One Health Advances

ARTICLE

Open Access

Codon usage and evolutionary dynamics of genetic diversity of novel imported porcine reproductive and respiratory syndrome virus in China

Chang-zhan Xie[1*†], Ping Zhang[1,3†], Yi-mo Tao[2], Qi Wang[1*], Ning-yi Jin[2*] and Hui-jun Lu[2*]

Abstract

Porcine reproductive and respiratory syndrome (PRRS) is a problem that has significant economic impact on the global pig industry. In recent years, there has been an increased importation of pork into China, contributing to the emergence of novely imported porcine reproductive and respiratory syndrome virus (PRRSV) sub-types. Nevertheless, codon usage patterns and their effects on the evolution and adaptation of these new input PRRSV sub-types in hosts remain elusive. To investigate this, we employed a Bayesian approach to analyze two novel imported PRRSV sub-types, namely, NADC30-like and NADC34-like viruses. These sub-types have different codon preferences. Besides, the Effective Number of Codon (ENC) analysis revealed that both NADC30-like and NADC34-like fall within the expected curve distribution, describing a balanced codon usage for both NADC30-like and NADC34-like virus. Based on the Codon Adaptation Index (CAI), NADC30-like showed the highest similarity to the host, aligning with the main prevalence trend of the host. In contrast, NADC34-like exhibited the highest frequency of optimal codon usage; this analysis is based on Frequency of Optimal Codons (FOP). Moreover, the Relative Codon Deoptimization Index (RCDI) indicates that NADC30-like sub-types have a greater degree of inverse optimization sub-type. These findings suggest that mutational pressure affects codon usage preferences of genes in newly imported PRRSV, and that natural selection plays a vital role in determining PRRSV gene codon preferences. Our study provides new insights into the disease, origin, evolutionary patterns, and host adaptation of these newly imported PRRSV sub-types in China. It also contributes to the development of theoretical frameworks for studying genetics and the evolution of PRRSV.

Keywords PRRSV, NADC30-like, NADC34-like, Codon usage preference

†Chang-zhan Xie and Ping Zhang contributed equally to this work.

*Correspondence:
Chang-zhan Xie
xiechangzhan2019@163.com
Qi Wang
qiwang@caas.cn
Ning-yi Jin
ningyik@126.com
Hui-jun Lu
huijun_lu@126.com
Full list of author information is available at the end of the article

Introduction

Porcine reproductive and respiratory syndrome virus (PRRSV) is a major issue affecting the global pig industry, resulting in substantial economic losses each year [1]. The infection of PRRSV destroys the whole lymphocyte population in the infected animals, thereby eliminating or destroying the adaptive immune response system and resulting in immunosuppression through various mechanisms. This makes the infected pigs more susceptible to other pathogens, leading to subsequent infections that are harmful to both the animals and PRRSV, which is genetically diverse [2]. Since the first report of PRRS in China in 1995, it has become one of the significant infectious diseases threatening the swine industry in this country. In recent years, several novel imported PRRSV strains have emerged, including the NADC30-like strains in 2013 [3] and the NADC34-like strains in 2017 [4]. Moreover, the NADC34-like strain, recognized as PRRSV/CN/FJGD01/2021, was first identified in 2022. This recombinant virus originates from the combination of NADC30-, NADC34-, and JXA1-like isolates. In 2006, a novel PRRSV variant (HP-PRRSV) emerged in China, inflicting devastating damage on the pig industry in China [3]. It is recognized to cause moderately serious respiratory symptoms and significant histopathological damage to the lungs in piglets [5].

A codon is the smallest functional unit within a DNA or RNA sequence responsible for encoding amino acids during the process of protein translation. Under ideal conditions, without selective pressure and neutral mutations, the probability of occurrence of synonymous codons should theoretically be perfectly even [6, 7]. However, numerous studies have demonstrated that the use of synonymous codons is a non-random process [8] and that some codons are employed more preferentially than others during translation [9–11]. This phenomenon of codon usage preference contains not only simple random mutations in base composition, but also corresponding functions that can explain the occurrence of different codon usage patterns in distinct species at a biological level [12]. It describes why similar codon selection strategies might be employed, thus complementing the molecular data with biological significance.

It is presumed that codon preference is the result of long-term evolution and is influenced by multiple factors, including the external environment and internal molecular mutations during the evolutionary process. Besides, studying codon preference and the factors leading to its formation is essential for understanding the characteristics of the genome, molecular evolution and ecological adaptation of novely imported PRRSV. In recent years, NADC30-like and NADC34-like strains have been spreading in China. To proactively prevent outbreaks of the two PRRSV subtypes, we conducted an examination of the codon usage of NADC30-like and NADC34-like strains identified in China. Our findings revealed that their codon usage preferences are significantly shaped by natural selection.

Results

Phylogenetic analysis of novel imported PRRSV sub-types

To assess the genetic relationships between two novel imported PRRSV sub-types, namely the NADC30-like and NADC34-like strains, we conducted an analysis. This analysis included HP-PRRSV strains that were fully sequenced. The results clearly illustrated that the NADC30-like and NADC34-like strains clustered in two distinct branches signified by different colors (Fig. 1).

Constructing of phylogenies based on the Bayesian Markov Chain method

The codon mutation rate of GP5 structural proteins, from these two novel PRRSV NADC30-like and NADC34-like, were calculated using the Bayesian Markov Chain method. The results showed that CP1 + 2:0.77 and CP3:1.461 (Fig. 2A) of the NADC30-like strain were distinct from CP1 + 2:0.619 and CP3:1.763 (Fig. 3A) of the NADC34-like strain. In addition, CP1 + 2:0.77 and CP3:1.461 (Fig. 3A) of the NADC30-like strain were different from CP1 + 2:0.619 and CP3:1.763 (Fig. 3A) of NADC34-like strains, indicating divergence. Since the third codon has the highest mutation rate, some mutations do not change the amino acids encoded in the protein, resulting in a high degree of homology between the emerging imported sub-types of PRRSV strains. Additionally, the geographical distribution of two novel imported PRRSV sub-types, namely, NADC30-like and NADC34-like strains has been progressively expanding. The Skyline plot illustrates that since 2018, the effective outbreak of the NADC30-like strain has declined in comparison with its peak in 2013 (Fig. 2B, C). Moreover, the NADC30-like strain was first detected in China in 2017 and has experienced an upward trend through 2021 (Fig. 3B, C).

Nucleotide bias of novel imported PRRSV sub-types

Among the two novel imported PRRSV sub-types and their antigenic variants, the highest %T ratios were determined in the NADC30-like and NADC34-like strains, with values of 29.560 ± 1.224 and 28.754 ± 0.701 (means ± SD), respectively (Table 1). Concerning terms of synonymous codons, the mean values of C3s (0.360 ± 0.021) and the third codon position (0.402 ± 0.022) were the highest. Respectively (Table 1), the NADC30-like and NADC34-like strains had comparable patterns of synonymous codon composition at the third position. Except for that, in the NADC30-like and NADC34-like strains, the GC content of distinct synonymous codon locations

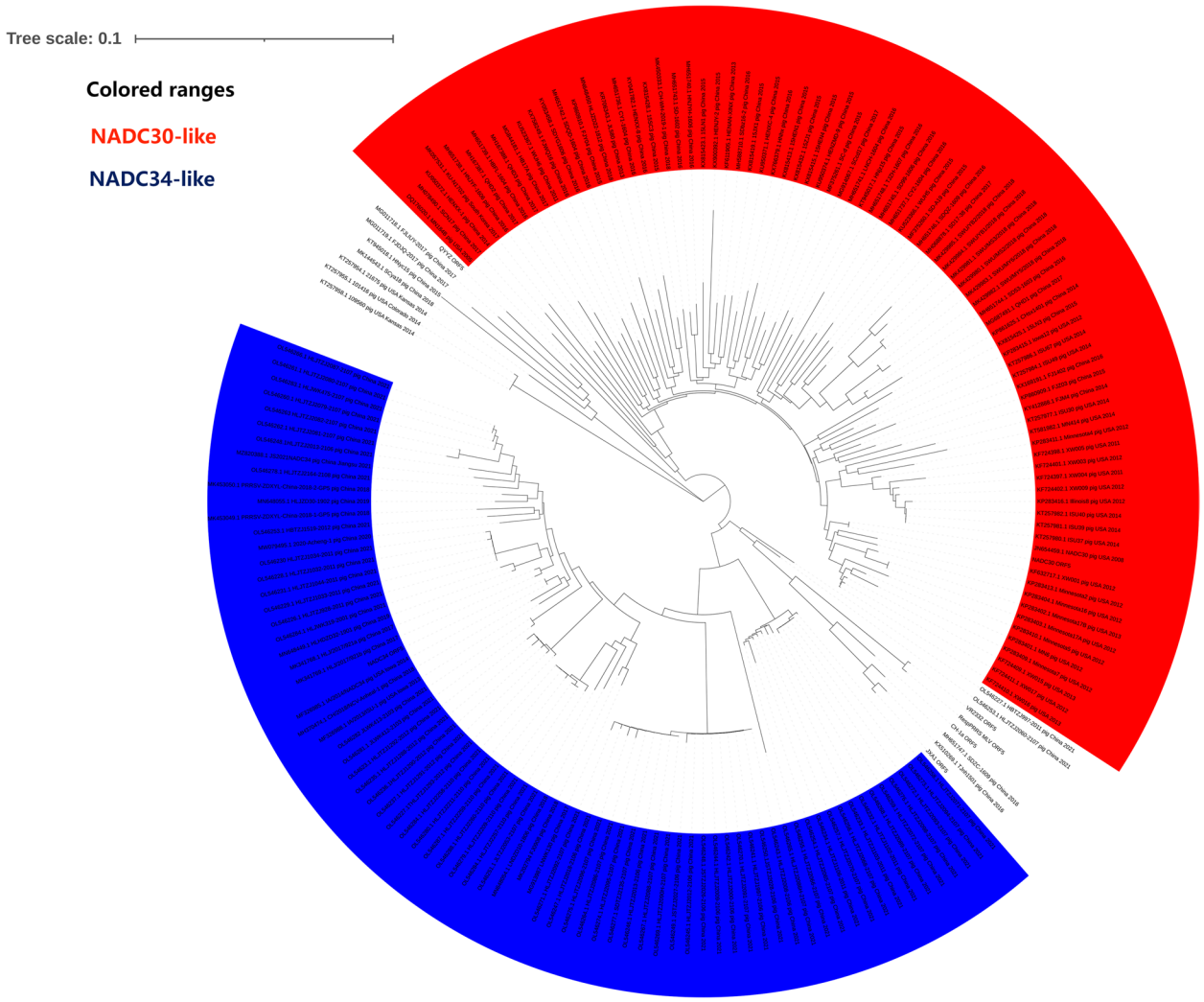

Fig. 1 PRRSV phylogenetic tree based on the nucleotide sequences of two novel imported PRRSV sub-types

was basically the same. GC3s in both strains were substantially higher than GC1s and GC2s, and across all species, GC3s > GC1s > GC2s.

Measurement of codon bias

The ENC values for the NADC30-like and NADC34-like strains were 59.223 ± 1.729 and 56.714 ± 2.387, respectively. All ENC values were higher than 35, indicating a balanced codon usage with low codon bias and balanced codon usage (Table 1).

Relative synonymous codon usage analysis of the two novel imported PRRSV sub-types genomes

Relative synonymous codon usage analysis (RSCU) is a common method for studying synonymous codon usage patterns. In our study, we noticed that 11 of the most commonly used synonymous codons are G/C-terminal codons (8 ending in C), and the number of T-terminal, A-terminal, and G-terminal codons is 4, 3, and 3, respectively (Table 2). Importantly, four of the 18 optional synonymous codons have RSCU values greater than 1.6 (RSCU < 0.6) (Fig. 4). In the following, we analyzed the tendency of synonymous codons in clades and found that G/C-terminal codons are more prevalent than A/T-terminal codons. We subsequently explored the relationship between the NADC30-like and NADC34-like strains and their natural host, pigs. Accordingly, we determined that the 18 most abundant host codons were essentially identical to those of the NADC30-like and NADC34-like strains. These data demonstrated that the two novel imported PRRSV sub-types are in a rich set of 18 synonymous codons, the majority of synonymous codon preferences are identical, which is crucial in exploring receptor binding in different sub-types of PRRSV.

A

Summary Statistic	CP1+2.mu	CP3.mu
Wean	0.7700	1.4610
95% HPD lower	0.6730	1.2814
95%HPD upper	0.8593	1.6539
Efective sample size (ESS)	2040.6	2040.6

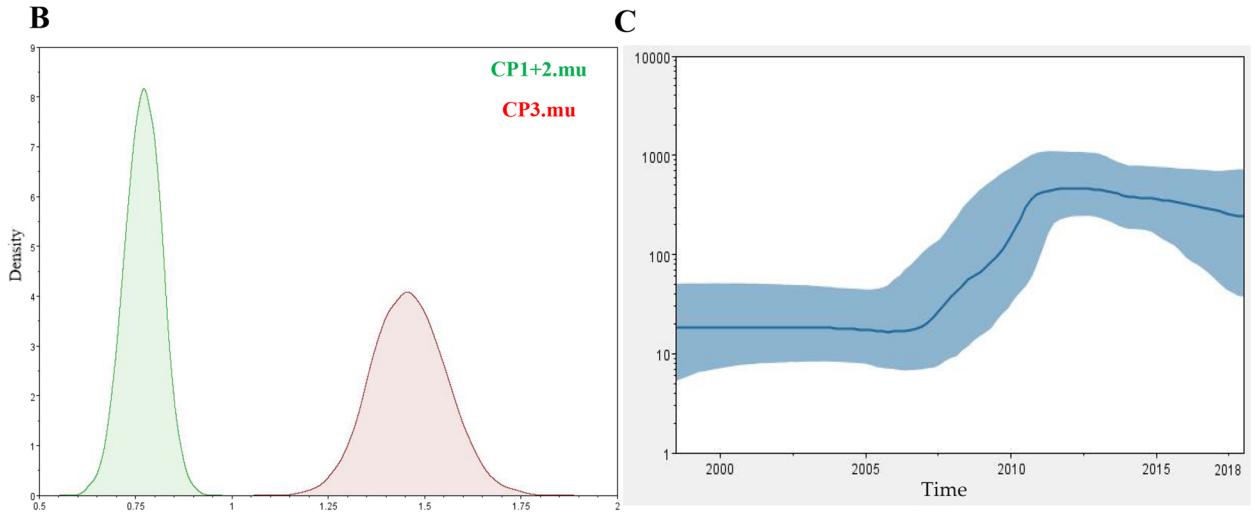

Fig. 2 Codon mutation rates and skyline plots for the structural protein of the novel input PRRSV NADC30-like strain (**A**, **B**). The Bayesian Markov Chain method (BEAST) was applied to estimate the codon mutation rate of the NADC30-like-GP5 structural protein gene (**C**)

Comparison of codon usage comparison between the virus and the host

Regarding the Codon Adaptation Index (CAI) values, both of the two novel imported PRRSV sub-type strains shared a CAI value converging to 0, indicating a low codon bias (Fig. 5A). The Relative Codon Deoptimization Index (RCDI) measures the translation rate of genes versus the general codon distribution. Among the virus genes tested, the NADC30-like strain had the highest RCDI value, implying that the higher the similarity to the host gene, the higher the translation rate (Fig. 5B).

Frequency of Optimal Codons (FOP) represents the most frequently utilized codon in a species' highly expressed genes. A comparison of the Fop values showed that the NADC34-like strain had a higher value than the NADC30-like strain, indicating a more efficient codon usage frequency as well as larger codon preference (Fig. 5C). As a PRRSV sub-type that has emerged in recent years, the NADC30-like strain's CAI, RCDI, and Fop values indicate its robust foreign gene expression ability and superior codon usage frequency compared to the novel imported PRRSV sub-type strains.

ENC–GC3s drawing analysis

The effect of GC3s on codon preference is studied using the ENC-GC3s plots. As shown in Fig. 6, the distribution of ENC-GC3s plots for the two novel imported PRRSV sub-types is relatively similar, closely fitting the expected curve. These data indicate that in the absence of any natural selection, the codon preference is merely subjected to the ideal state of mutational pressure.

G12s/GC3s neutrality plot analysis

Neutral analysis using GC12 and GC3 provided quantitative means for assessing the effects of stress mutation and natural selection. Through our assay, we observed a negative correlation between the coefficients of GC12s and GC3s in the NADC30-like strain. Conversely, the NADC34-like strain exhibited a positive correlation between the coefficients of GC12s and GC3s. Nonetheless, it is of great significance to note that these correlations were not determined to be statistically significant. This phenomenon suggested that the role of natural selection in shaping the primary factors of codon bias generation in two novel imported PRRSV sub-types, and in particular for NADC30-like (Slope of regression line −0.3562, R^2=0.0128) and NADC34-like (Slope of regression line 0.7706, R^2=0.0544) (Fig. 7). These observations suggest that mutational pressure has an effect on codon usage preferences of the two newly imported PRRSV sub-type genes, and that natural selection plays an extremely essential or even dominant role in the generation of codon preferences.

A

Summary Statistic	CP1+2.mu	CP3.mu
Wean	0.6190	1.7630
95% HPD lower	0.5006	1.5428
95% HPD upper	0.7286	1.9989
Efective sample size (ESS)	4279.2	4279.2

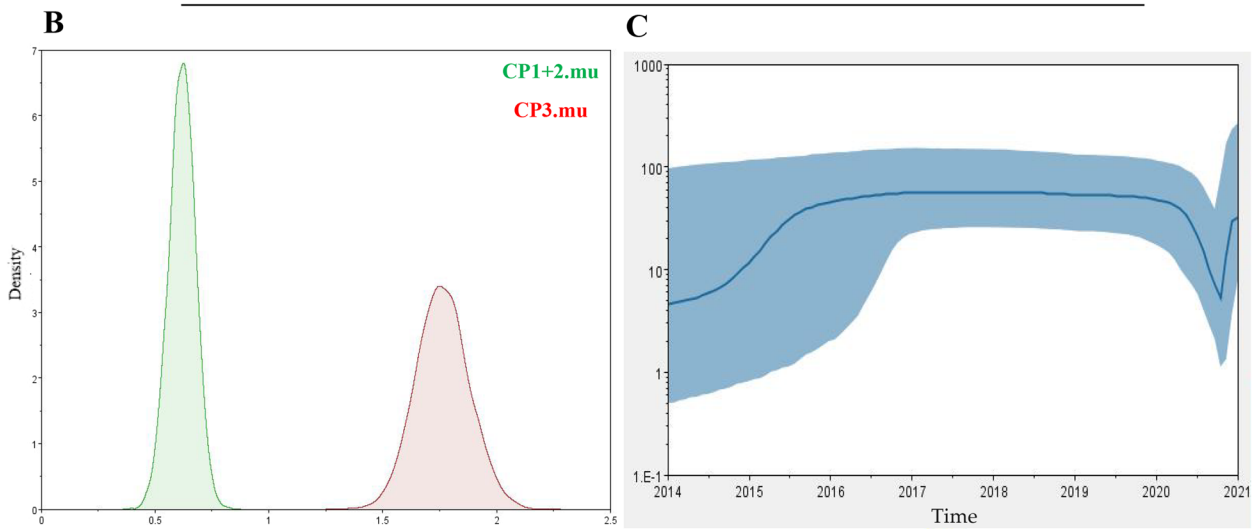

Fig. 3 Codon mutation rates and skyline plots for the structural protein of the novel input PRRSV NADC34-like strain (**A**, **B**). The Bayesian Markov Chain method (BEAST) was applied to estimate the codon mutation rate of the NADC34-like-GP5 structural protein gene (**C**)

Table 1 Properties of structural protein genes from PRRSV strains analyzed in this study (mean value ± SD)

Categories	NADC30-like	NADC34-like	All
%A	21.727 ± 3.237	21.466 ± 0.363	21.604 ± 2.361
%C	23.549 ± 0.743	24.756 ± 0.719	24.119 ± 0.947
%T	29.560 ± 1.224	28.754 ± 0.701	29.180 ± 1.086
%G	25.396 ± 0.574	25.023 ± 0.290	25.220 ± 0.497
$A3_S$	0.186 ± 0.014	0.188 ± 0.011	0.187 ± 0.013
$C3_S$	0.360 ± 0.021	0.402 ± 0.022	0.379 ± 0.030
$T3_S$	0.338 ± 0.020	0.312 ± 0.020	0.326 ± 0.024
$G3_S$	0.321 ± 0.014	0.300 ± 0.009	0.311 ± 0.016
%G + C	0.491 ± 0.008	0.500 ± 0.006	0.494 ± 0.008
$GC1_S$	0.453 ± 0.010	0.460 ± 0.008	0.455 ± 0.009
$GC2_S$	0.447 ± 0.011	0.444 ± 0.007	0.446 ± 0.009
$GC3_S$	0.570 ± 0.018	0.593 ± 0.002	0.581 ± 0.022
ENC	59.223 ± 1.729	56.714 ± 2.387	58.079 ± 2.402
CBI	0.079 ± 0.028	0.122 ± 0.029	0.099 ± 0.035

Discussion

PRRSV is a plus-stranded RNA virus with a genome approximately 15 kb in length and contains at least 10 open reading frames (ORFs), including ORF1a, ORF1b, ORF2a, ORF2b, ORF3, ORF4, ORF5, ORF5a, ORF6, and ORF7 [1]. The PRRSV-2 is widespread in China and has a subtantial impact on the pig farming industry. Chinese PRRSV-2 strains can be divided into four sub-types: VR-2332-like (Lineage 5), JXA1-like/CH-1a-like (lineage 8), QYYZ-Like (lineage 3) and NADC30-like (Lineage 1) [13, 14]. Recently, NADC34-like strain isolates appeared in the United States, and both sow herds and piglets demonstrated relatively high mortality [15]. The NADC34-like PRRSV strain was first reported in China (Shenyang) in 2017 and in Peru in 2019 [16], the NADC34-like PRRSV strain can be also isolated from MLV-inoculated pigs in South Korea. Besides, a recombinant porcine reproductive and respiratory syndrome virus with NADC30-, NADC34-, and JXA1-like strains was reported in 2022 [5]. For the purpose of exploring the evolutionary features and host adaptation of distinct imported PRRSV sub-types and analyzing codon usage, we conducted MCC phylogenetic analysis, identifing two major imported PRRSV sub-types, NADC30- and NADC34-like strains, and for the first time identified the evolutionary dynamics of a novel imported PRRSV sub-type.

Table 2 Properties of structural protein genes from two novel imported PRRSV sub-type strains, with relative synonymous codon usage analysis (Preferred codons, sub-types, and potential hosts are displayed in bold (mean value ± SD)

Categories	All	NADC30-like	NADC34-like	Sus scrofa
TTT(Phe)	**1.330 ± 0.161**	**1.322 ± 0.186**	**1.339 ± 0.127**	0.79
TTC(Phe)	0.670 ± 0.161	0.678 ± 0.186	0.661 ± 0.127	**1.21**
TTA(Leu)	0.441 ± 0.197	0.415 ± 0.193	0.470 ± 0.199	0.32
TTG(Leu)	**2.009 ± 0.280**	**2.143 ± 0.280**	**1.859 ± 0.193**	0.67
CTT(Leu)	0.965 ± 0.210	1.030 ± 0.217	0.893 ± 0.176	1.35
CTC(Leu)	0.865 ± 0.176	0.870 ± 0.199	0.860 ± 0.145	1.35
CTA(Leu)	0.383 ± 0.309	0.174 ± 0.225	0.616 ± 0.208	0.33
CTG(Leu)	1.335 ± 0.236	1.367 ± 0.280	1.300 ± 0.169	**2.68**
ATT(Ile)	1.170 ± 0.273	1.203 ± 0.313	1.132 ± 0.215	0.91
ATC(Ile)	**1.334 ± 0.249**	**1.322 ± 0.291**	**1.347 ± 0.191**	**1.67**
ATA(Ile)	0.496 ± 0.117	0.474 ± 0.138	0.520 ± 0.080	0.42
GTT(Val)	1.013 ± 0.190	0.971 ± 0.235	1.061 ± 0.104	0.57
GTC(Val)	**1.442 ± 0.247**	**1.311 ± 0.253**	**1.588 ± 0.134**	1.07
GTA(Val)	0.367 ± 0.175	0.453 ± 0.185	0.271 ± 0.097	0.34
GTG(Val)	1.175 ± 0.167	1.264 ± 0.177	1.076 ± 0.074	**2.03**
TCT(Ser)	0.724 ± 0.361	0.601 ± 0.353	0.862 ± 0.319	0.99
TCC(Ser)	1.556 ± 0.310	1.548 ± 0.332	1.564 ± 0.286	1.5
TCA(Ser)	0.952 ± 0.291	1.122 ± 0.204	0.762 ± 0.255	0.73
TCG(Ser)	0.428 ± 0.178	0.441 ± 0.215	0.413 ± 0.124	0.39
AGT(Ser)	0.389 ± 0.248	0.408 ± 0.196	0.368 ± 0.294	0.77
AGC(Ser)	**1.956 ± 0.257**	**1.885 ± 0.255**	**2.036 ± 0.237**	**1.62**
CCT(Pro)	**2.355 ± 0.624**	**1.957 ± 0.536**	**2.800 ± 0.361**	1.05
CCC(Pro)	0.471 ± 0.549	0.658 ± 0.619	0.261 ± 0.360	**1.46**
CCA(Pro)	0.728 ± 0.645	1.167 ± 0.528	0.238 ± 0.338	0.94
CCG(Pro)	0.448 ± 0.388	0.218 ± 0.401	0.706 ± 0.118	0.56
ACT(Thr)	1.114 ± 0.385	1.327 ± 0.249	0.877 ± 0.372	0.83
ACC(Thr)	**2.035 ± 0.309**	**1.888 ± 0.260**	**2.198 ± 0.277**	**1.68**
ACA(Thr)	0.511 ± 0.235	0.334 ± 0.152	0.709 ± 0.131	0.92
ACG(Thr)	0.341 ± 0.196	0.452 ± 0.144	0.218 ± 0.170	0.57
GCT(Ala)	0.583 ± 0.275	0.758 ± 0.230	0.387 ± 0.169	0.96
GCC(Ala)	1.430 ± 0.329	1.207 ± 0.216	1.679 ± 0.244	**1.8**
GCA(Ala)	0.477 ± 0.314	0.662 ± 0.279	0.271 ± 0.202	0.74
GCG(Ala)	**1.511 ± 0.266**	**1.372 ± 0.230**	**1.666 ± 0.213**	0.5
TAT(Tyr)	**1.211 ± 0.244**	**1.176 ± 0.203**	**1.249 ± 0.279**	0.73
TAC(Tyr)	0.789 ± 0.244	0.824 ± 0.203	0.751 ± 0.279	**1.27**
CAT(His)	**1.242 ± 0.286**	**1.321 ± 0.316**	**1.152 ± 0.216**	0.7
CAC(His)	0.758 ± 0.286	0.679 ± 0.316	0.848 ± 0.216	**1.3**
CAA(Gln)	**1.249 ± 0.271**	**1.294 ± 0.326**	**1.199 ± 0.182**	0.44
CAG(Gln)	0.751 ± 0.271	0.706 ± 0.326	0.801 ± 0.182	**1.56**
AAT(Asn)	0.896 ± 0.159	0.886 ± 0.183	0.907 ± 0.126	0.79
AAC(Asn)	**1.104 ± 0.159**	**1.114 ± 0.183**	**1.093 ± 0.126**	**1.21**
AAA(Lys)	**1.385 ± 0.250**	**1.273 ± 0.243**	**1.511 ± 0.192**	0.76
AAG(Lys)	0.615 ± 0.250	0.727 ± 0.243	0.489 ± 0.192	**1.24**
GAT(Asp)	0.846 ± 0.280	0.987 ± 0.232	0.689 ± 0.245	0.8
GAC(Asp)	**1.154 ± 0.280**	**1.013 ± 0.232**	**1.311 ± 0.245**	**1.2**
GAA(Glu)	0.624 ± 0.207	0.564 ± 0.233	0.690 ± 0.148	0.72
GAG(Glu)	**1.376 ± 0.207**	**1.436 ± 0.233**	**1.310 ± 0.148**	**1.28**
TGT(Cys)	0.838 ± 0.176	0.927 ± 0.159	0.738 ± 0.137	0.79

Table 2 (continued)

Categories	All	NADC30-like	NADC34-like	Sus scrofa
TGC(Cys)	**1.162 ± 0.176**	**1.073 ± 0.159**	**1.262 ± 0.137**	**1.21**
CGT(Arg)	1.268 ± 0.675	0.867 ± 0.417	1.716 ± 0.626	0.44
CGC(Arg)	0.785 ± 0.524	1.090 ± 0.386	0.444 ± 0.442	1.31
CGA(Arg)	0.178 ± 0.305	0.134 ± 0.308	0.227 ± 0.296	0.6
CGG(Arg)	1.340 ± 0.204	1.327 ± 0.218	1.354 ± 0.188	**1.29**
AGA(Arg)	**1.512 ± 0.402**	**1.677 ± 0.447**	**1.328 ± 0.235**	1.12
AGG(Arg)	0.918 ± 0.388	0.905 ± 0.385	0.933 ± 0.394	1.23
GGT(Gly)	1.281 ± 0.435	1.580 ± 0.322	0.947 ± 0.270	0.57
GGC(Gly)	**1.487 ± 0.416**	**1.182 ± 0.259**	**1.828 ± 0.266**	**1.46**
GGA(Gly)	0.272 ± 0.182	0.250 ± 0.201	0.297 ± 0.156	0.91
GGG(Gly)	0.960 ± 0.183	0.989 ± 0.215	0.929 ± 0.132	1.05

Studying codon usage preference is beneficial to enhance the efficiency of target gene expression in the receptor expression system, and the translation and expression of target genes in host cells are related to the expression system of the host cells.

If the target gene introduced into the host cell contains numerous rare codons that are not frequently utilized by the host cell's expression system, it can result in a decrease in the expression level of the target gene or premature termination of translation [17]. Optimization by codon modification has emerged as one of the most prosperous and efficient methods to enhance the expression efficacy when expressing ectopic genes in host cells [18]. The study of codon preference provides valuable insights into the evolutionary relationship between species and reflects the fundamental principles governing biological evolution. The degree of genetic relatedness between species and the extent of differences in codon preference are closely correlated. Analysis of mutational pressure and translational selection provides significant insights for further understanding of the characteristics of an organism's genes or gene sisters, molecular evolutionary pressure, and ecological adaptation. Furthermore, comprehending codon preference is crucial for several purposes, including identifying highly expressed genes, determining whether gene repression or gene level transfer is taking place, and discovering novel genes.

Previous studies have not only validated the bat origin of PCV3 with greater accuracy but have also successfully

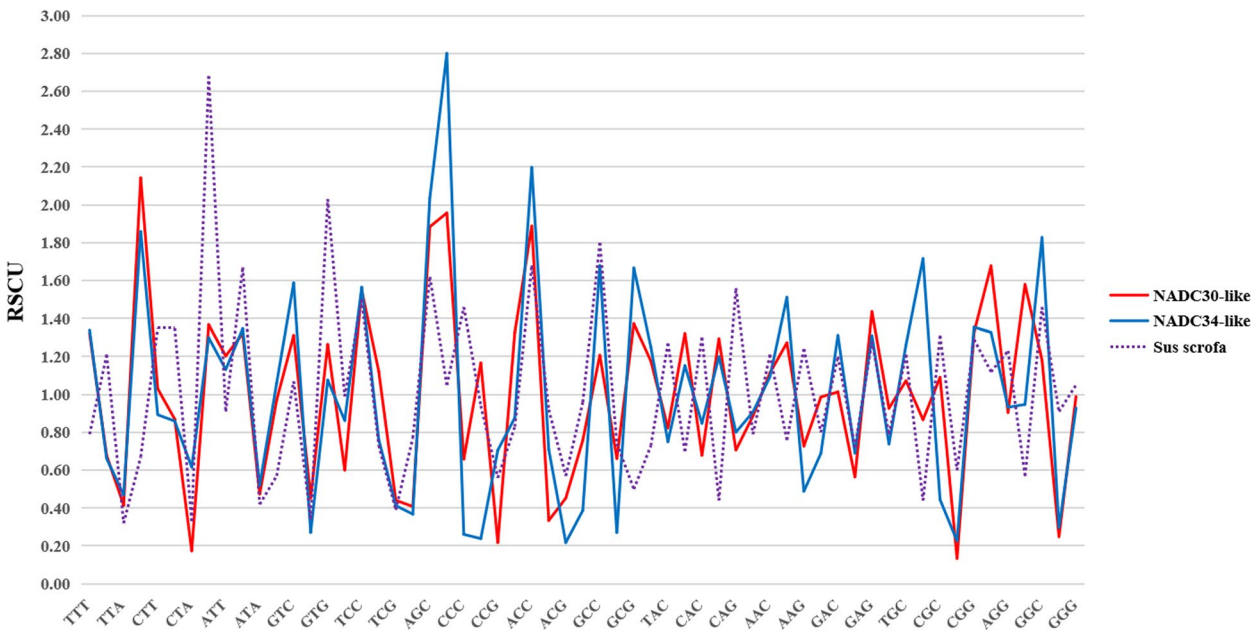

Fig. 4 Relative synonymous codon usage (RSCU) of PRRSV sub-type values. The colors of NADC30-like strain and NADC34-like strain are red and blue, respectively

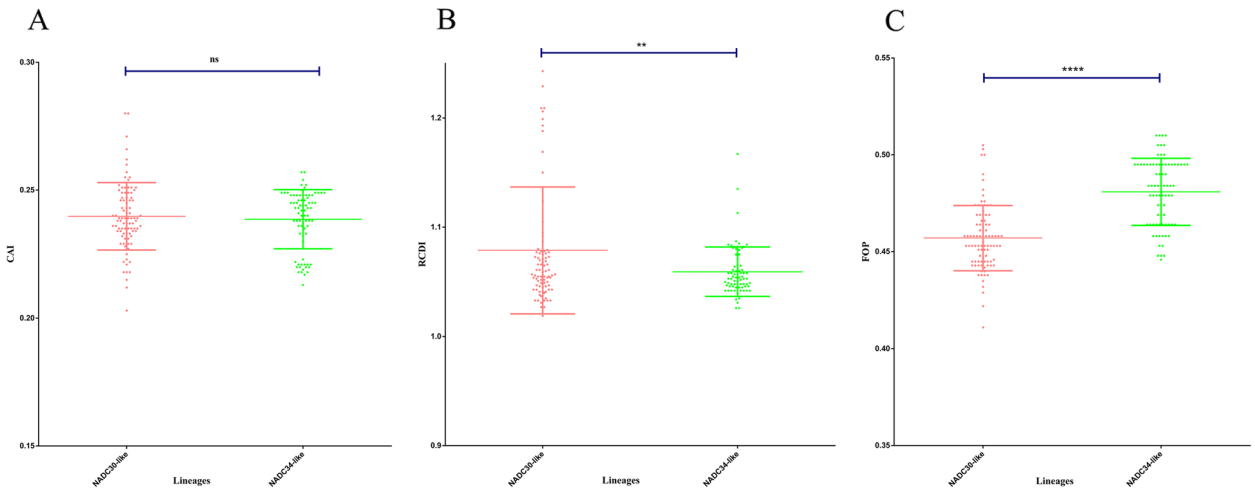

Fig. 5 **A** Scatter-plot of CAI of structural protein genes of two novel input PRRSV sub-types. **B** Scatter-plot of RCDI of structural protein genes of two novel input PRRSV sub-types. **C** Scatter-plot of FOP of structural protein genes of two novel input PRRSV sub-types. Asterisks indicate significant differences between two groups, 'ns' stands for non-statistical significance, $^{**}P < 0.01$ and $^{****}P < 0.0001$

identified various sub-types of PCV3. Consequently, PCV3 has replaced PCV2 and secured its position as the single-stranded DNA virus exhibiting the highest substitution rate known to date [19]. Natural selection is the main factor affecting the codon use preference of PCV2 clade [20]. In the analysis of synonymous codon use, 59 of the four sub-types of DENV indicated similar overall trends. Some codons, which are paired with low tRNA copy numbers in both primate species, tend to be more frequently in the translation start area compared to the open reading frame in DENV [21]. Codon usage bias was lower for the ICV HE gene and the main factor shaping codon usage was natural selection [22]. The mutation is the main factor in the codon use preference of the PPV gene, and natural selection plays a leading role in the codon use pattern [23].

In this study, we conducted a comprehensive analysis using CAI, RCID, and FOP to compare two newly imported sub-type strains of PRRSV. The results revealed a high expression capability of foreign genes and an optimal frequency of codon usage in both strains. These findings provide strong evidence of their adaptability in the host and suggest a potential correlation with virus-host infection [12]. While numerous studies look for similarities between virus and host genomes, most of these similarities are more closely related to a biological function rather than codon usage [12]. Whereas, it is not unusual for there to be different codon usage biases between viruses and host that correlate with the range of hosts infected, host translation and virulence, for example [24–26]. It has been demonstrated that the use of codons is more similar among related viruses than between viral and host genomes [27]. These insights are crucial for advancing our understanding of this virus and provided important guideline for studying the infectivity and pathogenicity of the new novel imported PRRSV sub-type strains. The codon preference of GC3s was explored using the ENc-GC3s map, and it was demonstrated that the codon preference of two novel imported PRRSV sub-type strains is merely affected by mutational pressure in the ideal state. Neutral analyses based on GC12 and GC3 could quantitatively assess the effects of mutation pressure and natural selection, showing that mutation stress affects the codon usage preference of the genes of the two novel PRRSV input sub-types, and natural selection plays a extremely essential or even dominant role in shaping codon preference. There are subtle differences in the two PRRSV sub-types, one with a higher CAI and the other with a higher FOP, and yet it is not clear if there is an evolutionary advantage for one or the other. Nonetheless, there is one sub-group that becomes more dominant over time also worth exploring and following up [28]. In conclusion, novel imported PRRSV sub-types are spreading in China, and the analysis of bias in codons indicates that the novel imported PRRSV sub-type strains are better adapted to the host for enhanced transmission.

To conclude, our study has successfully identified the impact of natural selection and mutational pressure on the codon usage preferences of two newly imported PRRSV sub-type strains. These findings shed light on the evolutionary dynamics and origins of these viruses, highlighting the crucial role of codon bias. Moreover, our study may serve as a valuable model for comprehending the evolution and codon preferences of novel imported PRRSV viruses, which is significant for predicting dynamic phylogenetic trends and establishing effective prevention strategies in clinical practice.

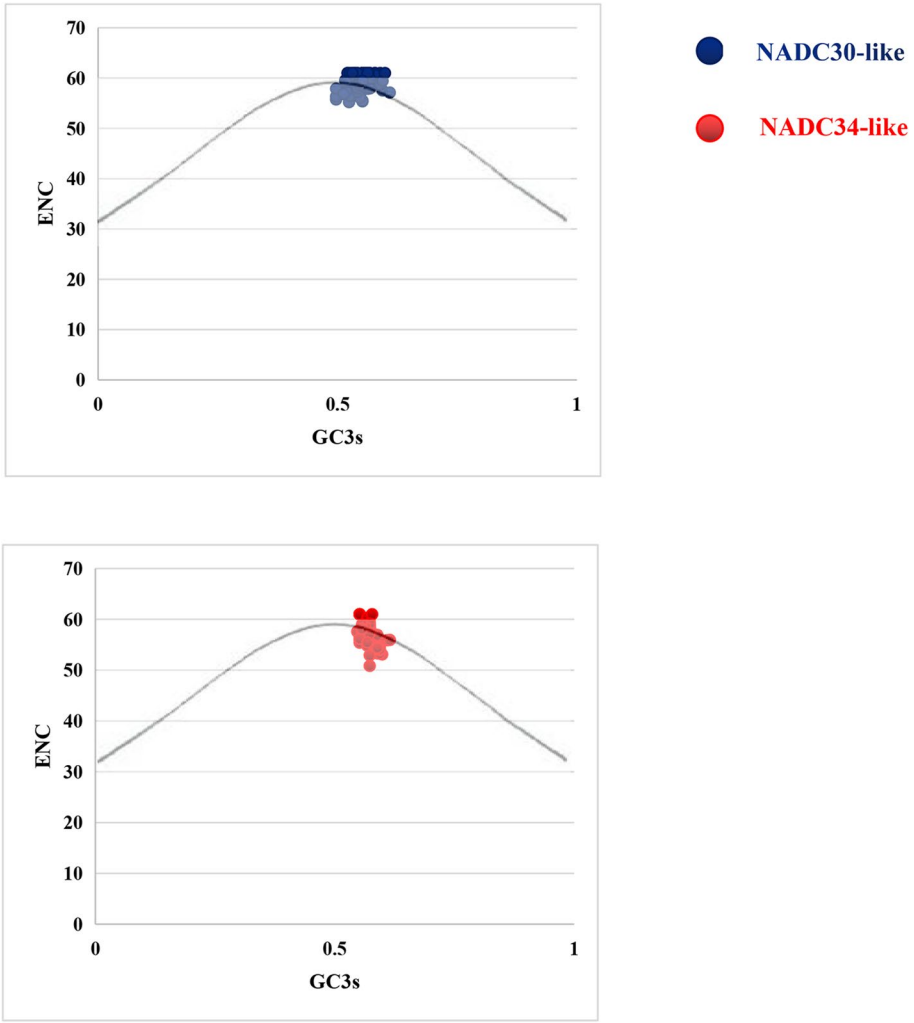

Fig. 6 Linear relationship between ENC and GC3 for two novel imported PRRSV sub-types

Materials and methods

Analysis of genome alignment and phylogenetic

Bootstrap analysis of nucleotide sequences with 1000 replicates was performed using iTOL v6 (https://itol.embl.de/). The reference sequence information is demonstrated in Table 3 (the nucleotide database was recorded in 2021).

Phylogeographic model

Phylogenetic and phylodynamic analysis based on GP5 genes of NADC30-like and NADC34-like, maximum cladistic confidence tree analysis using Markov Chain Monte Carlo (MCMC) method (computation: 10 million cycles). BEAST software (version 1.10.4) is used to calculate the time and evolution of the tMRCA rate. Best-fit models (GTR+I+G), relaxation clocks (lognormal), and the effective population size was evaluated by building a Coalescent Bayesian skyline model [29].

Codon usage index

Nucleotide composition

The nucleotide content (A%, T%, C% and G%), total GC and AT contents of PRRSV (NADC30-like-GP5 and NADC34-like-GP5) were calculated using Bioedit software (version 7.0.9.1). Besides, the frequency of A3s, C3s, T3s and G3s (i.e., the third nucleotide) was calculated using Codon W (version 1.4.2) (http://codonw.sourceforge.net/). Moreover, the GC content (%G+C) is the GC content at the first (GC1s), second (GC2s), and third (GC3s) codon locations and is calculated by Codon W (version 1.4.4). Furthermore, the GC content (%G+C) is the average frequency of GC1s and GC2s (GC12s) calculated by CAI [30].

Calculation of Effective Number of Codons (ENC)

Effective Number of Codons (ENC) is a quantitative value employed to define the average frequency of codon

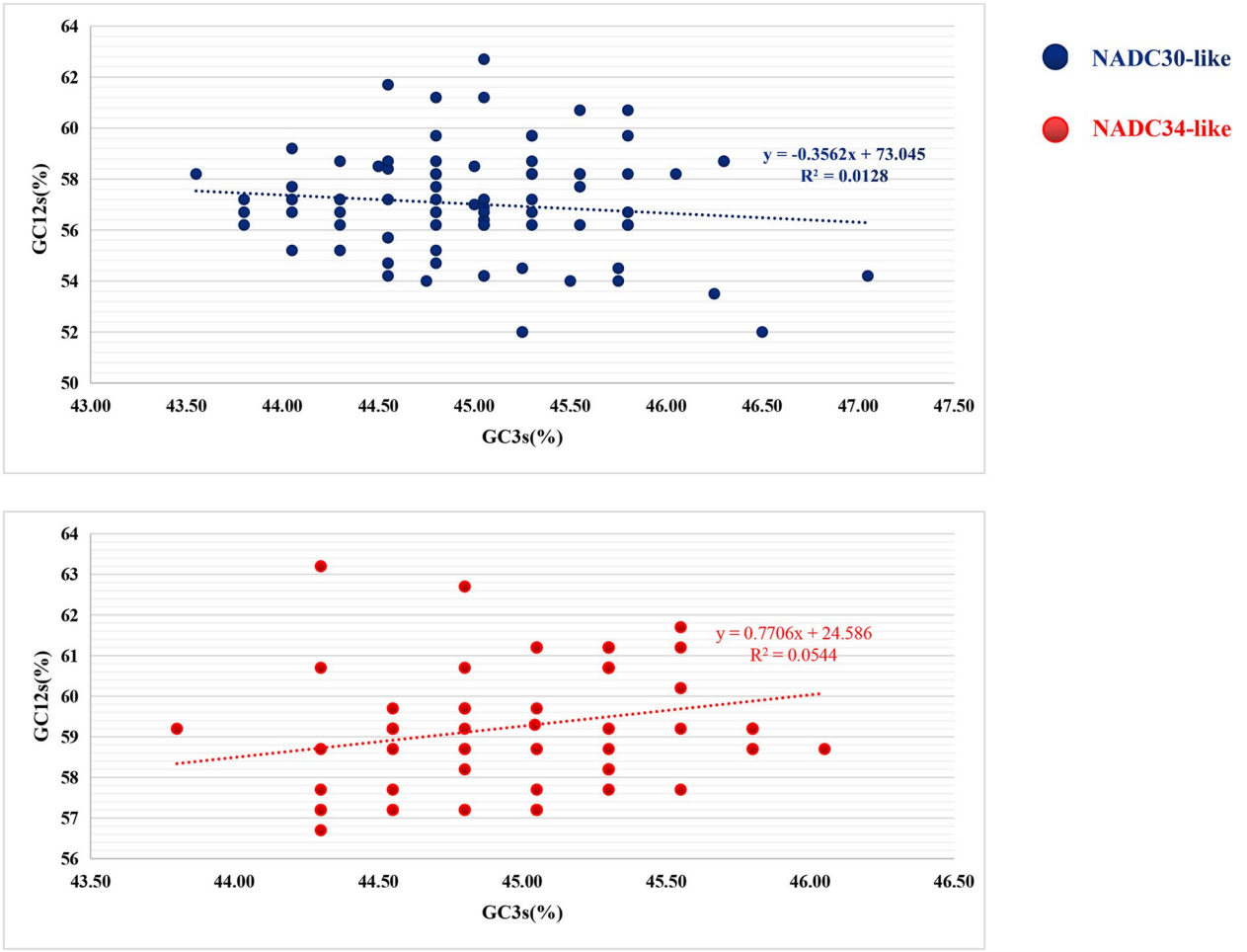

Fig. 7 Neutrality plot of two novel imported PRRSV sub-types

usage in genes that deviates from synonymous codons. It describes the extent to which codon usage deviates from random selection and better reflects the degree of preference for codon usage in genes [31].

Relative Synonymous Codon Usage Analysis (RSCU)

The RSCU is the ratio of the statistical observation of a synonymous codon to the original expectation of the number of occurrences of that codon, and is 1 if the codon has no preference, or greater than 1 if the codon is used preferentially compared to other synonymous codons, and vice versa [32] (http://www.bioinformatics.nl/cgi-bin/emboss/cusp) [23].

Relative Codon Deoptimization Index (RCDI)

The RCDI is a method for comparing general codon distributions. Translation rates and RCDI values are positively correlated with the similarity between viral and host genes (RCDI value close to 1), indicating the potential expression of certain genes or even a lower replication rate (http://genomes.urv.cat/CAIcal/RCDI) [23].

Measurement of Codon Adaptation Index (CAI)

The CAI value is adopted to measure codon usage preference, and the CAI value is the ratio of RSCU, which is evaluated by comparing the ideal RSCU value of a protein encoded by utilizing the optimal codon exclusively, weighted by applying its actual RSCU observations [33].

Analysis of Frequency of Optimal Codons (FOP)

The ratio of the best codon to the total number of codons of a gene, provided that the best codon in the highly expressed gene is obtained, is the Frequency of Optimal Codons (FOP) [23].

Calculation of Codon Bias Index (CBI)

The index used to calculate the extent of the use of the best codons is the Codon Bias Index (CBI), which

Table 3 Reference strains information

Strains	GenBank accession NO	Year	Area
MN414	KT581982.1	2005	USA
MN184B	DQ176020.1	2006	USA
NADC30	JN654459.1	2012	USA
XW017	KF724411.1	2013	USA
XW016	KF724410.1	2013	USA
XW015	KF724409.1	2013	USA
XW009	KF724402.1	2013	USA
XW005	KF724398.1	2013	USA
XW004	KF724397.1	2013	USA
XW003	KF724401.1	2013	USA
XW001	KF632717.1	2013	USA
MN6	KP283401.1	2014	USA
Minnesota7	KP283409.1	2014	USA
Minnesota5	KP283410.1	2014	USA
Minnesota4	KP283411.1	2014	USA
Minnesota2	KP283413.1	2014	USA
Minnesota17B	KP283402.1	2014	USA
Minnesota17A	KP283403.1	2014	USA
Minnesota16	KP283404.1	2014	USA
JL580	KR706343.1	2015	China
Iowa12	KP283415.1	2015	USA
Illinois8	KP283416.1	2015	USA
HNyc15	KT945018.1	2015	China
HNjz15	KT945017.1	2015	China
HENAN-XINX	KF611905.1	2015	China
FJZ03	KP860909.1	2015	China
FJY04	KP860910.1	2015	China
CHsx1401	KP861625.1	2015	China
WUH6	KU523367.1	2016	China
WUH5	KU523366.1	2016	China
TJnh1501	KX510269.1	2016	China
ISU67	KT257986.1	2016	USA
ISU49	KT257984.1	2016	USA
ISU40	KT257982.1	2016	USA
ISU39	KT257981.1	2016	USA
ISU37	KT257980.1	2016	USA
ISU30	KT257977.1	2016	USA
HNhx	KX766379.1	2016	China
HENZMD-9	KU950374.1	2016	China
HENXX-1	KU950372.1	2016	China
HENXC-4	KU950371.1	2016	China
FJ1402	KX169191.1	2016	China
109,560	KT257958.1	2016	USA
101,416	KT257955.1	2016	USA
21,675	KT257954.1	2016	USA
TJnh1501	KX510269.1	2017	China
SD-A19	MF375260.1	2017	China
SC-d	MF375261.1	2017	China
IA/2013/ISU-1	MF326988.1	2017	USA

Table 3 (continued)

Strains	GenBank accession NO	Year	Area
HENXX-8	KY041782.1	2017	China
HENJY-2	KX900392.1	2017	China
FJWQ16	KX758249.1	2017	China
FJM4	KY412888.1	2017	China
FJLIUY-2017	MG011718.1	2017	China
FJDJQ-2017	MG011719.1	2017	China
15ZJ1	KX815432.1	2017	China
15LN3	KX815428.1	2017	China
15LN1	KX815423.1	2017	China
15JX1	KX815419.1	2017	China
15HEN4	KX815415.1	2017	China
15HEN1	KX815413.1	2017	China
TJZH-1607	MH651748.1	2018	China
SDZC-1609	MH651747.1	2018	China
SDYG1606	KY053458.1	2018	China
SDQZ-1609	MH651746.1	2018	China
SDQD-1604	MH651742.1	2018	China
SD99-1606	MH651745.1	2018	China
SD53-1603	MH651744.1	2018	China
SD17-38	MH068878.1	2018	China
SD-1602	MH651743.1	2018	China
SCnj16	MF196906.1	2018	China
SCN17	MH078490.1	2018	China
SCcd17	MG914067.1	2018	China
QHD3	MH167388.1	2018	China
QHD2	MH167387.1	2018	China
QHD1	MG687491.1	2018	China
LNCH-1604	MH651741.1	2018	China
IA14737-2016	MF663706.1	2018	USA
HNJYH-1606	MH651740.1	2018	China
HNJYF-1606	MH651738.1	2018	China
HBFL-1604	MH651739.1	2018	China
HB17A	MG844181.1	2018	China
CY2-1604	MH651737.1	2018	China
CY1-1604	MH651736.1	2018	China
SWU/YB2/2018	MK429985.1	2019	China
SWU/YB1/2018	MK429984.1	2019	China
SWU/MY6/2018	MK429983.1	2019	China
SWU/MY5/2018	MK429982.1	2019	China
SWU/MS3/2018	MK429981.1	2019	China
SWU/MS2/2018	MK429980.1	2019	China
SDbz16-2	MH588710.1	2019	China
SCya18	MK144543.1	2019	China
KU-N1702	MK057531.1	2019	Korea
CH-WH-2019–1	MK450333.1	2019	China
LNWK96	MG860516.1	2018	China
LNWK130	MG913987.1	2018	China
PRRSV-ZDXYL-China-2018–1	MK453049.1	2018	China
PRRSV-ZDXYL-China-2018–2	MK453050.1	2018	China

Table 3 (continued)

Strains	GenBank accession NO	Year	Area
FJ0908	MK202794.1	2018	China
HLJZD22-1812	MN648450.1	2019	China
HLJZD30-1902	MN648055.1	2019	China
LNDZD10-1806	MN648054.1	2019	China
CH/2018/NCV-Anheal-1	MH370474.1	2018	China
NC/2015/ISU-10	KT257966.1	2016	USA
2020-Acheng-1	MW079495.1	2020	China
HLJ/2017/921a_glycoprotein_5_(GP5)	MH422084.1	2017	China
HLJ/2017/921b_glycoprotein_5_(GP5)	MH422085.1	2017	China
HLJTZJ2260-2110_GP5_(ORF5)	OL546288.1	2021	China
HLJTZJ2259-2110_GP5_(ORF5)	OL546287.1	2021	China
HLJTZJ2258-2110_GP5_(ORF5)	OL546286.1	2021	China
HLJTZJ2257-2110_GP5_(ORF5)	OL546285.1	2021	China
HLJWK319-2001_GP5_(ORF5)	OL546284.1	2021	China
HLJWK475-2107_GP5_(ORF5)	OL546283.1	2021	China
JLWK413-2103_GP5_(ORF5)	OL546282.1	2021	China
JLWK412-2103_GP5_(ORF5)	OL546281.1	2021	China
HLJTZJ2211-2110_GP5_(ORF5)	OL546280.1	2021	China
HLJTZJ2209-2110_GP5_(ORF5)	OL546279.1	2021	China
HLJTZJ2164-2108_GP5_(ORF5)	OL546278.1	2021	China
SDTZJ2135-2107_GP5_(ORF5)	OL546277.1	2021	China
HLJTZJ2098-2107_GP5_(ORF5)	OL546276.1	2021	China
HLJTZJ2096-2107_GP5_(ORF5)	OL546275.1	2021	China
HLJTZJ2095-2107_GP5_(ORF5)	OL546274.1	2021	China
HLJTZJ2094-2107_GP5_(ORF5)	OL546273.1	2021	China
HLJTZJ2093-2107_GP5_(ORF5)	OL546272.1	2021	China
HLJTZJ2092-2107_GP5_(ORF5)	OL546271.1	2021	China
HLJTZJ2091-2107_GP5_(ORF5)	OL546270.1	2021	China
HLJTZJ2090H-2107_GP5_(ORF5)	OL546269.1	2021	China
HLJTZJ2089-2107_GP5_(ORF5)	OL546268.1	2021	China
HLJTZJ2088-2107_GP5_(ORF5)	OL546267.1	2021	China
HLJTZJ2087-2107_GP5_(ORF5)	OL546266.1	2021	China
HLJTZJ2086H-2107_GP5_(ORF5)	OL546265.1	2021	China
HLJTZJ2086-2107_GP5_(ORF5)	OL546264.1	2021	China
HLJTZJ2082-2107_GP5_(ORF5)	OL546263.1	2021	China
HLJTZJ2081-2107_GP5_(ORF5)	OL546262.1	2021	China
HLJTZJ2080-2107_GP5_(ORF5)	OL546261.1	2021	China
HLJTZJ2079-2107_GP5_(ORF5)	OL546260.1	2021	China
HLJTZJ2072-2107_GP5_(ORF5)	OL546259.1	2021	China
HLJTZJ2071-2107_GP5_(ORF5)	OL546258.1	2021	China
HLJTZJ2070-2107_GP5_(ORF5)	OL546257.1	2021	China
HLJTZJ2068-2107_GP5_(ORF5)	OL546256.1	2021	China
HLJTZJ2066-2107_GP5_(ORF5)	OL546255.1	2021	China
HLJTZJ2065-2107_GP5_(ORF5)	OL546254.1	2021	China
HLJTZJ2060-2107_GP5_(ORF5)	OL546253.1	2021	China
JLTZJ2053-2107_GP5_(ORF5)	OL546252.1	2021	China
SDTZJ2039-2107_GP5_(ORF5)	OL546251.1	2021	China
JSTZJ2028-2106_GP5_(ORF5)	OL546250.1	2021	China
JSTZJ2027-2106_GP5_(ORF5)	OL546249.1	2021	China

Table 3 (continued)

Strains	GenBank accession NO	Year	Area
JSTZJ2026-2106_GP5_(ORF5)	OL546248.1	2021	China
HLJTZJ2018-2106_GP5_(ORF5)	OL546247.1	2021	China
HLJTZJ2013-2106_GP5_(ORF5)	OL546246.1	2021	China
HLJTZJ2012-2106_GP5_(ORF5)	OL546245.1	2021	China
HLJTZJ2009-2106_GP5_(ORF5)	OL546244.1	2021	China
HLJTZJ2008-2106_GP5_(ORF5)	OL546243.1	2021	China
HLJTZJ2000-2106_GP5_(ORF5)	OL546242.1	2021	China
HLJTZJ1997-2106_GP5_(ORF5)	OL546241.1	2021	China
HBTZJ1519-2012_GP5_(ORF5)	OL546240.1	2021	China
THLJTZJ1293-2012_GP5_(ORF5)	OL546239.1	2021	China
HLJTZJ1292-2012_GP5_(ORF5)	OL546238.1	2021	China
HLJTZJ1291-2012_GP5_(ORF5)	OL546237.1	2021	China
HLJTZJ1290-2012_GP5_(ORF5)	OL546236.1	2021	China
HLJTZJ1288-2012_GP5_(ORF5)	OL546235.1	2021	China
HLJTZJ1106-2011_GP5_(ORF5)	OL546234.1	2021	China
HLJTZJ1103-2011_GP5_(ORF5)	OL546233.1	2021	China
HLJTZJ1102-2011_GP5_(ORF5)	OL546232.1	2021	China
HLJTZJ1044-2011_GP5_(ORF5)	OL546231.1	2021	China
HLJTZJ1034-2011_GP5_(ORF5)	OL546230.1	2021	China
HLJTZJ1033-2011_GP5_(ORF5)	OL546229.1	2021	China
HLJTZJ1032-2011_GP5_(ORF5)	OL546228.1	2021	China
HBTZJ997-2011_GP5_(ORF5)	OL546227.1	2021	China
HLJTZJ928-2011_GP5_(ORF5)	OL546226.1	2021	China
HLHDZD32-1901	MN648449.1	2019	China
HLJZD30-1902	MN648055.1	2019	China
IA/2014/NADC34	MF326985.1	2017	USA
JXA1	EF112445.1	2016	China
VR2332	EF536003.1	2014	USA
CH-1a	AY032626.1	2016	China
QYYZ	JQ308798.1	2012	China
RespPRRS_MLV	AF066183.4	2016	USA

correlates well with the ENC values, and the expression of foreign genes in the target host is reflected. Access to the best codons in highly expressed genes is a prerequisite for the calculation of the CBI [34].

ENC–GC3s drawing

The linear relationship between ENC (vertical coordinate) and GC3s (horizontal coordinate) was explored to investigate whether factors other than mutational pressure are involved in codon usage pattern formation. Codon preference is merely affected by mutational pressure in the absence of natural selection, the ENC value will fall on or closer to the ideal curve as the ideal state. On the condition that codon preference is affected by natural selection and mutational pressure as well as other factors, the ENC value will fall below the desired curve, indicating that other factors also affect codon usage preference [35].

Neutrality plot analysis (G12s/GC3s)

The linear relationship between GC12 and GC3 mainly indicates the effect of mutation and natural selection pressure on codon usage bias. When the correlation between GC12 and GC3 is significant, mutation is the major factor and the effect of natural selection is a minor factor. Conversely, on the condition that the correlation between GC12 and GC3 is not significant, the effect of natural selection becomes the major factor and mutation is a minor factor [23].

Statistical analysis

The scores of CBI, CAI, RCDI, RSCU, FOP, ENC–GC3s and G12s/GC3s are not strictly normally distributed; $P \leq 0.001$ for highly significant relationships (****).

Acknowledgements
We are grateful to Dr. Chen Peng at China Agricultural University for his support in manuscript preparation.

Authors' contributions
C.X., P.Z., N.J., H.L. and Y.T. designed the study; P.Z. and Y.T. collected the data; C.X., P.Z. and Q.W. analyzed and interpreted the data; C.X. wrote the manuscript. All authors reviewed, revised, and approved the final report.

Funding
This study is supported by the Postdoctoral Science Foundation 72nd Grant of China (2022M723906) and Projects of Jilin Province Science and Technology Development Plan (20220508051RC).

Availability of data and materials
All the data supporting the conclusions of this article is included within the article.

Declarations

Consent for publication
Not applicable.

Competing interests
All authors declare that they have no conflicts of interest.

Author details
[1]Institute of Urban Agriculture, Chinese Academy of Agricultural Sciences, Chengdu National Agricultural Science & Technology Center, Chengdu 610213, China. [2]Chinese Academy of Agricultural Sciences, Changchun Veterinary Research Institute, Changchun, China. [3]Institute of Special Animal and Plant Sciences of Chinese Academy of Agricultural Sciences, Changchun, China.

Received: 18 June 2023 Revised: 18 September 2023 Accepted: 25 September 2023
Published online: 29 December 2023

References

1. Xie C, Ha Z, Nan F, Zhang Y, Zhang H, Li J, et al. Characterization of porcine reproductive and respiratory syndrome virus (ORF5 RFLP 1-7-4 viruses) in northern China. Microb Pathog. 2020;140:103941.
2. Jiang Y, Li G, Yu L, Li L, Zhang Y, Zhou Y, et al. Genetic diversity of porcine reproductive and respiratory syndrome virus (PRRSV) from 1996 to 2017 in China. Front Microbiol. 2020;11:618.
3. Tian K, Yu X, Zhao T, Feng Y, Cao Z, Wang C, et al. Emergence of fatal PRRSV variants: unparalleled outbreaks of atypical PRRS in China and molecular dissection of the unique hallmark. PLoS ONE. 2007;2(6):e526.
4. Zhang HL, Zhang WL, Xiang LR, Leng CL, Tian ZJ, Tang YD, et al. Emergence of novel porcine reproductive and respiratory syndrome viruses (ORF5 RFLP 1-7-4 viruses) in China. Vet Microbiol. 2018;222:105–8.
5. Liu J, Liu C, Xu Y, Yang Y, Li J, Dai A, et al. Molecular characteristics and pathogenicity of a novel recombinant porcine reproductive and respiratory syndrome virus strain from NADC30-, NADC34-, and JXA1-like strains that emerged in China. Microbiol Spectr. 2022;10(6):e0266722.
6. Supek F. The code of silence: widespread associations between synonymous codon biases and gene function. J Mol Evol. 2016;82(1):65–73.
7. Iriarte A, Lamolle G, Musto H. Codon usage bias: an endless tale. J Mol Evol. 2021;89(9–10):589–93.
8. Hershberg R, Petrov DA. Selection on codon bias. Annu Rev Genet. 2008;42:287–99.
9. Buchan JR, Aucott LS, Stansfield I. tRNA properties help shape codon pair preferences in open reading frames. Nucleic Acids Res. 2006;34(3):1015–27.
10. Hanson G, Coller J. Codon optimality, bias and usage in translation and mRNA decay. Nat Rev Mol Cell Bio. 2018;19(1):20–30.
11. Sun MY, Zhang JZ. Preferred synonymous codons are translated more accurately: proteomic evidence, among-species variation, and mechanistic basis. Sci Adv 2022;8(27):eabl9812.
12. Jitobaom K, Phakaratsakul S, Sirihongthong T, Chotewutmontri S, Suriyaphol P, Suptawiwat O, et al. Codon usage similarity between viral and some host genes suggests a codon-specific translational regulation. Heliyon 2020;6(5):e03915.
13. Shi M, Lam TT, Hon CC, Murtaugh MP, Davies PR, Hui RK, et al. Phylogeny-based evolutionary, demographical, and geographical dissection of North American type 2 porcine reproductive and respiratory syndrome viruses. J Virol. 2010;84(17):8700–11.
14. Guo Z, Chen XX, Li R, Qiao S, Zhang G. The prevalent status and genetic diversity of porcine reproductive and respiratory syndrome virus in China: a molecular epidemiological perspective. Virol J. 2018;15(1):2.
15. Alkhamis MA, Perez AM, Murtaugh MP, Wang X, Morrison RB. Applications of bayesian phylodynamic methods in a recent U.S. porcine reproductive and respiratory syndrome virus outbreak. Front Microbiol. 2016;7:67.
16. Ramirez M, Bauermann FV, Navarro D, Rojas M, Manchego A, Nelson EA, et al. Detection of porcine reproductive and respiratory syndrome virus (PRRSV) 1-7-4-type strains in Peru. Transbound Emerg Dis. 2019;66(3):1107–13.
17. Thanaraj TA, Argos P. Ribosome-mediated translational pause and protein domain organization. Prot Sci. 1996;5(8):1594–612.
18. Zelasko S, Palaria A, Das A. Optimizations to achieve high-level expression of cytochrome P450 proteins using *Escherichia coli* expression systems. Protein Expr Purif. 2013;92(1):77–87.
19. Li G, He W, Zhu H, Bi Y, Wang R, Xing G, et al. Origin, genetic diversity, and evolutionary dynamics of novel porcine circovirus 3. Adv Sci. 2018;5(9):1800275.
20. He W, Zhao J, Xing G, Li G, Wang R, Wang Z, et al. Genetic analysis and evolutionary changes of porcine circovirus 2. Mol Phylogenet Evol. 2019;139:106520.
21. Zhou JH, Zhang J, Sun DJ, Ma Q, Chen HT, Ma LN, et al. The distribution of synonymous codon choice in the translation initiation region of dengue virus. PLoS ONE. 2013;8(10):e77239.
22. Zhang W, Zhang L, He W, Zhang X, Wen B, Wang C, et al. Genetic evolution and molecular selection of the HE gene of influenza C virus. Viruses 2019;11(2):167.
23. Xie C, Tao Y, Zhang Y, Zhang P, Zhu X, Ha Z, et al. Codon usage for genetic diversity, and evolutionary dynamics of novel porcine parvoviruses 2 through 7 (PPV2-PPV7). Viruses 2022;14(2):170.
24. Tian L, Shen XJ, Murphy RW, Shen YY. The adaptation of codon usage of plus ssRNA viruses to their hosts. Infect Genet Evol. 2018;63:175–9.
25. Chen F, Wu P, Deng SY, Zhang H, Hou YT, Hu Z, et al. Dissimilation of synonymous codon usage bias in virus-host coevolution due to translational selection. Nat Ecol Evol. 2020;4(4):589–600.
26. Chen F, Yang JR. Distinct codon usage bias evolutionary patterns between weakly and strongly virulent respiratory viruses. Iscience 2022;25(1):103682.
27. Plant EP, Ye ZP. Bias at the third nucleotide of codon pairs in virus and host genomes. Sci Rep-Uk 2022;12(1):4522.
28. Zhao HZ, Wang FX, Han XY, Guo H, Liu CY, Hou LN, et al. Recent advances in the study of NADC34-like porcine reproductive and respiratory syndrome virus in China. Front Microbiol 2022;13:950402.
29. Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol. 2007;7:214.
30. Morton BR. Chloroplast DNA codon use: evidence for selection at the psb A locus based on tRNA availability. J Mol Evol. 1993;37(3):273–80.
31. Sharp PM, Tuohy TM, Mosurski KR. Codon usage in yeast: cluster analysis clearly differentiates highly and lowly expressed genes. Nucleic Acids Res. 1986;14(13):5125–43.

32. Wright F. The "effective number of codons" used in a gene. Gene. 1990;87(1):23–9.
33. 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(19):9687–96.
34. Sharp PM, Li WH. The codon adaptation index–a measure of directional synonymous codon usage bias, and its potential applications. Nucleic Acids Res. 1987;15(3):1281–95.
35. Majeed A, Kaur H, Bhardwaj P. Selection constraints determine preference for A/U-ending codons in Taxus contorta. Genome. 2020;63(4):215–24.

Publisher's Note