ORIGINAL ARTICLE

Comprehensive Analysis of Synonymous Codon Usage Bias for Complete Genomes and E2 Gene of Atypical Porcine Pestivirus

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

Abstract

Atypical porcine pestivirus (APPV) is an emerging novel pestivirus causing the congenital tremor (CT) in piglets. The worldwide distribution characteristic of APPV make it a threat to global swine health. E2 is the major envelope glycoprotein of APPV and the crucial target for vaccine development. Considering the genetic variability of APPV complete genomes and its E2 gene as well as gaps for codon analysis, a comprehensive analysis of codon usage patterns was performed. Relative synonymous codon usage (RSCU) and effective number of codon (ENC) analyses showed that a relatively instable change existed and a slight low codon usage bias (CUB) were displayed in APPV genomes. ENC-plot analysis and correlation analyses of nucleotide compositions and ENC showed that mutation pressure and natural selection both affected the codon usage bias of the APPV and natural selection had a more obvious influence for E2 gene compared with complete genomes. Principal component analysis (PCA) and correlation analyses confirmed the above results. Correlation analyses between Gravy and Aromaticity values and the codon bias showed that natural selection played an important role in shaping the synonymous codon bias. Furthermore, neutrality plot analysis showed that natural selection was the main force while mutation pressure was a minor force influencing the codon usage pattern of the APPV E2 gene and complete genomes. The results could illustrate the codon usage patterns of APPV genomes and provided valuable basic data for further fundamental research of evolution of APPV.

Keywords APPV · E2 · Codon usage bias · Mutation pressure · Natural selection

✉ Zhangyong Ning
ningzhyong@scau.edu.cn

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

Introduction

Atypical porcine pestivirus (APPV), first identified in 2015 in USA, is a novel pestivirus causing the congenital tremor (CT) type A-II and showed a worldwide distribution (Arruda et al. 2016; de Groof et al. 2016; Hause et al. 2015; Mosena et al. 2018). In China, our group first reported infection of APPV in swine herds in 2016 and subsequent investigations showed that the disease outbreak in China (Zhang et al. 2017). APPV is an enveloped, single-stranded and positive-sense RNA virus with the genome of 11 to 12 kb encoding a 3635 amino acid polyprotein and this polyprotein was cleaved into various mature viral proteins including four structural (C, Erns, E1 and E2) and eight non-structural proteins (Npro, P7, N2, NS3, NS4A, NS4B, NS5A and NS5B) (Hause et al. 2015; Zhang et al. 2017). Recently, a new uniform naming system for pestivirus was proposed with the format pestivirus X, where X represents a different capital letter for each species, and APPV belongs to pestivirus K (Smith et al. 2017).

E2 is the major envelope glycoprotein of APPV and it is highly immunogenic similar to other pestiviruses (Cagatay et al. 2019). APPV is a highly variable RNA virus with a high genetic variability (up to 21% genetic distance) between different clinical strains and E2 also showed similar characteristic (Postel et al. 2017). Although a subunit vaccine based on E2 of APPV could induce a strong humoral and cellular immune response in mice, there are still a lot of works to be done (Zhang et al. 2018a). A comprehensive analysis to substantiate the genetic characteristics of E2 of this emerging virus is necessary for further research of virus evolution and developing new control strategies.

Synonymous codons do not occur randomly in viruses, prokaryotes and eukaryotes, which is also called codon usage bias (CUB) (Butt et al. 2014). Mutation pressure, natural selection, gene length, tRNA abundance and RNA structure could affect codon usage bias, and the first two of the above factors were thought to be the two most important factors for RNA viruses (Butt et al. 2014). Considering the genetic variability of APPV complete genomes and its E2 gene as well as gaps for codon analysis, the codon usage bias was firstly investigated to understand the mechanism of codon distribution and evolution of the virus.

Material and Methods

Sequence Data and Recombination Analysis

After removing repetitive and cloning sequences, E2 gene coding sequences from 71 APPV isolates and 60 complete coding sequences were collected from GenBank database (http://www.ncbi.nlm.nih.gov) available up to November 2019. All coding sequences showed worldwide distribution and the detailed information of GenBank accession number, strain name, collection year and country were given in Supplementary Table S1. As possible recombination events could influence the

codon usage bias of genomes or genes, recombination analysis was performed by Recombination Detection Program 4.0 (RDP4) using RDP, Bootscan, MaxChi, GENECONV, Chimaera, SiScan and 3Seq methods to validate recombinations in these strains (Martin et al. 2015). The *P* value cut-off was set to 0.05 and the bonferroni correction was applied in all analyses. Genomes were considered to be recombinants if at least three of the above seven algorithms consistently conformed, and all recombinants would be excluded from further analysis.

Nucleotide Content and Codon Usage Composition

The overall nucleotide compositions (A%, C%, U% and G%) in each sequence were analyzed by Bio Edit (version 7.0.9.0). The nucleotide compositions at the third position (A3s, U3s, C3s and G3s) were computed using Codon W 1.4.4 and GC content at third codon positions (GC3s) were also calculated by Codon W.

Relative Synonymous Codon Usage (RSCU)

RSCU value refers to the ratio of the usage frequency of one specific codon to the usage frequency of all expected synonymous codons for the same amino acid, which is widely applied to standardize the codon usage bias between different sequences (Chen et al. 2014). If the RSCU value of a codon is 1, it represents that there is no codon bias and the synonymous codons are used equally for one amino acid. If the RSCU value > 1, it shows positive codon usage bias while when the RSCU value < 1, it shows negative codon usage bias. Moreover, if RSCU is more than 1.6 or less than 0.6, it is considered to be over-represented or under-represented (Singh and Pandey 2017; Taylor et al. 2017). AUG, UGG and the three stop codons (UAG, UAA and UGA), which have no synonymous codons, were excluded from the RSCU analysis (Butt et al. 2016). The RSCU index was calculated using the following equation:

$$RSCU = \frac{X_{ij}}{\sum_{j}^{n_i} X_{ij}} n_i$$

In this formula, X_{ij} is the number of the *i*th codon for the *j*th amino acid, and n_i represents the degenerate numbers of a synonymous codon which encode the jth amino acid. The RSCU value of each codon was calculated using the online software Emboss: cusp (http://emboss.toulouse.inra.fr/cgi-bin/emboss/cusp). Swine is the host of the APPV, and the RSCU values of swine were downloaded from the codon usage database (http://www.kazusa.or.jp/codon/).

Effective Number of Codon (ENC)

The ENC value is a simple and absolute measure of codon usage bias in genes and genomes, which can quantify the degree of CUB and reflect the extent of preference of synonymous codons. ENC values range from 20 to 61. The value of 61 indicates that there is no CUB, and the value of 20 indicates that the codon bias is at a

maximum while only one codon was used for each amino acid. It is generally considered that the gene has an obvious codon usage bias if the ENC value is equal to or less than 35.

The ENC value was calculated using the following equation:

$$\text{ENC} = 2 + \frac{9}{F_2} + \frac{1}{F_3} + \frac{5}{F_4} + \frac{3}{F_6}$$

where $F_{i\ (i\ =\ 2,\ 3,\ 4,\ 6)}$ is the average of the F_i values for i-fold degenerate codon families. The F_i value is calculated using the following equation:

$$F_i = \frac{n \sum_{j=1}^{i} \left(\frac{n_j}{n}\right)^2 - 1}{n - 1}$$

where n is the total number of the observed value of codons for one amino acid and n_j stands the numbers of the particular codon for that amino acid.

To determine the influencing factors of codon usage bias, an ENC-plot, between the GC3s and ENC values, was generated using Graph Pad Prism 6.0. The expected ENC for each GC3s was calculated as follows:

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

where s represents the frequency of G + C of synonymous codons in the third codon position (GC3s). In ENC-plot, if the codon usage is only constrained by the GC3s, the observed ENC values just lies on or around the expected curve, and it also indicates that mutation pressure is almost the only factor for evolution. Whereas if several factors constrain the codon usage, the observed ENC values lies far lower than the expected curve (Zhang et al. 2018b).

General Average Hydropathicity (Gravy) and Aromaticity (Aroma)

Codon W 1.4.4 was used to calculate the Gravy and Aroma values. Gravy and Aroma showed the frequencies of hydrophobic and aromatic amino acids which had much effect on codon usage pattern (Xu et al. 2015). They are also two major indexes on natural selection.

Principal Component Analysis (PCA)

PCA is a multivariate statistical method to identify major variation trends by analyzing the relationship between variables and samples (Bera et al. 2017; Chen et al. 2014). To further explore the trends of codon usage pattern among the different APPV strains, the PCA was performed with Statistic Package for Social Science 22 (SPSS). In the analysis, the RSCU values of each APPV strain were distributed into a 59-dimensional vector corresponding to the 59 synonymous codons (excluding the codons of AUG, UGG and the three-terminal codons). In other words, RSCU

values were transformed into uncorrelated variables. PCA plots were constructed with the first two axes which were accounting for most of the component influencing the codon usage variation among genes. Figures in PCA were drawn by Graph Pad Prism 6.0.

Neutrality Plot Analysis

The neutrality plot analysis was carried out to explore the effects of natural selection and mutation pressure in shaping the CUB by regression of GC12s on GC3s (Chakraborty et al. 2019). Figures were drawn using Graph Pad Prism 6.0. In the plot, each point represented an independent APPV strain. If the slope of the regression line was close to 1, it indicated that the main force was mutation pressure for the CUB, while a slope of 0 indicated that the dominant role was natural selection. Moreover, if the slope of the regression curve is close to ± 0.5, it was considered no or weak external selection pressure.

Correlation Analysis

Correlation analyses were conducted by Karl Pearson's method to investigate the relationships among the nucleotide compositions, codon compositions, Gravy, Aroma and principal axes.

Result

Recombination Analysis

Two recombination events of complete sequences, containing MG792803 and MH493896, were detected by the RDP software. Those two recombinant isolates were taken out and surplus coding sequences were carried on to the next analysis.

Nucleotide Compositions of the APPV E2 Gene and Complete Genomes

Analyses of the composition of the E2 coding sequences showed the results as follow: (1) The mean values of nucleotide from high to low were A% (30.75 ± 0.48), G% (27.42 ± 0.51), U% (21.45 ± 0.52) and C% (20.36 ± 0.48). (2) The mean values of the codon composition at the third position from high to low were A3s (39.91 ± 1.85), C3s (30.87 ± 2.04), G3s (28.95 ± 1.61), U3s (26.18 ± 2.10). (3) The mean compositions of AU% (52.21 ± 0.64) were above the GC% (47.79 ± 0.64). The same case was also identified with complete genomes. All the detailed information was listed in Supplementary Tables S2 and S3.

RSCU and ENC Analysis

RSCU values of 59 synonymous codons were calculated to investigate the codon usage bias of APPV complete genomes and E2 gene, and the effect of pig on codon usage pattern. As shown in Table 1, among the 18 frequently used synonymous codons, six were commonly used between the host and both APPV complete genomes and E2 gene. Moreover, twelve preferred codons for APPV E2 gene were A/U-ended (A-ended: 8; U-ended: 4) and six were G/C-ended (G-ended: 2; C-ended: 4), and the result for complete genomes was similar (A-ended: 8; U-ended: 3; G-ended: 3; C-ended: 4). Additionally, seven codons of APPV E2 gene were over-represented (mean RSCU value > 1.6) and up to thirteen codons were under-represented (mean RSCU value < 0.6). Meanwhile, only two codons of complete genomes were over-represented and seven codons were under-represented. The above findings showed that codon usage bias appeared in APPV genomes and the bias of E2 gene was more obvious compared with complete genomes.

The ENC values of E2 gene ranged from 49.15 to 61.00 with a mean value of 52.74 ± 2.69 (mean $\pm$ SD), and the values of complete genomes ranged from 53.85 to 55.14 with a mean value of 54.52 ± 0.25. The observed result revealed a slight low CUB compared with a high ENC value (> 40). Of note, the ENC mean value of APPV complete genomes was higher than the value of E2 gene and the standard deviation was much lower than the value of E2 gene, which suggested that CUB for APPV complete genomes was more stable and the bias was lower than E2 gene. The detailed information of ENC values was shown in Supplementary Tables S2 and S3.

ENC-Plot Analysis and Correlation Analyses of Nucleotide Compositions and ENC

ENC-plot was used to investigate the mutational pressure in shaping codon usage bias. As shown in Fig. 1, most of the points were lower than the standard curve no matter APPV complete genomes or E2 gene, which indicated that several factors, including natural selection, gene length and RNA structure, influenced the codon usage bias besides mutational pressure. In particular, one strain of E2 gene, which isolated from China, was slightly higher than the expected curve, and other a few values also lied around the curve, which revealed that several strains had an obviously low codon usage bias and a relatively instable change existed in all E2 gene.

To further determine the effect of mutational pressure on the codon usage bias, the correlation analyses between nucleotide compositions and ENC were carried out for APPV complete genomes and E2 gene (Table 2). The analyses showed that strong correlations ($r > 0.5$) were more than half among each other and the correlations were more frequently appeared in analyses for APPV complete genomes. Furthermore, the ENC value had a weak correlation with each content. Taken together, the above result suggested that mutation pressure and natural selection both affect the codon usage bias of the APPV and natural selection had a more obvious influence for E2 gene compared with complete genomes.

Table 1 Overall RSCU of collected sequences of the APPV E2 gene and complete genomes

Amino acid	Codon	RSCU/E2 gene	RSCU/APPV complete genomes	RSCU/swine	Amino acid	Codon	RSCU/E2 gene	RSCU/APPV complete genomes	RSCU/swine
Phe	UUU	0.98	0.95	0.7868	Ser	AGC	1.58	1.16	1.6238
	UUC	1.02	1.05	1.2132		AGU	0.78	1.13	0.7713
Leu	CUA	1.50	1.19	0.3311		UCA	1.62	1.32	0.7226
	CUC	0.79	0.68	1.3475		UCC	0.67	1.05	1.5021
	CUG	1.42	1.38	2.6776		UCG	0.63	0.38	0.3897
	CUU	0.61	0.61	0.6506		UCU	0.72	0.96	0.9905
	UUA	0.76	1.08	0.3195	Cys	UGC	1.45	1.06	1.2152
	UUG	0.91	1.06	0.6738		UGU	0.55	0.94	0.7848
Tyr	UAC	1.25	0.98	1.2685	Pro	CCA	1.57	1.48	0.9423
	UAU	0.75	1.02	0.7315		CCC	0.96	0.87	1.4563
His	CAC	0.86	0.93	1.2946		CCG	0.76	0.60	0.5536
	CAU	1.14	1.07	0.7054		CCU	0.72	1.05	1.0478

Table 1 (continued)

Amino acid	Codon	RSCU/E2 gene	RSCU/APPV complete genomes	RSCU/swine	Amino acid	Codon	RSCU/E2 gene	RSCU/APPV complete genomes	RSCU/swine
Gln	CAA	0.81	1.03	0.441	Arg	AGA	2.50	2.29	1.1155
	CAG	1.19	0.97	1.559		AGG	1.73	2.23	1.2347
Ile	AUA	2.20	1.40	0.4208		CGA	0.45	0.41	0.6065
	AUC	0.28	0.89	1.6697		CGC	0.26	0.28	1.3105
	AUU	0.52	0.71	0.9095		CGG	0.60	0.48	1.2888
Asn	AAC	0.62	0.97	1.2133		CGU	0.46	0.32	0.444
	AAU	1.38	1.03	0.7867	Thr	ACA	1.32	1.13	0.9145
Lys	AAA	1.24	1.16	0.7603		ACC	0.96	1.29	1.6803
	AAG	0.76	0.84	1.2397		ACG	0.25	0.42	0.5725
Val	GUA	0.89	1.05	0.3385		ACU	1.47	1.16	0.8327
	GUC	0.59	1.16	1.0646	Ala	GCA	1.70	1.36	0.7386
	GUG	2.12	1.35	2.0308		GCC	1.12	1.33	1.801
	GUU	0.39	0.45	0.5661		GCG	0.47	0.34	0.5057
Asp	GAC	1.52	1.21	1.1975		GCU	0.71	0.97	0.9546
	GAU	0.48	0.79	0.8025	Gly	GGA	0.20	0.90	0.9117
Glu	GAA	1.56	1.21	0.7256		GGC	1.58	0.91	1.46443
	GAG	0.44	0.79	1.2744		GGG	0.60	1.38	1.0541
						GGU	1.63	0.81	0.5698

RSCU values of the preferred codon for APPV complete genomes, E2 gene and swine are in bold italics
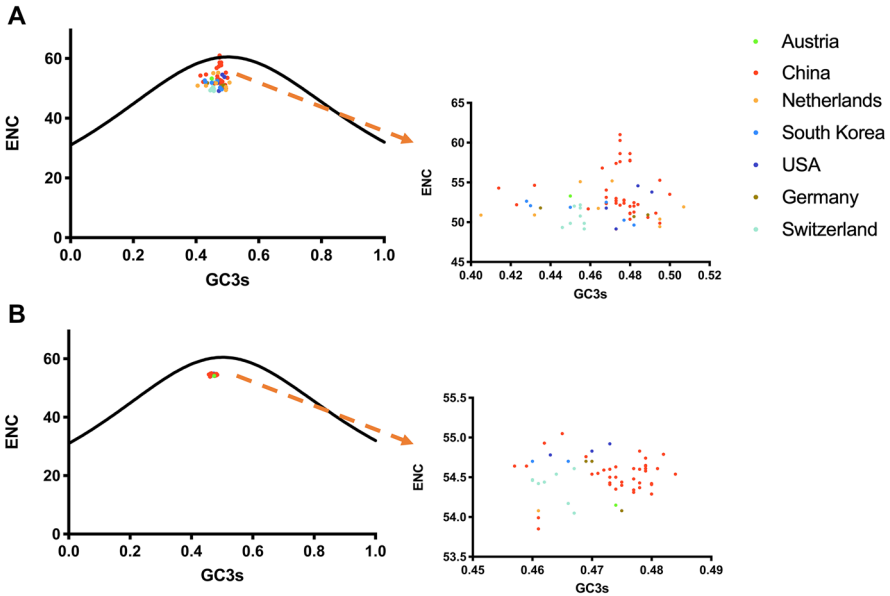
Fig. 1 The relationship between the ENC values and GC3s. **a** ENC plots for E2 gene showing the relationship between the ENC values and GC3s. The result showed that most of the points were lower than the standard curve, which indicates mutational pressure and other factors both influenced the codon usage bias. **b** ENC plots for APPV complete genomes. The result showed that all the points were lower than the standard curve. The larger version was indicated by the arrow

Table 2 correlation analyses of nucleotide compositions and ENC

	A3s	C3s	G3s	U3s	GC3s	ENC
APPV E2 gene						
A	0.6461**	0.0714	− 0.5759**	− 0.1172**	− 0.339**	− 0.2480**
C	− 0.0889**	0.8625**	0.0299**	− 0.7574**	0.6941**	− 0.2337**
G	− 0.5992**	− 0.0304**	0.7123**	− 0.0513**	0.4425**	0.3048**
U	0.0810**	− 0.8248**	− 0.1977**	0.8473**	− 0.7580**	0.1465**
GC	− 0.5571**	0.6176**	0.5981**	− 0.6006**	0.8749**	0.0574**
APPV complete genomes						
A	0.9301**	− 0.5051**	− 0.8099**	0.2740**	− 0.7504**	− 0.2900**
C	− 0.3246**	0.7793**	0.2978**	− 0.5237**	0.5834**	− 0.1982**
G	− 0.7759**	0.5740**	0.9318**	− 0.5500**	0.8696**	0.1049**
U	0.1622**	− 0.7779**	− 0.4608**	0.7909**	− 0.6926**	0.3658**
GC	− 0.6980**	0.7904**	0.7920**	− 0.6419**	0.8930**	− 0.0258**

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

Principal Component Analysis (PCA)

PCA with the RSCU values of the coding sequences were carried out to construct the distributions of each vector (Fig. 2a, b). For E2 gene, the first axis accounted for 23.23% of the data inertia and the second axis accounted for 17.86%. For APPV complete genomes, the first axis accounted for 34.67% and the second axis accounted for 19.69%. Based on the above results, the first two axes can explain 41.10% and 54.35% respectively, which suggested that they had reached the level to show characteristics of the overall codon usage trend of E2 gene and complete genomes. Therefore, the PCA plot of the first axis and the second axis was drawn according to the different countries (Fig. 2c, d). Points of different countries were dispersed, confirming that mutation pressure was one of the factors playing a role in shaping the CUB. In particular, the majority of Chinese strains were gathered into two groups with only several strains showing diversity, revealing that besides mutation pressure, the contribution of geographic selection (natural selection) in shaping the CUB in Chinese APPV strains is greater than that in other countries.

Correlation analyses were also performed to determine the correlations between the first two axes and codon compositions of E2 gene and complete genomes (Table 3). Whereas, the results indicated that the majority of correlations between codon compositions and axes were weak, which confirmed that mutational pressure was an unimportant factor contributing to the codon usage bias for the virus.
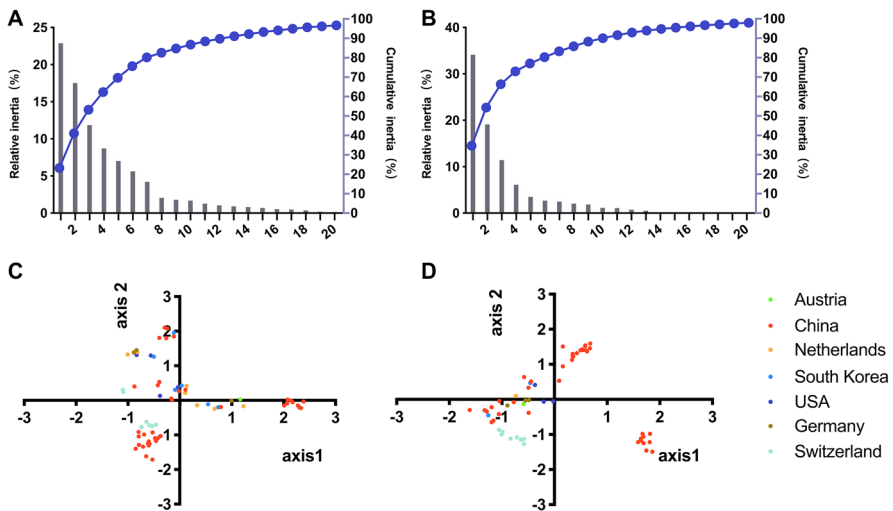

Fig. 2 Principal component analysis. **a**, **b** The distributions of the first 20 vectors by PCA for APPV E2 gene and complete genomes, respectively. Columns represent the relative inertia and the curve represents the cumulative inertia. **c**, **d** The PCA plot 1st axis against 2nd axis for E2 gene and complete genomes. Regarding E2 gene, the first axis accounted for 23.23% of the data inertia and the second axis accounted for 17.86%. Regarding APPV complete genomes, the first axis accounted for 34.67% and the second axis accounted for 19.69%. Different countries are represented by different colors

Table 3 Summary of correlation between the first two axes and nucleotide constraints

	APPV E2 gene		APPV complete genomes	
	Axis 1	Axis 2	Axis 1	Axis 2
A	− 0.0030*	− 0.0153*	− 0.3169*	− 0.6365*
C	− 0.3051	0.3043	0.1824	0.2489
G	0.0528*	− 0.0689*	0.5853	0.4625
U	0.2357	− 0.1947	− 0.4854	− 0.0467
A3s	0.4078**	0.3526**	− 0.1074**	− 0.6649**
C3s	− 0.0234*	0.3513*	0.4701*	0.1608*
G3s	− 0.3324*	− 0.1545*	0.6694*	0.3435*
U3s	− 0.0750*	− 0.4570*	− 0.7958*	0.1549*
GC	− 0.1966**	0.1720**	0.4907**	0.4449**
GC3s	− 0.2168**	0.1853**	0.6686**	0.2729**
ENC	0.7757**	− 0.1915**	0.0179**	0.3719**

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

The Role of Natural Selection in the Codon Usage Bias

To determine the role of natural selection, correlation analyses were applied for evaluating the relationship between Gravy and Aromaticity values and the codon bias (Axis1, Axis2, ENC, GC3s and GC). As shown in Table 4, most correlations were significant with the P values much below 0.01. Particularly, for E2 gene, strongly significant correlations were found between Gravy and Axis1 as well as Gravy and ENC. Above analyses suggested that the hydrophobicity and aromaticity of amino acid were associated with the codon usage variation in APPV genomes, and revealing that natural selection shaped the codon usage pattern with an important role.

Neutrality Plot Analysis

To determine the main factor shaping the codon usage pattern of the APPV E2 gene and complete genomes, neutrality plot analysis was performed by comparing

Table 4 Correlation analyses among AROMO, GRAVY, the first two axes, ENC, GC3s and GC

	Axis1	Axis2	ENC	GC3s	GC
APPV E2 gene					
Gravy	0.6174**	0.1327**	0.6185**	0.1908**	0.2003**
Aromo	0.0551	0.4896	− 0.1358**	− 0.1245**	− 0.2502**
APPV complete genomes					
Gravy	0.2361	− 0.0101	0.1027**	0.1650**	− 0.0174**
Aromo	0.3235	0.0396	0.0616**	0.2976**	0.1402**

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

the value of GC12s and GC3s (Fig. 3a, b). Regarding APPV E2 gene, the GC3s was significantly correlated with GC12s ($r = -0.243$, $P < 0.0001$) and the slope of the regression line indicated that relative neutrality (mutation pressure) was 5.34% and relative constraint (natural selection) was 94.66%. Regarding APPV complete genomes, a significant correlation was also found between GC3s and GC12s ($r = -0.495$, $P < 0.0001$) and the slope of the regression line indicated that relative neutrality (mutation pressure) was 10.30% and relative constraint (natural selection) was 89.70%. Therefore, natural selection was the main force while mutation pressure was a minor force influencing the codon usage pattern of the APPV E2 gene and complete genomes.

Discussion

APPV is an emerging virus showing worldwide distribution which regarded as a threat to global swine health. Due to eminent genetic differences and biological properties, APPV was thought to be different from established pestiviruses (Postel et al. 2017; Zhang et al. 2017). Though the variability and evolution of APPV had been investigated previously, a deep and systematic investigation should be given to fill the gaps (Pan et al. 2019). E2 is the major envelope glycoprotein of APPV and the crucial target for vaccine development while its codon usage bias during evolution is still obscure. Systematic analysis of the codon usage bias of E2 gene and complete genomes will help to clarify related questions.

Viruses were constantly evolving to adapt to the environment and host. Codon usage bias is an important manifestation of gene evolution and ENC is a simple measure of the degree of codon usage bias. Generally, the higher the ENC value, the lower the codon usage bias. For the E2 gene of APPV, ENC values ranged from 49.15 to 61.00, and the mean value of ENC was 52.74 ± 2.69, which was higher than that in other members of pestivirus, including CSFV (52.33) and BVDV (51.43) (Chen et al. 2017; Ma et al. 2018). It is worth note that the ENC value of one Chinese APPV strain even reached to maximum, which suggested that no codon usage
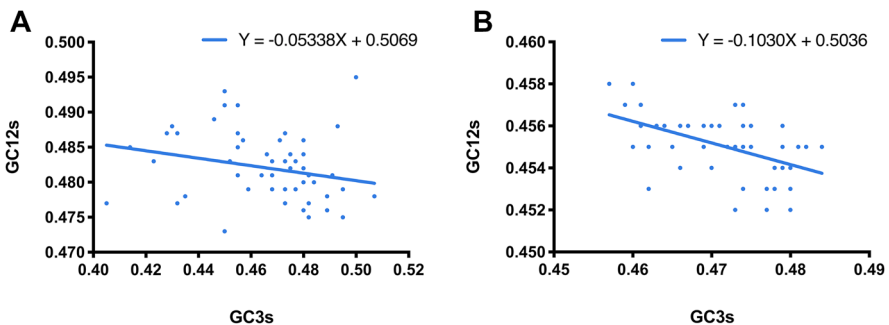

Fig. 3 Neutrality analysis with GC3s plotted against GC12s. **a**, **b** The Neutrality analysis for E2 gene and complete genomes, respectively. The regression line was represented by the straight line and the regression equation was showed on the plot

bias in this strain. Low codon usage bias allowed the virus to make use of several codons for each amino acid, which was a benefit for viral replication in the host cells and persistent infection (Chen et al. 2014; Zhang et al. 2018b). This theory may account for the phenomenon that adult sows and part of piglets did not display any symptoms of CT with APPV viremia (Munoz-Gonzalez et al. 2017).

Generally, mutational pressure was a more crucial factor compared with natural selection for RNA viruses (Peixoto et al. 2003; Tao and Yao 2019). In this study, although correlation analyses of nucleotide compositions and ENC showed the part of the strong correlations ($r > 0.5$), weak correlations were also detected especially between ENC values and each content. Secondly, almost points were below the standard curve no matter APPV complete genomes or E2 gene, but a few values also lied around the curve. Additionally, significant correlations were found with Gravy and Aromaticity values, which were two major indexes on natural selection. Finally, neutrality plot analysis was performed and the result showed that relative constraint (natural selection) was 89.70%. Overall, we can indicate that natural selection was the main force and mutational pressure played an important role in determining the CUB for APPV genomes and its E2 gene.

In summary, a slight low codon usage bias was displayed in E2 gene of APPV and natural selection was the main force while mutation pressure was a minor force influencing the codon usage pattern of APPV complete genomes and E2 gene. The results of codon usage pattern of APPV genomes will provide valuable basic data for explaining the evolution of this emerging virus.

Compliance with Ethical Standards

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

References

Arruda BL, Arruda PH, Magstadt DR, Schwartz KJ, Dohlman T, Schleining JA, Patterson AR, Visek CA, Victoria JG (2016) Identification of a divergent lineage porcine pestivirus in nursing piglets with congenital tremors and reproduction of disease following experimental inoculation. PLoS ONE 11:e0150104

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 Genomics 18:652

Butt AM, Nasrullah I, Tong Y (2014) Genome-wide analysis of codon usage and influencing factors in chikungunya viruses. PLoS ONE 9:e90905

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

Cagatay GN, Meyer D, Wendt M, Becher P, Postel A (2019) Characterization of the humoral immune response induced after infection with Atypical Porcine Pestivirus (APPV). Viruses 11:880

Chakraborty S, Deb B, Barbhuiya PA, Uddin A (2019) Analysis of codon usage patterns and influencing factors in Nipah virus. Virus Res 263:129–138

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

Chen Y, Li X, Chi X, Wang S, Ma Y, Chen J (2017) Comprehensive analysis of the codon usage patterns in the envelope glycoprotein e2 gene of the classical swine fever virus. PLoS ONE 12(9):e0183646

de Groof A, Deijs M, Guelen L, van Grinsven L, van Os-Galdos L, Vogels W, Derks C, Cruijsen T, Geurts V, Vrijenhoek M, Suijskens J, van Doorn P, van Leengoed L, Schrier C, van der Hoek L (2016) Atypical porcine pestivirus: a possible cause of congenital tremor type A-II in newborn piglets. Viruses 8:271

Hause BM, Collin EA, Peddireddi L, Yuan F, Chen Z, Hesse RA, Gauger PC, Clement T, Fang Y, Anderson G (2015) Discovery of a novel putative atypical porcine pestivirus in pigs in the USA. J Gen Virol 96:2994–2998

Ma X, Ma P, Chang Q, Li L, Zhou X, Zhang D, Li M, Cao X, Ma Z (2018) The analyses of relationships among nucleotide, synonymous codon and amino acid usages for e2 gene of bovine viral diarrhea virus. Gene 660:62–67

Martin DP, Murrell B, Golden M, Khoosal A, Muhire B (2015) RDP4: detection and analysis of recombination patterns in virus genomes. Virus Evol 1:vev003

Mosena ACS, Weber MN, da Cruz RAS, Cibulski SP, da Silva MS, Puhl DE, Hammerschmitt ME, Takeuti KL, Driemeie D, de Barcellos D, Canal CW (2018) Presence of atypical porcine pestivirus (APPV) in Brazilian pigs. Transbound Emerg Dis 65:22–26

Munoz-Gonzalez S, Canturri A, Perez-Simo M, Bohorquez JA, Rosell R, Cabezon O, Segales J, Domingo M, Ganges L (2017) First report of the novel atypical porcine pestivirus in Spain and a retrospective study. Transbound Emerg Dis 64:1645–1649

Pan S, Yan Y, Shi K, Wang M, Mou C, Chen Z (2019) Molecular characterization of two novel atypical porcine pestivirus (APPV) strains from piglets with congenital tremor in China. Transbound Emerg Dis 66:35–42

Peixoto L, Zavala A, Romero H, Musto H (2003) The strength of translational selection for codon usage varies in the three replicons of Sinorhizobium meliloti. Gene 320:109–116

Postel A, Meyer D, Cagatay GN, Feliziani F, De Mia GM, Fischer N, Grundhoff A, Milicevic V, Deng MC, Chang CY, Qiu HJ, Sun Y, Wendt M, Becher P (2017) High abundance and genetic variability of atypical porcine pestivirus in pigs from Europe and Asia. Emerg Infect Dis 23:2104–2107

Singh RK, Pandey SP (2017) Phylogenetic and evolutionary analysis of plant ARGONAUTES. Methods Mol Biol 1640:267–294

Smith DB, Meyers G, Bukh J, Gould EA, Monath T, Scott Muerhoff A, Pletnev A, Rico-Hesse R, Stapleton JT, Simmonds P, Becher P (2017) Proposed revision to the taxonomy of the genus Pestivirus, family Flaviviridae. J Gen Virol 98:2106–2112

Tao J, Yao H (2019) Comprehensive analysis of the codon usage patterns of polyprotein of Zika virus. Prog Biophys Mol Biol. https://doi.org/10.1016/j.pbiomolbio.2019pii:S0079-6107(18)30093-2

Taylor TL, Dimitrov KM, Afonso CL (2017) Genome-wide analysis reveals class and gene specific codon usage adaptation in avian paramyxoviruses 1. Infect Genet Evol 50:28–37

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

Zhang K, Wu K, Liu J, Ge S, Xiao Y, Shang Y, Ning Z (2017) Identification of atypical porcine pestivirus infection in swine herds in China. Transbound Emerg Dis 64:1020–1023

Zhang H, Wen W, Hao G, Chen H, Qian P, Li X (2018a) A subunit vaccine based on E2 protein of atypical porcine pestivirus induces Th2-type immune response in mice. Viruses 10:673

Zhang X, Cai Y, Zhai X, Liu J, Zhao W, Ji S, Su S, Zhou J (2018b) Comprehensive analysis of codon usage on rabies virus and other lyssaviruses. Int J Mol Sci 19:2397