GMR Genetics and Molecular Research
Online Journal
ISSN:1676-5680

Synonymous codon selection in the hepatitis B virus translation initiation region

M.-R. Ma, L. Hui, M.-L. Wang, Y. Tang, Y.-W. Chang, Q.-H. Jia, X.-P. Yang, X.-H. Wang and X.-Q. Ha

Key Laboratory of Stem Cells and Gene Drugs of Gansu Province, Experimental Center of Medicine, Lanzhou General Hospital, Lanzhou Military Area Command, Lanzhou, China

Corresponding author: L. Hui
E-mail: lrlmamingren@163.com

Genet. Mol. Res. 14 (3): 8955-8963 (2015)
Received July 21, 2014
Accepted March 23, 2015
Published August 7, 2015
DOI http://dx.doi.org/10.4238/2015.August.7.4

ABSTRACT. Hepatitis B virus (HBV) infection is a major health problem worldwide. This virus and its hosts are often fated to continual co-evolutionary interactions. Codon usage analysis has significance for studies of co-evolution between viruses, their hosts, and mRNA translation. Adaptation of the overall codon usage pattern of HBV to that of humans is estimated using the synonymous codon usage value (RSCU), and the synonymous codon usage biases for the translation initiation region (TIR) of HBV are analyzed by calculation of the usage fluctuation of each synonymous codon along the TIR (the first 50 codon sites of the whole coding sequence of HBV). With respect to synonymous codon usage, our results demonstrated that HBV had no significant tendency to select over-represented codons, but had a significant tendency to select certain under-represented codons in the viral genome. Within the three common HBV hosts, 14 of 59 codons had a similar usage pattern, suggesting that mutation pressure from this DNA virus played an important role in the formation of virus synonymous codon usage. In addition, there was no obvious trend for the codons with relatively low energy to be highly selected in the

 ©FUNPEC-RP www.funpecrp.com.br

TIR of HBV, suggesting that the synonymous codon usage patterns for the TIR might not be affected by the nucleotide sequence secondary structure; however, synonymous codon usage in the TIR of HBV was influenced by the overall codon usage patterns of the hosts to some degree. Our results suggest that mutation pressure from HBV plays an important role in the formation of synonymous codon usage of the viral genome, while translation selection from the hosts contributes to virus translational fine-tuning.

Key words: Hepatitis B virus; Synonymous codon usage value; Translation initiation region; Mutation pressure; Translation selection

INTRODUCTION

Hepatitis B virus (HBV) infection is an important worldwide public health problem. Some adaptive advantages (i.e., mutation pressure from the virus itself and translation selection from the hosts) favor the highly efficient dissemination of the virus by different modes of transmission resulting in a widespread distribution of HBV. HBV is the prototype member of the family *Hepadnaviridae*, and has a compact and circular DNA genome of about 3.2 kb in length, with four overlapping open reading frames, including a large S region (PreS/S), and PreC/C, X, and P regions (Zhang et al., 2009). Overlapping genetic regions are helpful for evolutionary studies on point mutations because the incidence of recombination is rare and any point mutation could influence the genetic characteristics of two genes that overlap the mutation. Therefore, the evolution of HBV is expected to be interactional and constrained by the overlap of genes (Mizokami et al., 1997). The evolution of one protein encoded in the overlapping reading frame might be constrained by negative selection, while the other might have evolved more rapidly, and the overlapping genes might be subject to different selective forces (Jordan et al., 2000; Pavesi, 2006). It has been observed that the various genetic diversities in the nucleotide composition of the HBV-coding sequence are selective rather than random, because natural selection from the host is responsible for selection of various strains shaped by mutation. Much attention has been recently paid to the question of preferential codon usage. Genes with high expression preferentially select a subset of all codons, whereas genes with low expression often utilize the whole complement. Ribosomal initiation on an mRNA is normally the rate-limiting step in the process of translation, and blockage of the initiation site can be avoided if the codons closest to this site allow fast translation by the ribosome. Different selective forces may act on the selection of synonymous codons in the initiation region than elsewhere on a given mRNA (Liljenström and von Heijne, 1987). In previous reports, translation selection and compositional constraints under mutational pressure were suggested to be the major factors accounting for codon usage variation among genomes in microorganisms (dos Reis and Wernisch, 2009). It has been reported that the codon usage patterns can control ribosome scanning speed along the coding sequence, and that the synonymous codon usage bias plays a role in influencing the translation initiation efficiency at the 5'-terminus of the target coding sequence (Tuller et al., 2010). Here, we analyzed the synonymous codon usage bias in the first 50 codons (translation initiation region; TIR) of the N-terminus of the complete coding sequence of HBV to evaluate the effect of usage of synonymous codons on the translation initiation of the polyprotein of this virus.

 ©FUNPEC-RP www.funpecrp.com.br

MATERIAL AND METHODS

Sequences and databases

The 59 complete RNA sequences of HBV were downloaded from the National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/Genbank/) and detailed information about the viruses were recorded. Sequence accession numbers were as follows: AF405706, X04615, AY741795, U87747, AY123041, AF282917, AY233287, AY233280, AY233283, AY233291, DQ448620, AY373432, AY373430, DQ448620, DQ448621, DQ448622, DQ448625, AY233273, DQ448628, DQ448627, AY233275, AY233278, AY233277, AY233274, AY233276, AY233274, DQ448623, AY233281, AY233279, AY233282, AY233290, AY233288, AY233289, AY233285, AY233284, AY233286, AF282918, AF068756, M57663, AF100308, AB033554, AY741797, U87746, AY741798, AY741796, AY741794, AF100309, GQ872210, AY23329, AY233294, AY233293, AY233296, GQ161799, U95551, GQ161805, AY796031, AY796032, GQ161818, and AY796030.

In addition, in order to evaluate the adaptation of the overall codon usage pattern of HBV to its natural hosts (*Homo sapiens*, *Gorilla gorilla*, and *Pan troglodytes*), the codon usage frequencies of the three hosts were obtained from the codon usage database (Nakamura et al., 2000).

Comparison of codon usage overrepresentation between HBV and its natural hosts

To investigate the codon usage pattern without the confounding influence of amino acid composition between different sequences, the relative synonymous codon usage (RSCU) values for synonymous codons in this study was calculated according to the equation given in Sharp et al. (1986). Five codons, including three stop codons (UAA, UAG, and UGA), AUG for Met, and UGG for Try, were not introduced into the RSCU calculation. For codon usage frequencies of the human genome, the RSCU values were calculated for the 59 synonymous codons by the approach mentioned above. In order to identify the usage bias of each synonymous codon, based on the standard of codon usage bias defined previously (Wong et al., 2010), a synonymous codon with an RSCU value < 0.6 or > 1.6 was regarded as biased in this study. For comparison of the synonymous codon usage pattern between HBV and humans, if both the RSCU values for a codon of HBV and those of the same codon for humans, at the same time, were >1.6 or <0.6, the codon usage was thought to represent a similar pattern. Here, a group of codons with RSCU values ranging from 0.6 to 1.6 needed to be divided deeply, namely, when both RSCU values of HBV and of human for the same codon, at the same time, ranged from 0.6 to 1.0 or from 1.0 to 1.6, the usage pattern of the specific codon between the virus and the host was thought to be similar.

Calculation of the synonymous codon usage in the TIR of HBV

To analyze any discrepancy in the synonymous codon usage preference between the specific TIR with six specific lengths (the first 10 codons, the first 20 codons, the first 30 codons, the first 40 codons, and the first 50 codons) and the whole coding sequence of HBV, we depended on a simple method based on a previous report (Zhou et al., 2013b).

$$R = \ln\left(\frac{f_n / F_n}{f / F}\right)$$

where f_n is the sum of a specific synonymous codon in the size that ranged from the initiation codon (AUG) to the n^{th} codon; F_n is the sum of the corresponding amino acids over the given region; f is the sum of this synonymous codon over the whole coding sequence; and F is the sum of the corresponding amino acid over the whole coding sequence.

RESULTS

Similarity of overall codon usage between HBV and its three hosts

Eight codons (ATA for Ile, TCG for Ser, CCC for Pro, ACC for Thr, GCC for Ala, GGT for Gly, and CGT and CGG for Arg) demonstrated under-representation, and one codon (TCT for Ser) demonstrated over-representation (Table 1) in the synonymous codon usage pattern of HBV, which suggested that the translation selection from the natural hosts influenced the synonymous codon usage pattern of HBV. Examination of the synonymous codon usage patterns between HBV and its three hosts demonstrated that 14 of 59 codons were similarly used. Among the codons with different usage patterns between HBV and the three primates, there was no codon with an RSCU value < 0.6 for HBV and > 1.6 for the primates, or with an RSCU value > 1.6 for HBV and < 0.6 for the primates at the same time (Table 1). These results might suggest that in the procession of interactions between HBV cycling and anti-viral host response, HBV, in spite of maintaining various genotypes and subgenotypes, has a strong tendency to adapt to the environment of the host cell to some degree.

Bias of synonymous codon usage in the TIR

Examination of the synonymous codon usage patterns of HBV in this study demonstrated that the synonymous codons fail to be selected in equal frequency in the TIR; some members in the same synonymous family were selected at low frequency, whereas others were highly utilized. For example, for the three amino acids Phe, Tyr, and His, CAC and TAT are generally selected for His and Tyr, respectively, in the target region, compared with the usage degree of other synonymous members for the same synonymous family. Compared with the usage of synonymous codons (CAC and TAT) in the first 30 codons site in the TIR, the synonymous codons (TAC and CAT) were not selected (**Figure S1**). For Gln, Asn, and Lys, the synonymous codons for Gln, and the codons AAA and AAC exist for Lys and Asn, respectively, in the TIR, whereas AAT and AAG for Asn and Lys, respectively, do not exist in the first 30 codons (**Figure S2**). For Asp, Glu, and Cys, the synonymous codons for Asp and Glu are selected in the TIR, whereas the synonymous codons for Lys are not selected by the first 40 codon sites of the target region (**Figure S3**). Compared with the usage of synonymous codons ATC and ATA for Ile, the usage of the synonymous codon ATT is generally higher in the TIR (**Figure S4**). The synonymous codon GCA for Ala has a stronger tendency to exist in the TIR than do the others (**Figure S5**). Compared with the usage in the TIR of the synonymous codons GGC and GGG for Gly, which are high energy codons, the synonymous codon GGA is selected more often in the first 10 codon sites and the GGT is more often selected in the 20th to the 50th codon sites (**Figure S6**). For Val, GTG has the strongest tendency to exist in the first 20 codons (R > 1.0) whereas GTT fails to be selected in the first 40 codons of the TIR (**Figure S7**). The synonymous codons CCC and CCT for Pro have a stronger tendency to exist in the TIR than do CCG and CCA (**Figure S8**).

 ©FUNPEC-RP www.funpecrp.com.br

Table 1. Relationship of the synonymous codon usage pattern between HBV and the hosts.

Codon/Amino acid	HBV	Human	*Gorilla gorilla*	Pan troglodytes
TTT(F)	1.06	0.87	0.85	0.77
TTC(F)	0.94	1.13	1.15	1.23
TTA(L)	0.67	0.39	0.42	0.35
TTG(L)	1.08	0.73	0.74	0.64
CTT(L)	1.11	0.73	0.74	0.70
CTC(L)	1.22	1.22	1.25	1.35
CTA(L)	0.85	0.40	0.44	0.41
CTG(L)	1.06	2.53	2.42	2.56
ATT(I)	1.27	1.04	1.04	0.95
ATC(I)	1.26	1.52	1.41	1.57
ATA(I)	0.48	0.44	0.55	0.48
GTT(V)	1.27	0.69	0.76	0.62
GTC(V)	0.91	1.00	0.96	0.99
GTA(V)	0.65	0.42	0.44	0.36
GTG(V)	1.17	1.90	1.84	2.03
TCT(S)	1.69	1.11	1.28	1.22
TCC(S)	1.48	1.39	1.34	1.44
TCA(S)	1.28	0.84	0.89	0.80
TCG(S)	0.58	0.33	0.27	0.31
AGT(S)	1.48	0.84	0.84	0.77
AGC(S)	1.01	1.50	1.38	1.45
CCT(P)	0.99	1.12	1.14	1.09
CCC(P)	0.51	1.35	1.35	1.42
CCA(P)	1.37	1.07	1.04	0.97
CCG(P)	1.38	0.46	0.47	0.52
ACT(T)	0.89	0.94	0.97	0.85
ACC(T)	0.37	1.52	1.50	1.70
ACA(T)	1.32	1.07	1.10	1.01
ACG(T)	1.24	0.46	0.42	0.44
GCT(A)	0.99	1.09	1.11	1.09
GCC(A)	0.45	1.64	1.58	1.57
GCA(A)	1.27	0.85	0.86	0.78
GCG(A)	0.73	0.42	0.45	0.56
TAT(Y)	1.05	0.84	0.90	0.77
TAC(Y)	0.95	1.16	1.10	1.23
CAT(H)	1.21	0.81	0.85	0.80
CAC(H)	0.79	1.19	1.15	1.20
CAA(Q)	1.08	0.51	0.56	0.46
CAG(Q)	0.92	1.49	1.44	1.54
AAT(N)	1.36	0.89	0.93	0.85
AAC(N)	0.64	1.11	1.07	1.15
AAA(K)	0.73	0.82	0.89	0.81
AAG(K)	1.27	1.18	1.11	1.19
GAT(D)	1.04	0.89	0.86	0.79
GAC(D)	0.96	1.11	1.14	1.21
GAA(E)	1.23	0.81	0.77	0.68
GAG(E)	0.77	1.19	1.23	1.32
TGT(C)	0.80	0.86	0.91	0.85
TGC(C)	1.06	1.14	1.09	1.15
CGT(R)	0.48	0.51	0.49	0.43
CGC(R)	0.78	1.20	1.11	1.21
CGA(R)	0.61	0.63	0.58	0.61
CGG(R)	0.37	1.20	0.99	1.16
AGA(R)	1.49	1.20	1.52	1.29
AGG(R)	1.39	1.26	1.31	1.30
GGT(G)	0.60	0.64	0.61	0.55
GGC(G)	0.81	1.40	1.26	1.40
GGA(G)	1.36	0.98	0.99	0.90
GGG(G)	1.22	0.98	1.13	1.15

 ©FUNPEC-RP www.funpecrp.com.br

For Thr, ACA has the strongest tendency (R > 1.0) to exist in the first 10 codons of the TIR, and ACG is not selected in the first 40 codons (**Figure S9**). For Arg, the synonymous codon CGG has the strongest tendency (R > 1.0) to exist in the TIR, while AGG failed to be selected in the target region (Figure 1). For Leu, it was noted that the synonymous codon CTA was selected in the first 10 codons, while the rest failed to be selected in the corresponding region (Figure 2). For Ser, the synonymous codon TCC was generally frequently selected in the TIR, while TCC was generally selected at low frequency in this region (Figure 3). These results suggested that the discrepancy of synonymous codon usage in the TIR might be influenced by translation selection from the hosts to some degree and be an adaptive result in order to ensure successful existence in the hosts.

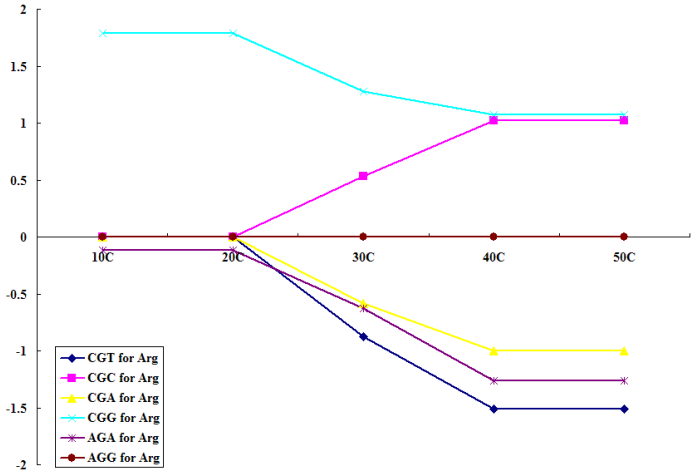

Figure 1. Usage bias of synonymous codons for Arg in the different lengths (the first 10 codons, the first 20 codons, the first 30 codons, the first 40 codons, and first 50 codons) of the translation initiation region in HBV ORF.

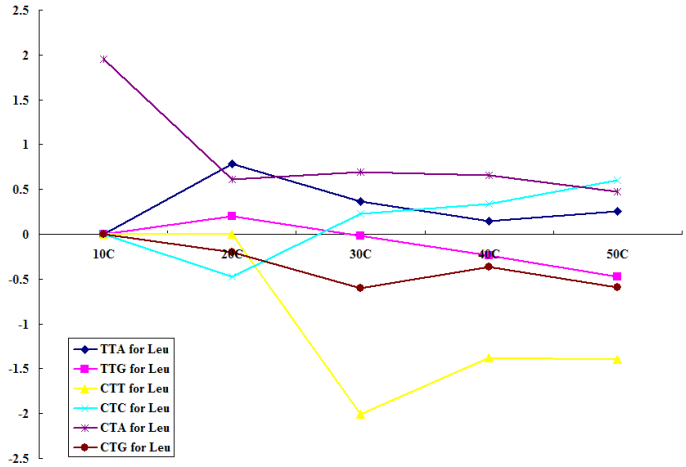

Figure 2. Usage bias of synonymous codons for Leu in the different lengths (the first 10 codons, the first 20 codons, the first 30 codons, the first 40 codons, and first 50 codons) of the translation initiation region in HBV ORF.

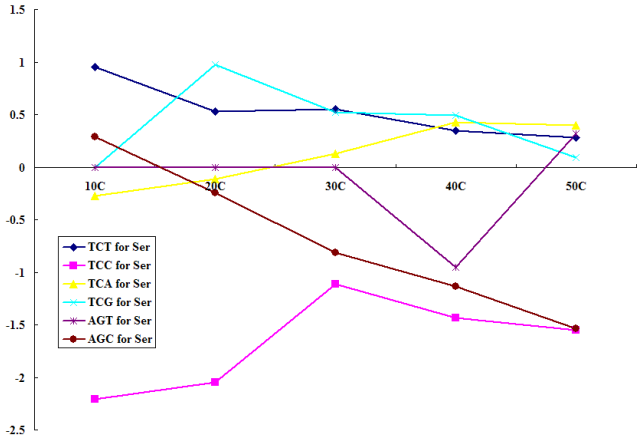

Figure 3. Usage bias of synonymous codons for Ser in the different lengths (the first 10 codons, the first 20 codons, the first 30 codons, the first 40 codons, and first 50 codons) of the translation initiation region in HBV ORF.

Adaptation of codon usage in the TIR to the overall codon usages of both HBV and its natural hosts

It is of interest that the synonymous codons CGG (Arg), GGT (Gly), CCC (Pro), and TCG (Ser) were selected in relatively low frequencies by the HBV ORF (Table 1), whereas these were highly selected by the viral TIR (Figures 1 and 3; **Figures S6** and **S8**). It was noted that the synonymous codons TCG (Ser) and GGT (Gly) were selected at low frequency by both HBV and its hosts (Table 1), suggesting that the translation selection from the hosts likely affected the codon usage bias in the viral TIR. In addition, the synonymous codon CTC (Ser) was frequently selected in the TIR of HBV (Figure 3), due to the over-represented usage of this codon (RSCU > 1.6) in the HBV ORF (Table 1). Although the synonymous codons AGG (Arg) and ATC (Ile) were selected in relatively high frequency by both the HBV ORF and the three natural hosts (Table 1), these codons failed to be selected or were selected at low frequency in the viral TIR (Figure 1 and **Figure S4**). These results also suggested that translation selection from the hosts might have influenced the synonymous codon usage bias in the viral TIR.

DISCUSSION

Viruses are generally ubiquitous cellular parasites and have a strong ability to replicate and evolve rapidly under the selection pressure deriving from their natural hosts (Bahir et al., 2009). This genetic characteristic may assist viruses to cater to the cellular environments and replicate in hosts (Zanotto et al., 1996; Lobo et al., 2009; Wong et al., 2010; Zhou et al., 2013a,b). In this study, it was found that the synonymous codon usage pattern of HBV had a low degree of similarity to that of its hosts (Table 1). This could be explained by the fact that as part of the viral replication process, the error-prone polymerase reverse transcriptase results in many genotypes and subgenotypes. Viral production is dependent upon maximization of the translation speed of viral protein synthesis, which necessitates impairment of the immune response inside virus-infected cells to prevent the large number of non-preferred codons, highly selected in viral genes, leading to low yield of viral proteins (Dupas et al., 2003; Sanchez et al.,

2003). HBV genotypes and subgenotypes have been increasingly associated with differences in clinical and virological features, such as severity of liver disease and response to antiviral therapies. Apart from the adaptation of overall codons usage of exogenous genes to their hosts, the initial rate of ORF elongation can play important roles in determining eventual levels of production. In this study, the usage degrees of 59 synonymous codons in the TIR of HBV were analyzed (Figures 1-3 and **Figure S1**-**S9**), and it was found that the usage patterns of certain synonymous codons were not similar to the overall usage patterns of the same codons in HBV, but were similar to the overall usage pattern of the same codons in the hosts. During the life cycle of viruses, production of viral proteins is often influenced by their host cell environments and strategies of translation initiation. It has been accepted that the bias of synonymous codon usage for the translation initiation region of genes is important for regulating the translation initiation efficiency (Eyre-Walker and Bulmer, 1993; Chen and Inouye, 1994; Stenstrom et al., 2001a,b). For HBV, certain synonymous codons are more highly selected in the TIR than in the whole coding sequence, whereas other synonymous codons are used to a lesser degree in the TIR than in the whole coding sequence. These codons with high or low usage degree might influence the translation initiation efficiency of HBV, and this phenomenon might suggest that other selection factors can act on the synonymous codon usage bias for the TIR in addition to mutation pressure. It has been reported that optimization of the first 5-17 codons of the human chorionic gonadotropin hormone gene contribute to 4- to 5-fold expression levels (Vervoort et al., 2000). In addition, it has been shown that the clustering of low-usage codons in the TIR can impair the rate of ribosomal scanning and thus can regulate the expression of genes (Zhang et al., 1994), suggesting that the synonymous codon usage bias for the TIR is important for the regulation of genes. Considering that the usage degree of some synonymous codons for the TIR of HBV is similar to the overall usage degree of the same codons of its hosts, these codons therefore likely reduce the probability of abortive translation initiation of the HBV genome by means of adaptation of codon usage pattern of this region to the host environment. The adaptation of codon usage patterns of exogenous genes to the overall codon usage pattern of the hosts is a readily available mechanism to produce functional proteins effectively.

ACKNOWLEDGMENTS

We would like to thank Mrs Xiao-xia Ma from Northwest University for Nationalities for her kind discussion and suggestions.

Supplementary material

REFERENCES

Bahir I, Fromer M, Prat Y and Linial M (2009). Viral adaptation to host: a proteome-based analysis of codon usage and amino acid preferences. *Mol. Syst. Biol.* 5: 311.

Chen GT and Inouye M (1994). Role of the AGA/AGG codons, the rarest codons in global gene expression in *Escherichia coli*. *Genes Dev.* 8: 2641-2652.

dos Reis M and Wernisch L (2009). Estimating translational selection in eukaryotic genomes. *Mol. Biol. Evol.* 26: 451-461.

Dupas S, Turnbull MW and Webb BA (2003). Diversifying selection in a parasitoid's symbiotic virus among genes involved in inhibiting host immunity. *Immunogenetics* 55: 351-361.

Eyre-Walker A and Bulmer M (1993). Reduced synonymous substitution rate at the start of enterobacterial genes. *Nucl. Acids Res.* 21: 4599-4603.

Jordan IK, Sutter BAT and McClure MA (2000). Molecular evolution of the Paramyxoviridae and Rhabdoviridae multiple-protein-encoding P gene. *Mol. Biol. Evol.* 17: 75-86.

Liljenström H and von Heijne G (1987). Translation rate modification by preferential codon usage: intragenic position effects. *J. Theor. Biol.* 124: 43-55.

Lobo FP, Mota BE, Pena SD, Azevedo V, et al. (2009). Virus-host coevolution: common patterns of nucleotide motif usage in Flaviviridae and their hosts. *PLoS One* 4: e6282.

Mizokami M, Orito E, Ohba K, Ikeo K, et al. (1997). Constrained evolution with respect to gene overlap of hepatitis B virus. *J. Mol. Evol.* 44 (Suppl 1): S83-S90.

Nakamura Y, Gojobori T and Ikemura T (2000). Codon usage tabulated from international DNA sequence databases: status for the year 2000. *Nucl. Acids Res.* 28: 292.

Pavesi A (2006). Origin and evolution of overlapping genes in the family *Microviridae*. *J. Gen. Virol.* 87: 1013-1017.

Sanchez G, Bosch A and Pinto RM (2003). Genome variability and capsid structural constraints of hepatitis a virus. *J. Virol.*77: 452-459.

Sharp PM, Tuohy TM and Mosurski KR (1986). Codon usage in yeast: cluster analysis clearly differentiates highly and lowly expressed genes. *Nucl. Acids Res.* 14: 5125-5143.

Stenstrom CM, Holmgren E and Isaksson LA (2001a). Cooperative effects by the initiation codon and its flanking regions on translation initiation. *Gene* 273: 259-265.

Stenstrom CM, Jin H, Major LL, Tate WP, et al. (2001b). Codon bias at the 3'-side of the initiation codon is correlated with translation initiation efficiency in *Escherichia coli. Gene* 263: 273-284.

Tuller T, Carmi A, Vestsigian K, Navon S, et al. (2010). An evolutionarily conserved mechanism for controlling the efficiency of protein translation. *Cell* 141: 344-354.

Vervoort EB, van Ravestein A, van Peij NN, Heikoop JC, et al. (2000). Optimizing heterologous expression in dictyostelium: importance of 5' codon adaptation. *Nucl. Acids Res.* 28: 2069-2074.

Wong EH, Smith DK, Rabadan R, Peiris M, et al. (2010). Codon usage bias and the evolution of influenza A viruses. Codon usage biases of influenza virus. *BMC Evol. Biol.* 10: 253.

Zanotto PM, Gould EA, Gao GF, Harvey PH, et al. (1996). Population dynamics of flaviviruses revealed by molecular phylogenies. *Proc. Natl. Acad. Sci. U. S. A.* 93: 548-553.

Zhang D, Chen J, Deng L, Mao Q, et al. (2009). Evolutionary selection associated with the multi-function of overlapping genes in the hepatitis B virus. *Infect. Genet. Evol.* 10: 84-88.

Zhang S, Goldman E and Zubay G (1994). Clustering of low usage codons and ribosome movement. *J. Theor. Biol.* 170: 339-354.

Zhou JH, Gao ZL, Zhang J, Ding YZ, et al. (2013a). The analysis of codon bias of foot-and-mouth disease virus and the adaptation of this virus to the hosts. *Infect. Genet. Evol.*14: 105-110.

Zhou JH, Su JH, Chen HT, Zhang J, et al. (2013b). Clustering of low usage codons in the translation initiation region of hepatitis C virus. *Infect. Genet. Evol.* 18: 8-12 .