* 73(6):4972-4982.