Infection, Genetics and Evolution 53 (2017) 212–217

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Research paper

Synonymous codon usage analysis of hand, foot and mouth disease viruses: A comparative study on coxsackievirus A6, A10, A16, and enterovirus 71 from 2008 to 2015

Weiheng Su [a,b,1], Xue Li [c,d,e,1], Meili Chen [c,d], Wenwen Dai [a], Shiyang Sun [a], Shuai Wang [a], Xin Sheng [c,d,e], Shixiang Sun [c,d,e], Chen Gao [c,d,e], Ali Hou [a,b], Yan Zhou [a,b], Bo Sun [a,b], Feng Gao [a,b], Jingfa Xiao [c,d,e], Zhewen Zhang [c,d,*], Chunlai Jiang [a,b,**]

[a] National Engineering Laboratory for AIDS Vaccine, School of Life Sciences, Jilin University, Changchun 130012, China
[b] Key Laboratory for Molecular Enzymology and Engineering, the Ministry of Education, School of Life Sciences, Jilin University, Changchun 130012, China
[c] BIG Data Center, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 100101, China
[d] CAS Key Laboratory of Genome Sciences and Information, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 100101, China
[e] University of Chinese Academy of Sciences, Beijing 100049, China

ARTICLE INFO

Article history:
Received 10 April 2017
Received in revised form 2 June 2017
Accepted 4 June 2017
Available online xxxx

Keywords:
Coxsackievirus A6
Coxsackievirus A10
Coxsackievirus A16
Enterovirus 71
Synonymous codon usage

ABSTRACT

Enterovirus 71 (EV71) and coxsackievirus A16 (CVA16) have been considered major pathogens of hand, foot and mouth disease (HFMD) throughout the world for decades. In recent years, coxsackievirus A6 (CVA6) and coxsackievirus A10 (CVA10) have raised attention as two other serious pathogens of HFMD. The present study focused on the synonymous codon usage of four viruses isolated from 2008 to 2015, with particular attention on P1 (encoding capsid proteins) and P2-P3 regions (both encoding non-structural proteins) in the genomic RNA. Relative synonymous codon usage, effective number of codons, neutrality and correspondence were analyzed. The results indicated that these viruses prefer A/T at the third position in codons rather than G/C. The most frequent codons of 4 essential and 2 semi-essential amino acids, as well as a key amino acid of metabolic junctions (Glu) used in the four viruses are also the most frequently used in humans. Effective number of codons (ENC) values indicated weak codon usage bias in all the viruses. Relatively, the force of mutation pressure in the P1 region was found to be stronger than that in the P2-P3 region, and this force in the P1 region of CVA6 and EV71 was stronger than that of CVA10 and A16. The neutrality analysis results implied that mutation pressure plays a minor role in shaping codon bias of these viruses. Correspondence analysis indicated that the codon usage of EV71 strains varied much more than that of other viruses. In conclusion, the present study provides novel and comparative insight into the evolution of HFMD pathogens at the codon level.

1. Introduction

Enterovirus (EV) is a genus of small, non-enveloped, RNA viruses in the family *picornaviridae*. EV has a 30 nM, icosahedral capsid, with an approximately 7.4 kb, single-stranded, positive-sense RNA genome. The EV genome encodes a polyprotein which can be divided into P1, P2 and P3 regions. The P1 region encodes capsid proteins VP1, VP2, VP3 and VP4, while the other two regions encode non-structural proteins such as 2A protease, 3C protease, 3D polymerase and VPg.

The EV genus includes four human EV species: A, B, C and D. Virus members within an EV species share >70% amino acid identity in the polyprotein and >60% amino acid identity in the P1 region (Smura et al., 2011). They also share a significant degree of compatibility in proteolytic processing, replication, encapsidation and genetic recombination (David et al., 2013). EV-A consists of 17 serotypes: coxsackievirus (CV) A2-A8, A10, A12, A14, A16, EV71, 76, 89–91 and 114 (David et al., 2013). These pathogens induce clinical symptoms such as rash, herpangina, paralysis, encephalitis, meningitis and myelitis (de Crom et al., 2016). Some of these symptoms also correspond to hand, foot and mouth disease (HFMD).

HFMD is caused generally by infection with CVA6, A10, A16 and EV71, and it only is sporadically caused by infection with other EV-A

* Correspondence to: Z. Zhang, BIG Data Center, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 100101, China.
** Correspondence to: C. Jiang, National Engineering Laboratory for AIDS Vaccine, School of Life Sciences, Jilin University, Changchun 130012, China.
E-mail addresses: zhangzw@big.ac.cn (Z. Zhang), jiangcl@jlu.edu.cn (C. Jiang).
[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.meegid.2017.06.004
1567-1348/

and some EV-B serotypes (Zhuang et al., 2015). EV71 infection has long been known to result in severe illness, such as aseptic encephalitis, meningitis, myelitis and pulmonary edema, which can ultimately lead to death (Wang and Liu, 2014). By contrast, clinical features of CVA16 infection are mild and self-limiting, such as papulovesicular rash and fever (Mao et al., 2014). CVA6 and CVA10 were reported as pathogens of HFMD initially in recent years, and the proportions of HFMD cases attributed by these viruses have increased sharply. Infection with CVA10 causes typical HFMD (Yang et al., 2015), while the clinical symptoms with CVA6 are different, usually including atypical herpes and onychomadesis (Bian et al., 2015). In addition, CVA6-associated HFMD presents high morbidity in adults. CVA6 has become the primary pathogen causing HFMD globally since 2009 (Bian et al., 2015). Variations in HFMD pathogens would lead to changes in the direction and strategy for prevention and treatment of this disease. In this context, mechanisms underlying the evolution and variability of pathogenic enterovirus serotypes need to be extensively explored.

Variations in the RNA virus genome during replication produce diverse populations. The populations will contain, or be able to quickly acquire, the necessary adaptive mutations to survive within the dynamic host environment. These mutations and selections are significant for fitness, evolvability and virulence of RNA viruses (Burch and Chao, 2000; Pfeiffer and Kirkegaard, 2005; Vignuzzi et al., 2006). Synonymous codon usage bias patterns are considered to reflect the action of natural selection, mutation and evolution of genomes. While synonymous mutation is often thought to be selectively neutral, the observed variation in codon usage across both viral and organismal taxa suggests the presence of mutational bias and/or selective pressure (Jenkins and Holmes, 2003; Plotkin and Kudla, 2011). Translational efficiency has been invoked as a selective pressure to explain the evolution of codon bias, codon pair bias and codon order (Cannarrozzi et al., 2010; Coleman et al., 2008; Tuller et al., 2010). In RNA viruses, constraints on RNA structures necessary for replication and packaging have been recognized as another important basis for codon bias (Goodfellow et al., 2000; Simmonds and Smith, 1999).

In recent years, more attention has been paid to codon usage bias patterns of enteroviruses which induce HFMD (Liu et al., 2011). At least three studies have evaluated the codon usage of complete RNA sequences of EV71 and described the influence of mutation pressure and natural selection in the evolution of this virus (Liu et al., 2011; Ma et al., 2014; Zhang et al., 2014). However, comparative research of codon bias patterns across several main HFMD pathogens in EV, which would imply potential information on prevalence and pathogenicity of dominant viruses, is still lacking.

The capsid proteins (i.e., structural proteins) of EV determine infectious specificity and prevalence, whereas the non-structural proteins determine replication features, pathogenicity and virulence. The codon usage patterns of structural and non-structural gene regions in HFMD viruses are also unclear. In the present study, synonymous codon usage patterns of major pathogenic serotypes, including CVA6, A10, A16 and EV71 isolated recently (from 2008 to 2015), were analyzed. Additionally, codon usages of structural and non-structural genes of these viruses were analyzed separately. A final aim was to understand the contribution of mutation pressure and natural selection to codon usage bias of HFMD pathogens.

2. Materials and methods

2.1. Genomes of viruses

Sixty-eight full-length genomes of HFMD virus strains isolated from 2008 to 2015 (15 for CVA6, 9 for CVA10, 18 for CVA16 and 26 for EV71) were chosen in this study (Table S1). Strains with duplicate information in the region, date of isolation and sub-genotype were avoided. Therefore, the chosen groups represented diversity in the above three aspects.

2.2. Calculation of relative synonymous codon usage (RSCU)

The RSCU value is defined as the ratio of the observed codon usage frequency to the frequency expected, and it is rarely influenced by the amino acid composition. RSCU values of P1 and P2-P3 regions in the CVA6, A10, A16 and EV71 genomes were calculated according to a previously described method (Sharp et al., 1986). A RSCU value of a codon near 1.00 means that the observed frequency is equal to the expected frequency, indicating that this codon lacks bias and is used randomly and equally. Correspondingly, a value of RSCU >1 (or RSCU <1) means that the codon is more (or less) frequently used than expected.

2.3. Analysis of codon usage bias

Effective number of codons (ENC) is generally used to quantify the codon usage bias for individual genes. ENC values of P1 and P2-P3 regions in CVA6, A10, A16 and EV71 genomes were calculated according to a previously described method (Wright, 1990). ENC values range from 20 to 61. An ENC value of 20 means that this gene uses only one synonymous codon for the corresponding amino acid, showing extreme bias, whereas a value of 61 indicates no bias and all synonymous codons are equally used.

2.4. Correspondence analysis (COA)

COA is a commonly used multivariate statistical analysis for studying the major trends in codon usage patterns. COA uses a mathematical procedure to transform RSCU values into uncorrelated 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 (Liu et al., 2011).

2.5. Neutral evolution analysis

The neutrality analysis is used to estimate the varying roles of mutational pressure and natural selection in the P1 and P2-P3 regions in the CVA6, A10, A16 and EV71 genomes. In the analysis, the GC12s value of the synonymous codon was plotted against its GC3s value (Sueoka, 1988), and then the slope of the regression line, correlation (r) and significance (p) were calculated.

2.6. Statistical analysis

Statistical analysis was performed using the two-tailed t-test. Data are expressed as the mean ± standard deviation (SD).

3. Results

3.1. Synonymous codon usage of P1 and P2-P3 regions in CVA6, A10, A16 and EV71 genomes

The RSCU values of the P1 region and P2-P3 region of the four viruses were calculated and shown in Tables S2 and S3. Notably, codon bias of 7 amino acids (Lys, Val, His, Glu, Asp, Asn and Gln) were found to be identical in full-length genomes of nearly all serotypes, while the other 11 (Arg, Ile, Leu, Phe, Thr, Tyr, Ala, Cys, Gly, Pro, Ser) were not (Table 1). In comparing the human codon bias of the 18 amino acids with that of the four viruses, they were shown to be the same at Lys, Val, His, Glu and Asn, and nearly the same at Ile, Thr, Arg and Tyr (Table 1). We observed an interesting phenomenon that Lys, Val, Ile and Thr are essential amino acids for humans, Arg and His are semi-essential amino acids, and Glu is a significant precursor of the deamination reaction, which produces α-ketoglutarate as an important intermediate of the tricarboxylic acid (TCA) cycle. Codon bias of an amino acid in humans is