On the continuing and structural evolution of porcine epidemic diarrhea virus

Yan Xiao
Henan Agricultural University
Wenying Zhao
national research center for veterinary medicine
Limin Xie
national research center for veterinary medicine
Guoqian Gu
City University of Hong Kong
Yiwen Yan
Xi'an Jiaotong-Liverpool University
Yunjing Zhang
national research center for veterinary medicine
Qin Zhao
Northwest A&F University: Northwest Agriculture and Forestry University
Baicheng Huang
national research center for veterinary medicine
Kegong Tian (vetvac@126.com)
National Research Center for Veterinary Medicine https://orcid.org/0000-0002-1420-6347

Research Article

Keywords: Porcine epidemic diarrhea virus, molecular evolution, positive selection, structural assay, haplotypes

Posted Date: April 18th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1536474/v1

Abstract

The coronavirus porcine epidemic diarrhea virus (PEDV) is still widespread in pigs rapidly due to its highly contagious, resulting in huge economic losses to the pig industry. Herein, we investigated the molecular divergence between PEDV and other related coronaviruses, the results showed that the spike (S) gene exhibited larger dS (synonymous substitutions per synonymous site) values than other genes. In the assay of the extent of positive selection, eight amino acid (aa) sites of S protein showed strong signals of positive selection, and seven aa sites of them were located in the surface of S protein (S1 domain), suggesting the high selection pressure of S protein during receptor binding. We analyzed the complete genomes of 647 strains retrieved from the GenBank database. Topologically, the high similarity between the complete genome and S gene indicated that the S gene is more representative of the evolutionary relationship at the genome-wide level than other genes. Structurally, the dominant of the highly mutated residues of S protein indicated that the evolutionary pattern is highly S1 domain related. We constructed the haplotype networks using the PEDV S gene, the results showed that the strains are obviously clustered geographically in the lineages corresponding to genotype GI and GII. The alignment analysis on representative strains in the main haplotypes revealed 3 distinguishable nucleic acid sites among those strains, suggesting a putative evolutionary mechanism of PEDV. These findings provide several new fundamental insights into the evolution of PEDV and the guidance for developing effective prevention countermeasures against PEDV.

1. Introduction

The gastroenteric disease porcine epidemic diarrhea (PED) in suckling piglets, caused by porcine epidemic diarrhea virus (PEDV) infection and first reported in Europe in the early 1970s [25], characterized by severe diarrhea, vomiting and dehydration. PEDV has caused severe epidemics since 1990s in Asia such as Japan [19] and Korea [7]. The devastating outbreaks of PEDV in China since 2006 [5, 22, 42] and in United States since 2013 [32, 37] lead to a serious threat to swine health. Until now, PEDV was still the primary viral pathogen causing porcine diarrhea in China. The mortality rate of suckling piglets caused by PED often reaches 80–100%, resulting in serious economic losses to swine industry.

PEDV is an enveloped, positive-sense, single-stranded RNA virus, belonging to the genus *Alphacoronavirus* in the family Coronaviridae of the order Nidovirales. Among the four PEDV structural proteins, namely spike (S), envelope (E), membrane (M), and nucleocapsid (N), the S protein dominates the surface of virus particles and mediates direct binding to cell receptors, is a major target to induce neutralizing antibodies [4, 15, 40], and also be the target in vaccine development [3]. In coronavirus, as the major surface protein that directly interacts with cellular receptors, the S protein bears the greatest evolutionary pressure and the S gene contains the most variable regions in the entire PEDV genome, such as the strains reported in the United States with insertions and deletions in the S gene (S-INDEL) in 2013–2014 [37, 38]. Thus, the diversity of the S gene is the phylogenetic marker in PEDV evolutionary analysis [13, 20, 36] like the subgroup distinction.

Despite the discoveries in the levels of the complete genome and the S gene, several fundamental issues related to the evolutionary patterns of PEDV remain unknown. In this study, we investigated the molecular divergence between PEDV and other related coronaviruses and carried out the genetic analyses of PEDV complete genomes and structural protein genes, particularly from the view of the crystal structure of the S protein. This work would provide new insights into the evolution of PEDV and its spread pattern in animals.

2. Materials And Methods

2.1. Data collection of PEDV and other related viruses

To analyze the viral evolution of PEDV clinically, part of the PEDV complete genome sequences in GenBank (https://www.ncbi.nlm.nih.gov/genbank) were not processed into analysis due to the high cell-adapted passages, or from the artificially modified mutation. So, a total of 647 qualified whole-genome nucleic acid sequences of PEDV were selected for analysis in this study. Details on the data set are summarized in Supplementary Table S1.

2.2. Phylogenetic analyses

The sequences of PEDV complete genome, S gene and other genes were aligned using MUSCLE v3.8 [10]. The alignments were trimmed and edited by trimAl [2]. The codon alignments of the conserved ORFs of ORF1ab, S, ORF3, E, M, and N were further concatenated for down-stream evolutionary analysis. To infer phylogenetic trees, ML approaches were applied by using RAxML [31]. The best-fit nucleotide substitution model was determined by ModuleTest-NG [9]. The model GTR + I + G4 was used in RAxML which was run in 1,000 bootstrap replicates.

2.3. Positive selection of amino acids

The ratio of dN (nonsynonymous substitutions per nonsynonymous site) to dS (synonymous substitutions per synonymous site) substitution rates was identified as value of ω (dN/dS). Positive selection was analyzed using EasyCodeML [12]. The M7 (beta) and M8 (beta and $\omega > 1$) models were compared. In the M7 model, ω follows a beta distribution ($0 \leq \omega \leq 1$), and in the M8 model, a proportion p0 of sites have ω drawn from the beta distribution, and the remaining sites with proportion p1 are positively selected and have $\omega 1 > 1$ [23]. After a likelihood-ratio test (LRT) for the pairwise comparisons of codon models using EasyCodeML, we used the Naive Empirical Bayes (NEB) and Bayes empirical Bayes (BEB) methods [43] to identify amino acid residues that have potentially evolved under selection. The threshold for identifying amino acid sites under selection is a posterior probability of 0.95 [30].

2.4. Codon usage bias analysis

We calculated the RSCU (Relative Synonymous Codon Usage) value of each codon in the PEDV CV777 genome (AF353511.1). The RSCU value for each codon was the observed frequency of this codon divided by its expected frequency under equal usage among the amino acid [50]. The codons with RSCU > 1 were defined as preferred codons, and those with RSCU < 1 were defined as unpreferred codons. The FOP (frequency of optimal codons) value of each gene was calculated as the number of preferred codons divided by the total number of preferred and unpreferred codons.

2.5. Amino acid alignment

We aligned amino acid sequences of S and N protein to reveal the sequence identity using ESPript 3.0 [28]. The positively selected sites in the S and N proteins from NEB analysis were labeled (*: P > 95%; **: P > 99%).

2.6. Protein spatial structure analysis

We generate aligned sequences of S protein to reveal the highly mutated regions using a sequence logo generator WebLogo 3 [8]. The highly mutated regions in the S protein were labeled in the structural model (spheres mode) of S protein based on the PDB 6vv5 [17], the brown transparent labeled region indicted the S1 domain.

2.7. Haplotype network

We used the software DnaSP v6 [29] to generate multi-sequence aligned haplotype data, and PopART v1.7 [21] was used to draw haplotype networks based on the haplotypes generated by DnaSP v6. Evolution pattern of the PEDV was analyzed based on the representative strains from dominant haplotypes using MEGA-X [18].

3. Results

3.1. Divergence between PEDV and other related coronaviruses

We concatenated six ORFs (ORF1ab, S, ORF3, E, M, N) and used CODEML in the PAMLX to calculate the pairwise dN, dS, and ω values between PEDV and other viruses (Table 1). The results showed that the dS value varied across genes in CV777 and the other viruses, and the S gene exhibited larger dS values than other genes (Table 1), which could be caused by a high mutation rate or by natural selection that favors synonymous substitutions. The result of codon usage bias analysis (Supplementary Table S2) suggests that the frequency of optimal codons (FOP) of S gene showed no difference with that of the genomic average (0.656 versus 0.659).

Table 1
The molecular divergence between PEDV and related viruses.

Gene	Aligned Length (nt)	JS2008	AH2012	AJ1102	OH851	GDS01	PDCoV-NH	SpDCoV_HKU17	TGEV_Purdue_P115	SADS-CoV/GDGL01/2016
Genome	28033	0.1257 (0.0075 0.0596)	0.0896 (0.0083 0.0926)	0.0889 (0.0085 0.0952)	0.0782 (0.0075 0.0965)	0.0932 (0.0090 0.0961)	0.3278 (0.7386 2.2532)	0.2554 (0.7279 2.8500)	0.1850 (0.4134 2.2343)	0.2039 (0.3713 1.8208)
ORF1a	12356	0.1493 (0.0082 0.0548)	0.0999 (0.0083 0.0833)	0.1081 (0.0091 0.0842)	0.0949 (0.0084 0.0890)	0.1158 (0.0100 0.0860)	0.5766 (1.0249 1.7774)	0.5470 (1.0125 1.8510)	0.3299 (0.6076 1.8419)	0.2999 (0.4547 1.5160)
ORF1ab	20346	0.1011 (0.0060 0.0597)	0.0727 (0.0063 0.0872	0.0742 (0.0066 0.0895	0.0668 (0.0061 0.0915)	0.0790 (0.0071 0.0902)	0.3579 (0.7918 2.2122)	0.3232 (0.7881 2.4389)	0.1909 (0.4191 2.1955)	0.1770 (0.3271 1.8478)
S (spike)	4152	0.2352 (0.0183 0.0777)	0.1646 (0.0207 0.1259)	0.1566 (0.0201 0.1282)	0.1378 (0.0180 0.1310)	0.1553 (0.0214 0.1378)	0.2411 (0.5682 2.3568)	/	0.1684 (0.4468 2.6537)	0.5395 (0.8751 1.6220)
ORF3	735	0.5474 (0.0577 0.1055)	0.2019 (0.0196 0.0970)	0.2528 (0.0245 0.0971)	0.2019 (0.0196 0.0970)	0.2528 (0.0245 0.0971)	NA	NA	/	0.4643 (0.7320 1.5766)
M	813	0.2456 (0.0064 0.0261)	0.0715 (0.0043 0.0597)	0.0714 (0.0043 0.0598)	0.0641 (0.0043 0.0666)	0.0714 (0.0043 0.0598)	/	/	0.1688 (0.4790 2.8373)	0.1437 (0.3316 2.3073)
E	303	1.0790 (1.3737 1.2731)	0.0512 (0.0069 0.1342)	0.0786 (0.0069 0.0873)	0.0508 (0.0069 0.1351)	0.1563 (0.0137 0.0880)	/	/	/	/
N	1425	0.0930 (0.0071 0.0762)	0.0918 (0.0142 0.1551)	0.1271 (0.0142 0.1120)	0.0791 (0.0114 0.1439)	0.1143 (0.0128 0.1120)	0.4675 (1.0071 2.1540)	0.4776 (0.9513 1.9918)	/	/

For each gene, the dN/dS (ω) ratio between PEDV CV777 strain and other virus are given, and the dN and dS values is given in the parenthesis (dN, dS). "/" means that the method of Nei and Gojobori (Nei & Gojobori, 1986) is inapplicable.

3.2. Positive selection of PEDV and related coronaviruses

The genome-wide ω value of 0.0782 to 0.3278 (Table 1) between PEDV CV777 and other viruses indicating a strong negative selection on the nonsynonymous sites, which means 67.22–92.18% of the nonsynonymous mutations were removed during viral evolution. In the assay of extent of positive selection, the S gene of all the viruses were analyzed using the M7 (beta: neutral and negative selection) and M8 (beta & $\omega > 1$: neutral, negative, and positive selection) model in CODEML. The M8 model (lnL = -18515.815, np = 22) was a significantly better fit than the M7 (lnL = -18531.752, np = 20) model (P = 1.20e-07), suggesting that some aa substitutions were favored by positive selection. Under the M8 model, 91.436% (p0) of the nonsynonymous substitutions were estimated under neutral evolution or purifying selection ($0 \leq \omega \leq 1$), and 8.564% (p1) of the nonsynonymous substitutions were under positive selection ($\omega = 1.503$). A Naive Empirical Bayes (NEB) analysis suggested that 8 aa sites showed strong signals of positive selection, 7 of these positively selected sites were in the N-terminal domain (NTD) of S protein (surface of the S protein), and one site (1101A) in the S2 domain (Fig. 1A and 1B, Supplementary Figure S1 and S2). Two sites (166R and 213R) of S protein with the highest values of "post mean +- SE for ω" in NEB analysis also identified as the positively selected sites in Bayes Empirical Bayes (BEB) analysis. The results of positive selection assay indicated these sites was responsible for the evolution of S protein sequences, deserving further functional studies. In the N gene, The M8 model (lnL = -5144.209, np = 22) was not better fit than the M7 (lnL = -5145.170, np = 20) model (P = 0.382), suggesting no favor of the positive selection. Only one aa site of N protein showed strong signal of positive selection (247V) in the NEB analysis (Fig. 2C), but not in the BEB analysis.

3.3. Molecular phylogenetic analysis of PEDV strains

We construct the phylogenetic trees of PEDV based on the sequence of the complete genome (represented by concatenated six ORFs, including ORF1ab, S, ORF3, E, M and N), spike, and ORF3-E-M-N, respectively. Topologically, as shown in Fig. 2, the similarity between the complete genome and spike is higher than that of the complete genome and ORF3-E-M-N. Genotypes of G1 and G2 showed clear differentiation in both the phylogenetic trees of the complete genome and spike, while the tree of ORF3-E-M-N was cross-connected topologically compared with the complete genome and spike, indicating that the spike gene is more representative of the evolutionary relationship at the genome-wide level than other genes. Geographically, the America strains in genotype GI were clustered with most of the strains of Japan and South Korean. The strains from China are distributed in both the genotypes GI and GII with a similar ratio in all the 3 phylogenetic trees.

3.4. PEDV mutation in the perspective of S protein spatial structure

We analyze the mutation sites in the S protein based on spatial structure. According to the result of mutation assay using 647 sequences of S protein (Fig. 3A) and lustrated in the spatial structural data of PEDV S protein (PDB: 6vv5), we refrained from modeling residues showed as gray because these are either missing from publicly available structures. It was found that most of the highly mutated residues of S protein were located in the S1 domain, the surface of the S protein (Fig. 3B), which was consistent with the result of positively selected sites assay (Fig. 1A), indicating that the evolutionary pattern of PEDV S protein is highly S1 domain related.

3.5. The evolutionary history of PEDV lineages

In the assay of PEDV lineages formation, the putative lineages of PEDV were founded when we constructed the haplotype networks using the PEDV S gene (297 representative strains out of 647 strains). The results showed that the strains are obviously clustered geographically in the lineages (Fig. 4A) corresponding to genotype GI and GII, which was consistent with the results in the phylogenetic analysis. We performed alignment analysis on representative strains in the main haplotypes (Fig. 4B), we found 3 distinguishable nucleic acid sites among those strains, a 12 nt insertion of the GII (AACCAGGGTGTC/T), a 3 nt insertion of the GII (G/AAT), and a 6 nt deletion of the GII (GGAAAA, or AATAGA), respectively. Interestingly, the linked strain TW/Yunlin550/2018 (MK673545.1) between GI and GII showed a completely different pattern in the 6 nt site (AATAGA, aa of Asn-Arg), which is in a highly mutated region of the S1 domain as shown in Fig. 3 (sites III).

4. Discussion

The S protein of coronavirus directly interacts with cellular receptors, the diversity of the S gene is often used as the phylogenetic marker in evolutionary analysis. The complete genomes of diverse strains to the global database promotes better understanding of evolutionary and phylogenetic relationships [14]. In the study of virus evolution, one method of testing for selection is to compute the ratio of nonsynonymous to synonymous substitution rates (ω), ω is expected to have a value of 1 under the assumption of neutral evolution. Positive and negative selection are indicated when $\omega > 1$ and $\omega < 1$, respectively [23]. The M8-M7 comparison model offers a very stringent test of positive selection [1]. In terms of S gene of different coronavirus, M8 model was a better fit than the M7 model, the favor of positive selection here and the high mutation rate of the S gene [13] make it the best target for evolutionary assays, which also was confirmed in the evolutionary comparison of the whole genome sequence with the S gene and other structural protein genes.

Previous study showed that the natural selection was the main force influencing the codon usage pattern of PEDV, while mutation pressure played a minor role [6, 41, 44]. Here, if positive selection is the driving force for the higher synonymous substation rate seen in spike, we expect the FOP of spike to be different from that of genome, the elevated synonymous substitution rate measured in S gene might be more likely caused by higher mutational rates, but the FOP of S gene showed no obvious difference with that of the genomic average, the underlying molecular mechanism remains unclear, deserving further studies. Synonymous substitutions may serve as another layer of genetic regulation, guiding the efficiency of mRNA translation by changing codon usage.

By the application of NEB analysis, we found that S and N protein were favor of mutation than other proteins in this study, which was consistent with antigenic study of possible antigenic differences emerged in both the spike and nucleocapsid proteins between different genogroups [16]. As we showed here, the S gene was considered to provide the maximal interpretative power in PEDV evolution for its highest phylogenetic signal with substitution rate and phylogenetic topology similar to those obtained from the complete genome [34, 35], which would facilitate the data analysis more lightly.

Variations in the S protein are important for revealing the genetic evolution and the pathogenicity of PEDV strains [11, 33, 39]. Structurally, we found that the highly mutated residues of S protein were S1 domain dominated, indicating a strong S1-related evolutionary pattern of PEDV, which might be induced by the antigenic drift as amino acid positions with significant variation among isolates from different regions and subgroups were found [13]. PEDV S protein may undergo a conformational change after receptor binding and cleavage by exogenous trypsin, which induces membrane fusion [24].

Selective pressures drive adaptive changes in the coronavirus S proteins directing virus-cell entry. The high hypervariability in the SARS-CoV-2 S protein appears to be driven by counterbalancing pressures for effective virus-cell entry and durable extracellular virus infectivity [27], which could be caused by the variation of amino acid positions. The binding domain of the PEDV cellular receptor APN was shown to reside within a domain in the C-terminal of S1 domain (residues 477–629), which is closed to one of the sites we found in haplotype networks for the potential differentiation of PEDV genotype. The strain of TW/Yunlin550/2018 (MK673545.1) that located on the edge of the genotypes of GI and GII showed a completely different pattern in the 6 nt site (AATAGA, aa of Asn-Arg, NR) compared with the GI strains (GGAAAA, aa of Gly-Lys, GK), which might represent a consequence for a better host adaption during virus evolution when facing the counterbalancing pressure, a further study for the evolution pattern is needed.

Geographically, the GI strains of Europe were evolutionarily separated from the GII strains in other global regions. In the genotype GII, the America strains were clustered with most of the strains in Japan and South Korea, and then strains in China are shown in a more compact branch. The increasingly international pig industry involves the trade of various breeding materials and animals, which may bring the risk of disease transmission. The global exchange of ingredients has created demand for products that prevent disease transmission from the feed, such as the use of the monoglyceride blend could mitigate and prevent PED transmission in piglets from contaminated feed [26].

Overall, we found that S protein showed strong signals of positive selection, and it is more representative of the evolutionary relationship at the genome-wide level than other genes. Structurally, the evolutionary pattern of S protein is highly S1 domain related, which also represents the marker for clustering lineages corresponding to genotype GI and GII geographically. These findings provide several fundamental insights into the evolution of PEDV and the guidance for developing effective prevention countermeasures against PEDV.

Declarations

Compliance with Ethical Standards:

Funding: This study was supported by the project of R&D and industrialization of genetically engineered vaccines for swine pseudorabies, swine ring and Mycoplasma hyopneumoniae (201200211200).

Conflict of interest: The authors declare that they have no conflict of interest related to this work.

Ethical approval: This article does not contain any studies with animals performed by any of the authors.

References

1. Anisimova M, Bielawski JP, Yang Z (2001) Accuracy and power of the likelihood ratio test in detecting adaptive molecular evolution. Mol Biol Evol 18:1585–1592
2. Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25:1972–1973
3. Chang CY, Wang YS, Wu JF, Yang TJ, Chang YC, Chae C, Chang HW, Hsu SD (2021) Generation and Characterization of a Spike Glycoprotein Domain A-Specific Neutralizing Single-Chain Variable Fragment against Porcine Epidemic Diarrhea Virus. Vaccines (Basel) 9:833
4. Chang SH, Bae JL, Kang TJ, Kim J, Chung GH, Lim CW, Laude H, Yang MS, Jang YS (2002) Identification of the epitope region capable of inducing neutralizing antibodies against the porcine epidemic diarrhea virus. Mol Cells 14:295–299
5. Chen J, Wang C, Shi H, Qiu H, Liu S, Chen X, Zhang Z, Feng L (2010) Molecular epidemiology of porcine epidemic diarrhea virus in China. Arch Virol 155:1471–1476
6. 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
7. Choi JC, Lee KK, Pi JH, Park SY, Song CS, Choi IS, Lee JB, Lee DH, Lee SW (2014) Comparative genome analysis and molecular epidemiology of the reemerging porcine epidemic diarrhea virus strains isolated in Korea. Infect Genet Evol 26:348–351
8. Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) WebLogo: a sequence logo generator. Genome Res 14:1188–1190
9. Darriba D, Posada D, Kozlov AM, Stamatakis A, Morel B, Flouri T (2020) ModelTest-NG: A New and Scalable Tool for the Selection of DNA and Protein Evolutionary Models. Mol Biol Evol 37:291–294
10. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797
11. Fan B, Jiao D, Zhao X, Pang F, Xiao Q, Yu Z, Mao A, Guo R, Yuan W, Zhao P, He K, Li B (2017) Characterization of Chinese Porcine Epidemic Diarrhea Virus with Novel Insertions and Deletions in Genome. Sci Rep 7:44209
12. Gao F, Chen C, Arab DA, Du Z, He Y, Ho SYW (2019) EasyCodeML: A visual tool for analysis of selection using CodeML. Ecol Evol 9:3891–3898
13. Guo J, Fang L, Ye X, Chen J, Xu S, Zhu X, Miao Y, Wang D, Xiao S (2019) Evolutionary and genotypic analyses of global porcine epidemic diarrhea virus strains. Transbound Emerg Dis 66:111–118

14. Jarvis MC, Lam HC, Zhang Y, Wang L, Hesse RA, Hause BM, Vlasova A, Wang Q, Zhang J, Nelson MI, Murtaugh MP, Marthaler D (2016) Genomic and evolutionary inferences between American and global strains of porcine epidemic diarrhea virus. Prev Vet Med 123:175–184
15. Kang KJ, Kim DH, Hong EJ, Shin HJ (2021) The Carboxy Terminal Region on Spike Protein of Porcine Epidemic Diarrhea Virus (PEDV) Is Important for Evaluating Neutralizing Activity. Pathogens 10:683
16. Kim SJ, Nguyen VG, Huynh TM, Park YH, Park BK, Chung HC (2020) Molecular Characterization of Porcine Epidemic Diarrhea Virus and Its New Genetic Classification Based on the Nucleocapsid Gene. Viruses 12:790
17. Kirchdoerfer RN, Bhandari M, Martini O, Sewall LM, Bangaru S, Yoon KJ, Ward AB (2021) Structure and immune recognition of the porcine epidemic diarrhea virus spike protein. Structure 29:385–392 e385
18. Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Mol Biol Evol 35:1547–1549
19. Kusanagi K, Kuwahara H, Katoh T, Nunoya T, Ishikawa Y, Samejima T, Tajima M (1992) Isolation and serial propagation of porcine epidemic diarrhea virus in cell cultures and partial characterization of the isolate. J Vet Med Sci 54:313–318
20. Lee DK, Park CK, Kim SH, Lee C (2010) Heterogeneity in spike protein genes of porcine epidemic diarrhea viruses isolated in Korea. Virus Res 149:175–182
21. Leigh JW, Bryant D (2015) popart: full-feature software for haplotype network construction. 6:1110–1116
22. Li W, Li H, Liu Y, Pan Y, Deng F, Song Y, Tang X, He Q (2012) New variants of porcine epidemic diarrhea virus, China, 2011. Emerg Infect Dis 18:1350–1353
23. Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol 3:418–426
24. Park JE, Cruz DJ, Shin HJ (2011) Receptor-bound porcine epidemic diarrhea virus spike protein cleaved by trypsin induces membrane fusion. Arch Virol 156:1749–1756
25. Pensaert MB, de Bouck P (1978) A new coronavirus-like particle associated with diarrhea in swine. Arch Virol 58:243–247
26. Phillips FC, Rubach JK, Poss MJ, Anam S, Goyal SM, Dee SA (2021) Monoglyceride reduces viability of porcine epidemic diarrhea virus in feed and prevents disease transmission to post-weaned piglets. Transbound Emerg Dis 69:121–127
27. Qing E, Kicmal T, Kumar B, Hawkins GM, Timm E, Perlman S, Gallagher T (2021) Dynamics of SARS-CoV-2 Spike Proteins in Cell Entry: Control Elements in the Amino-Terminal Domains. mBio 12:e0159021
28. Robert X, Gouet P (2014) Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Res 42:W320–324
29. Rozas J, Ferrer-Mata A, Sanchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, Sanchez-Gracia A (2017) DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets. Mol Biol Evol 34:3299–3302
30. Scheffler K, Seoighe C (2005) A Bayesian model comparison approach to inferring positive selection. Mol Biol Evol 22:2531–2540
31. Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313
32. Stevenson GW, Hoang H, Schwartz KJ, Burrough ER, Sun D, Madson D, Cooper VL, Pillatzki A, Gauger P, Schmitt BJ, Koster LG, Killian ML, Yoon KJ (2013) Emergence of porcine epidemic diarrhea virus in the United States: clinical signs, lesions, and viral genomic sequences. J Vet Diagn Invest 25:649–654
33. Sun J, Li Q, Shao C, Ma Y, He H, Jiang S, Zhou Y, Wu Y, Ba S, Shi L, Fang W, Wang X, Song H (2018) Isolation and characterization of Chinese porcine epidemic diarrhea virus with novel mutations and deletions in the S gene. Vet Microbiol 221:81–89
34. Sun M, Ma J, Wang Y, Wang M, Song W, Zhang W, Lu C, Yao H (2015) Genomic and epidemiological characteristics provide new insights into the phylogeographical and spatiotemporal spread of porcine epidemic diarrhea virus in Asia. J Clin Microbiol 53:1484–1492
35. Sung MH, Deng MC, Chung YH, Huang YL, Chang CY, Lan YC, Chou HL, Chao DY (2015) Evolutionary characterization of the emerging porcine epidemic diarrhea virus worldwide and 2014 epidemic in Taiwan. Infect Genet Evol 36:108–115
36. Tang X, Wu C, Li X, Song Y, Yao X, Wu X, Duan Y, Zhang H, Wang Y, Qian Z, Cui J, Lu J (2020) On the origin and continuing evolution of SARS-CoV-2. Natl Sci Rev 7:1012–1023
37. Vlasova AN, Marthaler D, Wang Q, Culhane MR, Rossow KD, Rovira A, Collins J, Saif LJ (2014) Distinct characteristics and complex evolution of PEDV strains, North America, May 2013-February 2014. Emerg Infect Dis 20:1620–1628
38. Wang L, Byrum B, Zhang Y (2014) New variant of porcine epidemic diarrhea virus, United States, 2014. Emerg Infect Dis 20:917–919
39. Wang X, Chen J, Shi D, Shi H, Zhang X, Yuan J, Jiang S, Feng L (2016) Immunogenicity and antigenic relationships among spike proteins of porcine epidemic diarrhea virus subtypes G1 and G2. Arch Virol 161:537–547
40. Wicht O, Li W, Willems L, Meuleman TJ, Wubbolts RW, van Kuppeveld FJ, Rottier PJ, Bosch BJ (2014) Proteolytic activation of the porcine epidemic diarrhea coronavirus spike fusion protein by trypsin in cell culture. J Virol 88:7952–7961
41. Xu X, Li P, Zhang Y, Wang X, Xu J, Wu X, Shen Y, Guo D, Li Y, Yao L, Li L, Song B, Ma J, Liu X, Xu S, Zhang H, Wu Z, Cao H (2019) Comprehensive analysis of synonymous codon usage patterns in orf3 gene of porcine epidemic diarrhea virus in China. Res Vet Sci 127:42–46
42. Yang X, Huo JY, Chen L, Zheng FM, Chang HT, Zhao J, Wang XW, Wang CQ (2013) Genetic variation analysis of reemerging porcine epidemic diarrhea virus prevailing in central China from 2010 to 2011. Virus Genes 46:337–344
43. Yang Z, Wong WS, Nielsen R (2005) Bayes empirical bayes inference of amino acid sites under positive selection. Mol Biol Evol 22:1107–1118
44. Yu X, Liu J, Li H, Liu B, Zhao B, Ning Z (2021) Comprehensive analysis of synonymous codon usage patterns and influencing factors of porcine epidemic diarrhea virus. Arch Virol 166:157–165

Figures

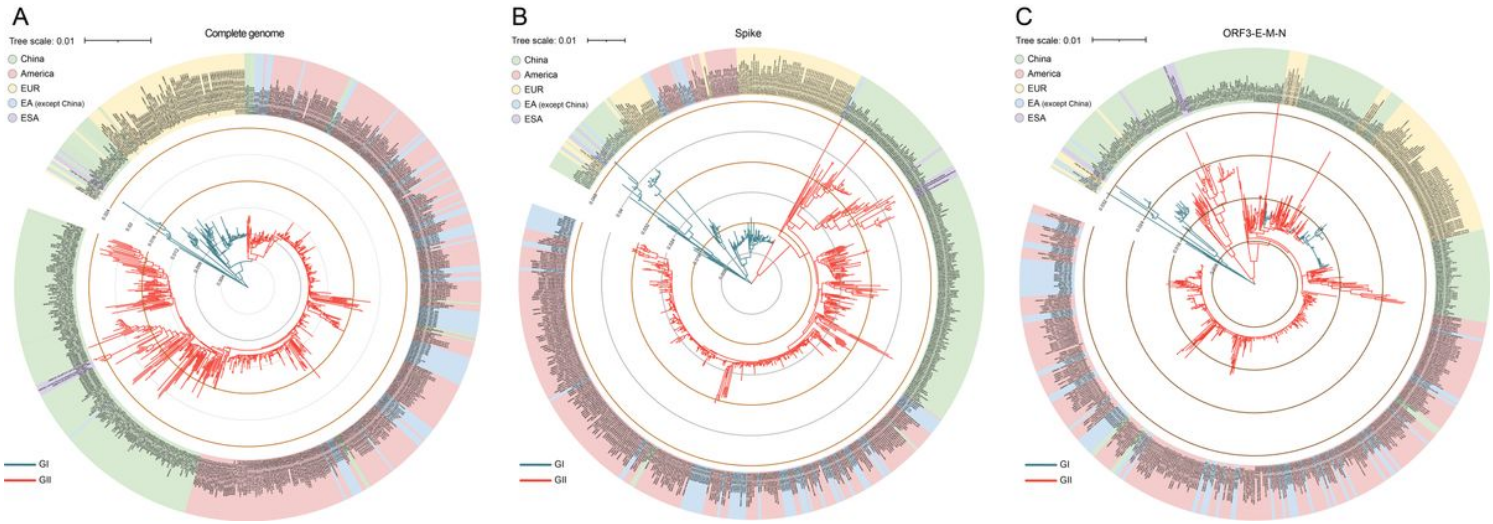

Figure 1

Selective pressures during the evolution of PEDV and related viruses.

(A) Eight positively selected sites in the spike (S) protein by NEB analysis, seven of them in the domain NTD and one in the domain S2 (*: P>95%; **: P>99%). (B) The NTD domain, in the surface of the S protein, showed as surface model based on the PDB file 6vv5. (C) One positively selected site in the N protein by NEB analysis (*: P>95%).

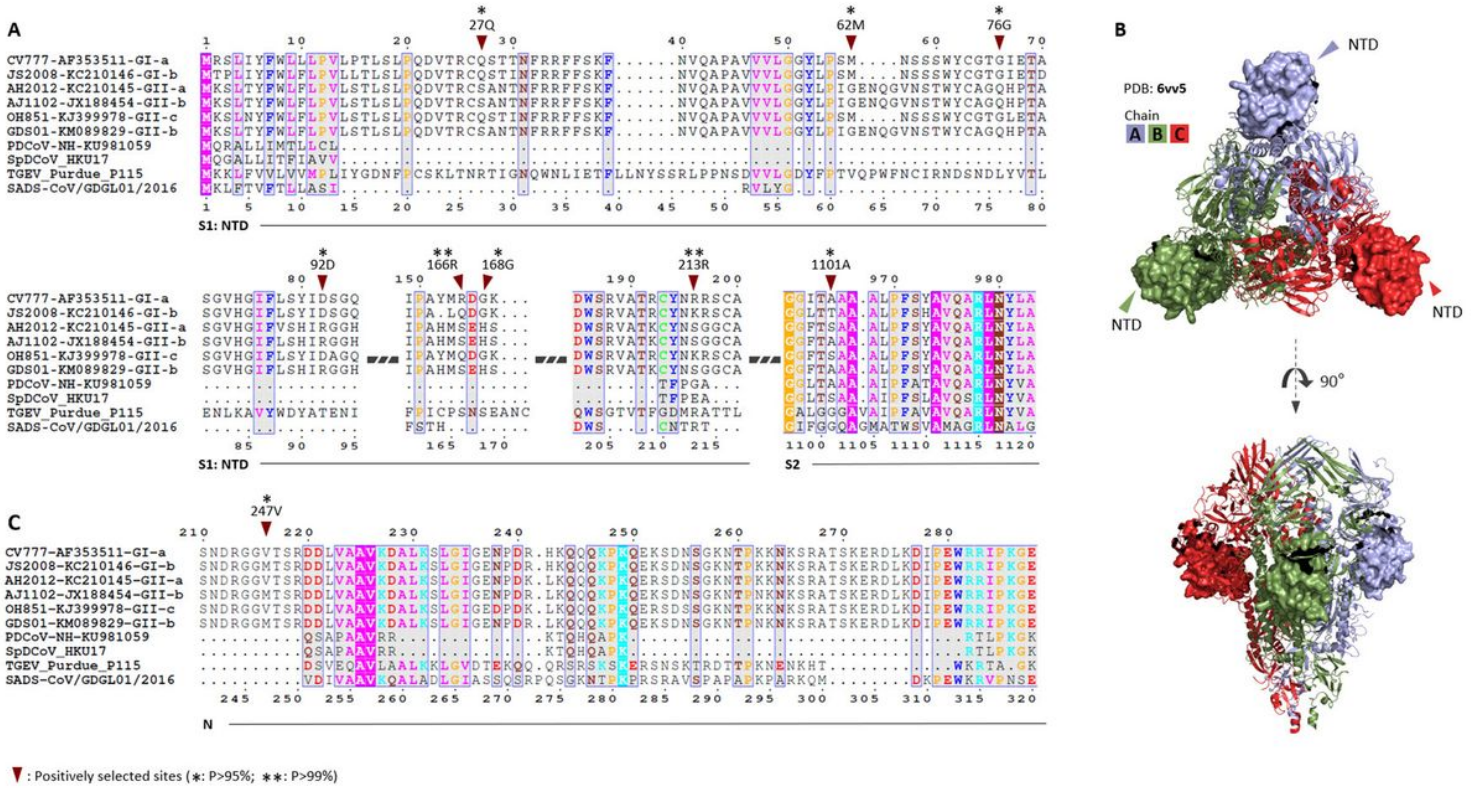

Figure 2

Genotyping and origin of the 647 PEDV strains based on different genes.

Phylogenetic trees were constructed by the maximum-likelihood (ML) method based on the (A) complete genome, (B) S gene, and (C) ORF3-E-M-N gene sequences, with 1,000 bootstrap replicates. Names of strains, isolation regions (EUR, Europe; EA, Japan and South Korean; ESA, Southeast Asia), GenBank accession numbers, genogroups are shown.

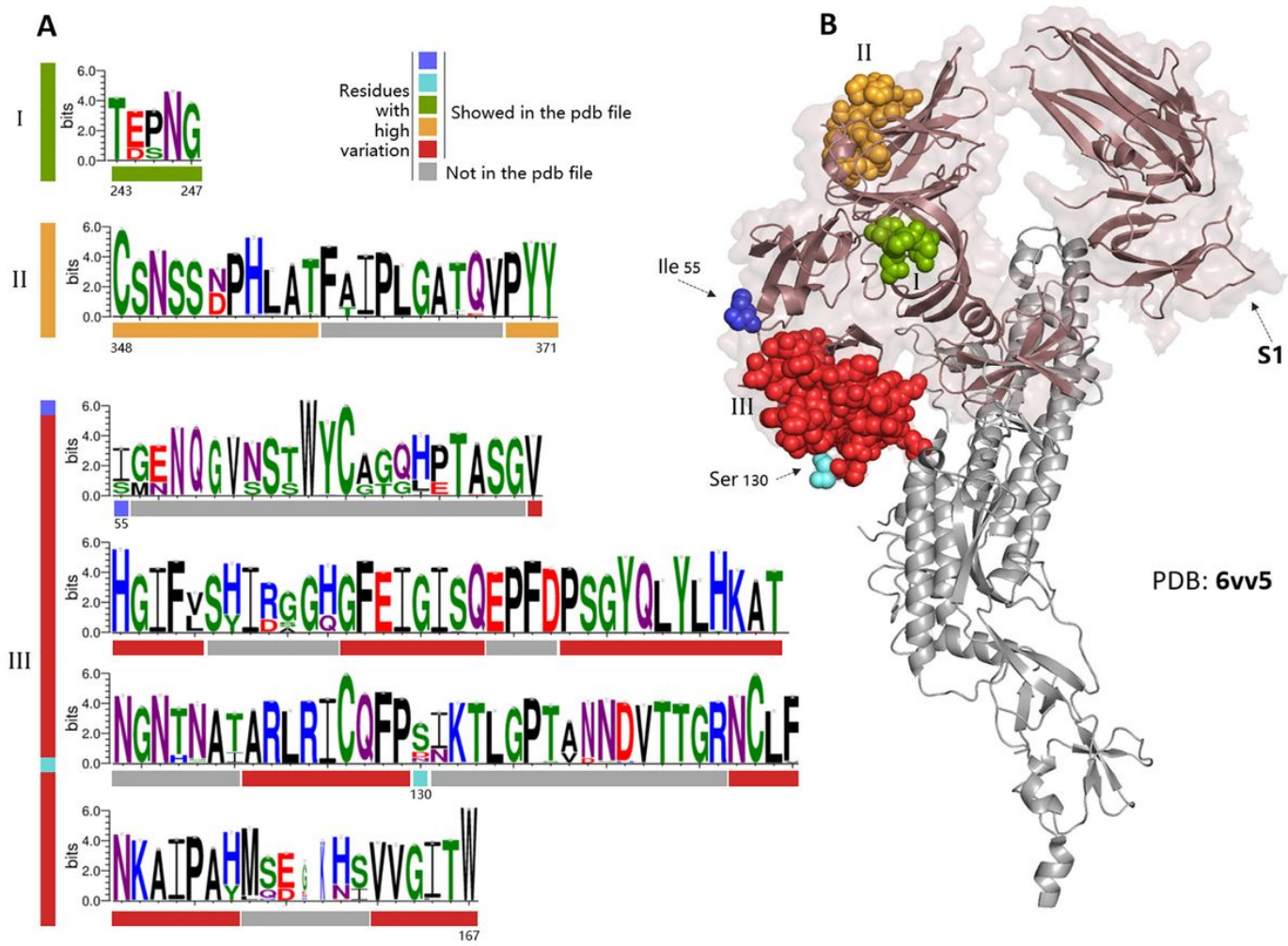

Figure 3

Schematic diagram of the highly mutated regions in PEDV S protein.

(A) Highly mutated regions in the S protein, analyzed by a sequence logo generator WebLogo 3, were marked with different color (I, II, and III), and the gray labeled sites indicted the uncontained sites in the PDB 6vv5. (B) Corresponding locations of the highly mutated regions (I, II, and III) in the S protein based on the crystal structure file of PDB 6vv5, the brown transparent labelled area indicted the S1 domain.

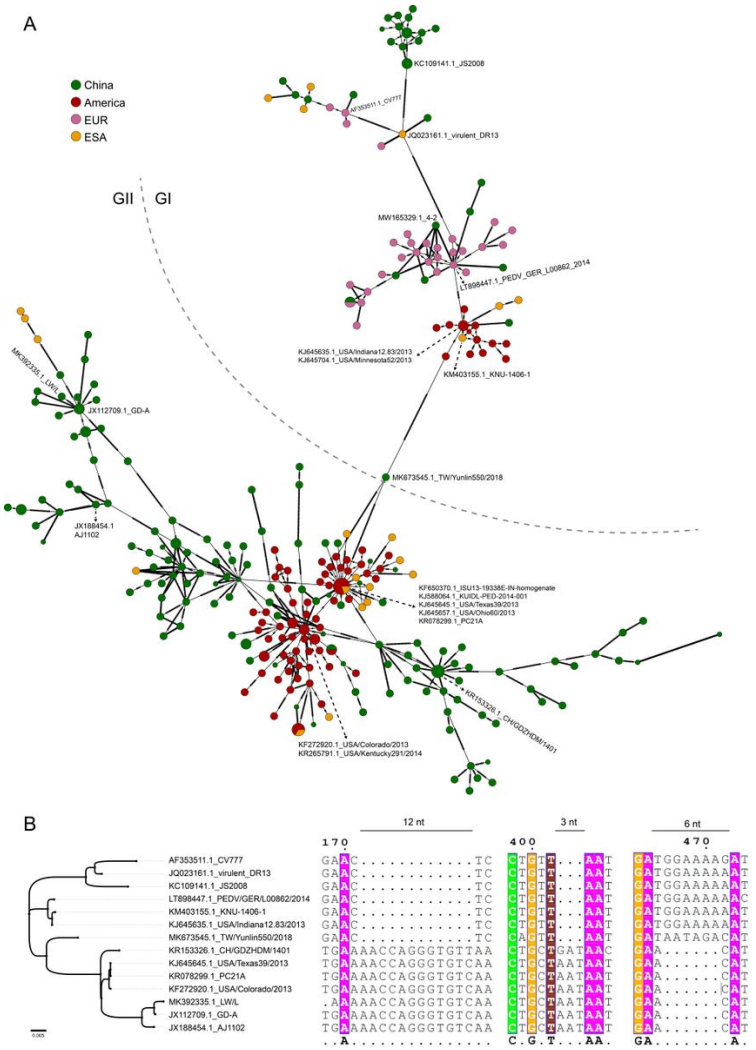

Figure 4

Haplotype analysis of PEDV S gene.

(A) The haplotype networks of PEDV based on the S gene. Different colors represent viruses in different regions. Green represents China, red represents the Americas, pink represents Europe (EUR), yellow represents East Asia (except China) and South Asia (ESA). (B) Evolution of the PEDV based on the representative strains from dominant haplotype. '.' represents the nucleotide sequence with the gap.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryFigureS1positiveselection.jpg
- SupplementaryFigureS2SNaaalin.pdf
- SupplementaryTableS1.xlsx
- SupplementaryTableS2.docx