ORIGINAL ARTICLE

Comprehensive analysis of synonymous codon usage patterns and influencing factors of porcine epidemic diarrhea virus

Xianglong Yu[1] · Jianxin Liu[1] · Huizi Li[1] · Boyang Liu[1] · Bingqian Zhao[1] · Zhangyong Ning[1]

Abstract

Porcine epidemic diarrhea virus (PEDV) is an enteric pathogen belonging to the family *Coronaviridae* that causes the porcine epidemic diarrhea, a highly contagious disease with high mortality in piglets and symptoms that include dehydration and severe diarrhea. Considering the high frequency of genetic mutations in PEDV and its potential for interspecies transmission, as it can infect and replicate in bat and human cells, a comprehensive analysis of its codon usage bias was performed. The effective number of codons (ENC) and the relative synonymous codon usage (RSCU) were determined, revealing codon usage bias in the PEDV genome. Principal component analysis (PCA), an ENC plot, and a parity rule 2 (PR2) plot showed that mutation pressure and natural selection have influenced the codon usage bias of the PEDV genomes. Correlation analysis with GRAVY and aromaticity values and neutrality plot analysis indicated that natural selection was the main force influencing the codon usage pattern, while mutation pressure played a minor role. This study provides valuable basic data for further fundamental research on evolution of PEDV.

Introduction

Porcine epidemic diarrhea virus (PEDV), first isolated in the early 1970s from pigs in Europe, is an enteric pathogen belonging to the genus *Alphacoronavirus*, family *Coronaviridae*, orders *Nidovirales* [1–3]. Porcine epidemic diarrhea (PED) caused by PEDV is a highly contagious disease with high mortality in piglets and symptoms of severe diarrhea and dehydration. PED has a worldwide distribution and is prevalent in many pig-raising countries in Europe, Asia, America, and Australia, causing serious damage to the pig farming industry [4–6]. PEDV is an enveloped, single-stranded and positive-sense RNA virus with a genome of about 28 kb, including 5' and 3' untranslated regions

(UTRs) and seven open reading frames (ORFs) [5]. More than two-thirds of the PEDV genome is occupied by ORF1a and ORF1b, which encode replicase polyproteins. In addition, four structural proteins, including the spike (S), envelope (E), membrane (M), and nucleocapsid (N) proteins, are encoded by ORF2, 4, 5, and 6, respectively, and one accessory protein is encoded by ORF3, which is located between the S and E coding regions.

Although a previous report demonstrated the evolutionary origin of PEDV is bats, it is of concern that this virus can replicate not only in porcine cells but also in bat and human cells, suggesting that it has the potential for interspecies transmission, and receptor homologs for this virus are present in various species [7]. Molecular and phylogenetic analysis of PEDV isolates has shown that insertions, deletions and point mutations are common and that strains can be divided into several separate clades corresponding to different geographical locations [8, 9]. Therefore, given the high frequency of genetic mutations in PEDV and the pandemic of coronavirus disease 2019 in humans, it is necessary to carry out further research on this virus. Here, a comprehensive analysis of codon usage bias (CUB) was performed to investigate the evolution of PEDV.

✉ Zhangyong Ning
ningzhyong@scau.edu.cn

1   College of Veterinary Medicine, South China Agricultural University, Guangzhou 510642, People's Republic of China

Materials and methods

Sequence data and compositional analysis

Nucleotide sequences of the coding regions of each PEDV strain were downloaded from NCBI (https://www.ncbi.nlm.nih.gov/genbank) and concatenated into a complete coding region. A total of 551 full-length PEDV sequences that were available up to March 2020 were obtained. The accession numbers and other detailed information including the strain name, collection year, and country are listed in Supplementary Table S1. Overall nucleotide compositions (A%, U%, G%, and C%) and G + C content (GC%) of each PEDV coding region were analyzed using BioEdit (version 7.0.9.0). The nucleotide composition at the third codon position (A3s, U3s, C3s and G3s) and the GC content at the third codon position (GC3s) of synonymous codons were determined using CodonW 1.4.4. The detailed information is shown in Supplementary Table S2.

Relative synonymous codon usage (RSCU)

The relative synonymous codon usage (RSCU) value is the ratio of the observed frequency of one specific synonymous codon to the expected frequency (no codon usage bias), which is an important measure of codon usage bias [10, 11]. If the RSCU value of a codon is 1.0, there is no codon usage bias, and there is equal usage of the codons for that amino acid. If the RSCU value is higher than 1.0, there is positive codon usage bias, and if the value is less than 1.0, there is negative codon usage bias. RSCU values higher than 1.6 and lower than 0.6 indicate overrepresented and underrepresented codons, respectively [12, 13]. In this study, the RSCU value of each codon was calculated using the CodonW program, and the corresponding values for the host of PEDV (swine) were downloaded from the codon usage database (https://www.kazusa.or.jp/codon).

Effective number of codons (ENC) and ENC plot analysis

The effective number of codons (ENC) indicates the degree of codon usage bias and reflects the extent of preference of synonymous codons [14]. ENC values range from 20 to 61. A value of 20 indicates a maximum level of codon bias, whereas a value of 61 indicates a complete lack of bias [15]. In general, if the ENC value is ≤ 35, the coding sequence is considered to have significant codon usage bias [16]. ENC values were calculated for each PEDV sequence using the CodonW program.

ENC plot analysis was used to identify factors influencing codon usage variation. An ENC-GC3s plot was generated using GraphPad Prism 8. The expected ENC value for each GC3s was calculated using the following equation:

$$ENC_{expected} = 2 + s + \left(\frac{29}{s^2 + (1 - s)^2} \right)$$

Where s is the GC3s value. If the codon usage is only constrained by mutation pressure, points will be on or around the expected curve. However, if multiple factors constrain the codon usage, the observed ENC values will lie below the expected curve [17].

Principal component analysis (PCA)

To identify major variation trends in codon usage patterns among the different PEDV strains, PCA was performed by analyzing the relationship between variables and samples [18]. In detail, the RSCU values of each strain were distributed into a 59-dimensional vector corresponding to the 59 synonymous codons (excluding the codons of AUG, UGG and the three stop codons), and they were then transformed into uncorrelated variables (principal components) [16]. The first two axes account for most of the components influencing the codon usage variation among genes, so PCA plots were constructed using the first two axes. PCA was performed using SPSS software (version 22), and the figures were drawn using Graph Pad Prism 8.0.

Hydropathicity and aromaticity analysis

General average hydropathicity and aromaticity are two major factors affecting translation and natural selection [19]. GRAVY and Aroma values are used to evaluate these factors and to represent the frequencies of hydrophobic and aromatic amino acids, respectively [20]. These values were calculated using the CodonW 1.4.4 program.

Parity rule 2 (PR2) bias plot analysis

Parity rule 2 (PR2) plot analysis was used to evaluate the effect of mutation pressure and natural selection at the third codon position for the four-codon amino acids. The PR2 plot distinguishes between AU bias [A3/(A3 + U3)] and GC bias [G3/(G3 + C3)]. Generally, if the effect of mutation pressure and natural selection are equal, the points will sit in the center of the plot, where A = T and G = C [21]. The PR2 plot was drawn using Graph Pad Prism 8.0.

Neutrality plot analysis

Neutrality plot analysis is a widely used method for investigating the effects of natural selection and mutation pressure on codon usage by plotting the GC12s values against GC3s values [22, 23]. Each point represents an independent PEDV strain, and a regression line is plotted. If the regression curve lies near the diagonal (slope = 1), this indicates that mutation pressure was the dominant cause of the codon usage bias, with weak external selection pressure. Alternatively, natural selection is considered the main force shaping codon usage if the slope of the regression curve tends toward 0. The neutrality plot was drawn using Graph Pad Prism 8.0.

Results

Compositional analysis and ENC analysis

The nucleotide U was the most abundant, with a mean value of $33.51 \pm 0.058\%$ (mean $\pm$ SD), followed by similar amounts of A ($24.85 \pm 0.024\%$), G ($22.63 \pm 0.048\%$) and C ($19.00 \pm 0.041\%$). The mean AU content ($58.40 \pm 0.077\%$) was higher than the GC content ($41.67 \pm 0.077\%$). Analysis of the nucleotides at the third position of synonymous codons showed that U3s ($54.41 \pm 0.171\%$) was more frequent than A3s ($23.95 \pm 0.076\%$), C3s ($22.86 \pm 0.121\%$) and G3s ($22.45 \pm 0.138\%$). The mean GC3s value was $35.10 \pm 0.170\%$, which was also lower than the AU3s value. The ENC values ranged from 47.81 to 48.49, and the mean ENC value was 48.04 ± 0.11 (mean $\pm$ SD).

Table 1 Overall RSCU of the 551 collected PEDV genomic sequences

Amino acid	Codon	RSCU/PEDV	RSCU/swine	Amino acid	Codon	RSCU/PEDV	RSCU/swine
Phe	UUU	*1.39*	0.79	Ser	AGC	0.67	*1.62*
	UUC	0.61	*1.21*		AGU	1.42	0.77
Leu	CUA	0.49	0.33		UCA	1.02	0.72
	CUC	0.60	1.35		UCC	0.67	1.50
	CUG	0.60	*2.68*		UCG	0.27	0.39
	CUU	*1.93*	0.65		UCU	*1.94*	0.99
	UUA	0.76	0.32	Cys	UGC	0.61	*1.22*
	UUG	1.61	0.67		UGU	*1.39*	0.78
Tyr	UAC	0.68	*1.27*	Pro	CCA	1.33	0.94
	UAU	*1.32*	0.73		CCC	0.53	*1.46*
His	CAC	0.60	*1.29*		CCG	0.21	0.55
	CAU	*1.40*	0.71		CCU	*1.93*	1.05
Gln	CAA	0.90	0.44	Arg	AGA	1.12	1.12
	CAG	*1.1*	*1.56*		AGG	0.93	1.23
Ile	AUA	0.58	0.42		CGA	0.35	0.61
	AUC	0.64	*1.67*		CGC	1.18	*1.31*
	AUU	*1.78*	0.91		CGG	0.24	1.29
Asn	AAC	0.72	*1.21*		CGU	*2.19*	0.44
	AAU	*1.28*	0.79	Thr	ACA	1.14	0.91
Lys	AAA	0.72	0.76		ACC	0.67	*1.68*
	AAG	*1.28*	*1.24*		ACG	0.32	0.57
Val	GUA	0.51	0.34		ACU	*1.88*	0.83
	GUC	0.74	1.06	Ala	GCA	1.11	0.74
	GUG	0.73	*2.03*		GCC	0.70	*1.80*
	GUU	*2.02*	0.57		GCG	0.26	0.51
Asp	GAC	0.75	*1.20*		GCU	*1.93*	0.95
	GAU	*1.25*	0.80	Gly	GGA	0.43	0.91
Glu	GAA	0.92	0.73		GGC	0.93	*1.46*
	GAG	*1.08*	1.27		GGG	0.16	1.05
					GGU	*2.49*	0.57

RSCU values of the preferred codon for PEDV and swine are in bold italics.

RSCU analysis

RSCU values were calculated for the 59 synonymous codons to determine the codon usage bias of the PEDV genome (Table 1). Ten codons, namely, CUU, CCU, AUU, CGU, ACU, GUU, GCU, UCU, UUG, and GGU were overrepresented (mean RSCU value > 1.6), and eleven codons, including CUA, CCC, CCG, AUA, CGA, CGG, ACG, GUA, GCG, GGA, and GGG, were underrepresented (mean RSCU value < 0.6). Among the 18 most abundantly used codons, three were G-ended and 15 were U-ended, indicating that codons ending with U were the most frequently used. PEDV and swine were found to have only three preferred codons in common (CAG [Gln], AAG [Lys], GAG [Glu]).

ENC plot analysis and correlation analysis

An ENC-GC3s plot was generated to investigate the role of mutational pressure in shaping codon usage bias. As shown in Fig. 1, all points in the plot, regardless of the country or continent from which the isolate originated, were lower than the standard curve. The correlation between ENC values and the relative amount of each nucleotide (A, C, G, U, and GC) was analyzed, and a strong correlation was found, with P-values much below 0.01 (Table 2). To further explore the effect of mutational pressure on codon usage, the correlation between nucleotide composition and codon composition (A3s, C3s, G3s, U3s, and GC3s) was also analyzed (Table 2). Significantly positive correlations were identified for all homologous regions, whereas significantly negative correlations were identified for some of the other regions.
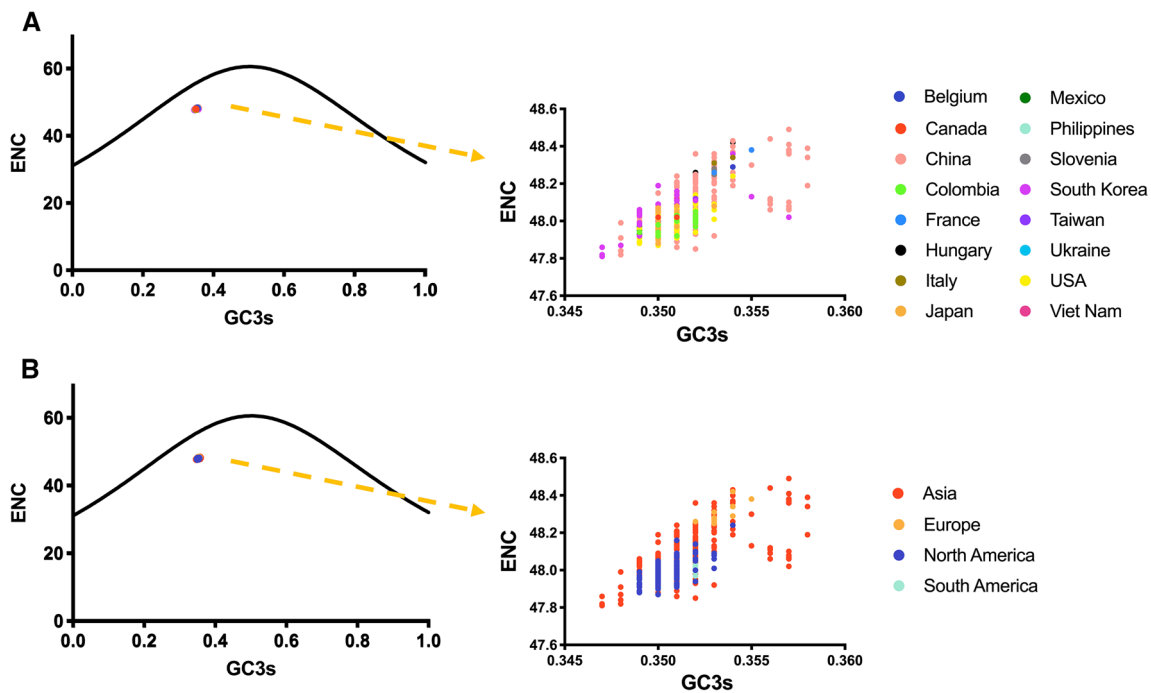

Fig. 1 ENC plots showing the relationship between ENC values and GC3s values. Each point represents one PEDV strain. (A) points classified by country. (B) points classified by continent. All of the points were below the standard curve, indicating that mutational pressure and other factors both play a role in PEDV codon usage.

Table 2 Correlation analysis of nucleotide composition and ENC

	A3s	C3s	G3s	U3s	GC3s	ENC
A	0.7262**	-0.1731**	-0.4249**	0.2060**	-0.3748**	-0.1727**
C	0.0095**	0.8236**	0.0055**	-0.6604**	0.5553**	0.7130**
G	-0.6978**	0.4362**	0.8888**	-0.6529**	0.7826**	0.4512**
U	0.2626**	-0.8594**	-0.5405**	0.9024**	-0.8654**	-0.7871**
GC	-0.4370**	0.7416**	0.5799**	-0.7976**	0.8157**	0.7132**

The numbers represent correlation coefficient "r" values (*, $P < 0.05$; **, $P < 0.01$)
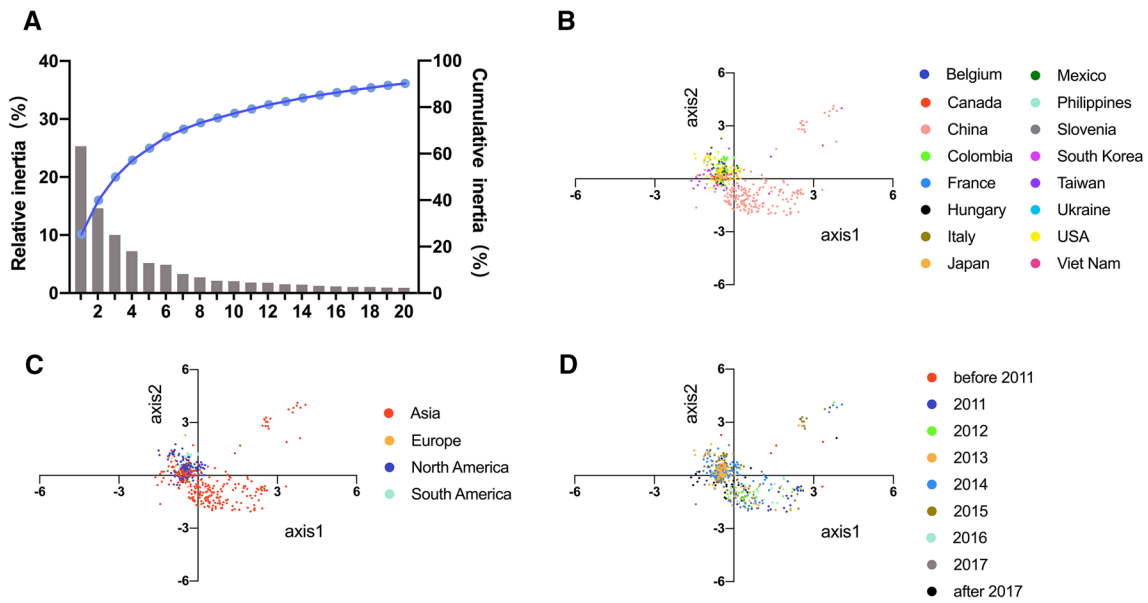
Fig. 2 Principal component analysis of PEDV. Each point represents one PEDV strain. (A) The distributions of the first 20 vectors by PCA. Columns represent the relative inertia, and the curve represents the cumulative inertia. (B) PCA plot constructed with the first two axes according to country. (C) PCA plot according to continent. (D) PCA plot according to date

Table 3 Correlation between the first two axes and nucleotide composition

	Axis 1	Axis 2
A	-0.1261**	0.3526**
C	0.2476**	0.2069**
G	0.4657**	-0.5086**
U	-0.5102**	0.1221**
A3s	-0.3574**	0.5085**
C3s	0.4917**	0.1148**
G3s	0.5204**	-0.4990**
U3s	-0.5453**	0.0513**
GC	0.4118**	-0.2179**
GC3s	0.6041**	-0.2113**

The numbers represent correlation coefficient "r" values (*, $P < 0.05$; **, $P < 0.01$)

Principal component analysis (PCA)

PCA was used to detect variations in codon usage and to construct the distributions of each vector. The first axis accounted for 25.33% of the total variation, while the

next three axes accounted for 14.63%, 10.02% and 7.23% (Fig. 2A). As the first two axes accounted for 39.96% in codon usage trend, PCA plots of the first and second axes were constructed based on different countries, continents, and dates (Fig. 2B, C, D). Subsequently, the correlation between the first two axes and nucleotide composition was analyzed (Table 3).

The role of natural selection in codon usage bias

The correlation between GRAVY, Aroma, axis1, axis2, ENC, and nucleotide composition was analyzed to identify the forces of natural selection (Table 4). Most of them had a significant correlation with P-values far below 0.01. Of note, there was no correlation between GRAVY and axis1 or axis2, demonstrating that amino acid usage plays a more prominent role for aromatic residues of PEDV proteins.

PR2 bias plot analysis

A PR2 bias plot for PEDV genomes showed that all points were located at the bottom right of the plot, indicating that C

Table 4 Correlation analysis for GRAVY, Aroma, the first two axes, ENC, and nucleotide composition

	Axis 1	Axis 2	ENC	GC3s	GC	A3s	C3s	G3s	U3s
GRAVY	-0.0738	0.0653	-0.0216**	-0.0597**	-0.1720**	0.0242**	-0.1134**	0.0413**	0.0332**
Aromo	-0.2201**	0.1098**	-0.1867**	-0.2232**	-0.2921**	0.2458**	-0.1568**	-0.1825**	0.2080**

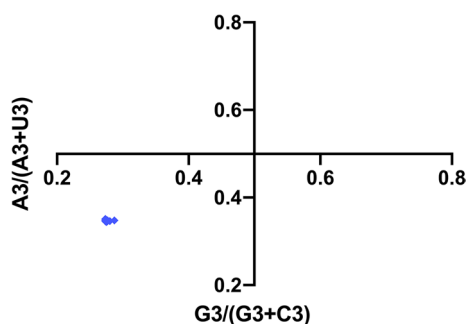The numbers represent correlation coefficient "r" values (*, $P < 0.05$; **, $P < 0.01$)

Fig. 3 PR2 plot analysis. The PR2 bias plot was calculated for the complete PEDV genome.
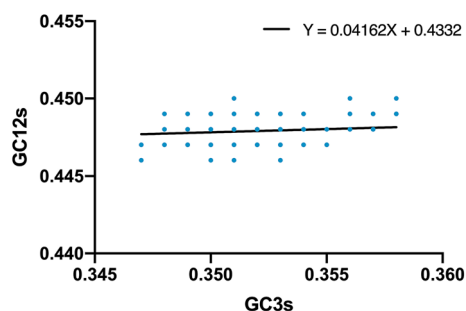

Fig. 4 Neutrality analysis performed by plotting GC12s values against GC3s values for the complete genome. The regression line is represented by the black straight line. The regression equation is also shown.

and U were used more frequently than G and A in the third codon position (Fig. 3).

Neutrality plot analysis

The main factors determining the codon usage pattern in PEDV genomes were identified by neutrality plot analysis (Fig. 4). A slight negative correlation was found between GC12s and GC3s values (r = -0.13, P < 0.01). The slope of the regression line was only 0.0416, suggesting that natural selection was the main force, while mutation pressure played a minor role in the codon usage pattern of the PEDV genome.

To explore whether natural selection acted equally on the structural and non-structural PEDV proteins, a neutrality plot analysis were also carried out for each gene (Fig. 5). Significant correlations were found between the GC12s and GC3s values of all proteins (P < 0.01), and there was only one strong correlation for the ORF3 gene (r = 0.67). The slope for the ORF1ab, S, ORF3, E, M, and N gene was 0.016, -0.230, -0.400, 0.100, 0.036, and 0.078, respectively. Thus, the contribution of natural selection was 98.4%, 77.0%, 60.0%, 90.0%, 96.4%, 92.2%, respectively,

demonstrating that natural selection played a dominant role in the codon usage for each PEDV protein.
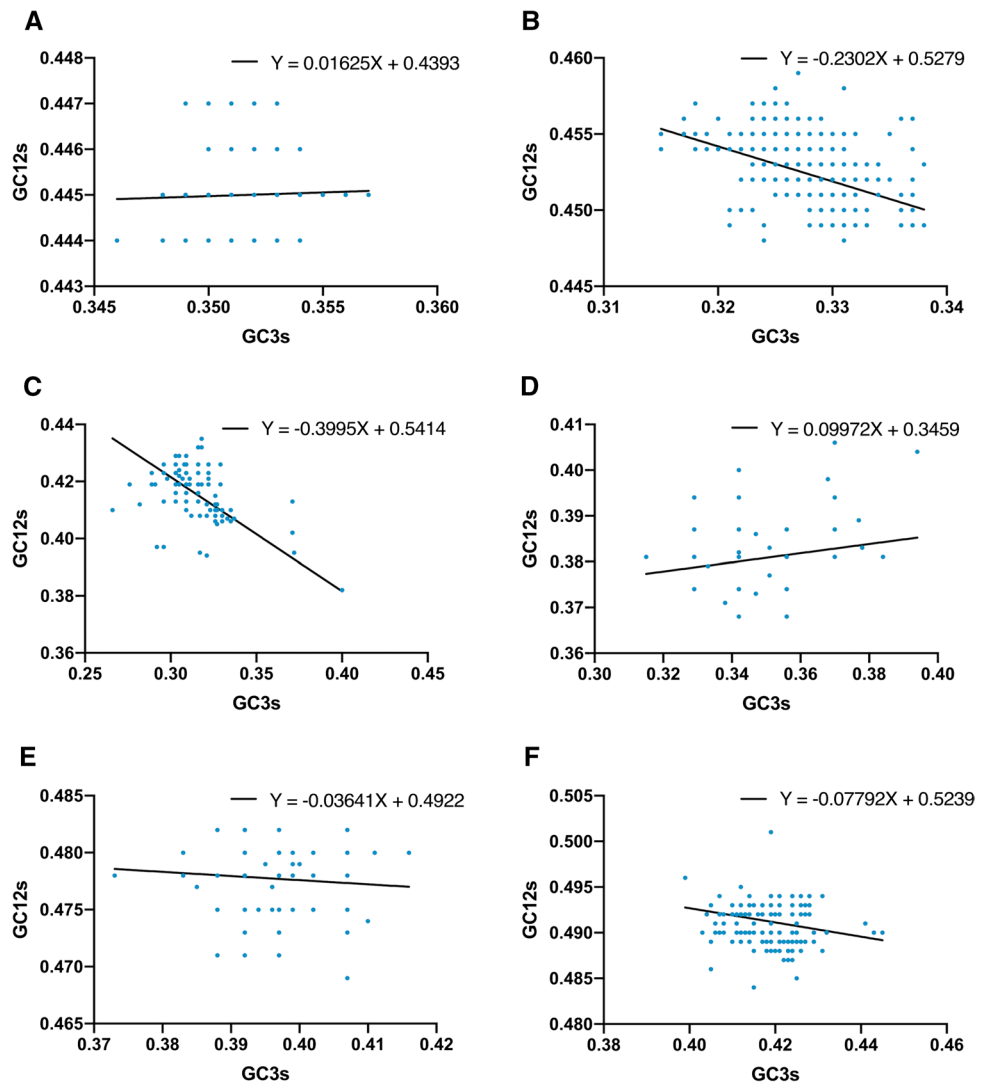
Discussion

PED is a highly contagious disease with worldwide distribution, and the high genetic variability of PEDV has been confirmed repeatedly [3, 24, 25]. Although the genetic diversity and evolution of PEDV had been investigated previously, a systematic analysis of the codon usage bias of the complete PEDV genome is still needed. A previous study reported the codon usage bias of PEDV with 43 strains collected up to 2014 [11], and two articles reported the CUB of individual regions of the genome (the N gene and the ORF3 gene) [26, 27], but in view of the large number of new complete PEDV genome sequences reported in the past few years, it was necessary to perform a new comprehensive analysis to fill the gaps. In this study, we used 551 PEDV sequences to determine the codon usage bias in the PEDV genome.

RSCU analysis is widely applied to standardize the analysis of codon usage bias. In this study, 10 codons were overrepresented and 11 codons were underrepresented, revealing considerable codon usage bias. In general, codons that are used less by the host are selected in the process of evolution of coronaviruses. Here, we found that PEDV and its host have only three preferred codons in common, implying that PEDV tends to use codons that are less used by the host in order to avoid competition with the host cell during gene translation.

The nucleotide content and codon usage compositions can reflect the effect of mutation pressure on CUB. For PEDV, the AU content was higher than the GC content, and likewise, there was a preference for AU in the third codon position. In RSCU analysis, most of the abundantly used codons were U-ended (15/18). Moreover, analysis of the correlation between nucleotide composition and codon composition indicated a significant correlation in most cases. The coding sequences of PEDV were found to be AU-rich, and mutational pressure was found to be an important force affecting the codon usage bias. In addition, an analysis of the correlation between the first two axes and nucleotide composition showed a weak correlation in more than half of the comparisons, suggesting that mutation pressure and other factors contribute to the codon usage in PEDV strains.

The ENC is a useful measure of the extent of the codon usage bias of the virus, and low codon usage bias might make it easier for the virus to overcome host defense mechanisms. For complete PEDV genome sequences, the ENC values ranged from 47.81 to 48.49, and the mean ENC value was 48.04 ± 0.11. For comparison, the mean ENC values for other coronaviruses are as follows: (1) bovine coronavirus (mean ENC = 43.78) [28], (2) SARS coronavirus (mean

Fig. 5 Neutrality analysis performed for each gene. (A) ORF1ab, (B) S, (C) ORF3, (D) E, (E) M, (F) N

A Y = 0.01625X + 0.4393

B Y = -0.2302X + 0.5279

C Y = -0.3995X + 0.5414

D Y = 0.09972X + 0.3459

E Y = -0.03641X + 0.4922

F Y = -0.07792X + 0.5239

ENC = 48.99) [29], (3) porcine deltacoronavirus (mean ENC = 52.85) [30]. Although the mean ENC value of PEDV was not the highest among them, it is also greater than 45, which shows that the codon usage bias is somewhat low. Compared with the mean ENC value for PEDV in the first report published six year ago (47.91 ± 0.13) [11], the latest data changed little, with a small standard deviation. Of note, the ENC mean value of SARS-CoV-2 has been reported to be 51.90 ± 2.59 (mean $\pm$ SD) [31], which is higher than that of most coronaviruses, suggesting that it is well adapted to its host and able to overcome its defense mechanisms.

ENC plot and PR2 bias plot analyses are widely applied to evaluate the influence of mutation pressure and natural selection [14, 21]. In our ENC plot analysis, all of the points were below the standard curve, revealing that mutational pressure and other forces, such as natural selection, gene length, tRNA abundance, or RNA structure, together shaped the PEDV codon usage pattern. PR2 bias plot analysis further indicated that natural selection and mutation pressure

influenced the codon usage bias of PEDV with unequal contributions.

PCA were performed to identify major variation trends. As shown in Fig. 2B, most of the points from one country were concentrated in the same region, confirming that natural selection is an important factor in shaping the codon usage bias. In particular, data from China and South Korea were comparatively disperse, suggesting that the contribution of mutation pressure shaping the CUB in these strains was greater than that in other countries. Furthermore, American strains formed two groups in the PCA plot, exactly corresponding to two genetically different PEDV strains isolated in the USA (U.S. PEDV prototype and S-INDEL-variant strains) [32]. Phylogenetic analysis indicated that the PEDV prototype strains emerging in USA originated from China [33], which accounts for the phenomenon that some of the data points from China and the USA were concentrated in the same region of the plot. As shown in Fig. 2C, the data from Asia were more

disperse than those from other continents, suggesting a stronger contribution of mutation pressure in Asian strains and a more conserved evolutionary process on other continents. Regarding the collection date, the data points tended to cluster together up to 2013 but tended to disperse later (Fig. 2D). This revealed that the impact of mutation pressure in shaping the CUB had a development process from weak to strong and then to weak. Due to the large number of complete PEDV genome sequences uploaded to NCBI in recent years, more analyses can be performed with high accuracy, allowing the whole evolutionary process of PEDV codon usage pattern to be studied.

Neutrality plot analysis is one of the most common methods for exploring the effects of natural selection and mutation pressure. The results showed that the relative constraint (natural selection) was 95.84%, strongly suggesting that natural selection was the main force in determining the CUB. This conclusion was also supported by other analyses. First, for correlation analyses, significant positive correlations between nonhomologous nucleotide comparisons were also observed, which implied that natural selection might play a considerable role in determining the codon usage pattern. Second, many significant correlations were observed with GRAVY and Aroma, which are two major indexes for natural selection. Third, the PCA plot showed that most of the data points, when classified by country, continent, or date, were relatively concentrated. Accordingly, we also compared our latest data with the first research concerning the CUB of PEDV. More significant correlations with GRAVY and Aroma were observed, which is also consistent with the results of the neutrality plot analysis (relative constraint rise). The PCA plot based on the country of origin showed that points were less clustered together than they had been previously, which is in agreement with the small decrease in the mean ENC value, both of which reveal a higher codon usage bias at the present time. In addition, we carried out a neutrality plot analysis for each gene. Compared with previous studies performed with the N gene (natural selection = 65.19%) and the ORF3 gene (natural selection = 76.32%), the result of our study showed that natural selection remained the main force for the codon usage pattern.

In conclusion, the results of this comprehensive analysis of the synonymous codon usage patterns in the PEDV genome revealed a low level of codon usage bias and suggested that natural selection was the primary force influencing codon usage. Given the growing PEDV epidemic situation, the latest data on the evolution of this virus will benefit further basic research.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Pensaert MB, de Bouck P (1978) A new coronavirus-like particle associated with diarrhea in swine. Arch Virol 58:243–247
2. Chasey D, Cartwright SF (1978) Virus-like particles associated with porcine epidemic diarrhoea. Res Vet Sci 25:255–256
3. Song D, Moon H, Kang B (2015) Porcine epidemic diarrhea: a review of current epidemiology and available vaccines. Clin Exp Vaccine Res 4:166–176
4. Chen J, Wang C, Shi H, Qiu H, Liu S, Chen X, Zhang Z, Feng L (2010) Molecular epidemiology of porcine epidemic diarrhea virus in China. Arch Virol 155:1471–1476
5. Chen Q, Li G, Stasko J, Thomas JT, Stensland WR, Pillatzki AE, Gauger PC, Schwartz KJ, Madson D, Yoon KJ, Stevenson GW, Burrough ER, Harmon KM, Main RG, Zhang J (2014) Isolation and characterization of porcine epidemic diarrhea viruses associated with the 2013 disease outbreak among swine in the United States. J Clin Microbiol 52:234–243
6. Stevenson GW, Hoang H, Schwartz KJ, Burrough ER, Sun D, Madson D, Cooper VL, Pillatzki A, Gauger P, Schmitt BJ, Koster LG, Killian ML, Yoon KJ (2013) Emergence of Porcine epidemic diarrhea virus in the United States: clinical signs, lesions, and viral genomic sequences. J Vet Diagn Invest 25:649–654
7. Liu C, Tang J, Ma Y, Liang X, Yang Y, Peng G, Qi Q, Jiang S, Li J, Du L, Li F (2015) Receptor usage and cell entry of porcine epidemic diarrhea coronavirus. J Virol 89:6121–6125
8. Park SJ, Kim HK, Song DS, Moon HJ, Park BK (2011) Molecular characterization and phylogenetic analysis of porcine epidemic diarrhea virus (PEDV) field isolates in Korea. Arch Virol 156:577–585
9. Gao Y, Kou Q, Ge X, Zhou L, Guo X, Yang H (2013) Phylogenetic analysis of porcine epidemic diarrhea virus field strains prevailing recently in China. Arch Virol 158:711–715
10. Singh NK, Tyagi A (2017) A detailed analysis of codon usage patterns and influencing factors in Zika virus. Arch Virol 162:1963–1973
11. Chen Y, Shi Y, Deng H, Gu T, Xu J, Ou J, Jiang Z, Jiao Y, Zou T, Wang C (2014) Characterization of the porcine epidemic diarrhea virus codon usage bias. Infect Genet Evol 28:95–100
12. Butt AM, Nasrullah I, Qamar R, Tong Y (2016) Evolution of codon usage in Zika virus genomes is host and vector specific. Emerg Microbes Infect 5:e107
13. Singh RK, Pandey SP (2017) Phylogenetic and evolutionary analysis of plant ARGONAUTES. Methods Mol Biol 1640:267–294
14. Liu X, Wu C, Chen AY (2010) Codon usage bias and recombination events for neuraminidase and hemagglutinin genes in Chinese isolates of influenza A virus subtype H9N2. Arch Virol 155:685–693
15. Grocock RJ, Sharp PM (2001) Synonymous codon usage in Cryptosporidium parvum: identification of two distinct trends among genes. Int J Parasitol 31:402–412
16. He Z, Gan H, Liang X (2019) Analysis of synonymous codon usage bias in potato virus M and Its adaption to hosts. Viruses 11:E752

17. Bera BC, Virmani N, Kumar N, Anand T, Pavulraj S, Rash A, Elton D, Rash N, Bhatia S, Sood R, Singh RK, Tripathi BN (2017) Genetic and codon usage bias analyses of polymerase genes of equine influenza virus and its relation to evolution. BMC Genom 18:652

18. Nasrullah I, Butt AM, Tahir S, Idrees M, Tong Y (2015) Genomic analysis of codon usage shows influence of mutation pressure, natural selection, and host features on Marburg virus evolution. BMC Evol Biol 15:174

19. Lobry JR, Gautier C (1994) Hydrophobicity, expressivity and aromaticity are the major trends of amino-acid usage in 999 *Escherichia coli* chromosome-encoded genes. Nucleic Acids Res 22:3174–3180

20. Xu Y, Jia R, Zhang Z, Lu Y, Wang M, Zhu D, Chen S, Liu M, Yin Z, Cheng A (2015) Analysis of synonymous codon usage pattern in duck circovirus. Gene 557:138–145

21. Sueoka N (1999a) Translation-coupled violation of Parity Rule 2 in human genes is not the cause of heterogeneity of the DNA G + C content of third codon position. Gene 238:53–58

22. Sueoka N (1999b) Two aspects of DNA base composition: G+C content and translation-coupled deviation from intra-strand rule of A = T and G = C. J Mol Evol 49:49–62

23. Choudhury MN, Uddin A, Chakraborty S (2018) Nucleotide composition and codon usage bias of SRY gene. Andrologia 50:e12787

24. Hou Y, Lin CM, Yokoyama M, Yount BL, Marthaler D, Douglas AL, Ghimire S, Qin Y, Baric RS, Saif LJ, Wang Q (2017) Deletion of a 197-amino-acid region in the N-terminal domain of spike protein attenuates porcine epidemic diarrhea virus in piglets. J Virol 91:e00227-17

25. Sun R, Leng Z, Zhai SL, Chen D, Song C (2014) Genetic variability and phylogeny of current Chinese porcine epidemic diarrhea virus strains based on spike, ORF3, and membrane genes. Sci World J 2014:208439

26. Sheikh A, Altaher AY, Alnazawi M, Almubarak AI, Kandeel M (2020) Analysis of preferred codon usage in the coronavirus N genes and their implications for genome evolution and vaccine design. J Virol Methods 277:113806

27. Xu X, Li P, Zhang Y, Wang X, Xu J, Wu X, Shen Y, Guo D, Li Y, Yao L, Li L, Song B, Ma J, Liu X, Xu S, Zhang H, Wu Z, Cao H (2019) Comprehensive analysis of synonymous codon usage patterns in orf3 gene of porcine epidemic diarrhea virus in China. Res Vet Sci 127:42–46

28. Castells M, Victoria M, Colina R, Musto H, Cristina J (2017) Genome-wide analysis of codon usage bias in Bovine Coronavirus. Virol J 14:115

29. Gu W, Zhou T, Ma J, Sun X, Lu Z (2004) Analysis of synonymous codon usage in SARS Coronavirus and other viruses in the Nidovirales. Virus Res 101:155–161

30. He W, Wang N, Tan J, Wang R, Yang Y, Li G, Guan H, Zheng Y, Shi X, Ye R, Su S, Zhou J (2019) Comprehensive codon usage analysis of porcine deltacoronavirus. Mol Phylogenet Evol 141:106618

31. Dilucca M, Forcelloni S, Georgakilas AG, Giansanti A, Pavlopoulou A (2020) Codon usage and phenotypic divergences of sars-cov-2 genes. Viruses 12:498

32. Chen Q, Thomas JT, Giménez-Lirola LG, Hardham JM, Gao Q, Gerber PF, Opriessnig T, Zheng Y, Li GW, Gauger PC, Madson DM, Magstadt DR, Zhang JQ (2016) Evaluation of serological cross-reactivity and cross-neutralization between the united states porcine epidemic diarrhea virus prototype and s-indel-variant strains. BMC Vet Res 12:70

33. Huang YW, Dickerman AW, Pieyro P, Li L, Meng XJ (2013) Origin, evolution, and genotyping of emergent porcine epidemic diarrhea virus strains in the united states. Mbio 4:5