



OPEN ACCESS

EDITED BY

Xue Bai,
Institute of Special Animal and Plant
Sciences (CAAS), China

REVIEWED BY

Lisanework Ayalew,
University of Prince Edward
Island, Canada
Yan Zeng,
Jilin Agriculture University, China
Zhaocai Li,
Lanzhou Veterinary Research Institute
(CAAS), China

*CORRESPONDENCE

Fu-qiang Xie
xiefq@lzu.edu.cn

†These authors have contributed
equally to this work and share first
authorship

SPECIALTY SECTION

This article was submitted to
Animal Nutrition and Metabolism,
a section of the journal
Frontiers in Veterinary Science

RECEIVED 14 June 2022

ACCEPTED 25 July 2022

PUBLISHED 12 August 2022

CITATION

Wang X, Sun J, Lu L, Pu F-y, Zhang D-r
and Xie F-q (2022) Evolutionary
dynamics of codon usages for peste
des petits ruminants virus.
Front. Vet. Sci. 9:968034.
doi: 10.3389/fvets.2022.968034

COPYRIGHT

© 2022 Wang, Sun, Lu, Pu, Zhang and
Xie. This is an open-access article
distributed under the terms of the
[Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is
permitted, provided the original
author(s) and the copyright owner(s)
are credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which does
not comply with these terms.

Evolutionary dynamics of codon usages for peste des petits ruminants virus

Xin Wang^{1†}, Jing Sun^{2†}, Lei Lu¹, Fei-yang Pu³,
De-rong Zhang³ and Fu-qiang Xie^{4*}

¹School of Stomatology, Lanzhou University, Lanzhou, China, ²Geriatrics Department, The Second Hospital of Lanzhou University, Lanzhou, China, ³Center for Biomedical Research, Northwest Minzu University, Lanzhou, China, ⁴Maxillofacial Surgery Department, The Second Hospital of Lanzhou University, Lanzhou, China

Peste des petits ruminants virus (PPRV) is an important agent of contagious, acute and febrile viral diseases in small ruminants, while its evolutionary dynamics related to codon usage are still lacking. Herein, we adopted information entropy, the relative synonymous codon usage values and similarity indexes and codon adaptation index to analyze the viral genetic features for 45 available whole genomes of PPRV. Some universal, lineage-specific, and gene-specific genetic features presented by synonymous codon usages of the six genes of PPRV that encode N, P, M, F, H and L proteins reflected evolutionary plasticity and independence. The high adaptation of PPRV to hosts at codon usages reflected high viral gene expression, but some synonymous codons that are rare in the hosts were selected in high frequencies in the viral genes. Another obvious genetic feature was that the synonymous codons containing CpG dinucleotides had weak tendencies to be selected in viral genes. The synonymous codon usage patterns of PPRV isolated during 2007–2008 and 2013–2014 in China displayed independent evolutionary pathway, although the overall codon usage patterns of these PPRV strains matched the universal codon usage patterns of lineage IV. According to the interplay between nucleotide and synonymous codon usages of the six genes of PPRV, the evolutionary dynamics including mutation pressure and natural selection determined the viral survival and fitness to its host.

KEYWORDS

peste des petits ruminants virus, information entropy, synonymous codon usage, evolutionary dynamics, PPRV

Introduction

Peste des petits ruminants (PPR) caused by peste des petits ruminants virus (PPRV) is a highly contagious, acute and febrile viral disease of wild and domestic small ruminants, and poses a great threat to the ruminant industry in the world (1). PPRV was classified under the genus *Morbillivirus*, family *Paramyxoviridae*, and order *Mononegavirales* (2). This is an enveloped virus containing a single negative-strand RNA genome of about 16,000 nt in length and has six transcription units encoding nucleocapsid (N), phosphoprotein (P), matrix (M), fusion (F), hemagglutinin (H) and polymerase (L) (3).

H and F proteins (the two surface glycoproteins) function in the attachment and entry into the host cell. M protein is located on the inner surface of the viral membrane stabilizing the virion. N, P and L proteins are required for viral RNA polymerase activity. Because PPR produces a high mortality of up to 100% in immunologically naïve populations, it has been listed as a big threat to the development of sustainable agriculture by the Food and Agriculture Organization (FAO) and the World Organization for Animal Health (OIE) for eradication with the aim to globally eliminate PPR by 2030 (4).

The PPR outbreaks can cause high morbidity and mortality, resulting in severe economic losses in the developing countries. Hence, the analysis of epidemic tendency and evolutionary dynamics of PPRV for prevention and control remains particularly important. A Bayesian phylogenetic analysis of all PPRV lineages (I–IV) identified an ancestral PPRV and individual lineages of Nigeria for PPRV and Senegal for lineage I, Nigeria/Ghana for lineage II, Sudan for lineage III and India for lineage IV (5). In addition, some reports have put forwarded about the host range expansion and cross-species infection of PPRV (6–8). PPRV might switch hosts and spread more easily after eradication of rinderpest virus (RPV), similar to that the eradication of smallpox virus created a niche for cowpox and monkeypox viruses to cross the species barrier into humans (9). Based on PPR epidemics in large scale and rapidly spreading in China in the absence of RPV, the evolutionary dynamics of PPRV might provide potential opportunities for expanding the host range of PPRV in China to a certain extent. Previous reports on the evaluation of molecular epidemiology of PPRV, which was based on nucleotide usage patterns of small regions of partial sequences or the whole genome, were carried out and figured out some genetic variations for PPRV (5, 10, 11). Compared with nucleotide usage variations of coding sequences, the genetic codons consisting of triplets of nucleotides are generally redundant, and most of the amino acids can be coded by more than one codon. This phenomenon is referred to as “synonymous codon usage bias”. This bias acts as a key factor in modulating the efficiency and accuracy of protein production and maintaining the same amino acid sequence of the protein. The analysis of synonymous codon usage patterns reflects several evolutionary and functional factors in shaping the synonymous codon usage bias, including translational/natural selection, mutation pressure, host, genetic drift, gene expression, secondary protein structure and fine-tuning translation kinetics selection (12–21). Based on the knowledge on codon usages, optimization of synonymous codon usage can be frequently required for the efficient expression of genes in heterologous host systems (22–25). For better achievement of the general viral

fitness, survival and evasion of the host immune system and evolution, the interplay between synonymous codon and amino acid usages of the virus and that of its host were thought to be the key evolutionary factors (26, 27). Therefore, knowledge of synonymous codon usage of PPRV enlightens PPRV molecular evolution and extends our insights into the regulation of viral gene expression.

Materials and methods

Information about full genome of PPRV

The 45 whole genome sequences of PPRV strains available were downloaded from the National Center for Biotechnology (NCBI) Genbank database, accessed on 1 September 2017 (Supplementary Table S1). Based on coding sequence annotations of the 45 PPRV strains, the six coding sequences (F, H, L, M, N and P) were extracted from the corresponding genome. According to the given coding sequences of PPRV, the following nucleotide contents were calculated for the coding sequences by MEGA 6.0 software: (1) the overall nucleotide usage patterns (T%, A%, C% and G%); and (2) the nucleotide usage patterns at the 1st, 2nd, and 3rd codon positions (T1, A1, C1, G1, T2, A2, C2, G2, T3, A3, C3, and G3%). Depending on the statistical test (One-way ANOVA), the overall nucleotide usage patterns and nucleotide usage patterns at the 1st, 2nd, and 3rd codon positions were described in each gene, respectively. To further investigate the genetic diversity of PPRV at nucleotide levels, a phylogenetic tree was constructed with all the full genome sequences by producing neighbor-joining trees with Kimura 2-parameter model of base substitution (Gamma distributed rate and between transitional and transversal substitutions) using MEGA 6.0 software (28).

Nucleotide usage bias by information entropy method

As for the nucleotide usage bias at gene levels of the 45 PPRV strains, the normalized information entropy over the frequencies of different nucleotides in a given gene was presented by the below formula (29):

$$Entropy = - \frac{\sum_i f_i \times \log_2(f_i)}{4}$$

$$f_i = \frac{F_i}{F(A) + F(T) + F(G) + F(C)}$$

where f_i is the probability of the specific nucleotide (F_i), and F_i is the total number of occurrences of the specific nucleotide

Abbreviations: PPR, Peste des petits ruminants; PPRV, Peste des petits ruminants virus; FAO, Food and Agriculture Organization; OIE, World Organization for Animal Health; NCBI, National Center for Biotechnology; RSCU, relative synonymous codon usage; CAI, codon adaptation index.

in the target gene ($i = A, T, G$ or C). The value of *Entropy* for nucleotide usage bias ranges from 0 to 1, representing how the dispersed contribution of these four types of nucleotides is: the higher the value is, the more uniform the nucleotide usage is; in contrast, a lower value reflects a more biased usage of nucleotides.

To further compare the nucleotide usage biases in the six coding sequences, the overall nucleotide usage biases, and the nucleotide usage biases at the 1st, 2nd, and 3rd codon positions were estimated by One-way ANOVA method in SPSS 16.0 software, respectively.

Relative synonymous codon usage calculation

The relative synonymous codon usage (RSCU) values for the given coding sequences of the PPRV strains were calculated to quantify synonymous codon usage bias without the confounding influence of amino acid usage patterns or the length of different gene samples (30). Of note, two-thirds of PPRV strains were isolated from China and were classified into lineage IV (Supplementary Table S1). To better investigate the synonymous codon usage patterns of each gene of epidemic PPRV in China, we classified these strains into six groups (groups I–VI) for RSCU calculation and presented varied extents of synonymous codon usage bias in each group. In detail, the groups I, II and III corresponded to lineages I, II and III, respectively, and the groups IV, V and VI corresponded to lineage IV of foreign countries, China (2007–2008) and China (2013–2015), respectively. To identify synonymous codons with over-representation or under-representation, the synonymous codons with RSCU value of more than 1.6 and <0.6 were considered as over-represented or under-represented ones, respectively (21).

Analysis for evolutionary distance between two different gene samples by RSCU data

To quantify the extent of similarity of codon usages between the two gene samples, a similarity index for $D(A,B)$ was introduced into this study (31).

$$R(A, B) = \frac{\sum_{i=1}^{59} a_i \times b_i}{\sqrt{\sum_{i=1}^{59} a_i \times \sum_{i=1}^{59} b_i}}$$

$$D(A, B) = \frac{1 - R(A, B)}{2}$$

where $R(A,B)$ is defined as a cosine value of an included angle between A and B special vectors, meaning that the evolutionary distance between gene A and gene B at the aspect of 59 RSCU values, a_i is defined as the RSCU value for a specific codon in 59 synonymous codons of gene A , and b_i is termed as the RSCU value for the same codon of gene B . Here, the lower the $D(A,B)$ value is, the higher the extent of similarity of codon usage patterns between gene A and gene B .

Codon adaptation index for PPRV genes

The codon adaptation index (CAI) analysis of PPRV coding sequences were carried out depending on the CAIcal server (32), which was considered to be an improved CAI calculation measure, and estimated the expression level of a coding sequence in the host cell. The CAIcal webserver, freely available at <http://genomes.urv.es/CAIcal>, calculated the CAI for a group of viral sequences using the specific host reference set and included a complete set of tools related with codon usage adaptation. The host reference set required to calculate the CAI can be introduced in the codon usage database of the host. The synonymous codon usage patterns of host as reference, the synonymous codon usage bias with high extent represented the highest relative adaptation to the host, and coding sequences with higher CAI values should be regarded preferred over those with lower ones (32). The synonymous codon usage frequencies of *Ovis aries* (natural host of PPRV) were selected as the reference set, and the related data was obtained from the Codon Usage Database (33).

Statistical analysis

To better estimate the role of nucleotide usage bias at different codon positions in the overall codon usage bias, the Davies-Bouldin index (R_{ij}) (34) was introduced in this work. This index represented a ratio of within-group and between-group distances, and was defined as:

$$R_{ij} = \frac{X_i + X_j}{M_{ij}}$$

$$M_{ij} = \sqrt{(A_i - A_j)^2 + (B_i - B_j)^2}$$

$$X_i = \sqrt{\frac{\sum_{i=1}^{T_i} (B_i - A_i)^2}{T_i}}$$

and

$$X_j = \sqrt{\frac{\sum_{j=1}^{T_j} (B_j - A_j)^2}{T_j}}$$

X_i is the standard deviation of the Euclidean distance between each point (B_i) in the i th group and the centroid (A_i) of the i th group; X_j means the standard deviation of the Euclidean distance between each point (B_j) in the j th group and the centroid (A_j) of the j th group, T_i and T_j standard for the total numbers of points in the i th group and in the j th group, respectively. M_{ij} is the Euclidean distance between the centroids (A_i and A_j) of the i th group and the j th group. The smaller the R_{ij} value is, the stronger the interaction between the two groups.

One-way ANOVA method was used to compare the means of two or more groups containing numerical response data using the software SPSS 16.0 for Windows, and significant difference can be identified when p -value was <0.05 . Linear regression was used for modeling the relationship between a scalar dependent variable and one independent variable using the software GraphPad Prism 6 for Windows.

Results

Nucleotide usages in different genes of PPRV

To quantify nucleotide composition of the six genes of PPRV, each base composition has been calculated for the 45 PPRV strains. The mean contents of A, A1, U2 and A3% were the highest in F gene (Figure 1A), the A, A1, U2, and U3% were the highest in H gene (Figure 1B), the A, A1, A2, and C3% were the highest in L gene (Figure 1C), the A, A1, U2, and C3% were the highest in M gene (Figure 1D), the A, G1, U2 and G3% were the highest in N gene (Figure 1E), and the A, G1, A2, and C3% were the highest in P gene (Figure 1F). As shown in Figure 1, there were significant differences in the overall nucleotide usages and in the first, second and third codon positions of each PPRV gene ($P < 0.0001$). Generally, the nucleotide usage patterns represented the gene-specific compositional trends rather than the similar compositional trends with the overall content of nucleotides (Supplementary Table S2).

Nucleotide usage bias of viral gene by information entropy

According to the variations of nucleotide content at different codon positions of PPRV genes (Figure 1), information entropy was performed to quantify nucleotide usage bias at gene levels and the bias at the three nucleotide positions

of codons in viral gene. There were significant differences in the nucleotide usage biases at gene levels, the 1st, 2nd, and 3rd codon positions, respectively (Figure 2). As shown in Figure 2, the overall nucleotide usage bias in H gene was the highest (Figure 2A), the nucleotide usage bias in the 1st position in L gene was the highest (Figure 2B), the nucleotide usage bias in the 2nd position in N gene was the highest (Figure 2C), and the nucleotide usage bias in the 3rd position in L gene was the highest (Figure 2D). Nucleotide usage biases at gene levels and at different codon positions quantified by information entropy showed that the nucleotide mutations in different viral coding sequences resulted in the evolutionary dynamics. Although nucleotide usage biases at gene levels and at different codon positions represented a gene-specific characteristic (Supplementary Table S3), information entropy can comprehensively quantify the trends of nucleotide usage bias caused by four nucleotide contents and confirm the low extent of nucleotide substitution rates in the six genes of PPRV.

Synonymous codon usage bias of PPRV coding sequences

The RSCU analysis were carried out to quantify the extent of synonymous codon usage bias in the six coding sequences of PPRV (Supplementary Tables S4–S9). All over-representative synonymous codons were not the codons associated with G/C end or A/U end, and most synonymous codons containing CpG had a weak tendency to be selected by PPRV genes (Table 1). Interestingly, compared to those synonymous codons of lineage I (Supplementary Table S8), the synonymous codons (UGU for Cys and CAU for His) were never selected by the N gene of PPRV lineages II, III and IV, indicating that the synonymous codon usage was one of the evolutionary dynamics associated with PPRV. PPRV coding sequences had a weak tendency to select synonymous codons containing CpG dinucleotides (RSCU < 1.0 , Supplementary Tables S4–S8), except CGU for Arg in P gene (Supplementary Table S9). The comparisons of RSCU data for the coding sequences of groups IV, V and VI represented a stable Synonymous usage pattern in lineage IV PPRV strains of other countries, China (2007–2008) and China (2013–2015) (Supplementary Table S4–S9), suggesting that the synonymous codon usage pattern could sustain the difference between lineages of PPRV rather than that of outbreaks or countries. In addition, we adopted RSCU method for analyzing the usage pattern of the three stop codons. Despite the three canonical stop codons with the same biological function (gene translation end), the stop codon UGA was only selected by H gene and UAA was selected by P gene. Moreover, F and L genes strongly tended to select UAG as stop codon, M and N genes strongly tended to select UAA as stop codon, and UGA was never selected by F, L, M, and N genes of PPRV. These phenomena indicate that the

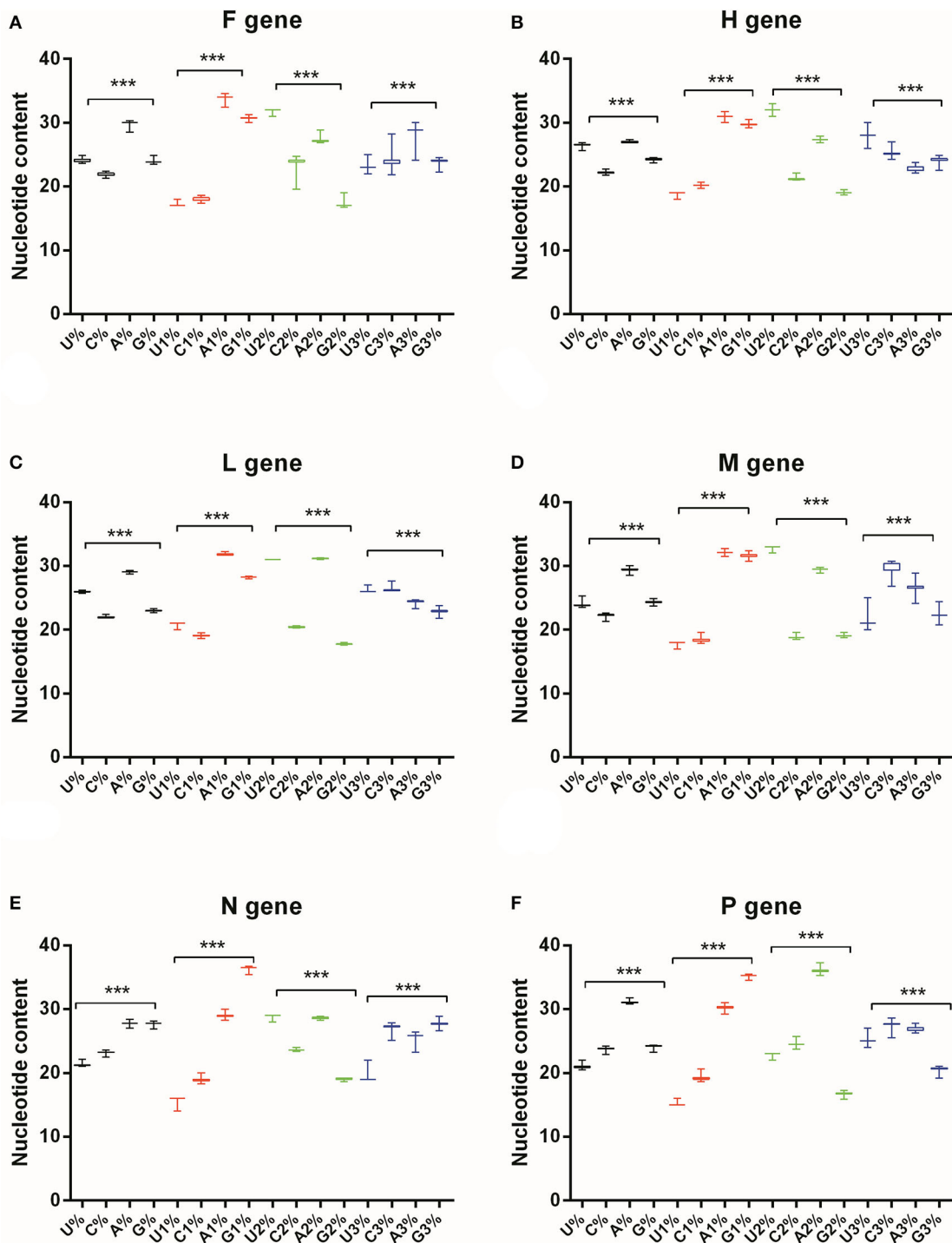


FIGURE 1
 Nucleotide content of PPRV coding sequences and different codon positions. U, C, A, and G% are the overall nucleotide content of coding sequence; U1, C1, A1, and G1% are nucleotide content in the first codon position; U2, C2, A2, and G2% are nucleotide content in the second codon position; U3, C3, A3, and G3% are nucleotide content in the third codon position. The One-way ANOVA method is used for estimating the differences of nucleotide usage patterns. **(A)** F gene, **(B)** H gene, **(C)** L gene, **(D)** M gene, **(E)** N gene, **(F)** P gene. When p -value < 0.05 , it means a significant difference between the given groups. Of note, because all analyses of One-Way ANOVA in this figure produce p -values < 0.001 , they are remarked as "****".

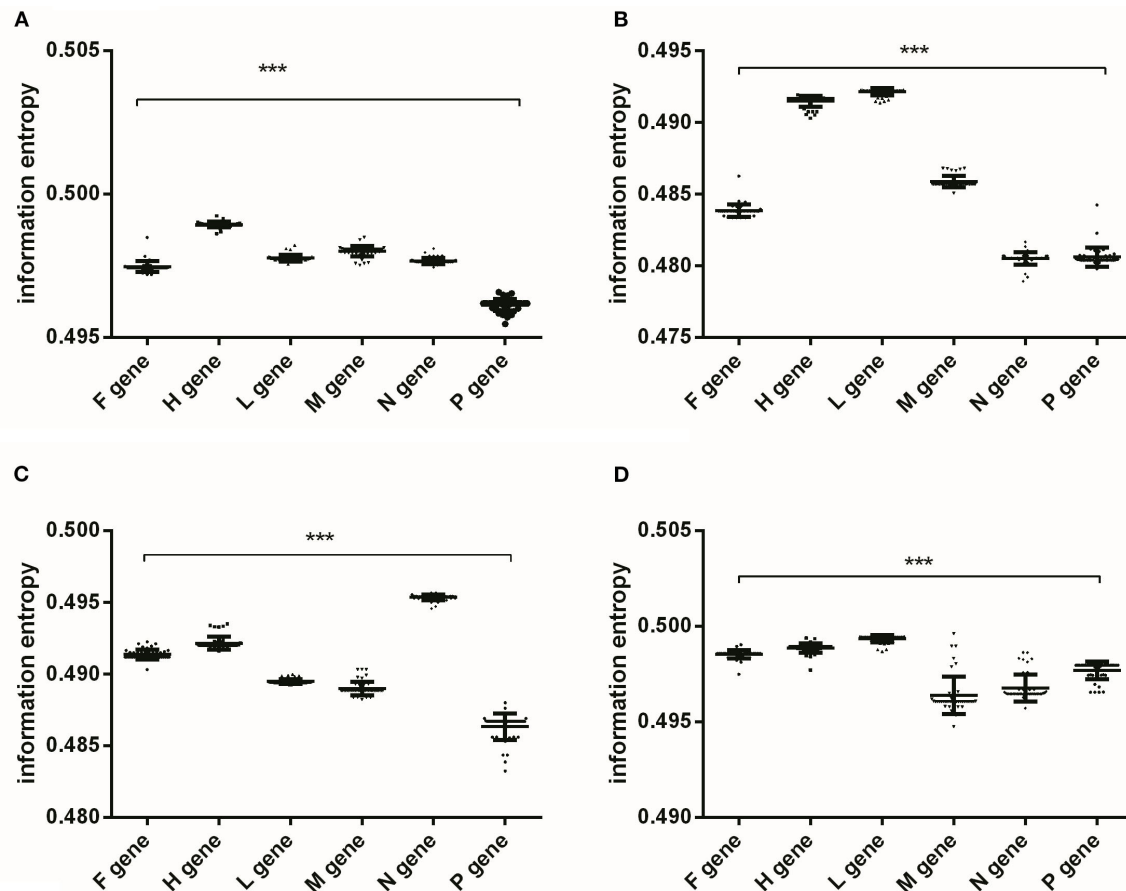


FIGURE 2

The analysis of nucleotide usage bias of PPRV coding sequences performed by information entropy. (A) the overall nucleotide usage bias at gene levels, and significant differences of the overall nucleotide usage biases among the six coding sequences; (B) the nucleotide usage bias at the first codon position, and significant differences of nucleotide usage bias at the first codon position among the six genes; (C) the nucleotide usage bias at the second codon position, and significant differences of nucleotide usage bias at the second codon position among the six genes; (D) the nucleotide usage bias at the third codon position, and significant differences of nucleotide usage bias at the third codon position among the six genes. The One-way ANOVA method is used for estimating the differences of nucleotide usage biases among the six genes, and p -value < 0.001 remarks as “***”.

PPRV coding sequences had evolved lineage- and gene-specific synonymous codon usage patterns.

Phylogenetic analysis

According to phylogenetic analysis *via* neighbor-joining model, although the lineage IV was not monophyletic, this lineage was able to be separated from other three lineages (Figure 3). Furthermore, the PPRV strains isolated from China (2013–2014) could be classified into a distinct clade in lineage IV, while the strains firstly emerged in China (2007–2008) were grouped into another distinct clade in lineage IV (Figure 3). To some degree, the genetic diversity of PPRV strains isolated from China displayed the geographic trace in comparison of

that of PPRV strains (KJ867542, NC_006383, KJ867541 and KC594074) isolated from non-Asia region. In addition, the PPRV strains with lineage III owned the isolated evolutionary pathway, compared with those of lineages I, II and IV. However, the two PPRV strains with lineage I displayed the obvious genetic divergence at the genome level (Figure 3).

Although the PPRV strains from China (2007–2008), China (2013–2015) and other countries generally shared a similar synonymous codon usage pattern (Supplementary Tables S4–S9), the differences of synonymous codon usage clarified the genetic divergence of the three groups [China (2007–2008), China (2013–2015) and other countries]. *D(A,B)* analysis also found highly similar extent of synonymous codon usage in viral genes of the three groups (Table 2) based on the data derived from the Supplementary Tables S4–S9.

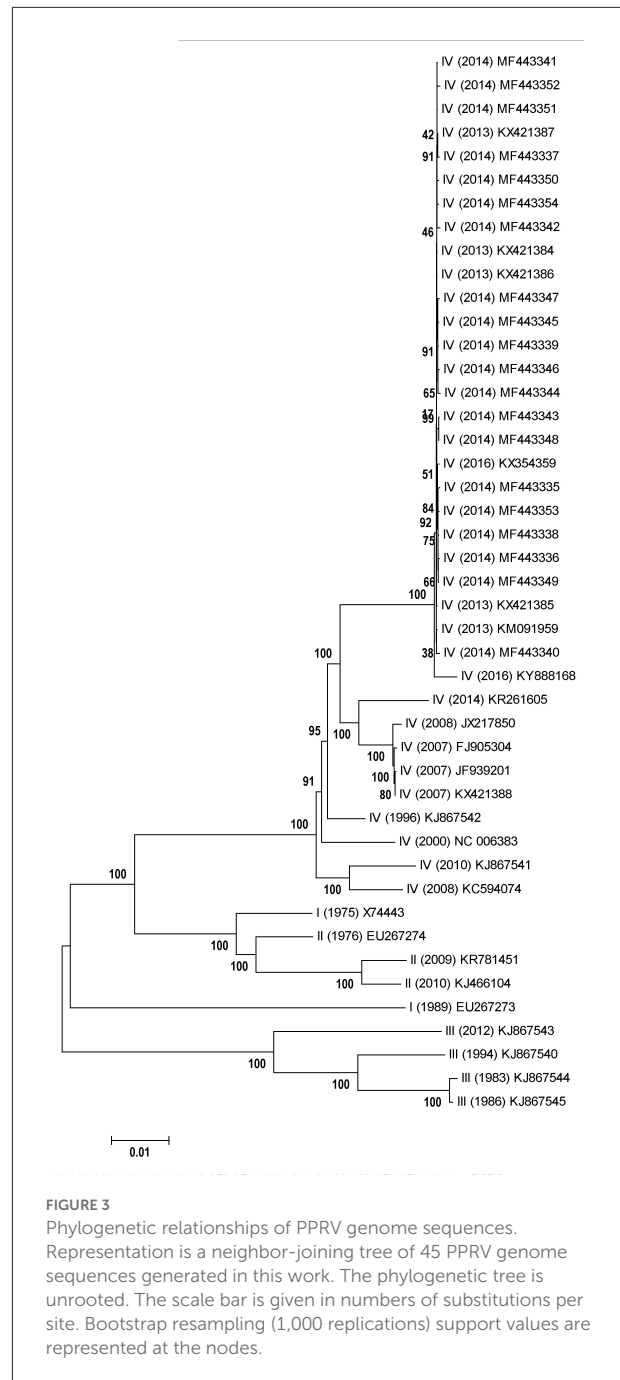
TABLE 1 The over-/under-representative synonymous codons in the six transcriptions of PPRV genome.

	Over-representative synonymous codons	Under-representative synonymous codons
F	UCA for Ser, CCA for Pro, GGG for Gly	CCG for Pro, ACG for Thr, GCG for Ala, GGU for Gly, CGC, CGA and CGG for Arg
H	UCA for Ser, AGA and AGG for Arg	UCG for Ser, ACG for Thr, GCG for Ala, CGU, CGC and CGA for Arg
L	AGA and AGG for Arg	UCG for Ser, ACG for Thr, GCG for Ala, CGU and CGC for Arg
M	UCA for Ser, CCC for Pro, UGC for Cys and AGA for Arg	UCC and UCG for Ser, CCG for Pro, ACG for Thr, GCG for Ala, UGU for Cys, CGU and CGG for Arg
N	UCA for Ser, AGA and AGG for Arg, and GGU for Gly	UUA for Leu, AGC for Ser, ACG for Thr, GCG for Ala, CGU, CGC, CGA and CGG for Arg
P	GUC for Val and AGA for Arg	GUA for Val, UCG for Ser, ACG for Thr, GCG for Ala, CGC, CGA and CGG for Arg

The extent of the overall codon usage similarity of lineage IV between other countries and China (2013–2015) was generally higher than that between other countries and China (2007–2008) or between China (2007–2008) and China (2013–2015) (Table 2). In addition, the extent of overall codon usage similarity of M and N genes between other countries and China (2007–2008) was less than that between China (2007–2008) and China (2013–2015) (Table 2). These results reflected that genetic divergence in synonymous codon usage of viral genes of PPRV lineage IV was complex and synonymous codon usage was able to alleviate the adverse effect by nucleotide mutation in viral genes.

PPRV presenting host-specific codon adaption patterns

The CAI analysis were performed to estimate the correlation between the synonymous codon usage bias and the expression efficiencies of gene samples of PPRV, implying that the strong codon adaptation of the viral genes fit the host (*Ovis aries*) cellular machinery. Based on the classification of gene types, a strong significant difference was found (Supplementary Table S3), implying that the six viral genes of PPRV might have different gene expression levels in the host cells. Generally, the viral gene (N gene) had the highest expression level, while F gene owned the lowest



one (Supplementary Figure S1), suggesting that PPRV had developed gene-specific codon usage patterns for adaption to the codon usage of host cellular environment. Based on the classification of lineages, significant differences were discovered among the four lineages of PPRV (Figure 4), suggesting that the six genes had developed lineage-specific codon usage patterns. To better compare the synonymous codon usage pattern between PPRV and its natural host (*Ovis aries*) based on RSCU data for *Ovis aries* (21), the six genes of PPRV generally shared

TABLE 2 Evolutionary distances among strains of PPRV lineage IV isolated from China (2007–2008 and 2013–2015) and other countries.

	<i>D(A,B)</i>	<i>D(A,C)</i>	<i>D(B,C)</i>
F gene	0.0027	0.0013	0.0027
H gene	0.0020	0.0028	0.0042
L gene	0.0003	0.0005	0.0007
M gene	0.0033	0.0011	0.0020
N gene	0.0050	0.0010	0.0022
P gene	0.0024	0.0018	0.0061

A means strains isolated from other countries.

B means strains isolated from Tibet in China (2007–2008).

C means strains isolated from other provinces in China (2013–2015).

a similar synonymous codon usage pattern with *Ovis aries*. However, some rare synonymous codons in *Ovis aries* were preferably selected by PPRV genes, including UCA (Ser) and AAU (Asn) in F gene, UCA (Ser) in H gene, UUG (Leu), UCA (Ser) and AAU (Asn) in L gene, CUA (Leu) and UCA (Ser) in M gene and UCA (Ser) in P gene. These synonymous codons likely mediated and regulated the relevant viral genes translation due to its low corresponding tRNA abundance in the host.

As shown in [Supplementary Figure S1](#), the high level of adaptation of viral synonymous codon usages to that of the host implied that PPRV coding sequences can be translated at relatively high efficiencies. The Davies-Bouldin index (*Rij* value) indicated that despite the obvious effects of nucleotide usage biases at different codon positions on the overall codon usage bias of PPRV coding sequences, nucleotide usage biases at the first and second codon positions played more vital roles than that at the third codon position in the codon usage bias for the coding sequences ([Table 3](#)).

Discussion

The high mutation rate of nucleotide usages in PPRV genome results in viral expansion both in geographical range and in the hosts it infects ([35](#)). It has been reported that some negative-sense single-stranded RNA viruses (such as Marburg virus and human metapneumovirus) contain all U- or A-ended codons to encode amino acids, because this synonymous codon usage pattern exhibited great association with its two nucleotides in high proportion in their genomes ([36, 37](#)). However, the nucleotide compositional patterns of the six coding sequences of PPRV were more complex than the generally analyzed AU and/or GC-rich compositions of most microorganisms. The gene-specific nucleotide usage pattern, the stable nucleotide usage patterns at the first and second codon positions and the various nucleotide usage patterns at the third codon position were obviously the genetic features of PPRV genes. This further suggests that the synonymous

codon usages can be considered as the evolutionary dynamics, alleviating the effects of nucleotide usage variation in viral genes on amino acid composition of viral protein. Previous studies have supported these PPRV genetic features and that the PPRV nucleotide variation throughout the complete genome proved genome plasticity, which might explain the viral ability to emerge and adapt in new geographic regions and hosts ([5, 38](#)). Although nucleotide usage variation finally influences the biological functions of viral proteins, synonymous codon usages play a non-negligible role in viral biological functions to achieve the viral evolutionary origins and adaptation to new hosts ([39–42](#)). For some microorganisms, including viruses, an AU-rich or GC-rich nucleotide composition was strongly correlated with their synonymous codon usage bias, in other words, an AU-rich genome tended to select synonymous codons with A/U ended, while a GC-rich genome strongly selected synonymous codons with G/C ended ([21, 36, 43–45](#)). If synonymous codon usage bias reflected such trends as mentioned above, the mutation pressure would play a dominant role in the codon usages. During the evolution of the negative-sense single-stranded RNA virus, mutation pressure remained the key factor that influenced the codon bias than natural selection in viral genes ([46, 47](#)). However, synonymous codon usage bias of PPRV coding sequences showed no AU or GC end, suggesting that the mutation pressure caused by nucleotide usage variation was not predominant in the PPRV evolutionary pathway. The frequencies of CpG and UpA dinucleotides played important roles in RNA virus replication and virulence, and nucleotide usage frequencies caused by dinucleotide usages meant selection pressures independent of coding capacity and profoundly influenced host-pathogen interactions ([48–50](#)). The rich CpG motif in genes can enhance immune response of the host against pathogens ([51–54](#)). Viral genes of PPRV should avoid selecting synonymous codons containing CpG dinucleotides. As for the respiratory syncytial virus, codon-optimized F gene with low level of CpG dinucleotides had higher expression of F, replicated more efficiently *in vivo*, and was more immunogenic ([55](#)). With poor CpG dinucleotides in the viral genes, PPRV can avoid stimulating strong immune responses of the host for immune escape. This suggests that apart from mutation pressure, other evolutionary dynamics related to natural selections played roles in the evolution of PPRV. Similar phenomena had been reported in influenza virus and foot-and-mouth disease virus ([21, 44](#)).

Flaviviridae family had a big epidemic around the world and their members had developed host- and vector-specific codon usage patterns to maintain successful replication and transmission chains within multiple hosts and vectors ([56, 57](#)). Some members of *Picornaviridae* family also demonstrated that the natural hosts played important roles in viral synonymous codon usages ([13, 21](#)). In the family *Paramyxoviridae*, codon usage patterns remained specific for each viral species and were markedly different among diverse hosts ([58](#)). CAI analysis for the six coding sequences of PPRV reflected good fitness of

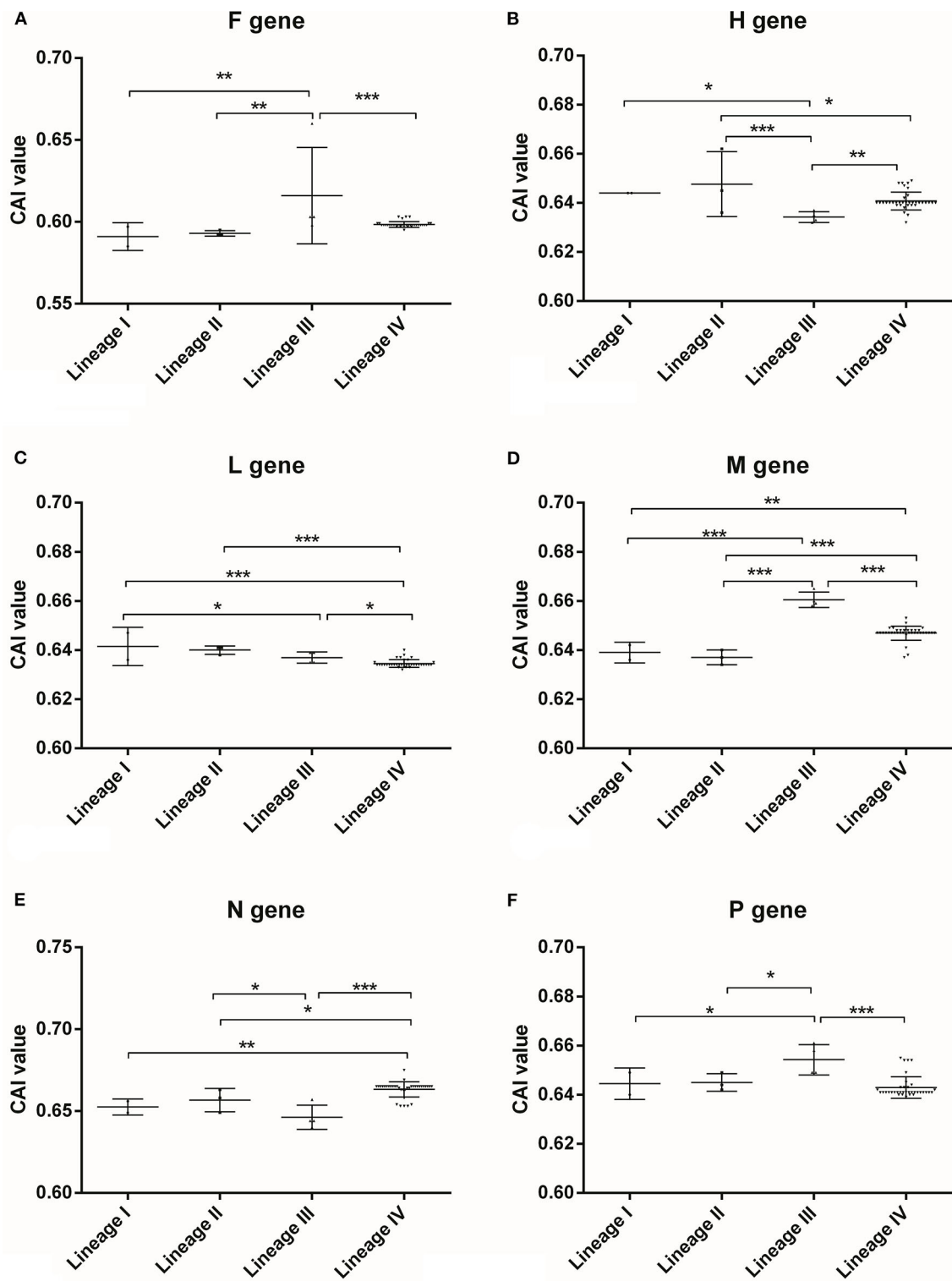


FIGURE 4

CAI analysis for PPRV coding sequences of different lineages in relation to its host performed by CAIcal server. CAI is frequently used as a measure of gene expression and to assess the adaptation of viral genes to their hosts, which indicates the influence of natural selection. The higher CAI value is, the more adaptation of synonymous codon usage of the target coding sequence is to its host. (A) F gene, (B) H gene, (C) L gene, (D) M gene, (E) N gene, (F) P gene. "*" means significant difference between the two different lineages with p -value <0.05 performed by One-way ANOVA method, "***" means significant difference between the two different lineages with p -value <0.01 performed by One-way ANOVA method, "****" means significant difference between the two different lineages with p -value <0.001 performed by One-way ANOVA method.

TABLE 3 Roles of nucleotide usage bias in the overall codon usage bias of PPRV genes.

	Rij(CAI,total)	Rij(CAI,N1)	Rij(CAI,N2)	Rij(CAI,N3)
F gene	0.096	0.087	0.092	0.098
H gene	0.037	0.037	0.037	0.038
L gene	0.021	0.021	0.020	0.022
M gene	0.040	0.038	0.040	0.045
N gene	0.045	0.042	0.044	0.048
P gene	0.038	0.037	0.040	0.040

The word 'total' means the overall nucleotide usage bias; 'N1' means nucleotide usage bias in the first codon position; 'N2' means nucleotide usage bias in the second codon position; 'N3' means nucleotide usage bias in the third codon position.

the virus to the host and high levels of viral gene expression in terms of codon usage pattern. Additional evidence has confirmed that usage of synonymous codons in protein coding sequences is necessarily biased and the overall codon usage pattern could match the tRNA pool of the host organism (27, 59–65). Since synonymous codon usage bias reflected tRNA abundance in host cells, and synonymous codon usage patterns of RNA virus, which was well fitted to its hosts, and might influence viral translation efficiencies (66–68). Even in the same genome, the synonymous codon usage patterns vary significantly among genes given their different expression levels, biological functions and tissue-specific patterns (69–71). Of note, some synonymous codons are preferentially selected over others at higher frequencies, resulting in synonymous codon usage bias, and is found in almost all available genomes. This biased synonymous codon usage is not neutral but involved in nucleotide usage bias (72, 73), mRNA stability (74, 75), translation accuracy, efficiency (76, 77), and protein folding formation (78).

In China, the first emergence of PPR occurred in Tibet China (2007) (79). Another outbreak of PPR occurred in wild small ruminants in Tibet China (2008) (79). PPR outbreak was not further reported until December 2013 in Xinjiang Yili (80). Although this epidemic was effectively controlled through a series of effective measures, PPR had widely and rapidly intruded into 21 provinces due to the movement of small ruminants (81). Although the two PPR epidemics in China presented genetic divergence in nucleotide usage bias and synonymous codon usage bias in viral genes, the two groups still had shared lineage-specific features in synonymous codon usage pattern. Since PPRV contains the gene for the RNA-dependent RNA polymerase NS5B in its genome and the polymerase does not have proofreading activity reading, PPRV has a high error rate leading to genetic heterogeneity and the formation of quasispecies. Synonymous codon is regarded as a linker between nucleotide and amino acid usages, resulting in enhancing fault tolerance of PPRV proteins caused by viral quasispecies to some degree. Moreover, synonymous codon usage bias

derived from the homeostasis between natural selection and mutation pressure is a universal phenomenon across the genomes of microorganisms and profoundly influences genomic evolution (36, 56).

Previously, the major bottleneck limiting in better understanding of the genetic features of PPRV was their dependence on nucleotide usage variation. Although nucleotide usage variation can be regarded as evolutionary dynamics of PPRV genome, synonymous codon usage patterns of PPRV coding sequences carry more genetic information, including viral adaptation to hosts, viral gene expression, and effects on the biological functions of viral protein.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary material.

Author contributions

Conceptualization: XW, F-yP, and F-qX. Methodology: F-yP, F-qX, XW, and D-rZ. Software: XW and F-yP. Formal analyses: F-yP, D-rZ, XW, and JS. Writing-review and editing: XW, JS, and F-qX. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by Department of Science and Technology of Gansu Province Nature and Science Fund [21JR1RA141], Cuiying Scientific and Technological Innovation Program of Lanzhou University Second Hospital [CY2021-BJ-A18, CY2019-MS07], Cuiying Postgraduate Student Supervisor Culture Plan [CYDSPY202005], Innovation Foundation for Higher Education Institution of Gansu Province [2020B-022], and Gansu Province Science and Technology Fund [20JR10RA737].

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those

of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Baron MD, Diallo A, Lancelot R, Libeau G. Peste des Petits Ruminants Virus. *Adv Virus Res.* (2016) 95:1–42. doi: 10.1016/bs.aivir.2016.02.001
- Sakalová A. [Occurrence and diagnosis of congenital enzymopathic hemolytic anemias in Slovakia]. *Cesk Pediatr.* (1978) 33:141–4.
- Munir M, Zohari S, Suluku R, Leblanc N, Kanu S, Sankoh FA-R, et al. Genetic characterization of peste des petits ruminants virus, Sierra Leone. *Emerg Infect Dis.* (2012) 18:193–5. doi: 10.3201/eid1801.111304
- Banyard AC, Parida S, Batten C, Oura C, Kwiatek O, Libeau G. Global distribution of peste des petits ruminants virus and prospects for improved diagnosis and control. *J Gen Virol.* (2010) 91:2885–97. doi: 10.1099/vir.0.025841-0
- Muniraju M, Munir M, Parthiban AR, Banyard AC, Bao J, Wang Z, et al. Molecular evolution of peste des petits ruminants virus. *Emerg Infect Dis.* (2014) 20:2023–33. doi: 10.3201/eid2012.140684
- Balamurugan V, Sen A, Venkatesan G, Bhanot V, Yadav V, Bhanuprakash V, et al. Peste des petits ruminants virus detected in tissues from an Asiatic lion (*Panthera leo persica*) belongs to Asian lineage IV. *J Vet Sci.* (2012) 13:203–6. doi: 10.4142/jvs.2012.13.2.203
- Khalafalla AI, Saeed IK, Ali YH, Abdurrahman MB, Kwiatek O, Libeau G, et al. An outbreak of peste des petits ruminants (PPR) in camels in the Sudan. *Acta Trop.* (2010) 116:161–5. doi: 10.1016/j.actatropica.2010.08.002
- Lembo T, Oura C, Parida S, Hoare R, Frost L, Fyumagwa R, et al. Peste des petits ruminants infection among cattle and wildlife in northern Tanzania. *Emerg Infect Dis.* (2013) 19:2037–40. doi: 10.3201/eid1912.130973
- de Swart RL, Duprex WP, Osterhaus ADME. Rinderpest eradication: lessons for measles eradication? *Curr Opin Virol.* (2012) 2:330–4. doi: 10.1016/j.coviro.2012.02.010
- Intisar KS, Ali YH, Haj MA, Sahar M a. T, Shaza MM, Baraa AM, Ishag OM, Nouri YM, Taha KM, Nada EM, et al. Peste des petits ruminants infection in domestic ruminants in Sudan. *Trop Anim Health Prod.* (2017) 49:747–54. doi: 10.1007/s11250-017-1254-3
- Zhou XY, Wang Y, Zhu J, Miao Q-H, Zhu LQ, Zhan SH, et al. First report of peste des petits ruminants virus lineage II in *Hydropotes inermis*, China. *Transbound Emerg Dis.* (2018) 65:e205–9. doi: 10.1111/tbed.12683
- Akashi H. Synonymous codon usage in *Drosophila melanogaster*: natural selection and translational accuracy. *Genetics.* (1994) 136:927–35. doi: 10.1093/genetics/136.3.927
- Aragonès L, Guix S, Ribes E, Bosch A, Pintó RM. Fine-tuning translation kinetics selection as the driving force of codon usage bias in the hepatitis A virus capsid. *PLoS Pathog.* (2010) 6:e1000797. doi: 10.1371/journal.ppat.1000797
- Gajbhiye S, Patra PK, Yadav MK. New insights into the factors affecting synonymous codon usage in human infecting *Plasmodium* species. *Acta Trop.* (2017) 176:29–33. doi: 10.1016/j.actatropica.2017.07.025
- Lafay B, Lloyd AT, McLean MJ, Devine KM, Sharp PM, Wolfe KH. Proteome composition and codon usage in spirochaetes: species-specific and DNA strand-specific mutational biases. *Nucleic Acids Res.* (1999) 27:1642–9. doi: 10.1093/nar/27.7.1642
- Ma X-X, Chang Q-Y, Ma P, Li L-J, Zhou X-K, Zhang D-R, et al. Analyses of nucleotide, codon and amino acid usages between peste des petits ruminants virus and rinderpest virus. *Gene.* (2017) 637:115–23. doi: 10.1016/j.gene.2017.09.045
- Saunders R, Deane CM. Synonymous codon usage influences the local protein structure observed. *Nucleic Acids Res.* (2010) 38:6719–28. doi: 10.1093/nar/gkq495
- Shah P, Gilchrist MA. Explaining complex codon usage patterns with selection for translational efficiency, mutation bias, and genetic drift. *Proc Natl Acad Sci USA.* (2011) 108:10231–6. doi: 10.1073/pnas.1016719108
- Sharp PM, Emery LR, Zeng K. Forces that influence the evolution of codon bias. *Philos Trans R Soc Lond B Biol Sci.* (2010) 365:1203–12. doi: 10.1098/rstb.2009.0305
- Wang Y-N, Ji W-H, Li X-R, Liu Y-S, Zhou J-H. Unique features of nucleotide and codon usage patterns in mycoplasmas revealed by information entropy. *Biosystems.* (2018) 165:1–7. doi: 10.1016/j.biosystems.2017.12.008
- Zhou J-H, Gao Z-L, Zhang J, Ding Y-Z, Stipkovits L, Szathmari S, et al. The analysis of codon bias of foot-and-mouth disease virus and the adaptation of this virus to the hosts. *Infect Genet Evol.* (2013) 14:105–10. doi: 10.1016/j.meegid.2012.09.020
- André S, Seed B, Eberle J, Schraut W, Bültmann A, Haas J. Increased immune response elicited by DNA vaccination with a synthetic gp120 sequence with optimized codon usage. *J Virol.* (1998) 72:1497–503. doi: 10.1128/JVI.72.2.1497-1503.1998
- Kane JF. Effects of rare codon clusters on high-level expression of heterologous proteins in *Escherichia coli*. *Curr Opin Biotechnol.* (1995) 6:494–500. doi: 10.1016/0958-1669(95)80082-4
- Smith DW. Problems of translating heterologous genes in expression systems: the role of tRNA. *Biotechnol Prog.* (1996) 12:417–22. doi: 10.1021/bp950056a
- Yadava A, Ockenhouse CF. Effect of codon optimization on expression levels of a functionally folded malaria vaccine candidate in prokaryotic and eukaryotic expression systems. *Infect Immun.* (2003) 71:4961–9. doi: 10.1128/IAI.71.9.4961-4969.2003
- Bahir I, Fromer M, Prat Y, Linial M. Viral adaptation to host: a proteome-based analysis of codon usage and amino acid preferences. *Mol Syst Biol.* (2009) 5:311. doi: 10.1038/msb.2009.71
- Novoa EM, Ribas de Pouplana L. Speeding with control: codon usage, tRNAs, and ribosomes. *Trends Genet.* (2012) 28:574–81. doi: 10.1016/j.tig.2012.07.006
- Tamura K, Stecher G, Peterson D, Filipinski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol.* (2013) 30:2725–9. doi: 10.1093/molbev/mst197
- Mioduser O, Goz E, Tuller T. Significant differences in terms of codon usage bias between bacteriophage early and late genes: a comparative genomics analysis. *BMC Genomics.* (2017) 18:866. doi: 10.1186/s12864-017-4248-7
- Sharp PM, Li WH. Codon usage in regulatory genes in *Escherichia coli* does not reflect selection for “rare” codons. *Nucleic Acids Res.* (1986) 14:7737–49. doi: 10.1093/nar/14.19.7737
- Zhou J, Zhang J, Sun D, Ma Q, Chen H, Ma L, et al. The distribution of synonymous codon choice in the translation initiation region of dengue virus. *PLoS ONE.* (2013) 8:e77239. doi: 10.1371/journal.pone.0077239
- Puigbò P, Bravo IG, Garcia-Valve S. CAIcal: a combined set of tools to assess codon usage adaptation. *Biol Direct.* (2008) 3:38. doi: 10.1186/1745-6150-3-38
- Nakamura Y, Gojobori T, Ikemura T. Codon usage tabulated from international DNA sequence databases: status for the year 2000. *Nucleic Acids Res.* (2000) 28:292. doi: 10.1093/nar/28.1.292
- Davies DL, Bouldin DW. A cluster separation measure. *IEEE Trans Pattern Anal Mach Intell.* (1979) 1:224–7.
- Nambulli S, Sharp CR, Acciardo AS, Drexler JF, Duprex WP. Mapping the evolutionary trajectories of morbilliviruses: what, where and whither. *Curr Opin Virol.* (2016) 16:95–105. doi: 10.1016/j.coviro.2016.01.019
- Nasrullah I, Butt AM, Tahir S, Idrees M, Tong Y. Genomic analysis of codon usage shows influence of mutation pressure, natural selection, and host features on Marburg virus evolution. *BMC Evol Biol.* (2015) 15:174. doi: 10.1186/s12862-015-0456-4

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fvets.2022.968034/full#supplementary-material>

37. Zhong Q, Xu W, Wu Y, Xu H. Patterns of synonymous codon usage on human metapneumovirus and its influencing factors. *J Biomed Biotechnol.* (2012) 2012:460837. doi: 10.1155/2012/460837
38. Wu X, Liu F, Li L, Zou Y, Liu S, Wang Z. Major mutation events in structural genes of peste des petits ruminants virus through serial passages in vitro. *Virus Genes.* (2016) 52:422–7. doi: 10.1007/s11262-016-1317-y
39. Bera BC, Virmani N, Kumar N, Anand T, Pavulraj S, Rash A, et al. Genetic and codon usage bias analyses of polymerase genes of equine influenza virus and its relation to evolution. *BMC Genomics.* (2017) 18:652. doi: 10.1186/s12864-017-4063-1
40. Chen Y, Li X, Chi X, Wang S, Ma Y, Chen J. Comprehensive analysis of the codon usage patterns in the envelope glycoprotein E2 gene of the classical swine fever virus. *PLoS ONE.* (2017) 12:e0183646. doi: 10.1371/journal.pone.0183646
41. He W, Zhang H, Zhang Y, Wang R, Lu S, Ji Y, et al. Codon usage bias in the N gene of rabies virus. *Infect Genet Evol.* (2017) 54:458–65. doi: 10.1016/j.meegid.2017.08.012
42. Roy A, Banerjee R, Basak S, HIV. Progression depends on codon and amino acid usage profile of envelope protein and associated host-genetic influence. *Front Microbiol.* (2017) 8:1083. doi: 10.3389/fmicb.2017.01083
43. Roychoudhury S, Mukherjee D, A. detailed comparative analysis on the overall codon usage pattern in herpesviruses. *Virus Res.* (2010) 148:31–43. doi: 10.1016/j.virusres.2009.11.018
44. Wong EHM, Smith DK, Rabadan R, Peiris M, Poon LLM. Codon usage bias and the evolution of influenza A viruses. Codon usage biases of influenza virus. *BMC Evol Biol.* (2010) 10:253. doi: 10.1186/1471-2148-10-253
45. Zhou J, Ding Y, He Y, Chu Y, Zhao P, Ma L, et al. The effect of multiple evolutionary selections on synonymous codon usage of genes in the *Mycoplasma bovis* genome. *PLoS ONE.* (2014) 9:e108949. doi: 10.1371/journal.pone.0108949
46. Kumar CS, Kumar S. Species based synonymous codon usage in fusion protein gene of Newcastle disease virus. *PLoS ONE.* (2014) 9:e114754. doi: 10.1371/journal.pone.0114754
47. Kumar CS, Kumar S. Synonymous codon usage of genes in polymerase complex of Newcastle disease virus. *J Basic Microbiol.* (2017) 57:481–503. doi: 10.1002/jobm.201600740
48. Atkinson NJ, Witteveldt J, Evans DJ, Simmonds P. The influence of CpG and UpA dinucleotide frequencies on RNA virus replication and characterization of the innate cellular pathways underlying virus attenuation and enhanced replication. *Nucleic Acids Res.* (2014) 42:4527–45. doi: 10.1093/nar/gku075
49. Burns CC, Campagnoli R, Shaw J, Vincent A, Jorba J, Kew O. Genetic inactivation of poliovirus infectivity by increasing the frequencies of CpG and UpA dinucleotides within and across synonymous capsid region codons. *J Virol.* (2009) 83:9957–69. doi: 10.1128/JVI.00508-09
50. Tulloch F, Atkinson NJ, Evans DJ, Ryan MD, Simmonds P. RNA virus attenuation by codon pair deoptimisation is an artefact of increases in CpG/UpA dinucleotide frequencies. *Elife.* (2014) 3:e04531. doi: 10.7554/eLife.04531.021
51. Antonialli R, Sulczewski FB, Amorim KN da S, Almeida B da S, Ferreira NS, Yamamoto MM, et al. CpG oligodeoxynucleotides and flagellin modulate the immune response to antigens targeted to CD8 α ⁺ and CD8 α ⁻ conventional dendritic cell subsets. *Front Immunol.* (2017) 8:1727. doi: 10.3389/fimmu.2017.01727
52. Ingelsson B, Söderberg D, Strid T, Söderberg A, Bergh A-C, Loitto V, et al. Lymphocytes eject interferogenic mitochondrial DNA webs in response to CpG and non-CpG oligodeoxynucleotides of class C. *Proc Natl Acad Sci USA.* (2018) 115:E478–87. doi: 10.1073/pnas.1711950115
53. Jenberie S, Thim HL, Sunyer JO, Skjødt K, Jensen I, Jørgensen JB. Profiling Atlantic salmon B cell populations: CpG-mediated TLR-ligation enhances IgM secretion and modulates immune gene expression. *Sci Rep.* (2018) 8:3565. doi: 10.1038/s41598-018-21895-9
54. Ma Y, Jiao Y-Y, Yu Y-Z, Jiang N, Hua Y, Zhang X-J, et al. A Built-In CpG Adjuvant in RSV F protein DNA vaccine drives a Th1 polarized and enhanced protective immune response. *Viruses.* (2018) 10:E38. doi: 10.3390/v10010038
55. Liang B, Ngwuta JO, Surman S, Kabatova B, Liu X, Lingemann M, et al. Improved prefusion stability, optimized codon usage, and augmented virion packaging enhance the immunogenicity of respiratory syncytial virus fusion protein in a vectored-vaccine candidate. *J Virol.* (2017) 91:e00189–17. doi: 10.1128/JVI.00189-17
56. Butt AM, Nasrullah I, Qamar R, Tong Y. Evolution of codon usage in Zika virus genomes is host and vector specific. *Emerg Microbes Infect.* (2016) 5:e107. doi: 10.1038/emi.2016.106
57. Lobo FP, Mota BEF, Pena SDJ, Azevedo V, Macedo AM, Tauch A, et al. Virus-host coevolution: common patterns of nucleotide motif usage in Flaviviridae and their hosts. *PLoS ONE.* (2009) 4:e6282. doi: 10.1371/journal.pone.0006282
58. Rima BK. Nucleotide sequence conservation in paramyxoviruses; the concept of codon constellation. *J Gen Virol.* (2015) 96:939–55. doi: 10.1099/vir.0.070789-0
59. Tian L, Shen X, Murphy RW, Shen Y. The adaptation of codon usage of +ssRNA viruses to their hosts. *Infect Genet Evol.* (2018) 63:175–9. doi: 10.1016/j.meegid.2018.05.034
60. Lyu X, Yang Q, Li L, Dang Y, Zhou Z, Chen S, et al. Adaptation of codon usage to tRNA I34 modification controls translation kinetics and proteome landscape. *PLoS Genet.* (2020) 16:e1008836. doi: 10.1371/journal.pgen.1008836
61. Frumkin I, Lajoie MJ, Gregg CJ, Hornung G, Church GM, Pilpel Y. Codon usage of highly expressed genes affects proteome-wide translation efficiency. *Proc Natl Acad Sci USA.* (2018) 115:E4940–9. doi: 10.1073/pnas.1719375115
62. Chan C, Pham P, Dedon PC, Begley TJ. Lifestyle modifications: coordinating the tRNA epitranscriptome with codon bias to adapt translation during stress responses. *Genome Biol.* (2018) 19:228. doi: 10.1186/s13059-018-1611-1
63. Sabi R, Tuller T. Modelling the efficiency of codon-tRNA interactions based on codon usage bias. *DNA Res.* (2014) 21:511–26. doi: 10.1093/dnares/dsu017
64. Pouyet F, Mouchiroud D, Duret L, Sémon M. Recombination, meiotic expression and human codon usage. *Elife.* (2017) 6:e27344. doi: 10.7554/eLife.27344
65. Paulet D, David A, Rivals E. Ribo-seq enlightens codon usage bias. *DNA Res.* (2017) 24:303–210. doi: 10.1093/dnares/dsw062
66. Burns CC, Shaw J, Campagnoli R, Jorba J, Vincent A, Quay J, et al. Modulation of poliovirus replicative fitness in HeLa cells by deoptimization of synonymous codon usage in the capsid region. *J Virol.* (2006) 80:3259–72. doi: 10.1128/JVI.80.7.3259-3272.2006
67. Mueller S, Papamichail D, Coleman JR, Skiena S, Wimmer E. Reduction of the rate of poliovirus protein synthesis through large-scale codon deoptimization causes attenuation of viral virulence by lowering specific infectivity. *J Virol.* (2006) 80:9687–96. doi: 10.1128/JVI.00738-06
68. Sánchez G, Bosch A, Pintó RM. Genome variability and capsid structural constraints of hepatitis A virus. *J Virol.* (2003) 77:452–9. doi: 10.1128/JVI.77.1.452-459.2003
69. dos Reis M, Wernisch L, Savva R. Unexpected correlations between gene expression and codon usage bias from microarray data for the whole *Escherichia coli* K-12 genome. *Nucleic Acids Res.* (2003) 31:6976–85. doi: 10.1093/nar/gkg897
70. Plotkin JB, Robins H, Levine AJ. Tissue-specific codon usage and the expression of human genes. *Proc Natl Acad Sci USA.* (2004) 101:12588–91. doi: 10.1073/pnas.0404957101
71. Lyu X, Liu Y. Nonoptimal codon usage is critical for protein structure and function of the master general amino acid control regulator CPC-1. *MBio.* (2020) 11:e02605–20. doi: 10.1128/mBio.02605-20
72. Affinito O, Palumbo D, Fierro A, Cuomo M, De Riso G, Monticelli A, et al. Nucleotide distance influences co-methylation between nearby CpG sites. *Genomics.* (2020) 112:144–50. doi: 10.1016/j.ygeno.2019.05.007
73. Das S, Das A, Bhattacharya DK, Tibarewala DN, A. new graph-theoretic approach to determine the similarity of genome sequences based on nucleotide triplets. *Genomics.* (2020) 112:4701–14. doi: 10.1016/j.ygeno.2020.08.023
74. Chen Y-H, Collier J, A. Universal code for mRNA stability? *Trends Genet.* (2016) 32:687–8. doi: 10.1016/j.tig.2016.08.007
75. Presnyak V, Alhusaini N, Chen Y-H, Martin S, Morris N, Kline N, et al. Codon optimality is a major determinant of mRNA stability. *Cell.* (2015) 160:1111–24. doi: 10.1016/j.cell.2015.02.029
76. Hanson G, Collier J. Codon optimality, bias and usage in translation and mRNA decay. *Nat Rev Mol Cell Biol.* (2018) 19:20–30. doi: 10.1038/nrm.2017.91
77. Mauro VP, Chappell SA, A. critical analysis of codon optimization in human therapeutics. *Trends Mol Med.* (2014) 20:604–13. doi: 10.1016/j.molmed.2014.09.003
78. Ling J, O'Donoghue P, Söll D. Genetic code flexibility in microorganisms: novel mechanisms and impact on physiology. *Nat Rev Microbiol.* (2015) 13:707–21. doi: 10.1038/nrmicro3568
79. Wang Z, Bao J, Wu X, Liu Y, Li L, Liu C, et al. Peste des petits ruminants virus in Tibet, China. *Emerg Infect Dis.* (2009) 15:299–301. doi: 10.3201/eid1502.080817
80. Bao J, Wang Q, Li L, Liu C, Zhang Z, Li J, et al. Evolutionary dynamics of recent peste des petits ruminants virus epidemic in China during 2013–2014. *Virology.* (2017) 510:156–64. doi: 10.1016/j.virol.2017.07.018
81. Wu X, Li L, Li J, Liu C, Wang Q, Bao JY, et al. Peste des Petits Ruminants Viruses Re-emerging in China, 2013–2014. *Transbound Emerg Dis.* (2016) 63:e441–446. doi: 10.1111/tbed.12308