Gene 637 (2017) 115–123

Contents lists available at ScienceDirect

Gene

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

Analyses of nucleotide, codon and amino acids usages between peste des petits ruminants virus and rinderpest virus

Xiao-xia Ma[1], Qiu-yan Chang[1], Peng Ma[1], Lin-jie Li, Xiao-kai Zhou, De-rong Zhang, Ming-sheng Li, Xin Cao*, Zhong-ren Ma*

Engineering & Technology Research Center for Animal Cell, Gansu College of Life Science and Engineering, Northwest Minzu University, Gansu 730030, PR China

ARTICLE INFO

Keywords:
Peste des petits ruminants virus
Rinderpest virus
Synonymous codon usage
Nucleotide pair
Mutation pressure
Translation selection
Host

ABSTRACT

Peste des petits ruminants virus (PPRV) and rinderpest virus (RPV) are two causative agents of an economically important disease for ruminants (i.e., sheep, cattle and goat). In this study, the nucleotide, codon and amino acid usages for PPRV and RPV have been analyzed by multivariate statistical methods. Relative synonymous codon usage (RSCU) analysis represents that ACG for Thr and GCG for Ala are selected with under-representation in both PPRV and RPV, and AGA for Arg in PPRV and AGG for Arg in RPV are used with over-representation. The usage of nucleotide pair (CpG) tends to be removed from viral genes of the two viruses, suggesting that other evolutionary forces take part in evolutionary processes for viral genes in addition to mutation pressure from nucleotide usage at the third codon position. The overall nucleotide usage of viral gene is not major factor in shaping synonymous codon usage patterns, while the nucleotide usages at the third codon position and the nucleotide pairs play important roles in shaping synonymous codon usage patterns. Although PPRV and RPV are closely related antigenically, the codon and amino acid usage patterns for viral genes represent a significant genetic diversity between PPRV and RPV. Moreover, the overall codon usage trends for viral genes between PPRV and RPV are mainly influenced by mutation pressure from nucleotide usage at the third codon position and translation selection from hosts. Taken together, this is first comprehensive analyses for nucleotide, codon and amino acid usages of viral genes of PPRV and RPV and the findings are expected to increase our understanding of evolutionary forces influencing viral evolutionary pathway and adaptation toward hosts.

1. Introduction

Peste des petits ruminants virus (PPRV) and rinderpest virus (RPV), which belong to members of the *Morbillivirus* genus of the *Paramyxoviridae* family, can cause highly contagious and devastating viral diseases of ruminants (Gibbs et al., 1979; Plowright, 1962). Like other morbiliviruses, PPRV and RPV are enveloped, single-stranded, nonsegmented, negative-sense RNA viruses, and these two viruses share similar genomes which own six genes in order 5′ L-H-F-M-P-N 3′. PPRV and RPV have a common with some immunological cross-reactions and relatively similar clinical syndromes (Kumar et al., 2014). Small ruminants are the major hosts of PPRV, and some other ruminant species can be also infected by this virus (Banyard et al., 2010). RPV can result in fatal diseases in cattle, yaks and buffaloes (Carrillo et al., 2010). Notably, transmission of PPRV from infected goats to cattle and wild animals has been recently focused (Lembo et al., 2013), and RPV also infects small ruminant species and wildlife species without clinical syndromes (Anderson, 1995), suggesting that PPRV and RPV can switch hosts. A successful viral life-cycle often needs several properties, for example, the virus has the capability to infect the host cells and controls cellular translation systems and directs them toward the efficient production of new viruses. All viruses are characterized by very high natural mutation rates, and RNA viruses often represent an exceptionally higher rate than DNA viruses (Drake and Holland, 1999; Jenkins and Holmes, 2003). Co-evolution and adaptation of viruses to their natural hosts have been mostly studied by comparing mutations at synonymous and non-synonymous coding site in specific genes and indicating a strong correlation between nucleotide usage patterns and synonymous codon usage patterns caused by multiple evolutionary forces (Bahir et al., 2009; Butt et al., 2014; Lobo et al., 2009; Wong et al., 2010; Zhou et al., 2013a; Zhou et al., 2013c).

In nature, 20 amino acids plus three stop codons are coded by 64 codons. Although this redundancy enhances the resistance of genes to mutation: the third codon letters can be interchanged without affecting

* Corresponding authors.
E-mail addresses: caoxin@xbmu.edu.cn (X. Cao), mzr@xbmu.edu.cn (Z.-r. Ma).
[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.gene.2017.09.045
Received 20 May 2017; Received in revised form 3 September 2017; Accepted 21 September 2017
Available online 22 September 2017
0378-1119/

the primary sequence of protein, synonymous codons are not selected in random and all synonymous codons might be integrated parts of the genetic code with equal importance in maintaining its functional integrity (Biro, 2008). This phenomenon of synonymous codon usage has been studied in a wide range of organisms, from prokaryotes to eukaryotes and viruses, because genetic codes are regarded as a link between nucleotide and amino acid and are involved in many biological functions (i.e. translation efficiency and protein structure). There are several factors in shaping synonymous codon usages, including secondary protein structure, replication and selective transcription, hydrophobicity, hydrophilicity of the protein, the external environment, mutation pressure and translational selection (Ma et al., 2015; Ma et al., 2013; Plotkin and Kudla, 2011; Rosano and Ceccarelli, 2009; Sharp et al., 2010; Zhang et al., 1994; Zhou et al., 2012; Zhou et al., 2013b; Zhou et al., 2006; Zouridis and Hatzimanikatis, 2008). As of yet, the comprehensive studies for genetic features of nucleotide, synonymous codon and amino acid usages for viral genes of PPRV and RPV have not been reported. Here, a better understanding the genetic diversity of each functional viral gene in the viral genome and the co-evolution between PPRV, RPV and their natural hosts is necessary to identify the characteristics of codon usage among the viral functional genes and the adaptation degree of virus to its host.

2. Materials and methods

2.1. Sequence data

The 9 genomes of PPRV and the 9 genomes of RPV were downloaded from the National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.gov/Genbank/). The sequence data of PPRV includes AJ849636: *Ovis aries* (sheep), FJ905304: goat, EU267273: goat, HQ197753: caprine goat, EU267274: goat, KC594074: alpine goat, JX217850: wild bharal, NC_006383: *Ovis aries* (sheep) and JF939201: goat. The sequence data of RPV includes AB547190: *Oryctolagus cuniculus* var. *domesticus* (rabbit), NC_006296: cattle, AB547189: *Bos Taurus* (cattle), X98291: cattle, JN234010: bovine, JN234009: bovine, JN234008: bovine and Z30697: vaccine strain. To estimate the synonymous codon usage patterns and the codon usage bias of viral genes, the six functional genes (L, P, F, H, M and N) from each viral genome were obtained by multiple sequence alignments with the Clustal W (1.7) computer program (Thompson et al., 1994). Additionally, codon usage frequencies of sheep (*Ovis aries*) and cattle (*Bos taurus*) were obtained from the codon usage database (Nakamura et al., 2000).

2.2. Analyses of nucleotide usages of viral genes

The nucleotide usages were calculated for viral genes of PPRV and RPV, including the total frequency of occurrence of each nucleotide (U %, C%, A% and G%), the frequency of occurrence of each nucleotide at different codon positions (U1%, U2%, U3%, C1%, C2%, C3%, A1%, A2%, A3%, G1%, G2% and G3%), the frequencies of occurrence of nucleotides GC at the first & second codon positions (GC12%) and GC at the third codon position (GC3%).

2.3. Calculation of nucleotide pairs in viral genes

To investigate effects of nucleotide pairs on synonymous codon usage patterns of viral genes, the relative dinucleotides abundance was utilised in the study. Based on the previous report (Karlin and Burge, 1995), the relative abundance of nucleotide pairs in viral genes was calculated. The odds ratio was calculated depending on the following formula:

$$P(xy) = \frac{F(xy)}{F(x)F(y)} \times \frac{n^2}{n-1}$$

where $F(xy)$ stands for the frequency of occurrence of nucleotide pair (xy), $F(x)$ means the frequency of occurrence of nucleotide (x), $F(y)$ means the frequency of occurrence of nucleotide (y), n means the total number of nucleotide in the sequence. Compared with a random association of mononucleotides, when $P(xy)$ is > 1.23, the nucleotide pair (xy) is regarded as the over-represented nucleotide pair; when $P(xy)$ is < 0.78, the nucleotide pair (xy) is regarded as the under-represented nucleotide pair.

2.4. Analyzing the synonymous codon usage patterns of viral genes

In order to avoid the effects of gene lengths and amino acid compositions on the synonymous codon usage patterns, the relative synonymous codon usage value (RSCU) was utilised in the study, based on the previous report (Sharp et al., 1986). Notably, three stop codons (UGA, UAA and UAG), UGG for Try and AUG for Met are not utilised into the RSCU analyses. For codon usage frequencies of each genomes of the sheep and cattle, the RSCU values were also calculated for the 59 synonymous codons by the formula mentioned above.

$$\text{RSCU} = \frac{gij}{\sum_{j=1}^{ni} gij} \cdot ni$$

where g_{ij} is the observed number of the jth codon for the ith amino acid (which has n_i synonymous codons). RSCU values show the ratio between the observed usage frequency of one codon in a viral gene and the expected usage frequency in the synonymous codon family given that all codons for the particular amino acid are used in random. When RSCU value is 1, the corresponding synonymous codon is no bias at codon usage.

In addition, to further investigate the genetic diversity of the specific viral gene between PPRV and RPV at the aspect of codon usage bias (CUB), the CUB values (CUB = RSCU − 1) of 59 synonymous codons from the target viral genes in PPRV and RPV were calculated by complete linkage clustering with Euclidean distance, and the heat maps associated with CUB values from the specific viral gene were drawn by software Java TreeView (http://jtreeview.sourceforge.net/).

2.5. Calculation for the overall trend of codon usage for viral gene

To investigate the overall trends of codon usage for the viral functional genes, the ‘effective number of codons’ (ENC), the useful estimator of absolute codon bias, was also employed to quantify the overall trends of codon usage for the specific genes (Wright, 1990). The plot of ENC value versus GC3% can be effectively applied to estimate the heterogeneity of the codon usages of different 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). Remarkably, when ENC value is less than or equal to 35, the gene of interest is regarded as the significant codon bias.

2.6. Calculating the relative amino acid usage value of viral genes

To better analyze the amino acid usage patterns for viral proteins, we referenced the formula for RSCU, and therefore a simple formula for the relative amino acid usage (RAAU) values of the target amino acid sequence was employed in the study.

$$RAAU_{ij} = \frac{M_{ij}/N_i}{1/20}$$

where $RAAU_{ij}$ denotes the relative magnitude of the amino acid j that is selected by the ith amino acid sequence, M_{ij} denotes the usage frequency of the amino acid j in the ith amino acid sequence, and N_i denotes the total number of amino acids in the ith amino acid sequence. Particularly, Met and Trp can be included in this RAAU value calculation.