Infection, Genetics and Evolution 14 (2013) 105–110

Contents lists available at SciVerse ScienceDirect

Infection, Genetics and Evolution

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

Short communication

The analysis of codon bias of foot-and-mouth disease virus and the adaptation of this virus to the hosts

Jian-hua Zhou [a,1], Zong-liang Gao [a,1], Jie Zhang [a], Yao-zhong Ding [a], Laszlo Stipkovits [b], Susan Szathmary [b], Zygmunt Pejsak [c,*], Yong-sheng Liu [a,*]

[a] State Key Laboratory of Veterinary Etiological Biology, National Foot-and-Mouth Disease Reference Laboratory, Lanzhou Veterinary Research Institute, Chinese Academy of Agricultural Sciences, Lanzhou, 730046 Gansu, PR China
[b] RT-Europe Center, Budapest, Hungary
[c] Department of Swine Diseases, National Veterinary Research Institute, 57 Partyzantow, 24-100 Puławy, Poland

ARTICLE INFO

Article history:
Received 23 May 2012
Received in revised form 23 September 2012
Accepted 25 September 2012
Available online 4 December 2012

Keywords:
Foot-and-mouth disease virus
Synonymous codon usage pattern
Genetic diversity
Mutation pressure
Translation selection

ABSTRACT

The synonymous codon usage patterns of open reading frame (ORF) in foot-and-mouth disease virus (FMDV), the similarity degree of the synonymous codon usage between this virus and the hosts and the genetic diversities of FMDV ORFs and the viral functional genes in viral ORF have been investigated by some simply statistical analyses. As for the synonymous codon usage of FMDV, some over-represented and under-represented codons have a similar usage in all seven serotypes. 33 out of 59 synonymous codons are similarly selected between FMDV ORF and the hosts. It is interesting that the overall codon usage pattern of the strains of serotype O isolated from pigs is different with that of strains of the same serotype isolated from non-pig origin, suggesting that the factor of pigs takes part in the formation of codon usage of FMDV serotype O. Projection of codon usage of nine viral functional genes onto the two-dimensional map represents that even though viral functional genes have various genetic diversities and each gene is not separated from each other based on seven serotypes, the codon usage patterns of VP2, 2C, 3A, 3C and 3D genes belonging to serotype O strains isolated from pigs are different with those of the same serotype strains from non-pig origin. In addition, the interaction between GC_{12}% and GC_3% of viral functional genes indicates that the codon usage patterns of VP1, VP2, 2B, 3A, 3C and 3D genes are influenced by mutation pressure from virus. Furthermore, distribution plots of ENC value vs. GC_3% for viral function genes indicate that mutation pressure is not the only factor in the formation of codon usage of these genes. The results suggest that both the mutation pressure from virus and the translation selection from the hosts take part in the evolution process of viral functional genes of FMDV.

1. Introduction

Synonymous codons are not selected randomly and equally in different viral genomes or in the functional genes in the same viral genome. Two conventional forces have been proposed to explain codon bias of genomes in viruses, namely mutational pressure from virus and translation selection from host. In nature, the RNA viruses are characterized by high mutation rates and high variation of synonymous codons usage. Because the viruses are simple intracellular parasites for the susceptible cells, they represent appreciable variation in the selectivity with which they infect the different hosts (Bahir et al., 2009). Co-evolution and adaptation of the viruses to the hosts were mostly studied by analyzing the synonymous codon usage bias in the whole coding sequence and the specific functional genes (Cutter et al., 2006; Duret, 2002; Grantham et al., 1980; Liu et al., 2010; Wong et al., 2010; Zhao et al., 2003; Zhou et al., 2012). Foot-and-mouth disease virus (FMDV) can cause a highly infectious disease of cloven-hoofed livestock including sheep, pig and cattle. As for the severe economic and social effects of a foot-and-mouth disease (FMD) outbreak, FMD has been recognized as the most important constraint on international trade in animals and animal products by the World Organization for Animal Health (OIE) (Leforban, 1999). FMDV is one of the members in the *Aphthovirus* genus which is classified into the *Picornaviridae* family. This virus exists in the form of seven different serotypes: O, A, C, Asia 1 and South African Territories 1 (SAT 1), SAT 2 and SAT 3, but many subtypes of the virus have evolved within each serotype (Knowles and Samuel, 2003). The genome of FMDV consists of a single open reading frame (ORF) flanked by a 5' untranslated region and a 3'untranslated region. The ORF encodes twelve viral proteins, from 5′ to 3′ termini of

* Corresponding authors. Tel.: +86 931 8342771; fax: +86 931 8340977 (Y.-s. Liu).
E-mail addresses: zpejsak@piwet.pulawy.pl (Z. Pejsak), liuyongshengvip8@163.com (Y.-s. Liu).
[1] These authors contributed equally to this work.

1567-1348/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.meegid.2012.09.020

the genome, including the leader protease (L^{pro}), four capsid proteins (VP4, VP2, VP3 and VP1) and seven nonstructural proteins 2A-C and 3A-D (Klein, 2009). The relationship and interaction between the nucleotide variation of FMDV and the environment of the hosts has been focused by some researchers in recent years. At the aspect of the synonymous codon usage patterns of FMDV, the translation rate of the structural protein (VP1) of FMDV could be improved by optimization of the codon usage pattern of FMDV VP1 gene (Song et al., 2004). Thereafter, Zhong et al. (2007); Zhou et al. (2010a,b) analyzed the synonymous codon usage patterns of seven serotypes of FMDV including 40 ORFs, and pointed out that mutation pressure plays an important role in the shaping codon usage pattern of this virus (Zhong et al., 2007; Zhou et al., 2010a), but they just focused on the role of mutation pressure deriving from virus and failed to estimate the effect of translation selection from the hosts of FMDV. Furthermore, the region flanked by the two start codons in L gene and the cleavage site in polyprotein of FMDV were analyzed by the synonymous codon usage bias, respectively (Zhou et al., 2011, 2010b), suggesting that the synonymous codon usage has a strong tendency to be selected in some specific region in FMDV ORF. Based on these previous reports, a better investigation into the genetic diversity of each functional viral gene along the FMDV ORF and the co-evolution between FMDV and the hosts is necessary to identify the characteristics of codon usage among the viral functional genes and the adaptation degree of this virus to the hosts.

2. Materials and methods

2.1. Sequences and other databases

The 144 open reading frames (ORF) of FMDV were downloaded from the National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/GenBank/), and the detailed information was listed in Table S1. To investigate the codon usage patterns of the viral functional genes within the FMDV ORF, the eleven functional genes including L, VP1-VP4, 2B-C (excluding the short nucleotide sequence of 2A) and 3A-D from each FMDV ORF were obtained by multiple sequence alignments with the Clustal W (1.7) computer programs (Thompson et al., 1994). Additionally, codon usage frequencies of pig (*Sus scrofa*), sheep (*Ovis aries*) and cattle (*Bos taurus*) were obtained from the codon usage database (Nakamura et al., 2000).

2.2. Comparative analysis of synonymous codon usage between FMDV ORF and its hosts

To investigate the variation without the confounding influence of amino acid composition of ORFs in each serotype of FMDV, respectively, the relative synonymous codon usage (RSCU) values for the ORF and the eleven viral functional genes were calculated according to the published equation (Sharp et al., 1986). Stop codons (UAA, UAG and UGA), AUG for Met and UGG for Try are not introduced into the RSCU analysis. For codon usage frequencies of each genomes of the sheep, cattle and pig, the RSCU values were also calculated for the 59 synonymous codons by the formula mentioned above. In order to identify the usage bias of the 59 synonymous codons, Wong et al., (2010) established a standard that codons with RSCU values > 1.6 were regarded as over-represented, while codons with RSCU values < 0.6 were thought to be under-represented (Wong et al., 2010). Therefore, a synonymous codon with RSCU value < 0.6 or > 1.6 can be defined as a bias one in this study. In comparison of synonymous codon usage pattern between FMDV and the hosts, if both RSCU values for a specific codon of FMDV and that of the corresponding codon for host are less than 0.6 or more than 1.6 or between 0.6 and 1.6, this pattern will be thought to be a similar codon usage pattern. In this study, a group of codons whose RSCU values range from 0.6 to 1.6 need to be define again, namely, when both RSCU value of FMDV and that of the hosts for the same codon range 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 is thought to be similar.

2.3. Analyzing the codon usage pattern of the viral functional genes of FMDV ORF

The analysis of the correlation between the percentage of the nucleotide GC at the synonymous 1st and 2nd codon position (GC_{12}%) and that of the nucleotide GC at the synonymous 3rd codon position (GC_3%) is useful to identify the effects of mutation pressure and translation selection on base composition, since the effects are present at all codon positions. To investigate degree of the consistent evolution along each viral functional gene among seven serotypes of FMDV, the GC_{12}% and the GC_3% of viral genes was analyzed based on Spearman's correlation method. The previous report indicated that the positive correlation between GC_{12}% and GC_3% is regarded as the hypothesis that the mutation pressure is the only determinant for the formation of codon usage (Jenkins and Holmes, 2003). In addition, to investigate the distributions of codon usage bias of the viral functional genes in FMDV, the 'effective number of codons' (ENC), the useful estimator of absolute codon bias, was also employed to quantify the codon bias for the coding sequence. The plot of ENC value against GC3% can be effectively applied to estimate the heterogeneity of the codon usage of FMDV genes. 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). If the codon usage pattern of a coding sequence is shaped only under the nucleotide composition caused by the mutation pressure, the ENC value can fall on a continuous curve (expected curve) (Gupta and Ghosh, 2001).

2.4. The genetic diversities of the viral functional genes and FMDV ORF by principle component analysis

To investigate the genetic diversity of FMDV strains in this study, principal component analysis (PCA) was employed, which reduces data dimensionality by performing a covariance analysis between factors (59 synonymous codons). This analysis gives a more convenient way to visualize the genetic characteristic of FMDV ORF. Based on the RSCU values, PCA was also carried out for the genetic characteristic of each viral gene with the size more than 300nt among seven serotypes of FMDV, including L, VP1-3, 2B, 2C, 3A, 3C and 3D genes.

3. Result

3.1. The synonymous codon usage pattern between FMDV ORF and its hosts

As for the overall codon usage patterns of seven serotypes of FMDV, it is interesting that some under-represented codons (UUA and CUA for Leu, AUA for Ile, GUA for Val, UAU for Tyr, CAU for His, AAU for Asn, GAU for Asp and CGA for Arg) and several over-represented codons (CUC for Leu, AUC for Ile, UAC for Tyr, CAC for His and AAC for Asn) are all existed in the seven serotypes of FMDV (Table 1). It is noted that the codon usage bias of the two synonymous codons for the amino acids Tyr, Asn and His are inversed. This phenomenon might serve as a genetic marker of FMDV. In addition, we found that the codons with CpG, CpC and