ORIGINAL ARTICLE

WILEY Transboundary and Emerging Diseases

Evolutionary and genotypic analyses of global porcine epidemic diarrhea virus strains

Jiahui Guo[1,2] | **Liurong Fang[1,2]** | **Xu Ye[1,2]** | **Jiyao Chen[1,2]** | **Shangen Xu[1,2]** |
Xinyu Zhu[1,2] | **Yimin Miao[1,2]** | **Dang Wang[1,2]** (iD) | **Shaobo Xiao[1,2]**

[1]State Key Laboratory of Agricultural Microbiology, College of Veterinary Medicine, Huazhong Agricultural University, Wuhan, China

[2]Key Laboratory of Preventive Veterinary Medicine in Hubei Province, The Cooperative Innovation Center for Sustainable Pig Production, Wuhan, China

Correspondence
Dang Wang and Shaobo Xiao, Laboratory of Animal Virology, College of Veterinary Medicine, Huazhong Agricultural University, 1 Shi-zi-shan Street, Wuhan 430070, Hubei, China.
Email: wangdang@mail.hzau.edu.cn (DW); vet@mail.hzau.edu.cn (SX)

Funding information
National Natural Science Foundation of China, Grant/Award Number: 31672569, 31672566; National Key R&D Program of China, Grant/Award Number: 2016YFD0500103; Natural Science Foundation of Hubei Province, Grant/Award Number: 2017CFA059; Special Project for Technology Innovation of Hubei Province, Grant/Award Number: 2017ABA138; Fundamental Research Funds for the Central Universities, Grant/Award Number: 2662016PY070

Abstract

Porcine epidemic diarrhea virus (PEDV), which re-emerged in China in October 2010, has spread rapidly worldwide. Detailed analyses of the complete genomes of different PEDV strains are essential to understand the relationships among re-emerging and historic strains worldwide. Here, we analysed the complete genomes of 409 strains from different countries, which were classified into five subgroup strains (i.e., GI-a, GI-b, GII-a, GII-b, and GII-c). Phylogenetic study of different genes in the PEDV strains revealed that the newly discovered subgroup GII-c exhibited inconsistent topologies between the spike gene and other genes. Furthermore, recombination analysis indicated that GII-c viruses evolved from a recombinant virus that acquired the 5′ part of the spike gene from the GI-a subgroup and the remaining genomic regions from the GII-a subgroup. Molecular clock analysis showed that divergence of the GII-c subgroup spike gene occurred in April 2010, suggesting that the subgroup originated from recombination events before the PEDV re-emergence outbreaks. Interestingly, *Ascaris suum*, a large roundworm occurring in pigs, was found to be an unusual PEDV host, providing potential support for cross-host transmission. This study has significant implications for understanding ongoing global PEDV outbreaks and will guide future efforts to develop effective preventative measures against PEDV.

KEYWORDS
genotyping, molecular epidemiology, porcine epidemic diarrhoea virus, recombination, spike gene

1 | INTRODUCTION

Porcine epidemic diarrhoea (PED) is a devastating enteric disease in pigs that results in severe diarrhoea, vomiting, and dehydration, with very high mortality observed in suckling pigs (Pensaert & de Bouck, 1978). Porcine epidemic diarrhoea virus (PEDV), which is the causative agent of PED, belongs to the genus *Alphacoronavirus* within the family *Coronaviridae* and is an enveloped single-stranded positive-sense RNA virus (Woo et al., 2012). The PEDV genome consists of seven open reading frames (ORFs) organized in the order ORF1a, ORF1b, spike (S) glycoprotein gene, ORF3 hypothetical protein gene,

envelope (E) gene, membrane (M) gene, and nucleocapsid (N) gene (Li et al., 2016). Among the proteins encoded by the ORFs, the S glycoprotein is located on the envelope of the virus in the large surface projections of the virion and plays an important role in the attachment of viral particles to host cell receptors (Lee, Park, Kim, & Lee, 2010). Thus, the S gene is considered important for understanding the genetic relatedness and epidemiological status of PEDV field isolates, as well as for advancing vaccine development (Chen, Liu, Lang, et al., 2013).

The PED disease was first discovered in pig farms in Belgium and the United Kingdom in 1976 (Pensaert & de Bouck, 1978), with

reports of its occurrence in China as early as the 1980s (Li et al., 2012). With the emergence of new PEDV strains, however, serious disease epidemics have been observed in China since October 2010 (Sun et al., 2012). Beyond China, the disease has rapidly spread to more than 38 states in the USA following its first outbreak in May 2013, affecting more than 4,000 farms accounting for more than 7 million piglets (Cima, 2014). Japan, Canada, Mexico, and Colombia have also experienced successive outbreaks, with considerable economic losses to the global pig industry (Lara-Romero et al., 2018; Ojkic et al., 2015; Takahashi, Okada, & Ohshima, 1983; Valko et al., 2017).

Major global outbreaks since 2012 have renewed concerns about the potential changes in the mode of PEDV transfer (Chen, Liu, Lang, et al., 2013; Li et al., 2012; Sun et al., 2012). Although increasing evidence suggests that PEDV routinely undergoes significant changes, especially in spike proteins (Lara-Romero et al., 2018; Stott et al., 2017), the prevalence and evolution of PEDV strains is not well-defined and limited knowledge is known regarding the ways in which PEDV subgroups circulate among themselves and how they might influence the evolution of PEDV. To better understand the molecular epidemiology and genetic diversity of PEDV field isolates, we investigated the genetic characterization, origin, and evolution of emergent PEDV strains worldwide, which will provide much needed information for the effective prevention and control of this disease.

2 | MATERIALS AND METHODS

2.1 | Sample collection

The complete PEDV genome was selected for genetic analysis. To clarify the evolution of PEDVs, we obtained four complete genome sequences of PEDV from our own lab (i.e., ZL29, AJ1102, HuB1-2017, and HuB7-2017) (Bi, Zeng, Xiao, Chen, & Fang, 2012) and 405 sequences (obtained 1 January 2018) from the National Center for Biotechnology Information (NCBI, https://www.ncbi.nlm.nih.gov) during the different time periods. Repeated sequences and culture-attenuated PEDV sequence were removed from the data set. The final data set contained 409 sequences originating from Asia, Europe, and America. Details on the data set are summarized in Supporting information Table S1.

2.2 | Phylogenetic analyses

Firstly, we performed multiple sequence alignment of the 409 complete PEDV genomes, as well as the ORF1ab, S, ORF3-E-M-N genes,

and applied the bat coronavirus BtCoV/512/2005 (GenBank accession no. DQ648858) sequence as an outgroup (Tang et al., 2006). A maximum-likelihood (ML) phylogenetic tree was constructed using IQ-TREE v.1.6.5 (Nguyen, Schmidt, von Haeseler, & Minh, 2015), with the best fitting evolutionary model suggested by the program following 1,000 bootstrap replicates. The phylogenetic tree was rooted against the PEDV-related bat coronavirus, with removal of the long-branch leading to greater resolution of the viruses of interest. Nucleotide and deduced amino acid sequences were aligned using MAFFT v.7.402 (Katoh & Standley, 2013). The resulting tree was visualized using iTOL v.4 (Interactive Tree of Life, http://itol.embl.de/).

2.3 | Recombinant analyses

We preliminarily screened the PEDV sequence data set for recombination using RDP, GENECONV, Chimaera, MaxChi, and 3Seq, followed by secondary scanning and recombination using BootScan and SiScan in Recombination Detection Program version.4.95 (RDP v.4.95) (Martin, Murrell, Golden, Khoosal, & Muhire, 2015). Sequences with significant signals for recombination determined by more than two methods were analysed in greater detail. Nucleotide sequence similarity was assessed by SimPlot v.3.5.1 (Lole et al., 1999), with a sliding window size of 500 bp, step size of 100 nucleotides, and 1,000 bootstrap replicates, using gap-stripped alignments and the F84 (ML) distance model. All data were analysed using GraphPad Prism software (v.5.03, San Diego, CA, USA).
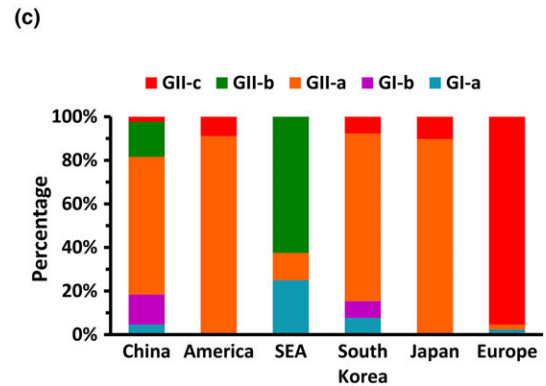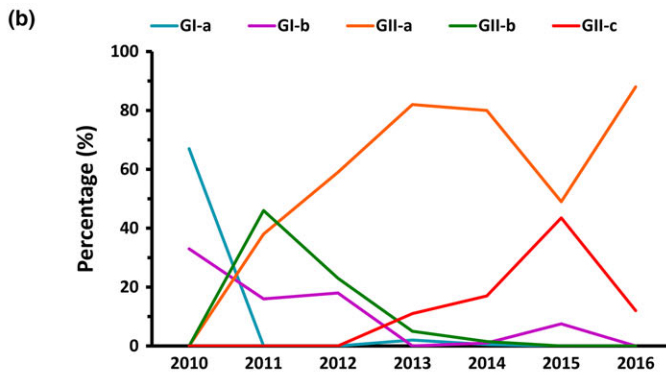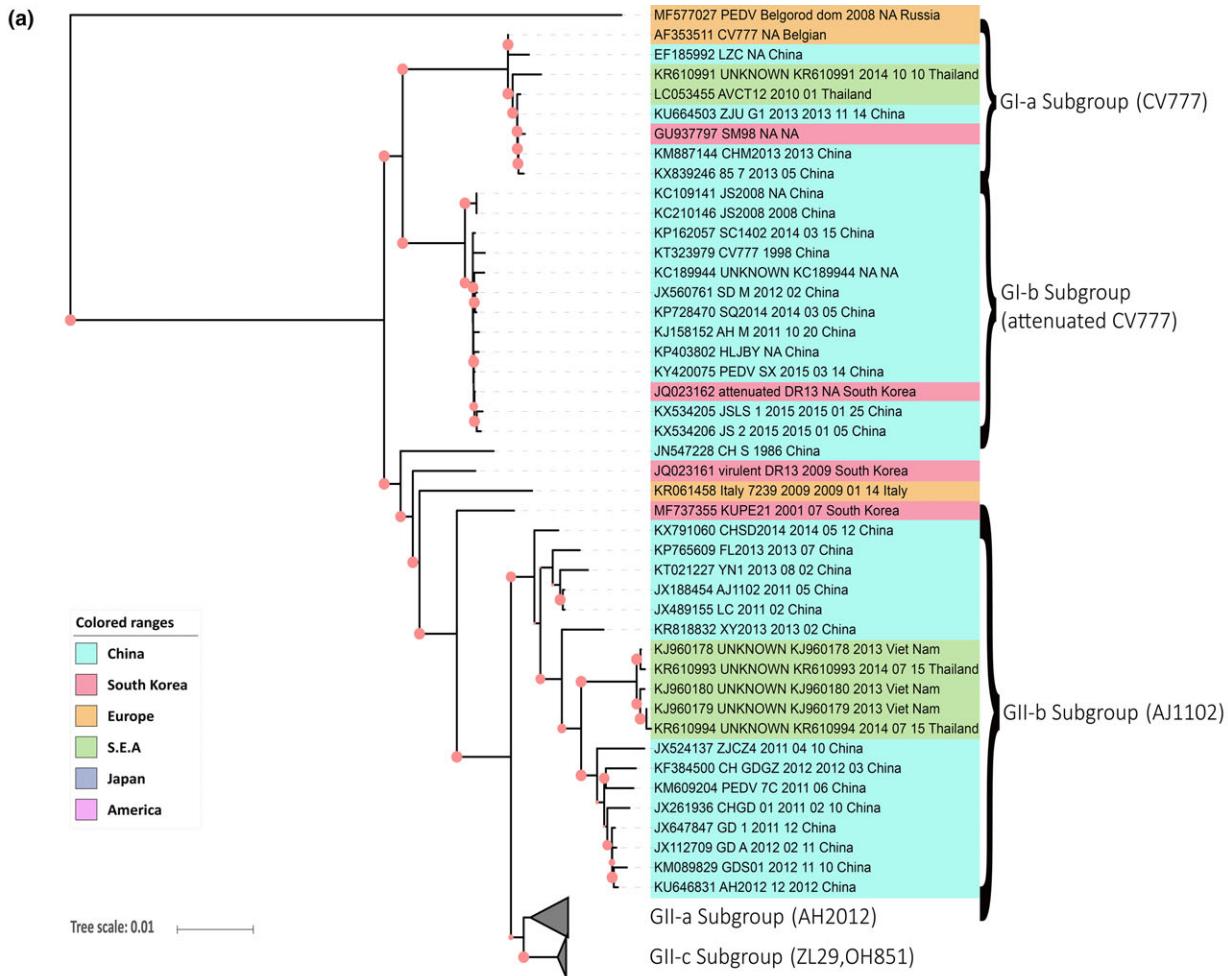
2.4 | Bayesian phylogeographic analyses

A Bayesian framework was applied to reconstruct the spatiotemporal diffusion history of the PEDVs. In brief, the spatial diffusion of the time-scaled genealogy was modelled as a standard continuous-time Markov chain (CTMC) process over discrete sampling locations. Analysis was performed using BEAST v.1.8.2 (Baele, Lemey, Rambaut, & Suchard, 2017) under the assumption of a molecular clock model with uncorrelated lognormal distribution. The analysis was run for 200 million chains, with sampling every 20,000 generations.

2.5 | Comparison of S protein sequences from different subgroups

All S protein sequences from the PEDV sample strains were analysed using the meta data-driven comparative analysis tool (meta-CATS) (Pickett et al., 2013), with a p-value threshold of 0.05 (this threshold

FIGURE 1 Genotyping and origin of the 409 PEDV strains based on full-length genomic sequence analyses. (a) Phylogram was tested by 1,000 bootstrap replicates, branch lengths were measured by the number of substitutions per site (see scale bars). Names of strains, years, places of isolation, GenBank accession numbers, genogroups, and subgroups are shown. (b) Line chart shows the number of PEDV sequences obtained by gene subgroup and year of sampling. Yearly percentages of samples positive for PEDV are indicated by different coloured lines respectively. Data are indicated below sampling years. (c) Subgroup distribution of all available complete or partial PEDV genome sequences from countries reporting PEDV infections (n.a., sequence not available). In the bar charts, counts are shown by country or region. Data are indicated below bar charts [Colour figure can be viewed at wileyonlinelibrary.com]

(a)

MF577027 PEDV Belgorod dom 2008 NA Russia
AF353511 CV777 NA Belgian
EF185992 LZC NA China
KR610991 UNKNOWN KR610991 2014 10 10 Thailand
LC053455 AVCT12 2010 01 Thailand
KU664503 ZJU G1 2013 11 14 China
GU937797 SM98 NA NA
KM887144 CHM2013 2013 China
KX839246 85 7 2013 05 China
KC109141 JS2008 NA China
KC210146 JS2008 2008 China
KP162057 SC1402 2014 03 15 China
KT323979 CV777 1998 China
KC189944 UNKNOWN KC189944 NA NA
JX560761 SD M 2012 02 China
KP728470 SQ2014 2014 03 05 China
KJ158152 AH M 2011 10 20 China
KP403802 HLJBY NA China
KY420075 PEDV SX 2015 03 14 China
JQ023162 attenuated DR13 NA South Korea
KX534205 JSLS 1 2015 2015 01 25 China
KX534206 JS 2 2015 2015 01 05 China
JN547228 CH S 1986 China
JQ023161 virulent DR13 2009 South Korea
KR061458 Italy 7239 2009 2009 01 14 Italy
MF737355 KUPE21 2001 07 South Korea
KX791060 CHSD2014 2014 05 12 China
KP765609 FL2013 2013 07 China
KT021227 YN1 2013 08 02 China
JX188454 AJ1102 2011 05 China
JX489155 LC 2011 02 China
KR818832 XY2013 2013 02 China
KJ960178 UNKNOWN KJ960178 2013 Viet Nam
KR610993 UNKNOWN KR610993 2014 07 15 Thailand
KJ960180 UNKNOWN KJ960180 2013 Viet Nam
KJ960179 UNKNOWN KJ960179 2013 Viet Nam
KR610994 UNKNOWN KR610994 2014 07 15 Thailand
JX524137 ZJCZ4 2011 04 10 China
KF384500 CH GDGZ 2012 2012 03 China
KM609204 PEDV 7C 2011 06 China
JX261936 CHGD 01 2011 02 10 China
JX647847 GD 1 2011 12 China
JX112709 GD A 2012 02 11 China
KM089829 GDS01 2012 11 10 China
KU646831 AH2012 12 2012 China

GI-a Subgroup (CV777)

GI-b Subgroup (attenuated CV777)

GII-b Subgroup (AJ1102)

GII-a Subgroup (AH2012)

GII-c Subgroup (ZL29,OH851)

Colored ranges
China
South Korea
Europe
S.E.A
Japan
America

Tree scale: 0.01

(b)

— GI-a — GI-b — GII-a — GII-b — GII-c

Percentage (%)

	≤2010	2011	2012	2013	2014	2015	≥2016	Total
GI-a	4	0	0	3	1	0	0	8
GI-b	2	2	3	0	2	4	0	13
GII-a	0	5	10	103	138	26	15	297
GII-b	0	6	4	6	3	0	0	19
GII-c	0	0	0	13	29	23	2	67
Total	6	13	17	125	173	53	17	404

(c)

GII-c GII-b GII-a GI-b GI-a

Percentage

	China	America	SEA	South Korea	Japan	Europe
GI-a	4	0	2	1	0	1
GI-b	12	0	0	1	0	0
GII-a	55	195	1	10	35	1
GII-b	14	0	5	0	0	0
GII-c	2	19	0	1	4	41
Total	87	214	8	13	39	43

was the maximum probability level for the likelihood that the position differed among groups simply by chance), to identify significantly different sites between the five subgroups.

3 | RESULTS AND DISCUSSION

3.1 | Phylogenetic analyses of complete genomes of global PEDV strains

To characterize the genetic diversity of PEDVs circulating globally, we constructed a phylogenetic tree using IQ-TREE based on the 409 complete PEDV genomes (see Materials and Methods 2.2). Consistent with our previous research (Wang, Fang, & Xiao, 2016a), the phylogenetic tree indicated that the complete PEDV genomes evolved into two separate genogroups, GI (classical) and GII (variant), as presented in Figure 1a. Furthermore, genogroup GI evolved into two subgroups (GI-a and GI-b) and genogroup GII evolved into three subgroups (GII-a, GII-b, and GII-c). The GI-a subgroup mainly included the earlier PEDV strains found in Europe and Belgium (virulent CV777) as well as some classical strains (LZC and SM98). The GI-b subgroup included most cell culture-adapted vaccine strains (attenuated CV777 and DR13) and other pandemic classical strains (AH-M, SD-M, and SQ2014) from China and South Korea. Since 2010, identification and sequencing analyses have shown that PEDV strain variants are highly prevalent worldwide. These prevalent variants, which have led to huge economic losses (Wang et al., 2016a), were mostly located within the GII genogroup. The GII-a subgroup included strains from the USA as well as from countries that reported US-like PEDVs (e.g., AH2012, HuB1-2017, and HuB7-2017) (Figure 1a). In contrast, the GII-b subgroup predominantly consisted of strains from China, South Korea, and Thailand (AJ1102 and CH/SD2014) (Supporting information Figure S1a). The phylogenetic tree further indicated a new GII subgroup (GII-c), which consisted of S-INDEL strains from America (e.g., OH851) and Europe (e.g., GER/L00862/2014) with high degrees of sequence similarity to the ZL29 strain from China (this study, Supporting information Figure S1b), suggesting that S-INDEL strains from Europe may have originated from a common ancestor with strain ZL29.

3.2 | Geographical and temporal analysis of PEDVs

We identified the geographical and temporal distributions of the PEDV strains to clarify the evolution of the virus. As shown in Figure 1b, only sporadic outbreaks of PEDV were reported before 2010, with the pathogens involved in these outbreaks found within the GI genogroup. The PEDVs were primarily located in the GI-a (virulent CV777) and GI-b (attenuated DR13) subgroups and were predominantly from Asia (Figure 1b). However, considerable PEDV outbreaks were reported in Asia and the United States after 2010, even for vaccinated piglets (Lin, Saif, Marthaler, & Wang, 2016). Based on our examination and assembly of public data, we identified that the PEDV strains were primarily from the GII genogroup.

Interestingly, GII-b subgroup strains were reported predominantly in 2011, whereas GII-a strains were reported more prevalently after 2011 and occupied a larger proportion of strains. Moreover, reports of the newly discovered GII-c strains showed a significant increase after 2012 due to further sequencing from Europe. The above results indicate that the epidemic strains from different periods were from the five different subgroups. Based on the geographical distribution of PEDVs (Figure 1c), the GI genogroup (classical and cell culture-adapted vaccine strains) largely originated from the earlier PEDV-threatened areas, such as China, South Korea, and Europe. The different subgroups from the GII genogroup also showed characteristic geographical distribution. While most GII-a subgroup strains were from the Americas, a small number were from China and Japan or from sporadic outbreaks in a few other isolated areas. The GII-b subgroup strains were mostly endemic to Asia, especially China, South Korea, and Japan. The GII-c subgroup strains were primarily from Europe, with some from the USA and China. Our study showed that PEDV strains from different subgroups were prevalent within the same areas, implying that the coincident "hot spots" in PEDV-endemic areas (e.g., China and South Korea, Figure 1c) are
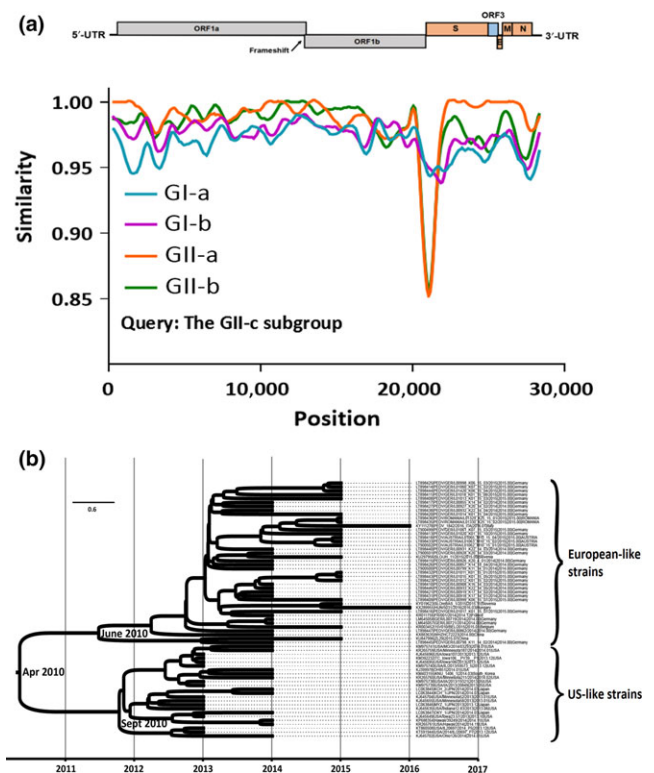
FIGURE 2 Recombination and origin of the PEDV GII-c subgroup. (a) SimPlot analysis for possible recombination events of the GII-c subgroup genome compared to the other four subgroups. (b) Time-scaled phylogeny of the PEDV S gene from the GII-c subgroup. Time-scaled Maximum Clade Credibility (MCC) trees were estimated from the complete S protein from the GII-c subgroup, as indicated in Figure 2c, with tip times reflecting time of sampling (x-axis). Node age estimates are shown, and the time line is indicated under the tree [Colour figure can be viewed at wileyonlinelibrary.com]

critical for determining the sources of some PEDV variations. These "hot spot" areas have the potential to be important reservoirs for the genetic variation of PEDVs, resulting in recombination between different PEDV subgroups.

We also examined the potential hosts of the 409 PEDV strains. Results showed an unusual PEDV strain (GenBank accession no. KX883635) hosted by *Ascaris suum* (Supporting information Table S1) (Shi et al., 2016), a large roundworm in pigs, thus providing novel insight into the possible epidemiology of PEDV infection. Indeed, parasites have long been regarded as a harmful factor to the pig industry as sources for a variety of infectious

agents (Jesudoss Chelladurai et al., 2017). For example, *Metastrongylus* larvae are considered a reservoir for various porcine viral pathogens, primarily swine fever virus and swine flu virus (Sen, Kelley, Underdahl, & Young, 1961). However, whether *Ascaris suum* plays a critical role as a PEDV reservoir requires further investigation.

3.3 | Global emergence of recombinant PEDV

Previous studies have reported on the occurrence of recombination in different PEDV strains (Chen et al., 2017, Li et al., 2018;
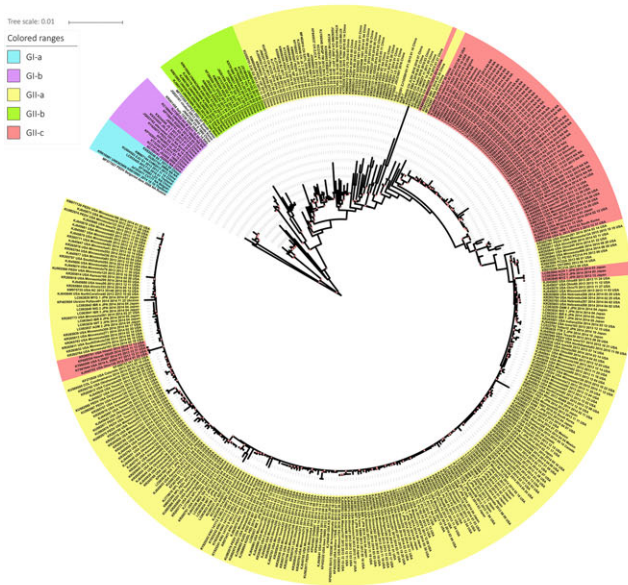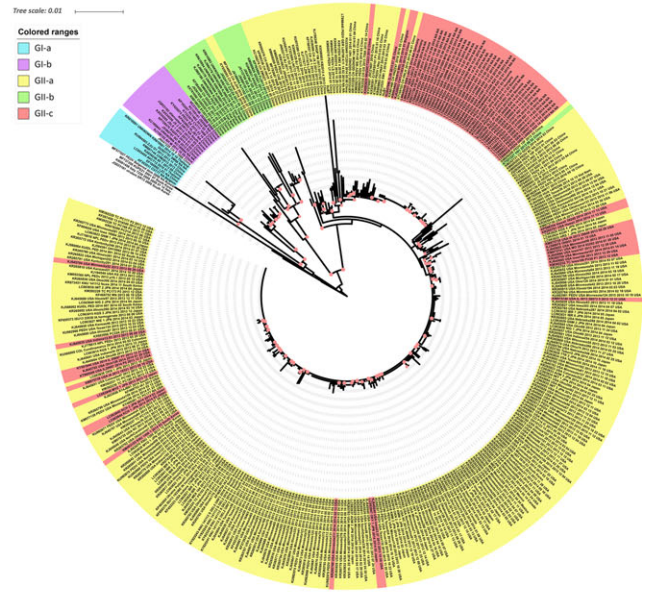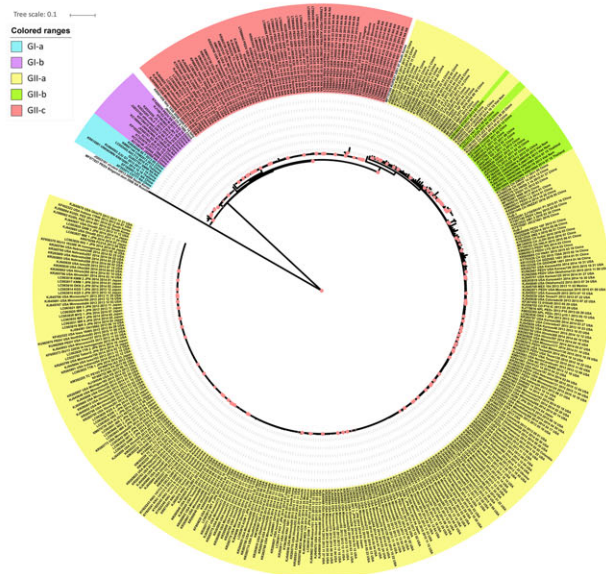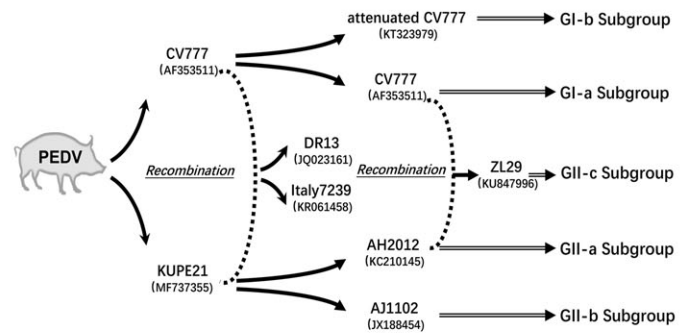
FIGURE 3 Genotyping and origin of the 409 PEDV strains based on different genes. (a–c) Phylogenetic trees were constructed by the maximum-likelihood (ML) method based on (a) ORF1ab, (b) ORF3-E-M-N, and (c) S gene sequences, with 1,000 bootstrap replicates. Names of strains, years, places of isolation, GenBank accession numbers, genogroups, and subgroups are shown. (d) Model for the recombination and origin of the S gene in the five different PEDV subgroups [Colour figure can be viewed at wileyonlinelibrary.com]

Boniotti et al., 2016). Here, the aligned complete genomes were all scanned for recombination using seven different algorithms implemented in RDP v.4.95. The RDP4 results were used to determine if the recombination parents of GII-c were from the GI-a or GII-a subgroups. To further characterize recombination events and accurately determine parents, we performed genome-scale similarity comparisons between the GII-c and other subgroups with SimPlot v.3.5.1, as demonstrated in Figure 2a. This analysis confirmed the chimeric nature of the GII-c subgroup. Furthermore, a statistically significant signal for phylogenetic incongruence in the GII-c subgroup defined two recombinant sources of the PEDV genome: that is, (i) positions 1 to 20,551 and 21,551 to the end from the GII-a subgroup, and (ii) positions 20,551 to 21,551 from the GI-a subgroup. Collectively, these phylogenies showed that the GII-c subgroup evolved from a recombination event, with the 5′ part of the S gene acquired from GI-a and the remaining genomic regions acquired from GII-a (Figure 2a). Molecular clock data analysis indicated that the GII-c subgroup S gene likely originated in April 2010 (Figure 2b). As the re-emergence of PEDV in China was reported in October 2010, the recombination event of the ancestral virus strain in the GII-c subgroup appears to have transpired before the PEDV re-emerging outbreaks.

3.4 | Phylogenetic analyses of ORF1ab, S, and ORF3-E-M-N genes

To further explore the evolution of PEDVs, we constructed three phylogenetic trees based on the ORF1ab, S, and ORF3-E-M-N gene sequences of the 409 PEDV strains. The ORF1ab and ORF3-E-M-N gene alignments confirmed that the GII-c subgroup was deeply nested within the GII-a subgroup (Figure 3a and b). Strikingly, the phylogeny of the S gene suggested an entirely different evolutionary history for the GII-c subgroup compared to the other subgroups (Figure 3c). All GII-c subgroup strains showed inconsistent topology in the S gene phylogenetic tree, differing from the ORF1ab gene phylogenetic tree. This inconsistent topology, in which outlier sequences were found between two well-defined subgroups in a phylogenetic tree, was attributed to genomic recombination. These results further demonstrated that the GII-c subgroup evolved from the recombination of different subgroups.

Several viruses, such as virulent DR13 (isolated in 2009), Italy/7239/2009 (isolated in 2009), and KUPE21/2001 (isolated in 2001), which showed inconsistent topologies in the subgenomic trees, suggesting a more varied recombination history, were not classified within the five main subgroups. After multiple sequence alignment of the complete PEDV genome, virulent DR13 showed strong similarity with KUPE21

Strains/Subgroups	Marker(difference between GI and GII gene-groups)												Unique in GII-c subgroup		
	116	356	365	549	594	719	724	959	1044	1173	1232	1298	67	83	113
CV777	I	I	E	T	G	N	N	A	S	G	S	R	I	S	T
The GI-a subgroup	I	I	E	T	G	N	N	A	S	G	S	R	I	S	T
attenuated CV777	I	I	E	T	G	N	N	A	S	G	S	R	I	S	T
The GI-b subgroup	I	I	E	T	G	N	N	A	S	G	S	R	I	S	T
Italy/7239/2009	T	T	Q	A	S	N	N	A	A	D	R	Q	I	S	T
ZL29	T	T	Q	S	S	S	S	V	A	D	R	Q	L	S	H
The GII-c subgroup	T	T	Q	S	S	S	S	V	A	D	R	Q	L	A/S	H/T
KUPE21/2001	T	T	Q	T	S	N	N	A	A	G	G	R	Y	G	T
AH2012	T	T	Q	S	S	S	S	V	A	D	R	Q	H	G	T
The GII-a subgroup	T	T	Q	S	S	S	S	V	A	D	R	Q	H	G	T
AJ1102	T	T	Q	S	S	S	S	V	A	D	R	Q	H	G	T
The GII-b subgroup	T	T	Q	S	S	S	S	V	A	D	R	Q	H	G	T

Strains/Subgroups	Epitopes domain(499-638)								Epitopes domain(764-771)	
	517	521	523	527	549	594	605	612	764	766
CV777	A	L	S	V	T	G	A	L	S	Y
The GI-a subgroup	A	L	S	V	T	G	A	L	P/S	Y
attenuated CV777	A	H	G	I	T	G	E	F	L	D
The GI-b subgroup	A	H	G	I	T	G	E	F	L/S	D/S
Italy/7239/2009	S	H	G	V	A	S	A	F	L	G
ZL29	S	H	G	I	S	S	E	F	S	S
The GII-c subgroup	S	H	G	I	S	S	E	F	S	S
KUPE21/2001	S	Y	G	I	T	S	E	F	S	S
AH2012	S	H	G	I	S	S	E	F	S	S
The GII-a subgroup	S	H	G	I	S	S	E	F	S	S
AJ1102	A	H	G	I	S	S	D	F	S	S
The GII-b subgroup	A	H	G	I	S	S	E	F	S	S

FIGURE 4 Amino acid variants of S proteins in different PEDV subgroups. (Upper) Amino acid changes in S proteins of the two gene groups, with 11 amino acid changes found to distinguish the two gene groups. (Lower) Amino acid substitutions in the S protein neutralization epitope regions of the five subgroups [Colour figure can be viewed at wileyonlinelibrary.com]

(95%–99% genetic identity to the ORF1ab and ORF3-E-M-N genes). In contrast, the genetic identity between the S gene sequences of these viruses was only 85% and showed strong similarity with virulent CV777 (95% gene identity). In the S gene phylogeny, virulent DR13 did not cluster with KUPE21 but instead with virulent CV777, suggesting that the PEDV S gene was subjected to relatively frequent recombination, even between divergent subgroups (Figure 3c and Supporting information Figure S2a). Moreover, the Italy/7239/2009 PEDV strain (Chen, Liu, Shi, et al., 2013) may have originated from similar recombination events as those between KUPE21 and virulent CV777 (Supporting information Figure S2b).

Based on the phylogenetic and recombination analyses, the evolution of the PEDV S gene through time was reconstructed (Figure 3d), showing the influence of recombination. This model highlighted the genetic diversity expansion of the S gene, initially through antigenic drift and more recently due to widespread intergenotypic recombination. Since the emergence of Italy/7239/2009 in 2009, recombination has influenced all subsequently identified strains in the GII-c subgroup. Intragenogroup recombination provides a mechanism for amalgamation among these distinct subgroups and increases the genetic repertoire of co-circulating PEDV strains.

3.5 | Analysis of amino acid sequences of complete S protein

The S protein attaches to the cellular receptors of a host, resulting in virus entry by membrane fusion, and contains the domain that stimulates the production of neutralizing antibodies. Variations in the S protein are important for understanding the genetic relatedness of PEDV field strains (Chang et al., 2002; Lara-Romero et al., 2018; Wang et al., 2016b). To determine the significant S protein sequences among the five different PEDV subgroups, meta-CATS analysis was performed for spike protein sequences of all 409 PEDV strains. We identified 257 amino acid positions with significant variation among the isolates from the five subgroups (Supporting information Table S2). As shown in Figure 4, the GII genogroup contained 11 distinct patterns of aa mutations (I116T, I356T, E365Q, T549S, G594S, N724S, A959V, S1044A, G1173D, S1232R, and R1298Q), distinguishing it from isolates in the GI genogroup. Although the GII-c subgroup S genes shared some aa substitutions with the GI-a and GII-a subgroups (two parents of recombination), they also exhibited three unique patterns (L76, A/S92, and H/T113), which clearly distinguished these isolates from those in the other subgroups, suggesting that the GII-c subgroup S gene evolved gradually through antigenic drift.

It is well established that the S protein of coronaviruses induces high levels of neutralizing antibodies. Four neutralizing epitopes (499–638, 748–755, 764–771, and 1,368–1,374 amino acids) have also been identified in the PEDV S protein (Chang et al., 2002). To explore whether the GII genogroup acquired substitutions in neutralizing epitopes characteristic of the GI genogroup, we mapped the significantly different positions to the equivalent positions in representative sequences (Figure 4). Results identified seven substitutions in the neutralizing epitopes of the S protein in the GII genogroup (L521H, S523G, V527I, T549S, G594S, A605E, and L612F), which

may explain why traditional inactivated vaccines and attenuated vaccines against the GI genogroup cannot effectively protect piglets threatened by a pandemic strain from the GII genogroup.

In summary, this study revealed the genetic diversity and evolutionary dynamics of PEDV strains. Our genetic analyses showed that the PEDV strains could be categorized into two groups, namely, GI (classical) and GII (variant). We also discovered a new subgroup (GII-c) with novel genetic, molecular, and phylogenic characteristics. The GII-c subgroup evolved from a recombination event between the GI-a and GII-a subgroups, and we further found recombination in two relatively early strains: virulent DR13 and Italy/7239/2009. These recombination events occurred prior to the re-emergence of PEDV in 2010. Additionally, to explore the potential link between S protein amino acid sequence variations and recombination, we performed a series of comparative analyses of the PEDV S protein sequences. We found 10 positions that were localized in a well-known neutralizing epitope and revealed several unique amino acids that could easily distinguish the different subgroups. This study provides critical information to help trace the sources of PEDV variants and identify the evolutionary mechanisms involved. Furthermore, this research will hopefully facilitate the development of diagnostic kits, vaccines, and new therapeutic strategies, which are expected to turn the tide in the prevention of pandemic outbreaks of PEDV.

CONFLICT OF INTEREST

None.

ORCID

Dang Wang http://orcid.org/0000-0003-3394-3702

REFERENCES

Baele, G., Lemey, P., Rambaut, A., & Suchard, M. A. (2017). Adaptive MCMC in Bayesian phylogenetics: An application to analyzing partitioned data in BEAST. *Bioinformatics*, 33, 1798–1805. https://doi.org/10.1093/bioinformatics/btx088

Bi, J., Zeng, S., Xiao, S., Chen, H., & Fang, L. (2012). Complete genome sequence of porcine epidemic diarrhea virus strain AJ1102 isolated from a suckling piglet with acute diarrhea in China. *Journal of virology*, 86, 10910–10911. https://doi.org/10.1128/JVI.01919-12

Boniotti, M. B., Papetti, A., Lavazza, A., Alborali, G., Sozzi, E., Chiapponi, C., … Marthaler, D. (2016). Porcine epidemic diarrhea virus and discovery of a recombinant swine enteric coronavirus, Italy. *Emerging infectious diseases*, 22, 83–87. https://doi.org/10.3201/eid2201.150544

Chang, S. H., Bae, J. L., Kang, T. J., Kim, J., Chung, G. H., Lim, C. W., … Jang, Y. S. (2002). Identification of the epitope region capable of inducing neutralizing antibodies against the porcine epidemic diarrhea virus. *Molecules and cells*, 14, 295–299.

Chen, J., Liu, X., Lang, H., Wang, Z., Shi, D., Shi, H., … Feng, L. (2013). Genetic variation of nucleocapsid genes of porcine epidemic diarrhea virus field strains in China. *Archives of virology*, 158, 1397–1401. https://doi.org/10.1007/s00705-013-1608-8

Chen, J., Liu, X., Shi, D., Shi, H., Zhang, X., Li, C., … Feng, L. (2013). Detection and molecular diversity of spike gene of porcine epidemic diarrhea virus in China. *Viruses*, 5, 2601–2613. https://doi.org/10.3390/v5102601

Chen, N., Li, S., Zhou, R., Zhu, M., He, S., Ye, M., … Zhu, J. (2017). Two novel porcine epidemic diarrhea virus (PEDV) recombinants from a natural recombinant and distinct subtypes of PEDV variants. *Virus Research*, 242, 90–95. https://doi.org/10.1016/j.virusres.2017.09.013

Cima, G. (2014). PED virus reinfecting U.S. herds. Virus estimated to have killed 7 million-plus pigs. *Journal of the American Veterinary Medical Association*, 245, 166–167.

Jesudoss Chelladurai, J., Murphy, K., Snobl, T., Bader, C., West, C., Thompson, K., & Brewer, M. T. (2017). Molecular epidemiology of Ascaris infection among pigs in Iowa. *The Journal of infectious diseases*, 215, 131–138. https://doi.org/10.1093/infdis/jiw507

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular biology and evolution*, 30, 772–780. https://doi.org/10.1093/molbev/mst010

Lara-Romero, R., Gomez-Nunez, L., Cerriteno-Sanchez, J. L., Marquez-Valdelamar, L., Mendoza-Elvira, S., Ramirez-Mendoza, H., & Rivera-Benitez, J. F. (2018). Molecular characterization of the spike gene of the porcine epidemic diarrhea virus in Mexico, 2013-2016. *Virus Genes*, 54, 215–224. https://doi.org/10.1007/s11262-017-1528-x

Lee, D. K., Park, C. K., Kim, S. H., & Lee, C. (2010). Heterogeneity in spike protein genes of porcine epidemic diarrhea viruses isolated in Korea. *Virus research*, 149, 175–182. https://doi.org/10.1016/j.virusres.2010.01.015

Li, W., Li, H., Liu, Y., Pan, Y., Deng, F., Song, Y., … He, Q. (2012). New variants of porcine epidemic diarrhea virus, China, 2011. *Emerging infectious diseases*, 18, 1350–1353. https://doi.org/10.3201/eid1808.120002

Li, R., Qiao, S., Yang, Y., Guo, J., Xie, S., Zhou, E., & Zhang, G. (2016). Genome sequencing and analysis of a novel recombinant porcine epidemic diarrhea virus strain from Henan, China. *Virus Genes*, 52, 91–98. https://doi.org/10.1007/s11262-015-1254-1

Li, K., Song, D., Zhang, F., Gong, W., Guo, N., Li, A., … Tang, Y. (2018). Complete genome sequence of a recombinant porcine epidemic diarrhea virus strain, CH/JXJA/2017, isolated in Jiangxi, China, in 2017. *Genome Announcements*, 6, e01590–17. https://doi.org/10.1128/genomea.01590-17

Lin, C. M., Saif, L. J., Marthaler, D., & Wang, Q. (2016). Evolution, antigenicity and pathogenicity of global porcine epidemic diarrhea virus strains. *Virus research*, 226, 20–39. https://doi.org/10.1016/j.virusres.2016.05.023

Lole, K. S., Bollinger, R. C., Paranjape, R. S., Gadkari, D., Kulkarni, S. S., Novak, N. G., … Ray, S. C. (1999). Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *Journal of virology*, 73, 152–160.

Martin, D. P., Murrell, B., Golden, M., Khoosal, A., & Muhire, B. (2015). RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evolution*, 1, vev003. https://doi.org/10.1093/ve/vev003

Nguyen, L. T., Schmidt, H. A., von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular biology and evolution*, 32, 268–274. https://doi.org/10.1093/molbev/msu300

Ojkic, D., Hazlett, M., Fairles, J., Marom, A., Slavic, D., Maxie, G., … Burlatschenko, S. (2015). The first case of porcine epidemic diarrhea in Canada. *The Canadian Veterinary Journal*, 56, 149–152.

Pensaert, M. B., & de Bouck, P. (1978). A new coronavirus-like particle associated with diarrhea in swine. *Archives of virology*, 58, 243–247. https://doi.org/10.1007/BF01317606

Pickett, B. E., Liu, M., Sadat, E. L., Squires, R. B., Noronha, J. M., He, S., … Scheuermann, R. H. (2013). Metadata-driven comparative analysis tool for sequences (meta-CATS): An automated process for identifying significant sequence variations that correlate with virus attributes. *Virology*, 447, 45–51. https://doi.org/10.1016/j.virol.2013.08.021

Sen, H. G., Kelley, G. W., Underdahl, N. R., & Young, G. A. (1961). Transmission of swine influenza virus by lungworm migration. *The Journal of experimental medicine*, 113, 517–520.

Shi, M., Lin, X. D., Tian, J. H., Chen, L. J., Chen, X., Li, C. X., … Zhang, Y. Z. (2016). Redefining the invertebrate RNA virosphere. *Nature*, https://doi.org/10.1038/nature20167

Stott, C. J., Temeeyasen, G., Tripipat, T., Kaewprommal, P., Tantituvanont, A., Piriyapongsa, J., & Nilubol, D. (2017). Evolutionary and epidemiological analyses based on spike genes of porcine epidemic diarrhea virus circulating in Thailand in 2008-2015. *Infection, Genetics and Evolution*, 50, 70–76. https://doi.org/10.1016/j.meegid.2017.02.014

Sun, R. Q., Cai, R. J., Chen, Y. Q., Liang, P. S., Chen, D. K., & Song, C. X. (2012). Outbreak of porcine epidemic diarrhea in suckling piglets, China. *Emerging infectious diseases*, 18, 161–163. https://doi.org/10.3201/eid1801.111259

Takahashi, K., Okada, K., & Ohshima, K. (1983). An outbreak of swine diarrhea of a new-type associated with coronavirus-like particles in Japan. *Nihon juigaku zasshi. The Japanese Journal of Veterinary Science*, 45, 829–832. https://doi.org/10.1292/jvms1939.45.829

Tang, X. C., Zhang, J. X., Zhang, S. Y., Wang, P., Fan, X. H., Li, L. F., … Guan, Y. (2006). Prevalence and genetic diversity of coronaviruses in bats from China. *Journal of virology*, 80, 7481–7490. https://doi.org/10.1128/jvi.00697-06

Valko, A., Biksi, I., Csagola, A., Tuboly, T., Kiss, K., Ursu, K., & Dan, A. (2017). Porcine epidemic diarrhoea virus with a recombinant S gene detected in Hungary, 2016. *Acta veterinaria Hungarica*, 65, 253–261. https://doi.org/10.1556/004.2017.025

Wang, X., Chen, J., Shi, D., Shi, H., Zhang, X., Yuan, J., … Feng, L. (2016b). Immunogenicity and antigenic relationships among spike proteins of porcine epidemic diarrhea virus subtypes G1 and G2. *Archives of virology*, 161, 537–547. https://doi.org/10.1007/s00705-016-2766-2

Wang, D., Fang, L., & Xiao, S. (2016a). Porcine epidemic diarrhea in China. *Virus research*, 226, 7–13. https://doi.org/10.1016/j.virusres.2016.05.026

Woo, P. C., Lau, S. K., Lam, C. S., Lau, C. C., Tsang, A. K., Lau, J. H., … Yuen, K. Y. (2012). Discovery of seven novel Mammalian and avian coronaviruses in the genus deltacoronavirus supports bat coronaviruses as the gene source of alphacoronavirus and betacoronavirus and avian coronaviruses as the gene source of gammacoronavirus and deltacoronavirus. *Journal of virology*, 86, 3995–4008. https://doi.org/10.1128/JVI.06540-11

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.