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

Overall codon usage pattern of enterovirus 71

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

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

Corresponding author: H. Ling
E-mail: mamingren@yahoo.com.cn

Genet. Mol. Res. 13 (1): 336-343 (2014)
Received August 8, 2013
Accepted October 28, 2013
Published January 21, 2014
DOI http://dx.doi.org/10.4238/2014.January.21.1

ABSTRACT. Hand, foot, and mouth disease (HFMD) is a systemic illness in children and is usually caused by enterovirus 71 (EV71). To provide new insights into the genetic features of EV71 and the relationship between the overall codon usage pattern of this virus and that of humans, values for relative synonymous codon usage (RSCU), effective number of codons (ENC), codon adaptation index (CAI), and nucleotide composition were calculated and analyzed. The relationship between ENC values and $(G+C)_3$% suggests that, although nucleotide composition plays an important role in shaping the overall codon usage pattern of this virus, other factors also affect this pattern. In addition, the negative correlation between the CAI value and $(G+C)_3$% suggests that the secondary structure of the EV71 coding sequence caused by its nucleotide composition can affect gene expression. Moreover, there was no significant correlation between ENC and CAI, suggesting that gene expression does not play a role in shaping the overall codon usage pattern of EV71. The overall codon usage pattern of the EV71 virus is only partly similar to the general codon pattern of human, suggesting that, although EV71 has co-evolved with humans for extended periods, mutation pressure played an important role in shaping the virus's overall codon

usage pattern. These results revealed that the EV71 virus has developed a subtle strategy during evolution for adapting to environmental changes in its host cells solely by means of mutation pressure.

Key words: Enterovirus 71; Relative synonymous codon usage value; Overall codon usage pattern; Co-evolve; Mutation pressure

INTRODUCTION

Hand, foot, and mouth disease (HFMD) is a systemic illness that occurs in children and is usually caused by a human enterovirus. Enterovirus 71 (EV71) is one of the group of viruses that can give rise to the HFMD clinical syndrome (Blomberg et al., 1974). EV71 is a member of the *Enterovirus* genus of the Picornaviridae family. It is a positive-strand RNA virus with a genome size of about 7500 bp and a typical genetic organization. Similar to the findings for other members of this family, the antigenic diversity of this virus mainly arises from variations in the four structural proteins, VP1 - VP4. The VP1 gene contains the most informative region of the enterovirus genome for understanding evolutionary relationships in that it possesses a neutralization determinant and also for understanding the genetic diversity associated with serotypes (Oberste et al., 1999). There have been indications that recombination may play a more important role than positive selection in the development of genetic diversity, even though evidence for positive selection was noted in VP1 (Chen et al., 2010). These results probably imply that the EV71 evolutionary process is comprehensive and systemic. For example, enterovirus neurovirulence is a complex phenotypic feature that is probably shaped by more than one region of the EV71 genome (McMinn, 2002). It is accepted that the nucleotide composition of ORFs is selective rather than random because the information in the protein sequence and the development of codon usage bias arise from this non-randomness of nucleotide composition. In general, translation selection and compositional constraints under mutational pressure are thought to be the major factors accounting for codon usage variation between genomes in microorganisms. In some RNA viruses, mutation pressure plays an important role in synonymous codon usage pattern relative to that of translation selection (Levin and Whittome, 2000; Jenkins and Holmes, 2003; Gu et al., 2004).

Traditional treatment for HFMD simply consists of palliative care for relieving the clinical symptoms to some degree; hence, an effective vaccine against EV71 would play a key role in protecting children against this virus. However, RNA viruses exhibit high variability in genes encoding their structural proteins; therefore, vaccines are of limited effectiveness in protecting children from EV71. This situation has made researchers aware of the importance of analyzing EV71 genetic diversity (Bible et al., 2008). In analyses of genetic diversity or evolution of a particular virus, the characteristics of the sequence encoding its structural proteins is a focus area, but the whole encoding sequence probably stores more genetic or evolutionary information. Thus, the codon usage pattern of the ORF may be important in determining the molecular mechanisms and evolutionary processes of EV71, separate from responses to the host cell, thus permitting effective vaccines to be developed. In our study, synonymous codon usage analytical methods were used to investigate the evolutionary characteristics of EV71 and to evaluate the key evolutionary determinants of codon usage bias in this virus.

MATERIAL AND METHODS

Coding sequence

The 74 open reading frames (ORF) of EV71 were downloaded from the National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/Genbank/), and the detailed information is listed in **Table S1**. The nucleotide content of ORFs for each EV71 strain was analyzed using DNAStar 7.0 for Windows.

Calculation of the relative synonymous codon usage (RSCU) for EV71 and humans

To investigate the characteristics of synonymous codon usage among different sequences without the confounding influence of amino acid composition, the relative synonymous codon usage (RSCU) values among different codons in each ORF were calculated. The RSCU value of the i^{th} codon for the j^{th} amino acid was calculated according to the published equation (Sharp et al., 1986).

$$RSCU = \frac{g_{ij}}{\frac{n_i}{\sum_j g_{ij}}} \cdot n_i$$

Where g_{ij} is the observed number of the i^{th} codon for the j^{th} amino acid that has n_i types of synonymous codons. The codons with RSCU values greater than 1.0 have positive bias, while those with values less than 1.0 have relative negative bias. When the RSCU value is equal to 1.0, it means that this codon is chosen equally and randomly. Additionally, to analyze the relationship between synonymous codon usage in EV71 and human, the human synonymous codon usage frequency was obtained from the codon usage database http://www.kazusa.or.jp/codon/ (Nakamura et al., 2000) and RSCU calculated using the formula above. The RSCU data were used to evaluate the differences in synonymous codon usage between this virus and humans.

Index for codon usage of EV71

The effective number of codons (ENC), a useful estimator of absolute codon usage bias, was used to quantify the codon usage bias of the whole EV71 coding sequence. The ENC value ranges from 20 (when only one synonymous codon is chosen for 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 in coding sequences for the proteins of EV71.

Codon Adaptation Index (CAI) was used to estimate the extent of bias towards codon usage; this is known to be higher in highly expressed genes. The CAI value lies between 0.0 and 1.0, with higher values indicating more bias (Sharp and Li, 1987). CAI has been shown to be the best gene expression value index and is extensively used as a measure of gene expression levels.

RESULTS

Overall codon usage, amino acid usage and nucleotide composition constraint analysis

The EV71 genome is a single positive RNA strand with an A+U content of 52.3%. Because it is AU rich, it was expected that codons ending in A or U would predominate in coding regions of this RNA virus. Indeed, the optimal codons for the ten AU-rich amino acids (Ala, Arg, Asp, Cys, Glu, Gly, Ile, Phe, Pro, and Ser) were specified by codons ending in A or U, while Asn, Glu, His, Leu, Lys, Thr, Tyr, and Val were specified by optimal ones ending with C or G (Table 1). This non-randomness in synonymous codon usage pattern could be attributed to mutational bias or to translation selection favoring specific codons. These results suggest that mutational pressure and translational selection from the host both play important roles in the evolutionary process of this RNA virus.

Table 1. Comparison between codon usage pattern of EV71 virus and that of human cells.

Amino acid	Codon	RSCU	Amino acid	Codon	RSCU
Ala[a]	GCA	1.22	Leu[b]	CUA	0.81
	GCC*	1.00		CUC	1.13
	GCG#	0.54		CUG*	0.98
	GCU	1.25		CUU	1.12
Arg[a]	AGA*	1.85		UUA#	0.81
	AGG	1.70		UUG	1.16
	CGA#	0.57	Lys[b]	AAA#	0.91
	CGC	1.02		AAG*	1.09
	CGG	0.39	Phe[a]	UUC*	0.98
	CGU	0.47		UUU#	1.02
Asn[b]	AAC*	1.09	Pro[a]	CCA	1.59
	AAU#	0.91		CCC*	0.88
Asp[a]	GAC*	0.90		CCG#	0.41
	GAU#	1.10		CCU	1.12
Cys[a]	UGC*	0.97	Ser[a]	AGC*	0.99
	UGU#	1.03		AGU	1.17
Gln[a]	CAA#	1.10		UCA	1.32
	CAG*	0.90		UCC	1.13
Glu[b]	GAA#	0.93		UCG*	0.37
	GAG*	1.07		UCU	1.02
Gly[a]	GGA	1.00	Thr[b]	ACA	1.20
	GGC*	0.50		ACC*	1.22
	GGG	0.99		ACG#	0.38
	GGU#	1.50		ACU	1.20
His[b]	CAC*	1.25	Tyr[b]	UAC*	1.09
	CAU#	0.75		UAU#	0.91
Ile[a]	AUA#	0.69	Val[b]	GUA#	0.47
	AUC*	1.05		GUC	0.89
	AUU	1.26		GUG*	1.68
				GUU	0.96

[a]Favored codon with A- or U-end; [b]Favored codon with C- or G-end; *Optimal codons used in human cells; #Rare codons used in human cells.

A plot (Nc plot) of effective number of codons (ENC) and G+C base composition at the third synonymous position, $(G+C)_3$%, has been used effectively to explore codon usage variation in the genome of an organism. Wright (1990) suggested that a comparison of the actual distribution of genes with the expected distribution under no selection could be indicative of codon usage bias in genes where there are influences other than compositional constraints.

It is interesting to note that for all EV71 isolates, all points with high Nc values lie well below the expected curve and toward the $(G+C)_3$%-rich area (Figure 1), suggesting that, apart from the dominant role of A+U compositional constraints, other factors might be influential in dictating codon usage variation. Interestingly, a significant correlation was found between $(G+C)_3$% and CAI for EV71 ($r = -0.39$, $P = 0.001$) (Figure 2). It indicates that the secondary structure of the coding sequence for EV71 caused by its nucleotide composition could play an important role in gene expression. In addition, there was no significant correlation ($r = 0.075$, $P > 0.05$) between ENC and CAI for all EV71 isolates, suggesting that there is no relationship between EV71 gene expression and its overall codon usage pattern and that nucleotide composition due to mutation pressure mainly determines codon usage pattern in this virus.

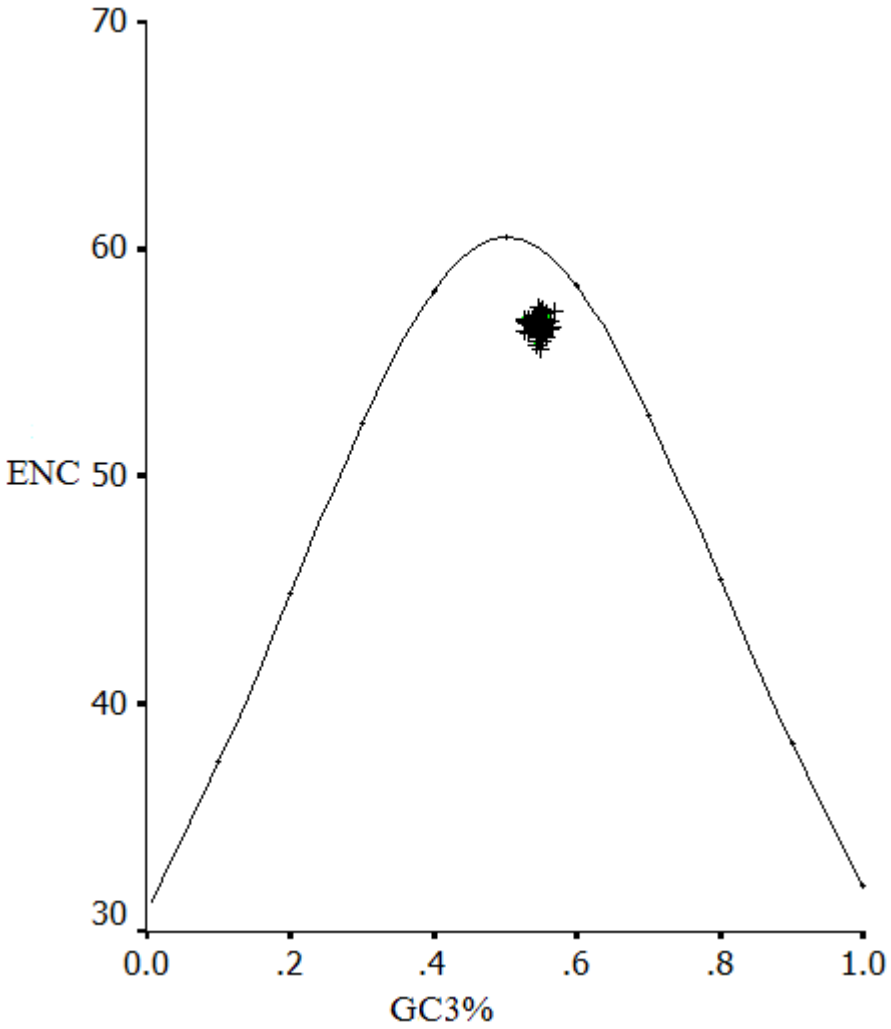

Figure 1. ENC value *vs* GC_3%. The continuous curve between ENC value and GC_3% during random codon usage. The dark squares represents each EV71 isolate.

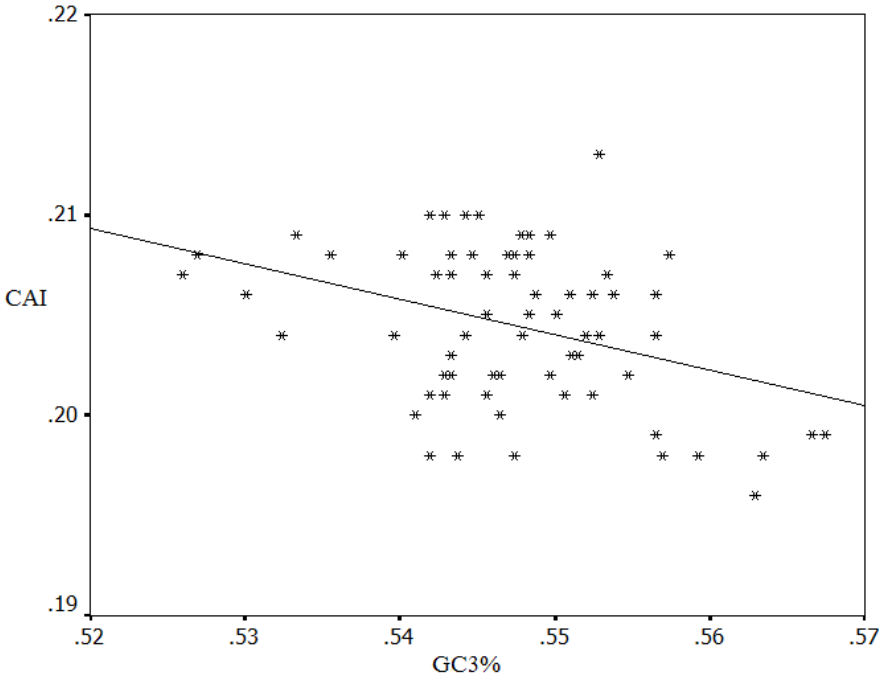

Figure 2. CAI value *vs* GC_3%. The dark squares represents each EV71 isolate. The oblique line represents the negative correlation between gene expression and nucleotide composition of EV71 coding sequence.

Coordination of synonymous codon usage between EV71 ORFs and its hosts

Twenty-eight of 59 codons are similarly used in EV71 and human cells, including GCG for Ala, CGA, CGC, and CGU for Arg; AAC and AAU for Asn; GAA and GAG for Glu; CAU and CAC for His; CCG and CCU for Pro; ACA, ACC, and ACG for Thr; GGA and GGG for Gly; CUC for Leu; UCG, UCC, and UCU for Ser; UAC and UAU for Tyr; GUG and GUA for Val; AUU for Ile; and AAA and AAG for Lys. However, some codons in the EV71 coding sequence, including CGG for Arg, GGC and GGU for Gly, CUG for Leu, and CCC for Pro, have a different codon usage from the corresponding codons of human cells (Table 1). This result suggests that, although EV71 has co-evolved with humans for extended periods, mutation pressure still drives its evolutionary process to a large extent.

DISCUSSION

In general, previous reports have indicated that many viruses, including foot-and-mouth disease viruses, influenza A virus subtype H5N1, severe acute respiratory syndrome Coronavirus (SARSCoV), and human bocavirus, preferentially use codons ending in C or G (Zhao et al., 2008; Zhou et al., 2010). However, in the case of the EV71 virus, only a few studies have been performed on the synonymous codon usage pattern influencing its evolution. There is no doubt that evolution has shaped the EV71 synonymous codon usage pattern to cater for its need to infect host cells, its replication, and challenges caused by the host cell

environment. In this study, interestingly, we found that for amino acids having at least two synonymous codons, some tend to be preferentially specified by codons ending in A or U, while others are predominantly specified by codons ending in C or G. These results suggest that the optimal codon usage pattern does not simply reflect the overall nucleotide composition, while composition limitations and mutation effects both play a role in the development of genetic variation in synonymous codon usage. One possible explanation to account for why EV71 has four types of optimal codons, compared with the optimal and rare codon usage patterns of host cells (Table 1), is that various patterns of synonymous codon usage are advantageous to EV71, which needs to replicate and express efficiently in host cells with potentially distinct codon preferences. The general view is that optimal codon usage increases the efficiency of translation (Andersson and Kurland, 1990; Zhou et al., 2011; Zhou et al., 2012). Optimal codons are probably translated more efficiently than non-optimal ones, in that ribosomes slide faster along mRNA regions consisting of many optimal codons, and the ribosomes can be quickly released to translate other mRNAs. It is possible that, when codon usage in the EV71 virus does not follow that of the host cell, and even when the virus uses codons that are rare in human cells, this avoids competition with favored codons in the host cell. At the same time, the process of adaptation to the host can require the use of host-preferred codons by EV71 at other locations in its genome. This subtle strategy might enable the virus to proliferate. Furthermore, optimal codon usage always enables more efficient ribosome utilization, leading to a faster growth rate. This is especially true for genes with high expression levels, encoding mRNAs that can be translated more often, implying that some viruses that act as parasites probably evolve a synonymous codon usage pattern that follows the tRNA population of the host cells, at least in the case of viruses occupying certain niches in the host cells. Furthermore, the negative correlation between codon bias and $(G+C)_3$% strongly implies that competition for tRNAs recognizing host cell-preferred codons could result in low yields of the EV71 virus before the virus can control host cells. This result also supports the idea that the virus adopts a variety of synonymous codon usage patterns.

In the current study, we analyzed the role of different evolutionary constraints that influence codon and amino acid usage patterns in EV71. We found that the gene expression level, mutation pressure, and translational selection operated in shaping codon usage variation in the isolates studied. Moreover, this study provides a comprehensive analysis of codon usage patterns and a basis for understanding the mechanisms determining codon usage bias that would aid in further study of the evolutionary mechanism, cloning, and heterologous expression of functionally important proteins.

ACKNOWLEDGMENTS

Research supported by grants from the Natural Science Foundation of Gansu Province (#1107RJZA106).

Supplementary material

REFERENCES

Andersson SG and Kurland CG (1990). Codon preferences in free-living microorganisms. *Microbiol. Rev.* 54: 198-210.
Bible JM, Iturriza-Gomara M, Megson B, Brown D, et al. (2008). Molecular epidemiology of human enterovirus 71 in the

United Kingdom from 1998 to 2006. *J. Clin. Microbiol.* 46: 3192-3200.
Blomberg J, Lycke E, Ahlfors K, Johnsson T, et al. (1974). Letter: New enterovirus type associated with epidemic of aseptic meningitis and-or hand, foot, and mouth disease. *Lancet* 2: 112.
Chen X, Zhang Q, Li J, Cao W, et al. (2010). Analysis of recombination and natural selection in human enterovirus 71. *Virology* 398: 251-261.
Gu W, Zhou T, Ma J, Sun X, et al. (2004). Analysis of synonymous codon usage in SARS coronavirus and other viruses in the Nidovirales. *Virus Res.* 101: 155-161.
Jenkins GM and Holmes EC (2003). The extent of codon usage bias in human RNA viruses and its evolutionary origin. *Virus Res.* 92: 1-7.
Levin DB and Whittome B (2000). Codon usage in nucleopolyhedroviruses. *J. Gen. Virol.* 81: 2313-2325.
McMinn PC (2002). An overview of the evolution of enterovirus 71 and its clinical and public health significance. *FEMS Microbiol. Rev.* 26: 91-107.
Nakamura Y, Gojobori T and Ikemura T (2000). Codon usage tabulated from international DNA sequence databases: status for the year 2000. *Nucleic Acids Res.* 28: 292.
Oberste MS, Maher K, Kilpatrick DR and Pallansch MA (1999). Molecular evolution of the human enteroviruses: correlation of serotype with VP1 sequence and application to picornavirus classification. *J. Virol.* 73: 1941-1948.
Sharp PM, Tuohy TM and Mosurski KR (1986). Codon usage in yeast: cluster analysis clearly differentiates highly and lowly expressed genes. *Nucleic Acids Res.* 14: 5125-5143.
Sharp PM and Li WH (1987). The codon Adaptation Index--a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res.* 15: 1281-1295.
Wright F (1990). The 'effective number of codons' used in a gene. *Gene* 87: 23-29.
Zhao S, Zhang Q, Liu X, Wang X, et al. (2008). Analysis of synonymous codon usage in 11 human bocavirus isolates. *Biosystems* 92: 207-214.
Zhou JH, Zhang J, Chen HT, Ma LN, et al. (2010). Analysis of synonymous codon usage in foot-and-mouth disease virus. *Vet. Res. Commun.* 34: 393-404.
Zhou JH, Zhang J, Chen HT, Ma LN, et al. (2011). The codon usage model of the context flanking each cleavage site in the polyprotein of foot-and-mouth disease virus. *Infect. Genet. Evol.* 11: 1815-1819.
Zhou JH, Gao ZL, Zhang J, Chen HT, et al. (2012). Comparative [corrected] codon usage between the three main viruses in *Pestivirus* genus and their natural susceptible livestock. *Virus Genes* 44: 475-481.