



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



Short communication

The characteristics of the synonymous codon usage in enterovirus 71 virus and the effects of host on the virus in codon usage pattern

Yong-sheng Liu^{a,1}, Jian-hua Zhou^{a,1}, Hao-tai Chen^a, Li-na Ma^a, Zygmunt Pejsak^b, Yao-zhong Ding^a, Jie Zhang^{a,*}^a State Key Laboratory of Veterinary Etiological Biology, Lanzhou Veterinary Research Institute, Chinese Academy of Agricultural Sciences, Lanzhou 730046, Gansu, PR China^b Department of Swine Diseases, National Veterinary Research Institute, 57 Partyzantow, 24-100 Putawy, Poland

ARTICLE INFO

Article history:

Received 27 October 2010

Received in revised form 23 February 2011

Accepted 24 February 2011

Available online 5 March 2011

Keywords:

Enterovirus 71

The relative synonymous codon usage value

Codon usage pattern

Effective number of codons

Mutation pressure

Natural selection

ABSTRACT

To give a new perspective on the evolutionary characteristics shaping the genetic diversity of enterovirus 71 (EV71) and the effects of natural selection from its host on the codon usage pattern of the virus, the relative synonymous codon usage (RSCU) values, codon usage bias (CUB) values, effective number of codons (ENCs) values and nucleotide contents were calculated to implement a comparative analysis to evaluate the dynamics of the virus evolution. The characteristics of the synonymous codon usage patterns and nucleotide contents of EV71 and the comparison between ENC values for the whole coding sequence of EV71 and that of coding sequences for viral proteins of EV71 all indicate that the interaction between mutation pressure from virus and natural selection from host exists in the processes of evolution of EV71. The synonymous codon usage pattern of EV71 is a mixture of coincidence and antagonism to that of host cell. In addition, the genetic diversity of EV71 strains and the preferential selection of some synonymous codons in EV71 strains based on the different epidemic areas were observed, suggesting that geographic and social factors may play roles in influencing the evolution of this virus.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

Hand-foot-and-mouth disease (HFMD) is a general illness in children which usually is caused by some human enteroviruses (Cherry, 1992). There were many reports which indicated some pandemics of HFMD were associated with EV71 infection in the Asia-Pacific area (AbuBakar et al., 1999; Chumakov et al., 1979; Fujimoto et al., 2002; Ho et al., 1999; Lin et al., 2005; Liu et al., 2000; Zheng et al., 1995). EV71 belongs to members of the *Enterovirus* genus of the *Picornaviridae* family and is a positive-strand RNA virus with a genome size of about 7500 bp. The two non-translated regions (5'-NTR and 3'-NTR) flank the single open reading frame (ORF) of EV71 virus genome. The coding sequence encodes one polyprotein that is cleaved by viral proteases to generate 11 proteins, namely VP4, VP2, VP3, VP1, 2A, 2B, 2C, 3A, 3B, 3C and 3D. The structural proteins VP1-3 are exposed on EV71 surface. The VP1 gene contained the major antigenic sites and genetic diversity associated with serotypes (Oberste et al., 1999a,b). Non-structural proteins were involved in polyprotein

processing, RNA replication and the shut-down of host cell protein synthesis. In addition, recombinations are well known to result in a genetic diversity and evolution of enteroviruses (Chan and AbuBakar, 2004; Chen et al., 2010; Yoke-Fun and AbuBakar, 2006). Due to various genetic diversities of EV71, the effect of the vaccine is limited to prevent children from EV71. This situation has made researchers aware of the importance of analysis of EV71 genetic diversity (Bible et al., 2008; Cardoso et al., 2003; Herrero et al., 2003; Huang et al., 2008; Lewis-Rogers et al., 2009; McMinn, 2002; Sanders et al., 2006). It is noticed that nucleotide composition comprising of EV71 coding sequence with various genetic diversities is selective rather than random, because the natural selection from host is responsible to select various strains shaped by mutation. In previous reports, translation selection and compositional constraints under the mutational pressure are thought to be the major factors accounting for codon usage variation among genomes in microorganisms (Gu et al., 2004; Karlin and Mrázek, 1996; Lesnik et al., 2000; Liu et al., 2010; Zhou et al., 2005, 2006, 2010). In some RNA viruses, compared with natural selection, mutation pressure plays a more important role in synonymous codon usage pattern (Jenkins and Holmes, 2003; Levin and Whittome, 2000). Although it is known that compositional constraints and translation selection are the more generally accepted mechanisms accounting for codon usage bias (Coleman

* Corresponding author. Tel.: +86 931 8342771; fax: +86 931 8340977.

E-mail address: liuyongshengvip8@163.com (Y.-s. Liu).¹ These authors contributed equally to this work.

et al., 2008; Karlin et al., 1990; Zhi et al., 2010; Zhou et al., 1999), other selection forces have also been proposed such as fine-tuning translation kinetics selection as well as escape of cellular antiviral responses (Aragones et al., 2008, 2010; Karlin et al., 1994; Sugiyama et al., 2005). Thus, the codon usage pattern may be important to disclose the molecular mechanism and evolutionary process of EV71 avoiding host cell response. To our knowledge, it is the first study that the synonymous codon usage pattern and evolutionary dynamics of EV71 were systemically analyzed and the relationship between codon usage pattern of EV71 and that of its host was also analyzed.

2. Materials and methods

2.1. Sequence data

The 74 complete RNA sequences of EV71 were downloaded from the National Center for Biotechnology Information (NCBI) (<http://www.ncbi.nlm.nih.gov/Genbank/>) and detailed information about the viruses were listed in Table S1. Each general nucleotide composition (U%, A%, C% and G%) and each nucleotide composition in the third site of codon (U₃%, A₃%, C₃% and G₃%) in EV71 coding sequence were calculated by biosoftware DNASTar 7.0 for windows.

2.2. Calculation of the relative synonymous codon usage (RSCU)

To investigate the characteristics of synonymous codon usage without the confounding influence of amino acid composition among different sequences, the relative synonymous codon usage (RSCU) values among different codons in the EV71 ORF was calculated according to the published equation (Sharp et al., 1986).

2.3. Analysis of codon usage bias

The 'effective number of codons' (ENCs), the useful estimator of absolute codon usage bias, was a measure quantifying the codon usage bias of the whole coding sequence of EV71. The ENC value ranges from 20 (when only one synonymous codon is chosen by the corresponding amino acid) to 61 (when all synonymous codons are used equally) (Wright, 1990). In this study, this measure was used to evaluate the degree of codon usage bias of coding sequences for proteins of EV71 and to calculate the degree of the codon bias for the whole coding sequence of EV71 and other picornaviruses.

Additionally, there is a simple method which is supposed that statistically equal and random usage of all available synonymous codons was the "neutral point" (RSCU₀ = 1.00) for the development of group-specific codon usage (Zhou et al., 2010). This method was introduced in the study to investigate the discrepancy of the synonymous codon usage pattern among of EV71 strains based on the different isolated areas.

2.4. Principal component analysis

Principal component analysis (PCA), which was a commonly used multivariate statistical method (Jolliffe, 2002; Mardia et al., 1979), was carried out to analyze the major trend in codon usage pattern among different strains of EV71. PCA involves a mathematical procedure that transforms some correlated variable (RSCU values) into a smaller number of uncorrelated variables called principal components. Each strain was represented as a 59 dimensional vector, and each dimension corresponded to the RSCU value of each sense codon, which only included several synonymous codons for a particular amino acid, excluding the codon of AUG, UGG and three stop codons. In addition, PCA was also

performed for analyzing the discrepancy between codon usage pattern of EV71 and that of host cell.

2.5. Correlation analysis

The relationship between each general nucleotide composition (U%, A%, C% and G%) and each nucleotide composition in the third site of codon (U₃%, A₃%, C₃% and G₃%) in EV71 coding sequence and the relationship between U₃%, A₃%, C₃%, G₃% and the codon usage pattern of EV71 were evaluated by the Pearson's rank.

All statistical processes were carried out by statistical software SPSS11.5 for windows.

3. Results

3.1. Synonymous codon usage in EV71

The A% and U% were higher than C% and G%, but A₃% and U₃% were lower than C₃% and G₃% in EV71 (Table S2). The overall nucleotide composition never affects the nucleotide contents in the third site of codon in EV71 coding sequence, suggesting that composition constraints may be one of the factors in affecting the codon usage pattern of EV71.

The optimal codons of Ala, Arg, Asp, Cys, Glu, Gly, Ile, Phe, Pro, Ser were A-ended or U-ended, while those of Asn, Glu, His, Leu, Lys, Thr, Tyr, Val were C-ended or G-ended (Table 1). EV71 does not depends on all optimal codons with either A/U-end or C/G-ended like influenza A virus subtype H5N1, sever acute respiratory syndrome *Coronavirus* or foot-and-mouth disease virus (FMDV)

Table 1
The synonymous codon usage patterns of EV71 virus and human cell.

AA ^a	Codon	RSCU ^d	AA ^a	Codon	RSCU ^d	
Ala ^b	GCA	1.223	Leu ^c	CUA	0.807	
	GCC ^f	0.996		CUC	1.131	
	GCG ^g	0.536		CUG ^f	0.976	
	GCU	1.245		CUU	1.119	
Arg ^b	AGA ^f	1.853	Lys ^c	UUA ^g	0.805	
	AGG	1.702		UUG	1.159	
	CGA ^g	0.568		AAA ^g	0.907	
	CGC	1.020		AAG ^f	1.093	
	CGG	0.388		Phe ^b	UUC ^f	0.979
	CGU	0.469			UUU ^g	1.020
Asn ^c	AAU ^g	1.091	Pro ^b	CCA	1.589	
	AAU ^g	0.909		CCC ^f	0.881	
Asp ^b	GAC ^f	0.900	Ser ^b	CCG ^g	0.405	
	GAU ^g	1.100		CCU	1.124	
Cys ^b	UGC ^f	0.972	Gln ^b	AGC ^f	0.994	
	UGU ^g	1.028		AGU	1.172	
Glu ^c	CAA ^g	1.103	Gly ^b	UCA	1.315	
	CAG ^f	0.897		UCC	1.133	
Gly ^b	GAA ^g	0.933	Thr ^c	UCC ^g	0.367	
	GAC ^f	1.067		UCU	1.019	
	GGA	0.997		ACA	1.201	
	GGC ^f	0.498		ACC ^f	1.221	
His ^c	GGG	0.994	Tyr ^c	ACG ^g	0.376	
	GGU ^g	1.499		ACU	1.202	
	CAC ^f	1.253		Ile ^b	UAC ^f	1.092
	CAU ^g	0.747			UAU ^g	0.908
AUA ^g	0.691	GUA ^g	0.471			
AUC ^f	1.048	GUC	0.893			
Ile ^b	AUU	1.261		GUG ^f	1.678	
				GUU	0.958	

^a Stands for amino acid.

^b Stands for amino acids whose optimal codons are A- or U-end.

^c Stands for amino acids whose optimal codons are C- or G-end.

^d Stands for the relative synonymous codon usage value.

^f Stands for the optimal codons used in human cells (Pintó et al., 2007; Sánchez et al., 2003).

^g Stands for the rare codons used in human cells (Pintó et al., 2007; Sánchez et al., 2003).

Table 2Summary of correlation analysis between the A%, U%, C%, G% and A₃%, U₃%, C₃%, G₃% in the whole coding sequences of 74 EV71 strains^a.

	A ₃ %	U ₃ %	C ₃ %	G ₃ %	(C ₃ + G ₃)%
A%	$r = 0.869^{**}$	$r = -0.346^{**}$	$r = -0.316^{**}$	$r = -0.316^{**}$	$r = -0.102^{NS}$
U%	$r = -0.307^{**}$	$r = 0.918^{**}$	$r = -0.703^{**}$	$r = -0.703^{**}$	$r = -0.882^{**}$
C%	$r = -0.467^{**}$	$r = -0.084^{NS}$	$r = 0.875^{NS}$	$r = 0.875^{NS}$	$r = 0.341^{**}$
G%	$r = 0.316^{**}$	$r = -0.751^{**}$	$r = -0.027^{NS}$	$r = 0.027^{NS}$	$r = 0.671^{**}$
(C + G)%	$r = -0.198^{NS}$	$r = -0.665^{**}$	$r = 0.832^{**}$	$r = 0.832^{**}$	$r = 0.884^{**}$

^a r value in this table is calculated in each correlation analysis.NS means non-significant ($p > 0.05$).^{**} Means $p < 0.01$.

(Gu et al., 2004; Zhao et al., 2008; Zhou et al., 2010), but shapes the optimal codons with any types of nucleotide-ended. It is noted that although Asn, Leu, Tyr, Glu, Lys and Thr possessed optimal codon with C- or G-ended, they also contained some favored codon with U- or A-ended. Similarly, Asp, Phe and Cys also had favored codons with C- or G-ended (Table 1). These amino acids which choose optimal codons with any nucleotide-ended are affected under both mutation pressure by itself and natural selection from host, since natural selection from host ultimately allows those strains with good-fitness to possess a special codon usage patterns.

3.2. Genetic relationship based on synonymous codon usage in EV71

The PCA detected the first principal component (f_1') which can account for 13.73% of the total synonymous codon usage variation, and the second principal component (f_2') for 11.81% of the total variation. It appeared to be a little complex with some overlapping plots representing different epidemic areas (Fig. S1). The plots for strains isolated from China-Mainland, compared with that of strains isolated from other areas, could aggregate highly, while the plots for strains isolated from Malaysia and China-Taiwan scattered largely, the plots for strains from Singapore, USA, Japan and Switzerland did not indicate the genetic diversity obviously, due to the limited samples. For strains circulating in China-Mainland, social factors (public health, interpersonal communication, etc.) may play a role in influencing genetic diversity of these strains. However, for strains in China-Taiwan and Malaysia, geographic factors likely influence genetic diversity of those strains except for social factors.

3.3. Compositional properties of the whole coding sequence of EV71

The nucleotide contents of the whole coding sequence of EV71 were analyzed. In Table 2, the significant positive correlations between A% and A₃%, U% and U₃%, C% and C₃% and significant negative correlations among most of heterogeneous nucleotide contents indicated that composition constraints play a role in codon usage pattern of EV71; however, significant positive correlation between G% and A₃%, C% and G₃% and no correlation between G% and G₃% might suggest that natural selection from host plays a role in codon usage pattern of EV71 as well. In addition, there were significant correlations between each nucleotide

Table 3

Summary of correlation analysis between the first two axes in principle and nucleotide contents in EV71.

Base compositions	f_1'	f_2'
A ₃ %	$r = 0.359^{**}$	$r = 0.542^{**}$
U ₃ %	$r = -0.513^{**}$	$r = -0.516^{**}$
C ₃ %	$r = 0.238^{**}$	$r = -0.355^{**}$
G ₃ %	$r = 0.238^{**}$	$r = -0.355^{**}$
(C ₃ + G ₃)%	$r = 0.439^{**}$	$r = 0.250^{**}$

^{*} Means $0.01 < p < 0.05$.^{**} Means $p < 0.01$.

content in the third site of codon and codon usage indices (f_1 and f_2) (Table 3). Although the positive and negative correlations existed between C₃% and f_1 , and between C₃% and f_2 , respectively, the positive correlation play an important role in affecting the codon usage pattern due to f_1 being the first principal component.

3.4. Qualitative evaluation of codon usage bias of the whole coding sequence of EV71

The strong discrepancy of the synonymous codon usage in strains based on the different isolated areas was observed. In details, in strains from China-Mainland, CGC for Arg, GAC for Asp, CUU for Leu, UUC for Phe were chosen by EV71 strains preferentially, while CGG for Arg, GAC for Asp, UUU for Phe were chosen poorly; in Singapore, AUU for Ile, CUG for Leu, UCA for Ser were chosen preferentially, while UUG for Phe and UCG for Ser were chosen poorly; in USA, AGA for Arg was used preferentially, while GGU for Gly was used poorly; in Japan, GCU for Ala, AAU for Asn, CAA for Gln, GAA for Glu were chosen preferentially, while GCG for Ala and CAG for Gln were poorly used; in Switzerland, AGG for Arg, UGC for Cys, CUC for Leu, UUG for Leu, AGU for Ser were chosen preferentially; while CGA for Arg, UGU for Cys, CUG for Leu, CUU for Leu, AGC for Ser were poorly used (Fig. S2). These results may suggest that with the development of evolution of EV71 strains, the discrepancy of some synonymous codon usage probably is formed in different epidemic regions.

3.5. Relationship between amino acids and codon usage pattern in EV71

In order to analyze whether the evolution of CUB was controlled by mutation effect or natural selection from host, the CUB values had been calculated based on data listed in Table 1. The transition from maximum-negative to maximum-positive values was smooth and there was no obvious or unambiguous border between the so-called dominant and prohibited codons (Fig. 1), namely, all synonymous codons were used. This result implied that the interaction between mutation pressure from EV71 and natural selection from host exists in the evolution of EV71.

3.6. Relationships between codon usage pattern of EV71 and that of the host cell

By comparing between the patterns of synonymous codon usage of human cell and that of EV71 virus, we found that the pattern of synonymous codon usage of EV71 strains is partially antagonistic to that of human cells. In detail, optimal codons of nine amino acids in EV71, including Ala, Asp, Cys, Gln, Gly, Leu, Phe, Pro, Ser, are the disfavored codons of the corresponding amino acids in its host. Among these non-coincidence patterns of synonymous codon usage of amino acids, the synonymous codon usage of Asp, Cys, Gln, Gly, Phe has evolved to be complementary to that of host cells (Table 1). In addition, the optimal and rare synonymous codon usage patterns of Arg, Asn, Glu, His, Lys, Thr,

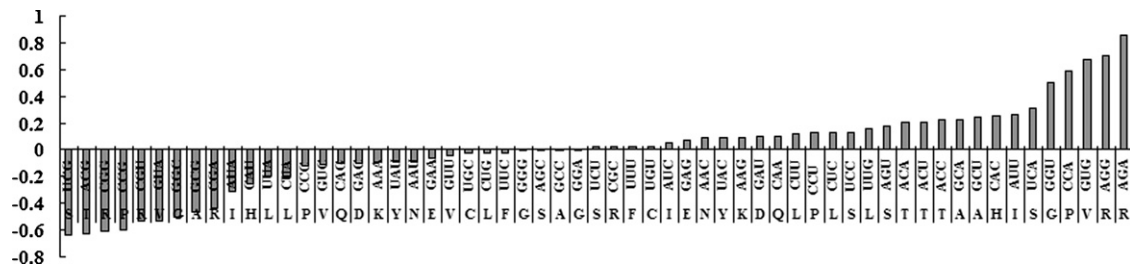


Fig. 1. Distribution of the CUB of a codon for each amino acid. CUB was taken from Table 1 and sorted in ascending order.

Tyr and Val of EV71 virus were in agreement with those of human cells (Table 1). Additionally, PCA was performed to examine the whole coding sequence of EV71 in this study. The method detected one major trend in the first axis (f_1') which can account for 11.79% of the total synonymous codon usage variation, and another major trend in the second axis (f_2') for 10.69% of the total variation. The plots for codon usage pattern of human are far from the plots for that of EV71 (Fig. S3).

The ENC values were calculated for FMDV, Cardiovirus, Hepatitis A virus (HAV), Poliovirus (PV) and compared with that of EV71 (Table S3). Among these virus examined, the ENC value for EV71 is highest, suggesting that EV71 has a most weak codon usage bias. In addition, we set up a plot which showed the relationship between $GC_3\%$ and ENC values of all viral proteins (excluding the very small 3B protein) of EV71 virus, and found that the plots of coding sequences for VP1, VP2, VP3, 2A, 2C, 3A, 3C and 3D aggregated around the expected curve, but the plots of coding sequences for VP4 and 2B scattered highly under the expected curve (Fig. S4a–4c). It may be explained that the codon usage bias of VP4 and 2B genes is influenced by their small size. In addition, there is no obvious geographic factor in influencing codon usage bias of the coding sequences of EV71, implying that the natural selection from the geographic factor does not affect the codon usage patterns of specific coding sequences of EV71, but shape the pattern of the whole sequence of this virus. Furthermore, we found that some specific non-optimal codons are preferentially chosen in some coding sequences of EV71. In details, three non-optimal codons (UUA, CUA and GUU) are chosen in the VP4 gene, UUG in the VP3 gene, CUU in the VP1 and 3A genes, UUU, CUU and GUU in the 3C gene. It is also found that all coding sequences of EV71 contain some preferential codons. In details, GUG, CCC, ACA, GCC and GAC are preferentially chosen in the VP4 gene, GUG, CCA and AGG in the VP2 gene, CUG, UCA, ACC and AGA in the VP3 gene, UCA in the VP1 gene, CUC, CCA and AGA in the 2A gene, CCU, AGA and AGG in the 2B gene, GUG, UCU, CCA, ACA and AGA in the 2C gene, CCA, ACU, AGU, AGC, AGA and AGG in the 3A gene, AUU, CCU, ACA, GCA, AGU and AGG in the 3C gene, GUG and AGA in the 3D gene. Taken together, there is no obvious relationship between the distribution of non-optimal codons and the deviation of ENC value from the theoretical value.

4. Discussion

The pattern of codon usage is a genetic characteristic of various organisms. Previous reports have been focused on viruses in Picornaviridae family, such as FMDV, HAV, Poliovirus (Aragones et al., 2008, 2010; Coleman et al., 2008; Zhong et al., 2007). Because A%, U%, $G_3\%$ and $C_3\%$ play roles in the formation of the different optimal codons with any nucleotide-ended, the codon usage pattern of EV71 is likely influenced by composition constraints. The codon usage pattern of PV is mostly coincident with that of its host, while the codon usage pattern of HAV is antagonistic to that of its host (Mueller et al., 2006; Sánchez

et al., 2003). The codon usage pattern of EV71 is a mixture of the two types of codon usage. The coincident portion of codon usage pattern of EV71 enable the corresponding amino acids to be translated efficiently, the other antagonistic portion of codon usage pattern of EV71 may enable viral proteins to be folded properly, although the translation efficiency of the corresponding amino acids decreased. In Epstein-Barr virus latent genes deoptimize codon usage in order to evade competition for host protein translation (Karlin et al., 1990) and attenuation of PV activity was performed by rare codon pairs inducing poor translation for sequences of viral proteins (Coleman et al., 2008). These results suggest that disfavored codons coding for amino acids may not be deleterious factor for viruses to adapt to host cells. For codon usage patterns of the coding sequences of EV71, the VP2, 2A, 2B, 2C and 3D genes possess only some preferential codons and none of non-optimal codons is preferentially used, implying that translation of the whole coding sequence of EV71 is possibly regulated under the translation selection. Furthermore, the alternative translation is the possibility of fine-tuning the kinetics of protein translation by a combination of rare and optimal codons (Aragones et al., 2010; Komar, 2009). For codon usage patterns of VP4, VP3, VP1, 3A and 3C genes of EV71, these genes possess combination of some non-optimal codons and optimal ones which are preferentially used, implying that translation of the coding sequences of EV71 is possibly regulated under fine-tuning translation kinetics selection.

The sequences 5'NTR and VP1 are often used to analyze the genetic diversity of EV71 (Hagiwara et al., 1984; Hsu et al., 2007; Li et al., 2005). By analyzing the codon usage pattern of the whole coding sequence of strains from different areas, genetic diversity resulting from geographic and social factors is likely observed. The genetic diversity of the most strains from China-Mainland could indicate that a relatively independent area with geographic, public health and personal communication that enables the genetic diversity of EV71 to be sustained with little outside influence. Compared with the genetic diversity of strains from China-Mainland, that of strains from China-Taiwan and Malaysia also indicated that social factors play an important role in shaping the codon usage patterns of these strains. Based on the genetic diversities of China-Taiwan and Malaysia, social factors may play important roles in shaping codon usage patterns of EV71 from the two areas. These genetic diversities of EV71 strains from different areas give a sign that geographic and social factors should be noticed at genetic diversity of virus from different areas.

The ENC values calculated for some picornaviruses indicate that a significantly lower bias of codon usage exists in EV71 than in the other viruses. As for RNA viruses, previous study reported that the major factor in shaping codon usage patterns appears to be mutation pressure rather than natural selection (Liu et al., 2010; Zhao et al., 2008; Zhong et al., 2007; Zhou et al., 2005, 2010). However, the genetic characteristics of EV71 suggest the interaction between mutation pressure and natural selection, although ENC values for the whole coding sequence of EV71

suggest mutation pressure is a factor in influencing codon usage pattern. Furthermore, in Fig. S4a–4c, the relationship between ENC data for EV71 proteins and CG₃% indicated that natural selection probably play roles in genetic diversity of EV71 strains except for mutation pressure in order to adapt to host. A general mutational pressure, which affects the whole genome would certainly account for the majority of the codon usage among some RNA viruses (Jenkins and Holmes, 2003).

The genetic diversity and codon usage patterns results we proposed here are useful to understand the processes influencing the evolution of EV71, especially the roles played by natural selection from host and mutation pressure from virus. Additionally, such information might be helpful to understand the roles of geographic and social factors in influencing genetic diversity of EV71.

Acknowledgements

This work was supported in parts by grants from National Science & Technology Key Project (2009ZX08007-006B) and International Science & Technology Cooperation Program of China (No. 2010DFA32640) and Science and Technology Key Project of Gansu Province (No. 0801NKDA034). This study was also supported by National Natural Science foundation of China (No. 30700597 and No. 31072143).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.meegid.2011.02.018.

References

- AbuBakar, S., Chee, H.Y., Al-kobaisi, M.F., Xiaoshan, J., Chua, K.B., Lam, S.K., 1999. Identification of enterovirus 71 isolates from an outbreak of hand, foot, and mouth disease (HFMD) with fatal cases of encephalomyelitis in Malaysia. *Virus Res.* 61, 1–9.
- Aragones, L., Bosch, A., Pinto, R., 2008. Hepatitis A virus mutant spectra under the selective pressure of monoclonal antibodies: codon usage constrains limit capsid variability. *J. Virol.* 82, 1688–1700.
- Aragones, L., Guix, S., Ribes, E., Bosch, A., Pinto, R.M., 2010. Fine-tuning translation kinetics selection as the driving force of codon usage bias in the Hepatitis A virus capsid. *PLoS Pathog.* 6, e1000797.
- Bible, J.M., Iturriza-Gomara, M., Megson, B., Brown, D., Pantelidis, P., Earl, P., Bendig, J., Tong, C.Y., 2008. Molecular epidemiology of human enterovirus 71 in the United Kingdom from 1998 to 2006. *J. Clin. Microbiol.* 46, 3192–3200.
- Cardosa, M.J., Perera, D., Brown, B.A., Cheon, D., Chan, H.M., Chan, K.P., Cho, H., McMinn, P., 2003. Molecular epidemiology of human enterovirus 71 strains and recent outbreaks in the Asia-Pacific region: comparative analysis of the VP1 and VP4 genes. *Emerg. Infect. Dis.* 9, 461–468.
- Chan, Y.F., AbuBakar, S., 2004. Human enterovirus 71 in hand, foot and mouth disease patients. *Emerg. Infect. Dis.* 10, 1468–1470.
- Chen, X., Zhang, Q., Li, J., Cao, W., Zhang, J.X., Zhang, L., Zhang, W., Shao, Z.J., Yan, Y., 2010. Analysis of recombination and natural selection in human enterovirus 71. *Virology* 398, 251–261.
- Cherry, J.D., 1992. Enteroviruses: polioviruses (poliomyelitis), coxsackieviruses, echoviruses and enteroviruses. In: Feign, R.D., Cherry, J.D. (Eds.), *Textbook of Pediatric Infectious Diseases*. 3rd ed. The W.B. Saunders Co., Philadelphia, pp. 1705–1752.
- Chumakov, M., Voroshilova, M., Shindarov, L., Lavrova, I., Gracheva, L., Koroleva, G.S., 1979. Enterovirus 71 isolated from cases of epidemic poliomyelitis-like disease in Bulgaria. *Arch. Virol.* 60, 329–340.
- Coleman, J.R., Papamichail, D., Skiena, S., Fitcher, B., Wimmer, E., Mueller, S., 2008. Virus attenuation by genome-scale changes in codon pair bias. *Science* 27, 1784–1787.
- Fujimoto, T., Chikahira, M., Yoshida, S., Ebira, H., Hasegawa, A., Totsuka, A., Nishio, O., 2002. Outbreak of central nervous system disease associated with hand, foot, and mouth disease in Japan during the summer of 2000: detection and molecular epidemiology of enterovirus 71. *Microbiol. Immunol.* 46, 621–627.
- Gu, W.J., Zhou, T., Ma, J.M., Sun, X., Lu, Z.H., 2004. Analysis of synonymous codon usage in SARS coronavirus and other viruses in the Nidovirales. *Virus Res.* 101, 155–161.
- Hagiwara, A., Yoneyama, T., Takami, S., Hashimoto, I., 1984. Genetic and phenotypic characteristics of enterovirus 71 isolates from patients with encephalitis and with hand, foot and mouth disease. *Arch. Virol.* 79, 273–283.
- Herrero, L.J., Lee, C.S., Hurrelbrink, R.J., Chua, B.H., Chua, K.B., McMinn, P.C., 2003. Molecular epidemiology of enterovirus 71 in peninsular Malaysia, 1997–2000. *Arch. Virol.* 148, 1369–1385.
- Ho, M., Chen, E.R., Hsu, K.H., Twu, S.J., Chen, K.T., Tsai, S.F., Wang, J.R., Shih, S.R., 1999. An epidemic of enterovirus 71 infection in Taiwan. *N. Engl. J. Med.* 341, 929–935.
- Hsu, B.M., Chen, C.H., Wan, M.T., 2007. Genetic diversity of epidemic enterovirus 71 strains recovered from clinical and environmental samples in Taiwan. *Virus Res.* 126, 69–75.
- Huang, S.C., Hsu, Y.W., Wang, H.C., Huang, S.W., Kiang, D., Tsai, H.P., Wang, S.M., Liu, C.C., Lin, K.H., Su, I.J., Wang, J.R., 2008. Appearance of intratypic recombination of enterovirus 71 in Taiwan from 2002 to 2005. *Virus Res.* 131, 250–259.
- Jenkins, G.M., Holmes, E.C., 2003. The extent of codon usage bias in human RNA virus and its evolutionary origin. *Virus Res.* 92, 1–7.
- Jolliffe, I.T., 2002. *Principal Component Analysis*, 2nd ed. Springer-Verlag, Inc., New York.
- Karlin, S., Mrázek, J., 1996. What drives codon choices in human genes? *J. Mol. Biol.* 262, 459–472.
- Karlin, S., Blaisdell, B.E., Schachtd, G.A., 1990. Contrasts in codon usage of latent versus productive genes of Epstein-Barr virus: data and hypotheses. *J. Virol.* 64, 4264–4273.
- Karlin, S., Doerfler, W., Cardon, L.R., 1994. Why is CpG suppressed in the genomes of virtually all small eukaryotic viruses but not in those of large eukaryotic viruses? *J. Virol.* 68, 2889–2897.
- Komar, A.A., 2009. A pause for thought along the co-translational folding pathway. *Trends Biochem. Sci.* 34, 16–24.
- Lesnik, T., Solomovici, J., Deana, A., Ehrlich, R., Reiss, C., 2000. Ribosome traffic in *E. Coli* and regulation of gene expression. *J. Theor. Biol.* 202, 175–185.
- Levin, D.B., Whittome, B., 2000. Codon usage in nucleopolyhedroviruses. *J. Gen. Virol.* 81, 2313–2325.
- Lewis-Rogers, N., Bendall, M.L., Crandall, K.A., 2009. Phylogenetic relationships and molecular adaptation dynamics of human rhinoviruses. *Mol. Biol. Evol.* 26, 969–981.
- Li, L., He, Y., Yang, H., Zhu, J., Xu, X., Dong, J., Zhu, Y., Jin, Q., 2005. Genetic characteristics of human enterovirus 71 and Coxsackievirus A16 circulating from 1999 to 2004 in Shenzhen, People's Republic of China. *J. Clin. Microbiol.* 43, 3835–3839.
- Lin, L., He, Y., Hong, Y., Zhu, J., 2005. Genetic characteristics of human enterovirus 71 and coxsackievirus A16 circulating from 1999 to 2004 in Shenzhen, People's Republic of China. *J. Clin. Microbiol.* 43, 3835–3839.
- Liu, C.C., Tsen, H.W., Wang, S.M., Wang, J.R., Su, I.J., 2000. An outbreak of enterovirus 71 infection in Taiwan, 1998: epidemiologic and clinical manifestations. *J. Clin. Virol.* 17, 23–30.
- Liu, Y.S., Zhou, J.H., Chen, H.T., Ma, L.N., Ding, Y.Z., Wang, M., Zhang, J., 2010. Analysis of synonymous codon usage in porcine reproductive and respiratory syndrome virus. *Infect. Genet. Evol.* 10, 797–803.
- Mardia, K.V., Kent, J.T., Bibby, J.M., 1979. *Multivariate Analysis*. Academic Press, New York.
- McMinn, P.C., 2002. An overview of the evolution of enterovirus 71 and its clinical and public health significance. *FEMS Microbiol. Rev.* 26, 91–107.
- Mueller, S., Papamichail, D., Coleman, J.R., Skiena, S., Wimmer, E., 2006. Reduction of the rate of poliovirus protein synthesis through large-scale codon deoptimization causes attenuation of viral virulence by lowering specific infectivity. *J. Virol.* 80, 9687–9696.
- Oberste, M.S., Maher, K., Kilpatrick, D.R., Pallansch, M.A., 1999a. Molecular evolution of the human enterovirus: correlation of serotype with VP1 sequence and application to picornavirus classification. *J. Virol.* 73, 1941–1948.
- Oberste, M.S., Maher, K., Kilpatrick, D.R., Flemister, M.R., Brown, B.A., Pallansch, M.A., 1999b. Typing of human enteroviruses by partial sequencing of VP1. *J. Clin. Microbiol.* 37, 1288–1293.
- Pintó, R.M., Aragonè, L., Costafreda, M.I., Ribes, E., Bosch, A., 2007. Codon usage and replicative strategies of hepatitis A virus. *Virus Res.* 127, 158–163.
- Sánchez, G., Bosch, A., Pintó, R.M., 2003. Genome variability and capsid structural constraints of Hepatitis A virus. *J. Virol.* 77, 452–459.
- Sanders, S.A., Herrero, L.J., McPhie, K., Chow, S.S., Craig, M.E., Dwyer, D.E., Rawlison, W., McMinn, P.C., 2006. Molecular epidemiology of enterovirus 71 over two decades in an Australian urban community. *Arch. Virol.* 151, 1003–1013.
- Sharp, P.M., Tuohy, T.M.F., Mosurski, K.R., 1986. Codon usage in yeast: cluster analysis clearly differentiates highly and lowly expressed genes. *Nucleic Acids Res.* 14 (13), 5125–5143.
- Sugiyama, T., Gursel, M., Takeshita, F., Coban, C., Jacqueline, C., Kaisho, T., Akira, S., Klinman, D.M., Ishii, K.J., 2005. CpG RNA: identification of novel single-stranded RNA that stimulates human CD14+ CD11c+ monocytes. *J. Immunol.* 174, 2273–2279.
- Wright, F., 1990. The 'effective number of codons' used in a gene. *Gene* 87, 23–29.
- Yoke-Fun, C., AbuBakar, S., 2006. Phylogenetic evidence for inter-typic recombination in the emergence of human enterovirus 71 subgenotypes. *BMC Microbiol.* 6, 74.
- Zhao, S., Zhang, Q., Liu, X.L., Wang, X.M., Zhang, H.L., Wu, Y., Jiang, F., 2008. Analysis of synonymous codon usage in 11 Human Bocavirus isolates. *Biosystems* 92, 207–214.
- Zheng, Z.H., He, P.J., Caueffield, D., Neumann, M., Specter, S., Baker, C.C., Bankowski, M.J., 1995. Enterovirus 71 isolated from China is serologically similar to the prototype EV71 BrCr strain but differs in the 5'-noncoding region. *J. Med. Virol.* 47, 161–167.

- Zhi, N., Wan, Z., Liu, X., Wong, S., Kim, D.J., Young, N.S., Kajigaya, S., 2010. Codon optimization of human parvovirus B19 capsid genes greatly increases their expression in nonpermissive cells. *J. Virol.* 84, 13059–13062.
- Zhong, J.C., Li, Y.M., Zhao, S., Liu, S., Zhang, Z., 2007. Mutation pressure shapes codon usage in the GC-Rich genome of foot-and-mouth disease virus. *Virus Genes* 35, 767–776.
- Zhou, J., Liu, W.J., Peng, S.H., Sun, X.Y., Frazer, I., 1999. Papillomavirus capsid protein expression level depends on the match between codon usage and tRNA availability. *J. Virol.* 73, 4972–4982.
- Zhou, J.H., Zhang, J., Chen, H.T., Ma, L.N., Liu, Y.S., 2010. Analysis of synonymous codon usage in foot-and-mouth disease. *Vet. Res. Commun.* 34, 393–404.
- Zhou, T., Gu, W.J., Ma, J.M., Sun, X., Lu, Z.H., 2005. Analysis of synonymous codon usage in H5N1 virus and other influenza A viruses. *Biosystems* 81, 77–86.
- Zhou, T., Sun, X., Lu, Z.H., 2006. Synonymous codon usage in environmental *Chlamydia* UWE25 reflects an evolution divergence from pathogenic *Chlamydiae*. *Gene* 368, 117–125.