



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



Contents lists available at ScienceDirect

Research in Veterinary Science

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

Comprehensive analysis of synonymous codon usage patterns in *orf3* gene of porcine epidemic diarrhea virus in China

Xin Xu^{a,c,d,1}, Pengfei Li^{e,1}, Yating Zhang^{a,c,1}, Xianhe Wang^{a,c}, Jiaxin Xu^{a,c}, Xuening Wu^{a,c}, Yujiang Shen^{a,c}, Dexuan Guo^{a,c}, Yuchang Li^{a,c}, Lili Yao^{a,c}, Liyang Li^{a,c}, Baifen Song^{a,c}, Jinzhu Ma^{a,c}, Xinyang Liu^{a,c}, Shuyan Xu^f, Hua Zhang^{a,b,c,*}, Zhijun Wu^{a,c,*}, Hongwei Cao^{a,c,*}

^a College of Life Science and Technology, HeiLongJiang BaYi Agricultural University, Daqing 163319, China

^b State Key Laboratory of Veterinary Biotechnology, Harbin Veterinary Research Institute of Chinese Academy of Agricultural Sciences, Harbin 150069, China

^c Biotechnology Center, HeiLongJiang BaYi Agricultural University, Daqing 163319, China

^d Branch of Animal Husbandry and Veterinary of HeiLongJiang Academy of Agricultural Sciences, Qiqihar 161005, China

^e Department of Nephrology, The Fifth Affiliated Hospital of Harbin Medical University, Daqing 163319, China

^f College of Science, HeiLongJiang BaYi Agricultural University, Daqing 163319, China

ARTICLE INFO

Keywords:

Porcine epidemic diarrhea virus
orf3 gene
Codon usage bias
Mutational bias
Natural selection

ABSTRACT

The ORF3 protein of porcine epidemic diarrhea virus (PEDV) is found to function as an ion channel which influences virus virulence and production. Taking consideration of the importance of PEDV *orf3* gene, we have performed comprehensive analysis to investigate its synonymous codon usage patterns. In this study, the results of base composition analysis showed A/T rich and G/C poor in PEDV *orf3* genes, and the most abundant base was nucleotide T. The relative synonymous codon usage value in each codon revealed that codon usage bias existed. The mean ENC value of each gene was 48.75, indicating a low codon usage bias, as well as a relatively instable change in PEDV *orf3* genes. The general correlation analysis between base composition and codon usage bias indicated that mutational bias has an impact on the PEDV codon usage bias. Neutral analysis suggested that natural selection pressure takes a more important influence than mutational bias in shaping codon usage bias. Moreover, other factors including hydrophobicity and aromaticity have been also found to influence the codon usage variation among the PEDV *orf3* genes. This study not only represents the most systematic analysis of codon usage patterns in PEDV *orf3* genes, but also provides a basic shaping mechanism of the codon usage bias.

As a highly contagious and acute enteric viral disease, porcine epidemic diarrhea (PED) is characterized by watery vomiting, diarrhea and severe dehydration, resulting into > 80% mortality in neonatal piglets (Song et al., 2015). The first PED outbreak was recognized in England in the early 1970s and then has been continually reported in other European, American and Asian countries, including China (Song and Park, 2012; Sun et al., 2012). The causative agent of PED is porcine epidemic diarrhea virus (PEDV), which belongs to the member of the *Coronaviridae* family, *Coronavirinae* subfamily, and *Alphacoronavirus* genus, including some other swine, bat and human coronaviruses (Chen et al., 2008). PEDV is a large, single-stranded positive-sense RNA enveloped virus, whose genome is approximately 28 knt encoding at least seven open reading frames (ORF1a, ORF1b, and ORF2–6), a 3' polyadenylated tail and a 5' untranslated region (5'-UTR) (Lee et al., 2015). Replicase proteins are encoded by ORF1a and ORF1b, and the viral

proteins are encoded by the next five ORFs, including the spike protein (S), the ORF3 protein (ORF3), the small membrane proteins (E), the membrane proteins (M), and the nucleocapsid protein (N) (Chen et al., 2014). As one of important viral gene, its product of *orf3* is the only accessory protein in PEDV and found to function as an ion channel to influence virus virulence and production (Song et al., 2003; Wang et al., 2012). In the majority of PEDV strains, *orf3* gene is widely used for diagnosis of PEDV infection because of its highly conserved characteristics (Wang et al., 2016). The differences in *orf3* genes between the attenuated-strain and wild-strain can also be served as a marker of the viral adaption to host and used as a potential method to study molecular evolutionary. Previous studies of PEDV *orf3* genes have been mainly limited to phylogenetic analysis (Huang et al., 2013), and few synonymous codon usage analyses have been performed (Chen et al., 2014). Especially, the factors influencing nucleotide composition and

* Corresponding authors at: College of Life Science and Technology, HeiLongJiang BaYi Agricultural University, Daqing 163319, China.

E-mail addresses: byndhzh@163.com (H. Zhang), zhjwu@tju.edu.cn (Z. Wu), byndcaohongwei@163.com (H. Cao).

¹ These authors equally contributed to this work.

<https://doi.org/10.1016/j.rvsc.2019.09.012>

Received 5 March 2019; Received in revised form 24 September 2019; Accepted 26 September 2019

0034-5288/ © 2019 Elsevier Ltd. All rights reserved.

synonymous codon usage bias in PEDV *orf3* genes have been studied only to a limited extent.

Except for tryptophan and methionine, other amino acids are encoded by 2–4 codons because the amino acids types are less than the genetic codes. This phenomenon is defined as synonymous codon usage (Chen et al., 2017). It is well known that synonymous codons for each amino acids are not used randomly in the genomes of organisms, but some codons are used more frequently than others, which is referred as synonymous codon usage bias (Marin et al., 1989). Many studies have determined codon usage bias in viruses, bacteria, fungi, and so on (D'Andrea et al., 2011). For example, the rotavirus and rubella show the strong codon usage bias among viral genome, whose degree of deviation are dependent on the identity of the virus (Belov and Lukashev, 2013). On the contrary, other virus display weak codon usage bias, such as classical swine fever virus (CSFV) (Tao et al., 2009), enterovirus 71 (EV71) (Zhang et al., 2014), and newcastle disease virus (Cao et al., 2014). Up to date, codon usage in RNA virus was also testified to be related to mutation bias, translational selection, dinucleotide bias, and other factors (Zhou et al., 2005; Sharp et al., 2010; Hussain et al., 2019). Elucidating the extent and causes of codon usage biases is beneficial for the understanding viral molecular evolution (Shackleton et al., 2006). Considering the highly contagious features of PEDV and significance of *orf3* gene, it is need to analyze the codon usage patterns of PEDV *orf3* gene during its evolution, which can provide important information about virus evolution, regulation of gene expression and protein synthesis, and further aid in vaccine design that may require high levels of viral antigen expression to produce immunity (Butt et al., 2014).

In this present study, a total of 518 coding sequences (CDS) of *orf3* gene (> 99% sequence identities excluded) of PEDV strains isolated from China were retrieved from GenBank database (<https://www.ncbi.nlm.nih.gov/nucleotide/>). The clustal X software (Thompson et al., 1997) was used for alignment of the *orf3* gene sequences. The program codonW program (version 1.4.2) (<http://codonw.sourceforge.net/>) was applied for calculating the effective number of codons (ENC), total G + C genomic content, as well as G + C content at first, second and third codon positions. The detailed information of the 518 *orf3* gene sequences is provided in supplemental data. The results showed that the T (38.22% with a SD of 0.25%) was the most abundant base, and the A (23.77% ± 0.17%), G (19.86% ± 0.23%) and C (17.09 ± 0.33%) were subsequently the second, third and fourth abundant base through base composition analysis. The average GC content of all PEDV *orf3* was 36.95% (from 36.16% to 37.95%, with a SD. of 0.29%), and the average GC3s content in codons was 33.21% (from 31.36% to 35.91%, with a SD. of 0.65%), indicating all of the PEDV *orf3* genes were A/T rich and G/C poor.

It is first proposed that the relative synonymous codon usage (RSCU) value of each codon can be calculated to directly reflect the characteristics of codon usage in 1986 (Sharp and Li, 1986). RSCU value represents the frequency of codon usage bias, whose value is 1.0 indicating no bias. In contrast, if RSCU deviates 1.0, indicating there exists a negative or positive codon usage bias (Ma et al., 2002). To gain insight into characteristics of synonymous codon usage in PEDV *orf3* genes, RSCU values were calculated using program GCUA (version 1.2) (<ftp://ftp.nhm.ac.uk/pub/gcu>), and the RSCU values of all 61 codons were displayed in Table 1. These results showed that the preferentially used codons were U-ended (11 ones), C-ended (4 ones), A-ended (3 ones), and G-ended (3 ones) codons. It was worth noting that the most preferentially used U-ended codons among the synonymous codons were similar with the result of the above T base. These results supported the evidence that T was the most abundant base content and was most preferentially used among the third position of the four kinds of nucleotides, suggesting that codon usage bias exists in the synonymous codon usage pattern in the PEDV *orf3* gene, which is influenced by compositional constraints.

The ENC value of a gene is usually performed to determine the

Table 1

Overall RSCU of the 518 collected sequences of the *orf3* gene of the PEDV.

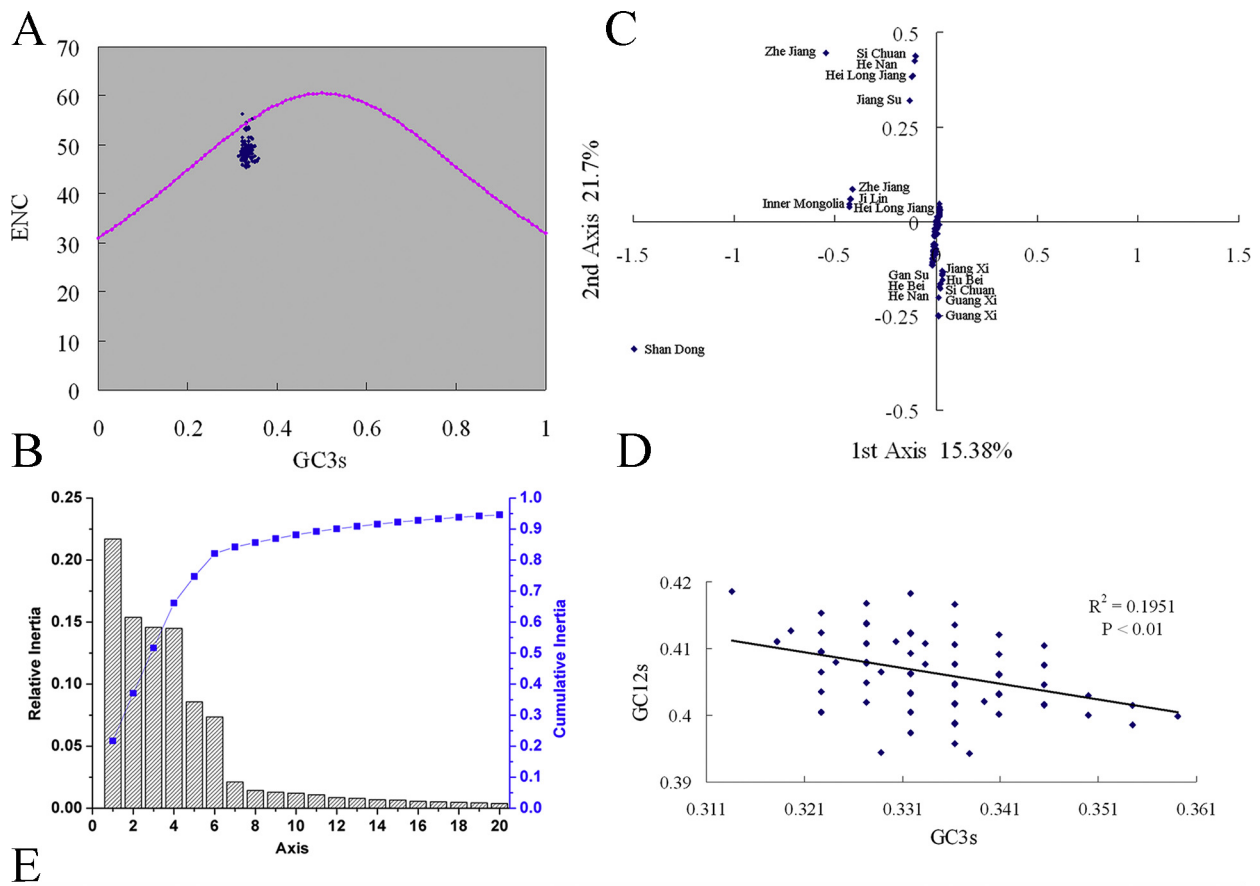
AA	Codon	N	RSCU	AA	Codon	N	RSCU
Phe	<i>UUU</i>	8073	1.61	Ser	<i>UCU</i>	3240	2.32
	UUC	1947	0.39		UCC	987	0.71
Leu	UUA	1848	0.72	Cys	UCA	1472	1.06
	<i>UUG</i>	3548	1.38		UCG	590	0.42
Tyr	<i>UAU</i>	5140	1.42	ter	UGU	1503	0.99
	UAC	2100	0.58		<i>UGC</i>	1547	1.01
ter	UAA	1	0.00	UGA	519	0.00	
ter	UAG	0	0.00	Trp	UGG	520	1.00
Leu	<i>CUU</i>	5664	2.20	Pro	CCU	7	0.03
	CUC	1565	0.61	CCC	6	0.02	
	CUA	1770	0.69	<i>CCA</i>	1042	3.95	
	CUG	1034	0.40	CCG	1	0.00	
His	<i>CAU</i>	1080	1.97	Arg	CGU	1035	1.34
	CAC	14	0.03		<i>CGC</i>	1043	1.35
Gln	<i>CAA</i>	3602	1.77	Thr	CGA	508	0.66
	CAG	476	0.23		CGG	517	0.67
Ile	<i>AUU</i>	7797	2.40	Met	<i>ACU</i>	2721	1.61
	AUC	1425	0.44		ACC	521	0.31
	AUA	522	0.16		ACA	2068	1.23
Met	AUG	1047	1.00	Asn	ACG	1431	0.85
	<i>AUU</i>	3632	1.54		Ser	<i>AGU</i>	1557
Lys	AAC	1092	0.46	Arg	AGC	520	0.37
	<i>AAA</i>	2585	1.25		AGA	495	0.64
Val	AAG	1554	0.75	Ala	<i>AGG</i>	1047	1.35
	<i>GUU</i>	4045	1.84		<i>GCU</i>	4857	1.97
	GUC	3169	1.44		GCC	334	0.14
	GUA	533	0.24		GCA	3120	1.27
Asp	GUG	1043	0.47	Gly	GCG	1550	0.63
	GAU	2043	0.72		GGU	2577	1.66
	<i>GAC</i>	3598	1.28		<i>GGC</i>	2600	1.67
Glu	GAA	1569	0.86	Glu	GGA	513	0.33
	<i>GAG</i>	2059	1.14		GGG	523	0.34

The preferentially used codons and RSCU values for *orf3* gene of the PEDV are in bold and italic. AA Amino acids, N number of codons, RSCU cumulative relative synonymous codon usage.

extent of codon usage bias. The ENC values fluctuate from 20 to 61. If the value is 20, indicating biased gene, but the value of 61 indicates the unbiased gene (Comeron and Aguade, 1998). In order to investigate the variation of codon usage bias in PEDV *orf3* genes, the ENC values of 518 genes were calculated. The results showed that ENC values varied from 45.44 to 56.37, with an average ± SD of 48.75 ± 1.29, which represented a relatively low codon usage bias and an instable change. In addition, we have performed the same analysis as the above *orf3* gene, which included a total of 294 coding sequences (CDS) of M gene of PEDV strains collected from China. The results showed that the ENC values of M gene varied from 47.45 to 60.47, with an average ± SD of 56.29 ± 1.74, which represented a comparatively stable change and a lower codon usage bias than *orf3*.

Mutational pressure and translational selection are thought to be two major factors influencing usage variation in RNA virus genome (Belov and Lukashev, 2013). The plot of ENC versus GC3s can be used to analyze synonymous codon usage bias of viral genes (Wright, 1990). Genes represented by the spots in the ENC-GC3s plot will locate above or below the predicted curve when codon usage is constrained only by a G + C mutational bias (Zhang et al., 2014). As shown in Fig. 1A, the ENC-GC3s plot showed that most points lay below the considerably predicted curve, revealing that the G + C mutational bias might play a major role in PEDV *orf3* codon usage. While some points located above the expected curve, suggesting that codon bias is also related to translational selection combined with other factors.

Subsequently, we performed a correspondence analysis (COA) to investigate the trends in 59 codon usage variation among PEDV *orf3* genes according to the previous method (Chen et al., 2014). Based on the relative and cumulative inertia of the first 20 factors, we used the Origin software (version 8.0) to display the distributions of each vector, respectively. The 21.7% of the total variation was accounted on the first



		A	C	G	U	Gravy	Aromo	ENC	Axis1	Axis2
A3s	<i>r</i>	0.786**	-0.569**	0.388**	-0.023	0.366**	-0.544**	0.401**	0.060	0.423**
	<i>P</i>	< 0.01	< 0.01	< 0.01	0.596	< 0.01	< 0.01	< 0.01	0.173	< 0.01
C3s	<i>r</i>	-0.508**	0.922**	-0.477**	-0.543**	-0.478**	0.525**	-0.215**	-0.077	-0.470**
	<i>P</i>	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	0.081	< 0.01
G3s	<i>r</i>	0.223**	-0.404**	0.829**	-0.229**	0.322**	-0.579**	0.362**	-0.041	0.305**
	<i>P</i>	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	0.350	< 0.01
U3s	<i>r</i>	0.050	-0.574**	-0.174**	0.882**	0.182**	0.036	-0.200**	0.092*	0.148**
	<i>P</i>	0.259	< 0.01	< 0.01	< 0.01	< 0.01	0.409	< 0.01	0.037	< 0.01
GC3s	<i>r</i>	-0.436**	0.790**	-0.037	-0.744**	-0.342**	0.241**	-0.024	-0.111*	-0.343**
	<i>P</i>	< 0.01	< 0.01	0.396	< 0.01	< 0.01	< 0.01	0.589	0.012	< 0.01
GC12s	<i>r</i>	0.006	-0.239**	0.477**	-0.013	0.184**	-0.517**	0.074	0.045	0.290**
	<i>P</i>	0.892	< 0.01	< 0.01	0.763	< 0.01	< 0.01	0.091	0.311	< 0.01
GC	<i>r</i>	-0.479**	0.733**	0.243**	-0.833**	-0.269**	-0.041	0.018	-0.096*	-0.207**
	<i>P</i>	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	0.357	0.678	0.029	< 0.01
Axis 1	<i>r</i>	-0.067	-0.204**	0.168**	0.137**	0.274**	-0.128**	-0.297**		
	<i>P</i>	0.127	< 0.01	< 0.01	0.002	< 0.01	0.004	< 0.01		
Axis 2	<i>r</i>	0.279**	-0.476**	0.411**	0.171**	0.286**	-0.334**	0.531**		
	<i>P</i>	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01		
ENC	<i>r</i>	0.381**	-0.214**	0.328**	-0.137**	0.245**	-0.245**			
	<i>P</i>	< 0.01	< 0.01	< 0.01	0.002	< 0.01	< 0.01			

(caption on next page)

Fig. 1. (A) ENC used in PEDV *orf3* genes plotted against the GC3s. The dotted red line is composed of the predicted ENC values. Blue dots show the results obtained for ENC values of the PEDV *orf3* genes. (B) The first 20 axes are used to display the tendency of codon usage bias of PEDV *orf3* genes. The plot is drawn according to the relative and cumulative inertia of the first 20 factors, respectively. The relative inertia is represented by the bar chart and the cumulative inertia is indicated by the curve chart based on principal component analysis. (C) Positions of the PEDV *orf3* gene in the plot of the first two major axes by COA of RSCU values. The first and second axes account for 21.7% and 15.38% of the total variation, respectively. (D) Neutrality analysis in relation to GC3s and GC12s displays the key role between mutational pressure and natural selection. (E) Summary of correlation analysis nucleotide composition, Axis1, Axis2, Gravy, Aroma, GC3s, GC12s and ENC. * P value $\leq .05$; ** P value $\leq .01$. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

principal axis. The next three axes accounted for 15.38%, 14.58%, and 14.50% of the variation, respectively, revealing that the first four axes accounted for 66.16% of the total variation (Fig. 1B). At the same time, COA was carried out on the RSCU values for each gene and its distribution in the plane defined by the first two principal axes of COA were displayed (Fig. 1C). The results showed that the vast majority of virulent genes were distributed around the origin of coordinate axis and did not distance too much from one other. Meanwhile, we also found that some genes were located at different positions in the plane, which were dispersed and far away from the origin. These strains mainly collected from southern China distributed more widespread than that of other region of China. In addition, most of the studied strains were isolated from southern China, and whose ENC values (51.73 ± 2.94) were higher than average ENC, as well as the strains belonging to other region of China. These data reflected the relatively low codon usage bias among different strains, indicating mutational bias might contribute to the codon usage bias of PEDV genome.

These above results revealed that both mutation pressure and natural selection contribute to the codon usage bias of the *orf3* gene of PEDV. Thus, to distinguish which one plays a more important role in shaping codon usage bias, the GC3s values were plotted against the GC12s values (Chen et al., 2014). The neutrality plot showed that the directional mutation pressure vs natural selection that shapes codon usage in the *orf3* gene of PEDV (Fig. 1D). We found that GC3s was significantly correlated with GC12s ($r = -0.442$, $P < .01$), with a correlation coefficient of -0.2368 , indicating that relative neutrality was 23.68%, conversely, natural selection was 76.32%. These results demonstrated that compared to mutational pressure, natural selection plays a major role in influencing the codon usage bias of *orf3* gene of PEDV.

To further analyze the possible effects of mutational pressure on the codon usage bias in PEDV *orf3* genes, we performed the correlation analysis among the nucleotide compositions (A%, T%, G%, C%, and GC%), codon compositions (A3s, T3s, G3s, C3s, and GC3s) and the ENC values. Furthermore, correlation analysis and regression analysis was performed using the values of the first two axes of this COA (Chen et al., 2014) and the Spearman's rank correlation analysis method (Tsai et al., 2007). We have conducted all statistical analyses using the statistical analysis software SPSS (Version 17.0). The nucleotide compositions were correlated with most of the codon compositions (Fig. 1E). Furthermore, there was a significant correlation between the ENC values and the most of nucleotide compositions, and all of P values were < 0.01 , which indicated that mutational bias shapes the synonymous codon usage pattern of the PEDV *orf3* gene. At last, we have evaluated the correlation between the Gravy and Aroma values and the codon contents. The results showed that Gravy value was correlated with the A3s, G3s, C3s, U3s, GC3s, GC12s and ENC. The Aroma value was correlated with the A3s, G3s, C3s, GC3s, GC12s and ENC. Both Gravy value and Aroma value were correlated with Axis 1 and Axis 2, indicating that natural selection influences the codon usage bias of PEDV *orf3* genes.

In conclusion, the codon usage pattern of PEDV *orf3* gene is comparatively low. Two main factors, mutational bias and natural selection pressure, contribute to the codon usage pattern with the latter playing a more critical role. Moreover, other factors, such as dinucleotide composition and aromaticity also influence codon usage bias. This study not only represents the most comprehensive analysis of PEDV *orf3* codon usage patterns, but also provides a basic understanding of the

mechanisms for codon usage bias. However, this study only applies to PEDV isolates from China, and our future direction of this work will focus on the comparison of PEDV isolates from other parts of the world to extensively examine the factors that cause the outbreak and evolution of this virus.

Declaration of Competing Interest

There is no conflict of interest among the contributors of this paper.

Acknowledgements

This work is supported by State Key Laboratory of Veterinary Biotechnology Foundation, Harbin Veterinary Research Institute of Chinese Academy of Agricultural Sciences (SKLVBF201913), National Natural Science Foundation of China (NSFC, Grant No. 31570159 and 31200121), Program for Young Scholars with Creative Talents in HeiLongJiang BaYi Agricultural University (CXRC2016-12), Doctor's Research Foundation, HeiLongJiang BaYi Agricultural University (XDB2015-16 and XDB2015-18), Postdoctoral Scientific Research Start-up Foundation of HeiLongJiang Province (LBH-Q17135), Joint Guidance Project of Natural Science Foundation of HeiLongJiang Province (LH2019C046), Key Project of Cultivation Foundation, HeiLongJiang BaYi Agricultural University (XA2017-02), and Earmarked Fund for China Agriculture Research System (No. CARS35). We are grateful to Prof. Songdong Meng (Professor of Institute of Microbiology, Chinese Academy of Sciences) and Prof. Paul Chu (Guest Professor of Institute of Microbiology, Chinese Academy of Sciences) for critical reading of the manuscript.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.rvsc.2019.09.012>.

References

- Belakov, I.S., Lukashev, A.N., 2013. Causes and implications of codon usage bias in RNA viruses. *PLoS One* 8, e56642.
- Butt, A.M., Nasrullah, I., Tong, Y., 2014. Genome-wide analysis of codon usage and influencing factors in Chikungunya viruses. *PLoS One* 9, e090905.
- Cao, H.W., Zhang, H., Li, D.S., 2014. Analysis of synonymous codon usage in Newcastle disease virus hemagglutinin-neuraminidase (HN) gene and fusion protein (F) gene. *Virus Diseases* 25, 132–136.
- Chen, J.F., Sun, D.B., Wang, C.B., Shi, H.Y., Cui, X.L., Liu, S.W., Qiu, H.J., Feng, L., 2008. Molecular characterization and phylogenetic analysis of membrane protein genes of porcine epidemic diarrhea virus isolates in China. *Virus Genes* 36, 355–364.
- Chen, Y., Shi, Y.Z., Deng, H.J., Gu, T., Xu, J., Ou, J.X., Jiang, Z.G., Jiao, Y.R., 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.X., Chi, X.J., Wang, S., Ma, Y.M., Chen, J.L., 2017. Comprehensive analysis of the codon usage patterns in the envelope glycoprotein E2 gene of the classical swine fever virus. *PLoS One* 12, e0183646.
- Comeron, J.M., Aguade, M., 1998. An evaluation of measures of synonymous codon usage bias. *J. Mol. Evol.* 47, 268–274.
- D'Andrea, L., Pinto, R.M., Bosch, A., Musto, H., Cristina, J., 2011. A detailed comparative analysis on the overall codon usage patterns in hepatitis A virus. *Virus Res.* 157, 19–24.
- Huang, Y.W., Dickerman, A.W., Pineyro, P., Li, L., Fang, L., Kiehne, R., Opriessnig, T., Meng, X.J., 2013. Origin, evolution, and genotyping of emergent porcine epidemic diarrhea virus strains in the United States. *MBio* 4 (e00737-13).
- Hussain, S., Rasool, S.T., Asif, A.H., 2019. A detailed analysis of synonymous codon usage in human bocavirus. *Arch. Virol.* 164, 335–347.
- Lee, S., Kim, Y., Lee, C., 2015. Isolation and characterization of a Korean porcine

- epidemic diarrhoea virus strain KNU-141112. *Virus Res.* 208, 215–224.
- Ma, J.M., Zhou, T., Gu, W.J., Sun, X., Lu, Z.H., 2002. Cluster analysis of the codon use frequency of MHC genes from different species. *Biosystems* 65, 199–207.
- Marín, A., Bertranpetit, J., Oliver, J.L., Medina, J.R., 1989. Variation in G + C-content and codon choice: differences among synonymous codon groups in vertebrate genes. *Nucleic Acids Res.* 17, 6181–6189.
- Shackleton, L.A., Parrish, C.R., Holmes, E.C., 2006. Evolutionary basis of codon usage and nucleotide composition bias in vertebrate DNA viruses. *J. Mol. Evol.* 62, 551–563.
- Sharp, P.M., Li, W.H., 1986. Codon usage in regulatory genes in *Escherichia coli* does not reflect selection for rare codons. *Nucleic Acids Res.* 14, 7737–7749.
- Sharp, P.M., Emery, L.R., Zeng, K., 2010. Forces that influence the evolution of codon bias. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 365, 1203–1212.
- Song, D., Park, B., 2012. Porcine epidemic diarrhoea virus: a comprehensive review of molecular epidemiology, diagnosis, and vaccines. *Virus Genes* 44, 167–175.
- Song, D.S., Yang, J.S., Oh, J.S., Han, J.H., Park, B.K., 2003. Differentiation of a Vero cell adapted porcine epidemic diarrhoea virus from Korean field strains by restriction fragment length polymorphism analysis of ORF 3. *Vaccine* 21, 1833–1842.
- Song, D., Moon, H., Kang, B., 2015. Porcine epidemic diarrhoea: a review of current epidemiology and available vaccines. *Clin. Exp. Vaccine Res.* 4, 166–176.
- Sun, R.Q., Cai, R.J., Chen, Y.Q., Liang, P.S., Chen, D.K., Song, C.X., 2012. Outbreak of porcine epidemic diarrhoea in suckling piglets, China. *Emerg. Infect. Dis.* 18, 161–163.
- Tao, P., Dai, L., Luo, M.C., Tang, F.Q., Tien, P., Pan, Z.S., 2009. Analysis of synonymous codon usage in classical swine fever virus. *Virus Genes* 38, 104–112.
- Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The CLUSTAL X Windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 25, 4876–4882.
- Tsai, C.T., Lin, C.H., Chang, C.Y., 2007. Analysis of codon usage bias and base compositional constraints in Iridovirus genomes. *Virus Res.* 126, 196–206.
- Wang, K., Lu, W., Chen, J., Xie, S., Shi, H., Hsu, H., Yu, W., Xu, K., Bian, C., Fischer, W.B., Schwarz, W., Feng, L., Sun, B., 2012. PEDV ORF3 encodes an ion channel protein and regulates virus production. *FEBS Lett.* 586, 384–391.
- Wang, E.Y., Guo, D.H., Li, C.Q., Wei, S., Wang, Z.H., Liu, Q.J., Zhang, B., Kong, F.Z., Feng, L., Sun, D.B., 2016. Molecular characterization of the ORF3 and S1 genes of porcine epidemic diarrhoea virus non S-INDEL strains in seven regions of China, 2015. *PLoS One* 11, e0160561.
- Wright, F., 1990. The effective number of codons used in a gene. *Gene.* 87, 23–29.
- Zhang, H., Cao, H.W., Li, F.Q., Pan, Z.Y., Wu, Z.J., Wang, Y.H., Cui, Y.D., 2014. Analysis of synonymous codon usage in enterovirus 71. *Virus Disease.* 25, 243–248.
- Zhou, T., Gu, W.J., Ma, J.M., Sun, X., Lu, Z.H., 2005. Analysis of synonymous codon usage in H5N1 virus and other influenza A viruses. *Biosystems* 81, 77–86.