Infection, Genetics and Evolution 35 (2015) 27–33

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

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

Short communication

Evolutionary characterization of Tembusu virus infection through identification of codon usage patterns

Hao Zhou [a,1], Bing Yan [a,1], Shun Chen [a,b,c,*,1], Mingshu Wang [a,b,c], Renyong Jia [a,b,c], Anchun Cheng [a,b,c,*]

[a] Institute of Preventive Veterinary Medicine, Sichuan Agricultural University, Chengdu, Sichuan 611130, PR China
[b] Avian Disease Research Center, College of Veterinary Medicine of Sichuan Agricultural University, Chengdu, Sichuan 611130, PR China
[c] Key Laboratory of Animal Disease and Human Health of Sichuan Province, Sichuan Agricultural University, Chengdu, Sichuan 611130, PR China

ARTICLE INFO

Article history:
Received 10 March 2015
Received in revised form 15 July 2015
Accepted 20 July 2015
Available online 20 July 2015

Keywords:
Tembusu virus
Codon usage pattern
Mutation pressure
Cross-species transmission

ABSTRACT

Tembusu virus (TMUV) is a single-stranded, positive-sense RNA virus. As reported, TMUV infection has resulted in significant poultry losses, and the virus may also pose a threat to public health. To characterize TMUV evolutionarily and to understand the factors accounting for codon usage properties, we performed, for the first time, a comprehensive analysis of codon usage bias for the genomes of 60 TMUV strains. The most recently published TMUV strains were found to be widely distributed in coastal cities of southeastern China. Codon preference among TMUV genomes exhibits a low bias (effective number of codons (ENC) = 53.287) and is maintained at a stable level. ENC-GC3 plots and the high correlation between composition constraints and principal component factor analysis of codon usage demonstrated that mutation pressure dominates over natural selection pressure in shaping the TMUV coding sequence composition. The high correlation between the major components of the codon usage pattern and hydrophobicity (Gravy) or aromaticity (Aromo) was obvious, indicating that properties of viral proteins also account for the observed variation in TMUV codon usage. Principal component analysis (PCA) showed that CQW1 isolated from Chongqing may have evolved from GX2013H or GX2013G isolated from Guangxi, thus indicating that TMUV likely disseminated from southeastern China to the mainland. Moreover, the preferred codons encoding eight amino acids were consistent with the optimal codons for human cells, indicating that TMUV may pose a threat to public health due to possible cross-species transmission (birds to birds or birds to humans). The results of this study not only have theoretical value for uncovering the characteristics of synonymous codon usage patterns in TMUV genomes but also have significant meaning with regard to the molecular evolutionary tendencies of TMUV.

1. Introduction

Tembusu virus (TMUV), a single-stranded, positive-sense RNA virus with a genome length of approximately 11 kb, was first identified in mosquitoes from Malaysia in 1955 (Platt et al., 1975). TMUV is a member of the *Flavivirus* genus in the family *Flaviviridae* and contains a unique open reading frame (ORF) encoding three structural proteins, including core (C), pre-membrane (prM), and envelope (E), and seven nonstructural (NS) proteins, including NS1, NS2A, NS2B, NS3, NS4A, NS4B, and NS5 (Chambers et al., 1990; Li et al., 2012a; Tang et al., 2012). TMUV has recently attracted increasing attention, the main reasons for which are its association with serious illness, heavy declines in avian egg production, and severe neurological symptoms in avian species, with ducks being particularly susceptible (Cao et al., 2011; Yan et al., 2011). This virus can have devastating effects on poultry farming and cause serious economic losses in the waterfowl industry. In addition, as a novel member of *Flavivirus*, TMUV should receive more attention because it represents a potential threat to public health. Indeed, cell-adapted duck Tembusu virus (DTMUV) presenting antibody-dependent enhancement (ADE) was able to replicate in a mouse model (Liu et al., 2013). With regard to humans, TMUV antibodies (71.9%) and RNA (47.7%) were detected in samples collected from duck farm workers in Shandong, China (Tang et al., 2013b). Nevertheless, the infectious route and evolutionary mechanisms of TMUV infection remain largely unknown; as such, the possibility of zoonotic transmission of TMUV from birds to humans should not be overlooked.

* Corresponding authors at: Institute of Preventive Veterinary Medicine, Sichuan Agricultural University, No. 211 Huimin Road, Wenjiang District, Chengdu, Sichuan Province 611130, PR China.
E-mail addresses: sophia_cs@163.com (S. Chen), chenganchun@vip.163.com (A. Cheng).
[1] These authors contributed equally as co-first authors of this work.

http://dx.doi.org/10.1016/j.meegid.2015.07.024
1567-1348/

For viruses, little information exists regarding the extent and origin of synonymous codon variation. Recent efforts to elucidate codon usage biases in *Flaviviridae* have concentrated mainly on Dengue virus (DGV) (Zhou et al., 2013), West Nile virus (WNV) (Moratorio et al., 2013), and Hepatitis C virus (HCV) (Hu et al., 2011). In contrast, investigations of TMUV have primarily focused on viral isolation and identification (Huang et al., 2013; Tang et al., 2013a), the establishment and application of diagnostic methods (Li et al., 2012b; Yun et al., 2012), and genetic analyses of complete genome sequences (Tang et al., 2012; Wan et al., 2012; Zhu et al., 2012), whereas information regarding codon usage patterns is not available. Therefore, we examined the codon usage bias of TMUV to clarify the factors influencing codon usage patterns of the TMUV genome, which may provide a unique and valuable foundation for a better understanding of the molecular evolutionary process of TMUV. Furthermore, a detailed understanding of the extent and causes of codon usage biases is critical for exploring the interplay between mutation pressure and natural selection. To our knowledge, this is the first comprehensive study to systematically investigate the codon usage bias of TMUV.

2. Materials and methods

2.1. Sequence data

A total of 60 genomic TMUV sequences were extracted from the National Center for Biotechnology Information (NCBI) database. One completed genome of TMUV was isolated from a duck in Chongqing and sequenced by our laboratory (Accession number KM233707). After removing redundant and repeated sequences, a coding sequence (CDS) analysis of all TMUV strain genomes was performed to investigate the characteristics of codon usage variation. Detailed information about these strains is listed in Supplementary Table S1, and the distribution of these strains in the different provinces of China is shown in Fig. 1.

2.2. Relative synonymous codon usage

Relative synonymous codon usage (RSCU) can measure the degree of synonymous codon usage bias, avoiding the unnecessary influence of amino acid composition in certain genes. Furthermore, an RSCU value of 1 shows that a codon has a random translation selection; an RSCU value greater than 1 means that a codon has a high frequency, and vice versa.

2.3. The ENC-GC3s plot

The effective number of codons (ENC) of sequences varies from 20 to 61, with a lower value indicating a stronger codon usage bias. Moreover, larger ENC values are associated with weaker codon preference. The GC contents of the third codon positions (GC3s) in TMUV were also calculated. The expected ENC-GC3 curve has been widely utilized to determine whether codon usages of given genes are affected by mutation only or also by other factors such as natural selection (Wright, 1990).

2.4. Principal component analysis

Principal component analysis (PCA) is a widely used multivariate statistical method that was employed in this study to analyze the major trends among different TMUV strains in a codon usage model. Each strain of TMUV was represented as a 59 dimensional vector (RSCU value of each codon), excluding the codons ATG and TGG and the three stop codons. The first principal component and the second principal component, namely f1 and f2, were extracted to visually determine the genetic relationship of each TMUV.

2.5. Data processing

The primary indices mentioned above and the calculation procedures were performed using the Codon W program and SPSS 20.0 software. The Spearman's rank correlation analysis and cluster analysis by Euclidean distance were performed with SPSS 20.0 software.

3. Results

3.1. Codon usage bias and synonymous codon usage among TMUV strains

To investigate the degree of codon usage variation in TMUV, we determined the details of the RSCU values of all strains (Table 1). The results showed a preference for codons ending with A/T versus C/G, at a ratio of 11:7. Four preferred codons end with G, whereas three end with C; in addition, the use of codons ending with A was more frequent than the use of codons ending with T (6:5). Furthermore, the data show that the frequency of a preferred codon ending with a particular nucleotide was essentially stable in TMUV genes.

3.2. Compositional properties of TMUV strains

The overall base composition of the 60 TMUV strains evaluated was nonrandom. There was no remarkable difference between A and G, with mean values of 28.59% and 28.96%, respectively. However, the mean value of the T content (22.52%) was found to be higher than that of the C content (19.93%). Evidently, the G + C value fluctuated from 48.66% to 49.32%, with a mean value of 48.88%, and an S.D. of 0.001, whereas the mean value of GC3s was 47.56%, with an S.D. of 0.001 (Table S2). The differences in nucleotide content suggest that composition constraints are determinants of the codon usage pattern of TMUV. The values of ENC among the TMUV strains were found to be similar, varying from 52.85 to 53.54, with a mean of 53.287 and an S.D. of 0.149 (Table S2), suggesting that the extent of codon preference in the TMUV genomes is not strongly biased (ENC > 40) and is maintained at a stable level.

3.3. The main determining factor of codon usage bias in TMUV

In general, mutation pressure and natural selection are two main factors that shape codon usage bias. A plot of actual ENC values against both GC3s (%) and the expected ENC values provides a useful display of trends in codon usage. In this study, all of the points lie below the expected curve (Fig. 2), indicating that although the TMUV genome is principally influenced by mutational pressure, other factors may be responsible for shaping the codon usage bias of TMUV. To further confirm whether the mutation pressure is due to natural selection pressure or viral mutation pressure, a correlation analysis was implemented to analyze relationships among the G + C content at the first and second codon positions (GC12s) and that at the synonymous third codon positions (GC3s). A significant correlation was observed ($r = -0.264$, $p = 0.042$), demonstrating that mutation pressure dominates over natural selection pressure in shaping the TMUV nucleotide composition.