



# Time-calibrated phylogenomics of the porcine epidemic diarrhea virus: genome-wide insights into the spatio-temporal dynamics

Jisung Jang<sup>1</sup> · Sook Hee Yoon<sup>2</sup> · Wonseok Lee<sup>2</sup> · Jihyun Yu<sup>2</sup> · Joon Yoon<sup>1</sup> · Seunghyun Shim<sup>3</sup> · Heebal Kim<sup>1,2</sup>

Received: 7 August 2017 / Accepted: 16 March 2018 / Published online: 21 April 2018  
© The Genetics Society of Korea and Springer Science+Business Media B.V., part of Springer Nature 2018

## Abstract

Porcine epidemic diarrhea virus (PEDV), the causative agent of porcine epidemic diarrhea (PED), has led to tremendous economic losses in the global swine industry. Although the phylogeny of PEDV has been investigated extensively at the molecular level, there was no time-calibrated phylogenomic study on the virus. To improve insight into this topic, we analyzed 138 published genome sequences using the Bayesian coalescent analyses as well as Bayesian inferences and maximum likelihood methods. All of the global PEDV isolates were divided into six groups, except for one unclassified isolate. Of the six groups, Groups 1–5 comprised pandemic viruses while the remaining Group 6 contained classical isolates. Interestingly, the two clades, both pandemic and classical, consisted of clade-specific amino acid sequences in five genes: ORF1a, ORF1b, S, ORF3, and N. Within the pandemic clade, Group 1 and Group 2 originated from North America, whereas Group 3–Group 5 were derived from Asia. In Group 2, there was a common origin of S INDEL isolates. Within each group, there was no apparent association between temporal or geographic origin and heterogeneity of PEDVs. Our findings also showed that the PEDV virus evolved at a rate of  $3.38 \times 10^{-4}$  substitutions/site/year, and the most recent common ancestor of the virus emerged 75.9 years ago. Our Bayesian skyline plot analysis indicated that the PEDV had maintained constant effective population size excluding only a short period, around 2012, when a valley shaped decline in the effective number of infections occurred.

**Keywords** Porcine epidemic diarrhea virus · Phylogenetics · Population size · The most recent common ancestor

## Introduction

Porcine epidemic diarrhea (PED) is an acute enteric tract infectious disease of swine characterized by vomiting, watery diarrhea, dehydration, and death (Debouck et al.

1981). These symptoms especially lead to high mortality rate in suckling piglet (90–100%), which is the main reason for the huge economic losses in the global swine industry (Li et al. 2012b; Sun et al. 2012). Recently, there were one of the most severe outbreaks in the USA; about 7 million piglets (10% of total domestic pigs) died from the disease within a single year (Jung and Saif 2015; Paarlberg 2014).

The causative agent, PED virus (PEDV), is classified as a member of the order *Nidovirales*, family *Coronaviridae*, and genus *Alphacoronavirus*. PEDV is an enveloped, single-stranded, and positive-sense RNA virus. Its genome is approximately 28 kb in length and consists of a 5' untranslated region (UTR), a 3' UTR, and seven open reading frames (ORFs): ORF1a, ORF1b, S, ORF3, E, M, and N (Kocherhans et al. 2001). The first two ORFs (ORF1a and ORF1b) encode two replicase polyproteins, and the remaining five code for an accessory protein (ORF3) and the following structural proteins: spike (S), envelope (E), membrane (M), and nucleocapsid (N) (Duarte et al. 1994).

The PEDV was first observed in domesticated pigs in England in 1971 (Oldham 1972), and then spread to European

Jisung Jang and Sook Hee Yoon have contributed equally to this work.

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s13258-018-0686-0>) contains supplementary material, which is available to authorized users.

✉ Heebal Kim  
heebal@snu.ac.kr

<sup>1</sup> Interdisciplinary Program in Bioinformatics, Seoul National University, Seoul 151-747, Republic of Korea

<sup>2</sup> Department of Agricultural Biotechnology and Research Institute for Agriculture and Life Sciences, Seoul National University, Seoul 151-921, Republic of Korea

<sup>3</sup> LAS Inc., 16, Arayuk-ro, Gochon-eup, Gimpo-si, Gyeonggi-do, Republic of Korea

countries such as Belgium (Pensaert and De Bouck 1978), England (Pritchard et al. 1999), Czech (Smíd et al. 1992) during the 1970s and 1980s. In Asia, the first case of the disease was reported in China in 1973 (Sun et al. 2016), and subsequently occurred in other Asian countries such as South Korea (Kweon et al. 1993), Japan (Takahashi et al. 1983), Viet Nam (Toan et al. 2011), and Thailand (Puranaveja et al. 2009). In late 2010, highly pathogenic PEDV strains appeared in China and were considered as the first pandemic strains (Sun et al. 2012). In April 2013, the highly virulent PEDV was also first reported (Stevenson et al. 2013) in the USA and spread across the nation (31 states) (Cima 2013). Since then, the notorious strains emerged continuously in many other countries including South Korea (Lee and Lee 2014), Japan (Suzuki et al. 2015), Taiwan (Chiou et al. 2014), Canada (Pasick et al. 2014), Mexico (Vlasova et al. 2014). The PEDV strains detected in these countries were classified as US-like PEDV strains. In December 2013, a new PEDV strain OH851 (later designated as S INDEL PEDV) causing lower mortality rate which has specific insertions and deletions in N-terminal of the S protein (S1 region) was detected in the USA (Oka et al. 2014; Vlasova et al. 2014). The other S INDEL PEDVs have also been reported in South Korea, Philippine, France, Germany and Portugal. (Grasland et al. 2015; Hanke et al. 2015; Kim et al. 2016; Lee and Lee 2014; Mesquita et al. 2015).

To date, there have been many investigations of the PEDV phylogenetics on the basis of particular genes—S (Li et al. 2012a), ORF3 (Li et al. 2014), E (Park et al. 2013), M (Chen et al. 2008), and N (Li et al. 2013)—and genome (Chen et al. 2014; Huang et al. 2013; Lin et al. 2016; Stevenson et al. 2013) sequences. Nevertheless, there was no time-calibrated phylogenomic study on this virus. In order to explore this topic, we analyzed 138 available public genome sequences that collected during the 38 years (1978–2015). The main focus of this study are (1) to characterize sequences of PEDV coding region; (2) to reconstruct the genome-wide phylogeny of the global PEDVs using two different analyses methods, Bayesian inference (BI) and maximum likelihood (ML); and (3) to examine the various evolutionary mechanisms including selection pressure, substitution rate, divergence times, and effective population size changes using Bayesian coalescent approach.

In this paper, the terms “classical” and “pandemic” PEDV isolates respectively correspond to the PEDV prototype CV777-like isolates that have appeared since the 1970s and other PEDV isolates globally detected after 2010.

## Materials and methods

### Data collection and sequence analysis

We analyzed 138 PEDV coding genome sequences available on NCBI, collected around the world from 1978 to

2015. These viruses were composed of 43 Asian (27 from China, eight from South Korea, four from Thailand, three from Viet Nam, and one from Japan), five European (two from Belgium, two from Germany, and one from France), and 90 American (87 from USA, two from Mexico, and one from Canada) viruses (Table S1). At the first stage of making our data matrix, all known recombinant sequences were removed. Namely, all sequences were screened for the presence of recombination events via the RDP 3.0b41 package (Martin et al. 2015) with the default parameters. This package consists of six different recombination detection programs: RDP, GENECONV, MaxChi, BOOTSCAN, Chimeara, and SiScan. To avert the possibility of detecting false-positive recombinants, we considered only putative recombinant sites detected by at least 3 out of 6 methods. In all analyses,  $p < 0.05$  was considered to indicate statistical significance.

Both nucleotide and amino acid sequences of coding regions were aligned with MAFFT 7 (Katoh and Standley 2013). The final alignment was composed of 27,436 nucleotide positions and can be obtained from the authors. The alignment of coding genome sequences was also divided into the corresponding individual genes: ORF1a–ORF1b–S–ORF3–E–M–N. All gene positions quoted here are with respect to PEDV genome of prototype CV777 (GenBank accession no. AF353511). For both nucleotide and amino acid sequences of the seven individual genes, as well as the entire coding region, identical sequence were filtered primarily and estimated transition–transversion ratio were calculated by Tree-Puzzle 5.3 (Schmidt et al. 2002). The transition–transversion ratio were estimated as 2.63 and it was used as input value in following recombinant detection model. After primal filtering, following measures were estimated using BioEdit 7.2.5 (Hall 1999), and Modeltest 3.7 (Posada 2003): total sites (including gaps), conserved sites, and average identities, base frequencies, substitution matrix, and evolutionary models.

In addition, we plotted the number of nucleotide and amino acid variations at each position throughout the 138 PEDV whole coding region alignment. We first estimated the nucleotide differences by counting the number of minor nucleotides at each position across the sequence alignment. Next, we defined the major nucleotide at each site as the most frequent nucleotide (A, T, C, or G), and the rest of the nucleotides were considered as minor nucleotides. For example, at a specific position of the alignment, if the nucleotide C was major in 120 samples, the nucleotides A, T, and G were considered as minor nucleotides in the remaining 18 samples; the nucleotide difference was 18. We applied the same principle for amino acid differences.

## Phylogenomic reconstructions

Phylogenomic trees for the PEDV coding genomes were reconstructed using two different analytical methods: Bayesian inference (BI) and maximum likelihood (ML). We chose the best-fit model of nucleotide substitution with the standard ModelTest PAUP block in PAUP 4.0b10 (Swofford 2003) and Akaike's information criterion (AIC) in ModelTest 3.7 (Posada 2003); GTR + I + G was selected as the best evolutionary model. The unpartitioned BI analysis was carried out using MrBayes 3.2.5 (Ronquist and Huelsenbeck 2003) with the GTR + I + G model. For the partitioned model approach of BI analysis, 7 genes were concatenated and the following models were applied: GTR + I + G for the ORF1a, ORF1b, and S gene regions; TrN + I for the ORF3 and M gene regions; TrN for the E gene region; GTR + G for the N gene region (Table 1). Each analysis was performed with the following parameters: number of substitution rate, 6; rates, gamma; the number of generation, 100,000,000; sample frequency, 500; the number of chains, 1; burn-in generation, 25% of the number of generations. Bayesian posterior probability (BPP) values (Erixon et al. 2003) shown on respective internal nodes indicated the robustness of the phylogenomic analysis. Bayesian posterior probability (BPP) values are shown on internal nodes to indicate the robustness of the phylogenomic analysis.

ML analysis was also conducted using PhyML 3.1 (Guindon et al. 2009) with the following options: model of nucleotide substitution, GTR; nonparametric bootstrap analysis, yes; number of replicates, 500; proportion of invariable site (pinvar), estimated; nst, 6; number of substitution rate categories, 6; gamma shape parameter, estimated by program; optimize tree topology, yes. A bootstrap test (with 1000 pseudoreplicates) (Felsenstein 1985) was performed to determine the statistical support for each node of the ML tree.

## Selection pressure analysis

The selection pressure is important indicators of evolutionary biology. To evaluate the selective pressure that drives PEDV evolution, we estimated the relative rates of non-synonymous and synonymous substitution ( $\omega = dN/dS$ ) by using ClustalX 1.81 (Thompson et al. 1997), PAL2NAL (Suyama et al. 2006) and the CodeML program in the PAML 4.7 package (Yang 2007). Ratio values equal to 1 indicate neutral evolution; < 1, negative selection; and > 1, positive selection.

## Co-estimation of substitution rates, time of the most recent common ancestor, and population size changes

To explore the evolutionary history of the global PEDVs, the substitution rates, time of the most recent common ancestor (tMRCA), and changes in population size were co-estimated using the Bayesian coalescent approaches as implemented in BEAST 2.3.1 (Bouckaert et al. 2014). We combined three molecular clock models (strict, relaxed uncorrelated exponential, and relaxed uncorrelated lognormal) with three different demographic models (constant size, exponential growth, and Bayesian skyline) to make nine datasets. These nine datasets were simulated following generalised time reversible (GTR) substitution model. Next, each dataset was run for 550,000,000 generations to ensure convergence of all parameters (ESS > 100) with discarded burn-in of 10%. By comparison of Bayesian factors ( $\log_{10}$  Bayes factors > 2 in all cases), based on the relative marginal likelihoods of the nine models, the relaxed uncorrelated exponential clock and exponential population size model were selected as the best fit for the data set. The resulting convergence was analyzed using Tracer 1.6 (Rambaut and Drummond 2013a), and the statistical uncertainties were summarized in the 95% highest posterior density (HPD) intervals. Trees were summarized as maximum clade credibility (MCC) tree using TreeAnnotator

**Table 1** The best fit evolutionary models estimated for PEDV genomic regions with Modeltest

Genomic region	Model	−lnL-	Base frequencies (A, C, G)	Substitution matrix (A–C, A–G, A–T, C–G, C–T)	Pinvar	Nst
Coding region	GTR+I+G	74448.13	0.25, 0.19, 0.23	0.78, 2.06, 0.54, 0.54, 5.62	0.69	6
ORF1a	GTR+I+G	29562.00	0.24, 0.18, 0.24	0.74, 2.02, 0.36, 0.28, 5.06	0.67	6
ORF1b	GTR+I+G	18720.92	0.26, 0.19, 0.22	0.45, 2.56, 0.62, 0.24, 9.04	0.75	6
S	GTR+I+G	14662.38	0.25, 0.20, 0.21	1.29, 1.91, 0.69, 1.04, 4.55	0.46	6
ORF3	TrN+I	1699.44	0.24, 0.19, 0.19	1.00, 3.26, 1.00, 1.00, 6.24	0.64	6
E	TrN	563.85	0.25, 0.19, 0.18	1.00, 0.84, 1.00, 1.00, 6.26	0.00	6
M	TrN+I	1422.62	0.22, 0.22, 0.23	1.00, 2.31, 1.00, 1.00, 8.02	0.76	6
N	GTR+G	3976.63	0.30, 0.23, 0.24	0.49, 1.47, 0.57, 0.64, 4.15	0.00	6

−lnL log likelihood scores, Pinvar proportion of invariable sites

1.7.5 (Rambaut and Drummond 2013b) and visualized using Figtree 1.4.2 (Rambaut 2012). The change in effective population size over the time was traced using Bayesian skyline plot (BSP) analysis (Drummond et al. 2005).

## Results

### Sequence analysis

For a total of 138 global PEDVs, the features of the entire coding region and their individual gene sequences are summarized in Tables 2, 3 and Fig. 1. The overall length of the alignment (including gaps) was 27,436 bps, and the deduced amino acid sequences were 9131 residues in length. The coding genome sequences showed a high degree of genetic similarities; 23,670 (86.2%) of the nucleotides and 7694 (84.3%) of the amino acids were conserved. The average identities among the complete polyprotein region sequences were 98.9% for the nucleotide sequences and 99.1% for the amino acid sequences. Of the seven individual genes, S was the most variable one (average sequence identities: 97.9% for nucleotides and 97.7% for amino acid) while M was the most

conserved (average sequence similarities: 99.5% for nucleotides and 99.4% for amino acids). Our plotting analyses also illustrated higher sequence variabilities for both nucleotides and amino acids in the S region, though both nucleotide and amino acid alterations were evenly distributed throughout the coding regions (Fig. 1). In addition, the group-specific conserved amino acid residues were investigated to observe difference between classical group and pandemic group. As a result, within multiple alignment, 27 specific conserved amino acid residues were discovered in classical isolates and 31 for pandemic isolates through five genes: ORF1a, ORF1b, S, ORF3, and N (Table 3). On the other hand, there were no group-specific conserved amino acid on E and M genes.

### Phylogenomic reconstructions

BI (both uniformed and partitioned model approaches) and ML methods produced identical tree topologies, and supported the features of the maximum clade credibility (MCC) tree (Fig. 2 and Fig. S1). 138 isolates of PEDV are classified as six groups except for one unclassified isolate (KM242131). Group 1–5 were clustered with the

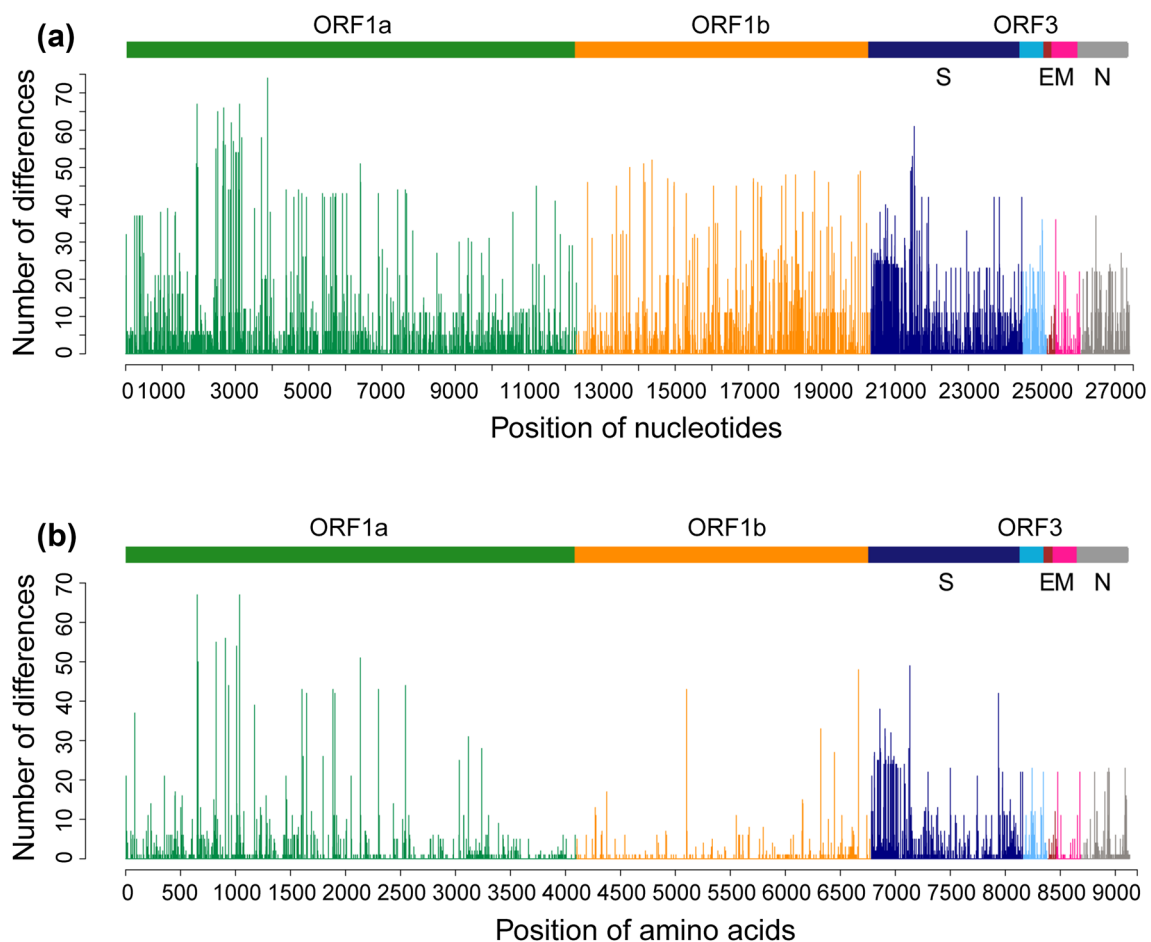
**Table 2** Summary of genomic regions of entire PEDV

Genomic region	Total sites including gaps, nt/aa	Conserved sites (%), nt/aa	Average identities (%), nt/aa	$\omega$ value (dN/dS)
Coding region	27,436/9131	23,670 (86.2%)/7694 (84.3%)	98.9/99.1	0.167
ORF1a	12,309/4103	11,047 (89.7%)/3586 (87.4%)	99.1/99.2	0.188
ORF1b	8035/2678	7326 (91.2%)/2475 (92.4%)	99.1/99.7	0.077
S	4179/1383	2900 (69.4%)/868 (62.8%)	97.9/97.7	0.243
ORF3	672/224	508 (75.6%)/164 (73.2%)	98.8/98.9	0.248
E	228/76	174 (76.3%)/58 (76.3%)	99.0/99.1	0.215
M	690/226	623 (90.3%)/205 (90.7%)	99.5/99.4	0.272
N	1323/441	1085 (82.0%)/338 (76.6%)	98.6/98.8	0.271

**Table 3** Conserved specific amino acid sequences between classical and pandemic PEDV clades

Genomic region	Classical clade (no = 11)	Pandemic clade (no = 127)
ORF1a	V-813 <sup>a</sup> , E-962, E-1021, V-1596, G-1894, A-2306, S-3074, A-3075, G-3299	A-813, D-962, D-1019, D-1021, N-1037, L-1596, D-1894, I-2138, S-2306, G-3074, V-3075, S-3299
ORF1b	A-4276, V-6512, V-6740	S-4276, M-5552, I-6512, I-6740
S	I-6785, I-6847, I-7136, E-7145, T-7329, G-7374, N-7504, A-7739, S-7824, G-7953, Y-7974, S-8012, R-8078	T-7136, Q-7145, S-7329, S-7374, S-7504, S-7546, V-7739, F-7743, A-7824, D-7953, R-8012, Q-8078
ORF3		F-8255
N	H-8931, Q-9086	L-8931, L-9086

<sup>a</sup>All positions were referred to PEDV coding genome sequences of strain CV777 (GenBank accession no. AF353511)



**Fig. 1** Plotting the nucleotide (a) and amino acid (b) sequence differences throughout the complete coding genomes of 138 global PEDVs. The number of differences at each site represents the number

of variable isolates estimated with multiple sequence alignment. Each color indicates a different genomic region. (Color figure online)

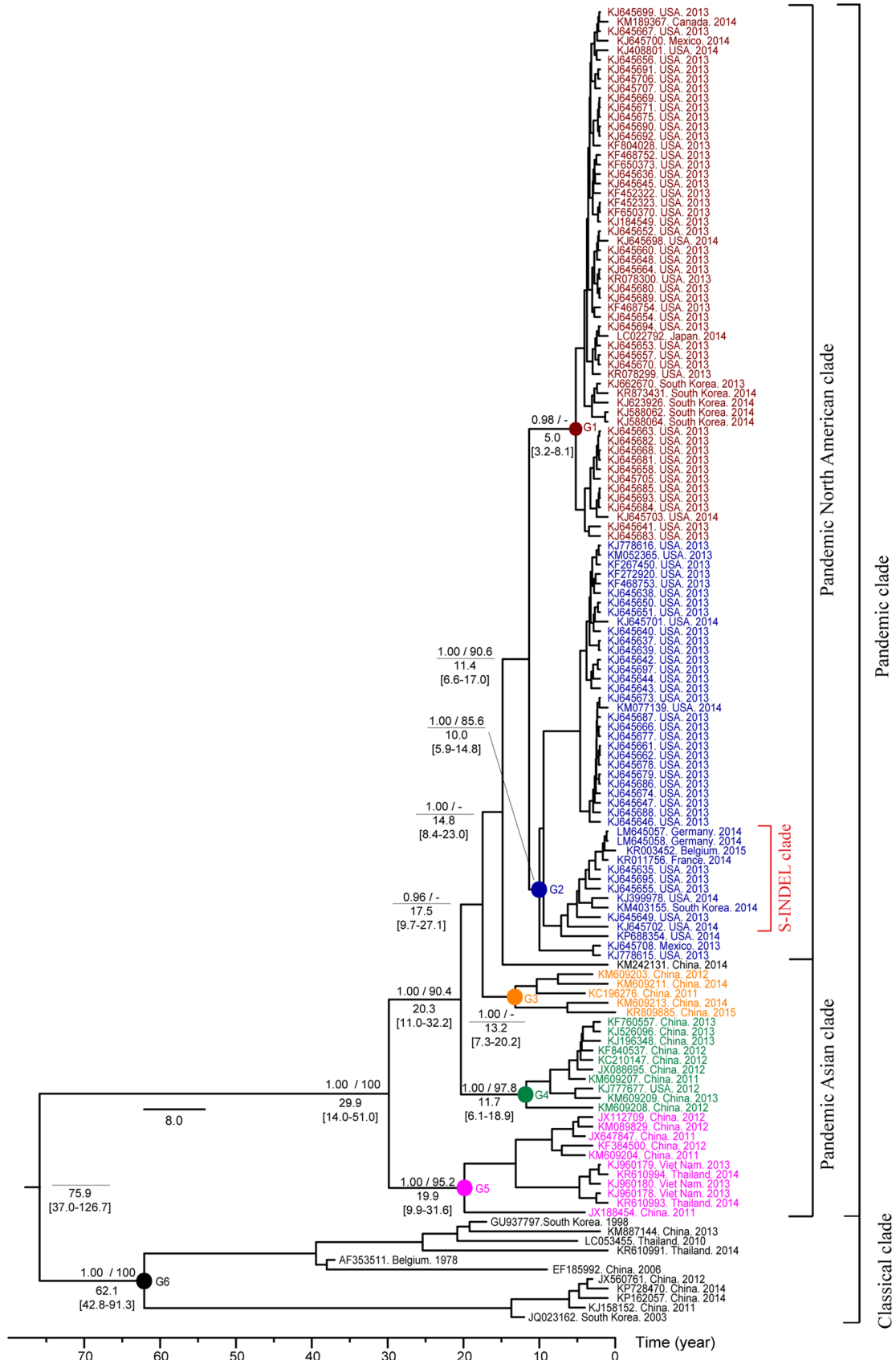
pandemic isolates, while Group 6 was grouped with the classical ones. Within the pandemic clade, the viruses of both Group 1 and Group 2 had North American origin. Group 1 was composed of 56 isolates sampled from South Korea, Japan, USA, Canada, and Mexico during 2013–2014. Group 2 consisted of 44 isolates from South Korea, USA, Mexico, Germany, Belgium, and France between 2013 and 2015, of which 11 were S INDEL variants that share specific insertions and deletions in the S gene. The remaining pandemic Groups 3–5 had Asian origin. Here Group 3 members were grouped of five Chinese viruses during 2011–2015, while ten isolates belonged to Group 4 were selected from the USA as well as China during 2011–2013. Group 5 includes isolates from China, Viet Nam, and Thailand between 2011 and 2014. Finally, Group 6 consisted of 11 classical isolates from China, Thailand, South Korea, and Belgium which were isolated from 1978 to 2014.

### Selection pressure analysis

The nonsynonymous/synonymous substitution ratio ( $\omega = dN/dS$ ) values derived from the entire data set were estimated and indicated in Table 2. The  $dN/dS$  value of complete coding sequences was 0.167, and all the values for each component gene were also lower than 1.0. Among the seven PEDV genes, the  $dN/dS$  value of M gene had the highest value (0.272), while the corresponding value for ORF1b was lowest (0.077).

### Co-estimation of substitution rate, divergence times, and population size changes

The complete coding genome sequences of 138 PEDVs collected during the past 38 years (1978–2015) were also analyzed using a Bayesian coalescent approach. Under the relaxed uncorrelated exponential clock and exponential growth population size model, the



**Fig. 2** Bayesian maximum clade credibility phylogenetic tree derived from the complete coding genome sequences of 138 PEDVs. The data set (27,436 bps) was also analyzed phylogenetically using Bayesian inference (BI) and maximum likelihood (ML) methods, and both of them produced identical topology. Divergence times (in years) are positioned below the nodes, and the 95% HPD intervals are in brackets. The credibility of the phylogenetic analysis is presented above the nodes: the left numbers represent Bayesian posterior probabilities (>0.80) and the right ones represent ML bootstrap values (>60%). Groups are indicated above the corresponding nodes using colored circles. (Color figure online)

evolutionary rate estimated was  $3.38 \times 10^{-4}$  (95% HPD  $2.75 \times 10^{-4}$ – $6.72 \times 10^{-4}$ ) substitutions/site/year, and the tMRCA calculated was 75.9 (95% HPD 37.01–126.65) years ago. Group 6 first emerged 62.1 (95% HPD 42.8–91.3) years ago, followed by chronological divergence of Group 5 (19.9 years ago; 95% HPD 9.9–31.6), Group 3 (13.2 years ago; 95% HPD 7.3–20.2), Group 4 (11.7 years ago; 95% HPD 6.1–18.9), Group 2 (10.0 years ago; 95% HPD 5.9–14.8) and Group 1 (5.0 years ago; 95% HPD 3.2–8.1). In Bayesian skyline plot, visualized in Fig. 3, the effective population size of PEDV keeps consistent level except for a short period around 2012. In the early phase in this period, the effective population size decreased rapidly. Moreover, in the latter phase of this period, the effective population size recovered its original level after a rapid increase.

## Discussion

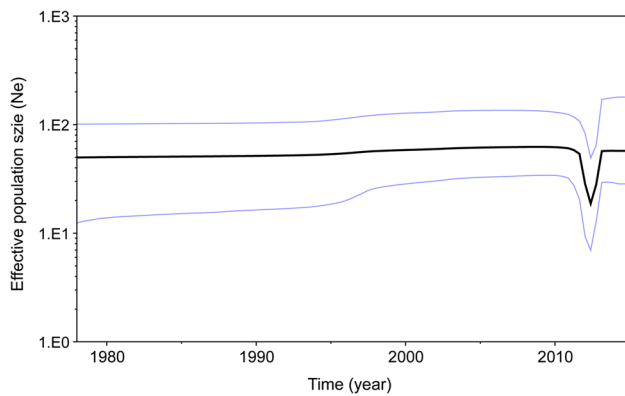
This study aimed to characterize the PEDV genome and explore its evolutionary features using protein coding genome sequences. Our pairwise comparison, of the complete coding genome sequences of 138 PEDVs, revealed that this virus had a high degree of genetic similarity compared to other porcine RNA viruses (average nucleotide identities, 77.8% for porcine reproductive and respiratory syndrome virus, 85.1% for foot-and-mouth disease virus, and 89.1% for classical swine fever virus) (Kwon et al. 2015; Yoon et al. 2013, 2011). Among the seven genes, our analyses showed that S was identified as the most variable, and the result was consistent with previous study (Huang et al. 2013). Like other coronavirus S proteins, the PEDV S protein is a primary target for vaccination; it plays key roles in receptor binding for viral entry and neutralizing antibody induction for protective immunity (Huang et al. 2013; Makadiya et al. 2016). An additional finding of the present study is that, for each of classical and pandemic clade, there were a number of amino acid residues that were conserved through five genes: ORF1a, ORF1b, S, ORF3, and N. Here, as expected, S gene had the largest proportion of specific amino acid sequences in both classical and pandemic viruses. Since only a few amino acid substitutions in its sequence can alter

the biological features as well as the antigenic properties of the virus, these specific amino acid residues might be regarded as useful markers for the classification and diagnosis of PEDV.

Although the phylogeny of PEDV has been studied in recent years using individual gene (S, ORF3, E, M, and N) or genome sequences, the present study provides further details on genome-wide phylogeny. Our findings revealed that there were two clearly defined top-level clades for PEDV, one containing the classical isolates and the other comprising the pandemic viruses. This point is consistent with the previous reports (Lin et al. 2016; Suzuki et al. 2015). The classical isolates were prototype CV777-like isolates that have appeared worldwide since the first report in England in 1971 (Oldham 1972). The pandemic PEDVs have also spread worldwide since the first emergence in China in late 2010 (Sun et al. 2012), and have caused great economic losses to the global swine industry due to its fatal mortality and morbidity. Within the pandemic clade, our phylogenomic trees showed two subclades (Asian and North American), and this supported the viewpoints of previous authors (Lee and Lee 2014; Suzuki et al. 2015; Vlasova et al. 2014; Yamamoto et al. 2016). Here, North American isolates that were first detected in the USA in April 2013, had a Chinese origin, which is in line with suggestions of other investigators (Huang et al. 2013; Kim et al. 2016; Lin et al. 2016; Suzuki et al. 2015). In addition, our results show that the North American pandemic clade were divided into two groups. The former consisted of only non-S INDEL PEDVs, while the latter was composed of both non-S INDEL and lower virulent S INDEL PEDVs containing insertions and deletions in the N-terminal region of the spike protein. This is in accord with the viewpoints of some PEDV researchers (Chen et al. 2016; Lin et al. 2016; Suzuki et al. 2015; Vlasova et al. 2014).

Next, our phylogenomic results postulated that there was no immediate relationship between time and/or country of collection and the evolution of global PEDVs within each clade. The phylogenetic groupings were consistent with the views of previous works (Lee 2015; Lee and Lee 2014; Lin et al. 2016; Yamamoto et al. 2016), and this may be largely due to the rapid expansion and diversification of PEDVs over a relatively short period and their rapid spread via frequent international trade in livestock. Such complicated population structure can make vaccine strategies and local regulation more difficult. Thus, it is essential to consistently monitor for changes in the mixed population structure of this virus.

In this study, we first analyzed the selection pressure on the genomes of PEDVs. Our analyses showed that the influence of purifying selection is acting on the coding genome of the global PEDVs; the mean nonsynonymous/synonymous substitution ratio ( $\omega = dN/dS$ ) values for individual genes as well as complete genomes were low in all cases



**Fig. 3** Bayesian skyline plot on the basis of the entire genome sequences of 138 PEDV isolates. The bold line shows the effective population size estimated through time. The upper and lower lines indicate the 95% HPD confidence intervals for this estimation

(all,  $\omega < 1$ ). Among them, the M gene showed the highest dN/dS value, while the ORF1b presented the lowest one. The purifying selections of other porcine RNA viruses were also postulated in our previous publications, which were much lower than that of PEDV (0.167); 0.068 for foot-and-mouth disease (Yoon et al. 2011) and 0.067 for classical swine fever virus (Kwon et al. 2015).

Finally, in order to explore the evolutionary mechanism of PEDV, we first co-estimated an overall evolutionary rate and population size changes as well as divergence times for the virus. The evolutionary rate of PEDV was computed at  $3.38 \times 10^{-4}$  substitutions/site/year for its complete coding genome, which was within the range of  $10^{-2}$ – $10^{-5}$  nucleotide substitutions/site/year for nearly all RNA viruses (Duffy et al. 2008). These fast evolutionary rates of RNA viruses including PEDV generally arise from several factors such as short generation times, small genome size, rapid mutation, and lack of polymerase proof-reading (Elena and Sanjuán 2005). As a result, it is possible that RNA virus raises viral population adaptation, survival, and fitness, allowing them to spread to novel environments and hosts rapidly (Lauring and Andino 2010). Our molecular dating indicates that the tMRCA of PEDV was 75.9 (95% HPD 37.01–126.65) years ago; they were derived approximately in 1940, which was about 9000 years later than the domestication of wild boar (Larson et al. 2005; Zeder 2008). This configuration revealed that the tMRCA of PEDV is similar to that of transmissible gastroenteritis virus (TGEV) and porcine respiratory coronavirus (PRCV) which are phylogenetically related porcine RNA viruses with similar genome size (28 kb). The tMRCA of TGEV and PRCV appeared in 1941 and the substitution rate was  $7.5 \pm 2 \times 10^{-4}$  (substitutions/site/year) (Sánchez et al. 1992). On the other hand, the origin of PEDV were

relatively younger than other porcine viruses with small genome size; 786 for tMRCA of porcine reproductive and respiratory syndrome virus (15 kb) (Yoon et al. 2011); 481 for tMRCA of foot-and-mouth disease virus (8 kb) (Yoon et al. 2011); 2770 for tMRCA of classical swine fever virus (12 kb) (Kwon et al. 2015). As above, the evolutionary rate and tMRCA of porcine RNA virus did not correspond to relationship between evolution rate and genome size. These conflicting result suggest that the evolution rate of porcine RNA virus was affected by other variables rather than genome size.

Regarding the effective population size changes of PEDVs, our Bayesian skyline plot analyses (BSP) indicated that the PEDV had maintained constant effective population size only excluding a short period, around 2012. This drastic change is visualized as a valley shape in our plot showing the effective number of infection occurrence. This feature was consistent with the history of PEDV prevalence as follows. Since the first report of PEDV in England in 1971, the virus prevailed until the 1980s in Europe. From the 1990s to early 2010s, the major outbreaks were also reported in East Asia. Moreover, most of them occurred in China between 2010 and 2012 (Gao et al. 2013; Sun et al. 2012). In 2012, the global outbreaks of PEDV actually seemed to settle down. However, unfortunately, there were recurrences of a new strain in April 2013 in the USA, and the virus spread to worldwide by 2015. The additional factor of the sharp decline of PEDV population size might be the decrease in host population resulting from the outbreaks of other porcine diseases such as foot-and-mouth disease; there was a culling of > 3 million pigs in South Korea during the 2010–2011 foot-and-mouth disease outbreaks. The effective population size change was helpful to understand PEDV evolution by complementing continuous information to the discontinuous information such as tMRCA and substitution rate on each node of phylogenetic tree.

PEDV is still one of the most acute pathogens in the global swine industry. Accordingly, global policies are necessary to prevent and control this acute disease. The extensive information on evolutionary dynamics we provide from this study might be very useful for the prevention and control of this virus as well as for improving our understanding of its epidemiology and evolution. Our study is the first genome-wide phylogenomic study on the temporal and spatial dynamics of PEDV.

**Author contributions** Conceived and designed the experiments: HK and SHY. Analyzed the data: JJ, WL, and JY. Wrote the manuscript: JJ and SHY. Help in revising manuscript: SS.

**Funding** Funding was provided by Rural Development Administration (Grant No. PJ01111501) and National Research Foundation of Korea (Grant No. 2012M3A9D1054622).



## References

- Bouckaert R et al (2014) BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput Biol* 10:e1003537
- Chen J-F et al (2008) Molecular characterization and phylogenetic analysis of membrane protein genes of porcine epidemic diarrhea virus isolates in China. *Virus Genes* 36:355–364
- Chen Q et al (2014) Isolation and characterization of porcine epidemic diarrhea viruses associated with the 2013 disease outbreak among swine in the United States. *J Clin Microbiol* 52:234–243
- Chen Q et al (2016) Pathogenesis comparison between the United States porcine epidemic diarrhoea virus prototype and S-INDEL-variant strains in conventional neonatal piglets. *J Gen Virol* 97:1107–1121
- Chiou M-T, Yu C-H, Chang C-C, Chung W-B, Wu H-Y, Lin C, Lin C (2014) Molecular characterization of the porcine epidemic diarrhea virus TW4/2014 in Taiwan. *Austin Virol Retrovirology* 1:4
- Cima G (2013) Viral disease affects US pigs: porcine epidemic diarrhea found in at least 11 states. *J Am Vet Med Assoc* 243:30–31
- Debouck P, Pensaert M, Coussement W (1981) The pathogenesis of an enteric infection in pigs, experimentally induced by the coronavirus-like agent, CV 777. *Vet Microbiol* 6:157–165
- Drummond AJ, Rambaut A, Shapiro B, Pybus OG (2005) Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol Biol Evol* 22:1185–1192
- Duarte M, Tobler K, Bridgen A, Rasschaert D, Ackermann M, Laude H (1994) Sequence analysis of the porcine epidemic diarrhea virus genome between the nucleocapsid and spike protein genes reveals a polymorphic ORF. *Virology* 198:466–476
- Duffy S, Shackleton LA, Holmes EC (2008) Rates of evolutionary change in viruses: patterns and determinants. *Nat Rev Genet* 9:267–276
- Elena SF, Sanjuán R (2005) Adaptive value of high mutation rates of RNA viruses: separating causes from consequences. *J Virol* 79:11555–11558
- Erixon P, Svennblad B, Britton T, Oxelman B (2003) Reliability of Bayesian posterior probabilities and bootstrap frequencies in phylogenetics. *Syst Biol* 52:665–673
- Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39:783–791
- Gao Y, Kou Q, Ge X, Zhou L, Guo X, Yang H (2013) Phylogenetic analysis of porcine epidemic diarrhea virus field strains prevailing recently in China. *Arch Virol* 158:711–715
- Grasland B et al (2015) Complete genome sequence of a porcine epidemic diarrhea S gene indel strain isolated in France in December 2014. *Genome Announc* 3:e00535–e00515
- Guindon S, Delsuc F, Dufayard J-F, Gascuel O (2009) Estimating maximum likelihood phylogenies with PhyML. In: Posada DD (ed) *Bioinformatics for DNA sequence analysis*. Humana Press, Totowa, pp 113–137
- Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. In: *Nucleic acids symposium series*, pp 95–98
- Hanke D et al (2015) Comparison of porcine epidemic diarrhea viruses from Germany and the United States, 2014. *Emerg Infect Dis* 21:493–496
- Huang Y-W et al (2013) Origin, evolution, and genotyping of emergent porcine epidemic diarrhea virus strains in the United States. *MBio* 4:e00737–e00713
- Jung K, Saif LJ (2015) Porcine epidemic diarrhea virus infection: etiology, epidemiology, pathogenesis and immunoprophylaxis. *Vet J* 204:134–143
- Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 30:772–780
- Kim YK, Cho Y-Y, An B-H, Lim S-I, Lim J-A, Cho I-S, An D-J (2016) Molecular characterization of the spike and ORF3 genes of porcine epidemic diarrhea virus in the Philippines. *Arch Virol* 161:1323–1328
- Kocherhans R, Bridgen A, Ackermann M, Tobler K (2001) Completion of the porcine epidemic diarrhoea coronavirus (PEDV) genome sequence. *Virus Genes* 23:137–144
- Kweon C et al (1993) Isolation of porcine epidemic diarrhea virus (PEDV) in Korea. *Korean J Vet Res* 33:249–254
- Kwon T, Yoon SH, Kim K-W, Caetano-Anolles K, Cho S, Kim H (2015) Time-calibrated phylogenomics of the classical swine fever viruses: genome-wide bayesian coalescent approach. *PLoS ONE* 10:e0121578
- Larson G et al (2005) Worldwide phylogeography of wild boar reveals multiple centers of pig domestication. *Science* 307:1618–1621
- Lauring AS, Andino R (2010) Quasispecies theory and the behavior of RNA viruses. *PLoS Pathog* 6:e1001005
- Lee C (2015) Porcine epidemic diarrhea virus: an emerging and re-emerging epizootic swine virus. *Virol J* 12:1
- Lee S, Lee C (2014) Outbreak-related porcine epidemic diarrhea virus strains similar to US strains, South Korea, 2013. *Emerg Infect Dis* 20:1223–1226
- Li W et al (2012a) New variants of porcine epidemic diarrhea virus, China, 2011. *Emerg Infect Dis* 18:1350
- Li Z-L et al (2012b) Molecular characterization and phylogenetic analysis of porcine epidemic diarrhea virus (PEDV) field strains in south China. *Virus genes* 45:181–185
- Li Z et al (2013) Sequence and phylogenetic analysis of nucleocapsid genes of porcine epidemic diarrhea virus (PEDV) strains in China. *Arch Virol* 158:1267–1273
- Li R, Qiao S, Yang Y, Su Y, Zhao P, Zhou E, Zhang G (2014) Phylogenetic analysis of porcine epidemic diarrhea virus (PEDV) field strains in central China based on the ORF3 gene and the main neutralization epitopes. *Arch Virol* 159:1057–1065
- Lin C-M, Saif LJ, Marthaler D, Wang Q (2016) Evolution, antigenicity and pathogenicity of global porcine epidemic diarrhea virus strains. *Virus Res* 226:20–39
- Makadiya N, Brownlie R, van den Hurk J, Berube N, Allan B, Gerds V, Zakhartchouk A (2016) S1 domain of the porcine epidemic diarrhea virus spike protein as a vaccine antigen. *Virol J* 13:1
- Martin DP, Murrell B, Golden M, Khoosal A, Muhire B (2015) RDP4: detection and analysis of recombination patterns in virus genomes. *Virus Evol* 1:vev003
- Mesquita J, Hakze-van der Honing R, Almeida A, Lourenco M, Poel W, Nascimento M (2015) Outbreak of porcine epidemic diarrhea virus in Portugal, 2015. *Transbound Emerg Dis* 62:586–588
- Oka T et al (2014) Cell culture isolation and sequence analysis of genetically diverse US porcine epidemic diarrhea virus strains including a novel strain with a large deletion in the spike gene. *Vet Microbiol* 173:258–269
- Oldham J (1972) Letter to the editor. *Pig Farming* 10:72–73
- Paarlberg PL (2014) Updated estimated economic welfare impacts of porcine epidemic diarrhea virus (PEDV). Purdue University, Department of Agricultural Economics, Working Papers 14:1–38
- Park S-J, Song D-S, Park B-K (2013) Molecular epidemiology and phylogenetic analysis of porcine epidemic diarrhea virus (PEDV) field isolates in Korea. *Arch Virol* 158:1533–1541
- Pasick J et al (2014) Investigation into the role of potentially contaminated feed as a source of the first-detected outbreaks of porcine epidemic diarrhea in Canada. *Transbound Emerg Dis* 61:397–410
- Pensaert M, De Bouck P (1978) A new coronavirus-like particle associated with diarrhea in swine. *Arch Virol* 58:243–247
- Posada D (2003) Using MODELTEST and PAUP\* to select a model of nucleotide substitution. *Curr Protoc Bioinf*. <https://doi.org/10.1002/0471250953.bi0605s00>

- Pritchard C, Paton D, Wibberley G, Ibata G (1999) Transmissible gastroenteritis and porcine epidemic diarrhoea in Britain. *Vet Rec* 144:616–618
- Puranaveja S et al (2009) Chinese-like strain of porcine epidemic diarrhoea virus, Thailand. *Emerg Infect Dis* 15:1112–1115
- Rambaut A (2012) FigTree version 1.4.0. <http://tree.bio.ed.ac.uk/software/figtree>
- Rambaut A, Drummond A (2013a) Tracer 1.6. University of Edinburgh, Edinburgh
- Rambaut A, Drummond A (2013b) TreeAnnotator v1.7.5
- Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572–1574
- Sánchez CM, Gebauer F, Suñé C, Mendez A, Dopazo J, Enjuanes L (1992) Genetic evolution and tropism of transmissible gastroenteritis coronaviruses. *Virology* 190:92–105
- Schmidt HA, Strimmer K, Vingron M, von Haeseler A (2002) TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* 18:502–504
- Smíd B, Valíček L, Rodák L, Kudrna J, Musilová J (1992) Detection of porcine epidemic diarrhoea virus using electron microscopy in the Czech Republic. *Vet Med* 38:333–341
- Stevenson GW et al (2013) Emergence of porcine epidemic diarrhoea virus in the United States: clinical signs, lesions, and viral genomic sequences. *J Vet Diagn Invest* 25:649–654
- Sun R-Q, Cai R-J, Chen Y-Q, Liang P-S, Chen D-K, Song C-X (2012) Outbreak of porcine epidemic diarrhoea in suckling piglets, China. *Emerg Infect Dis* 18:161–163
- Sun D, Wang X, Wei S, Chen J, Feng L (2016) Epidemiology and vaccine of porcine epidemic diarrhoea virus in China: a mini-review. *J Vet Med Sci* 78:355–363
- Suyama M, Torrents D, Bork P (2006) PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res* 34:W609–W612
- Suzuki T et al (2015) Molecular characterization of pig epidemic diarrhoea viruses isolated in Japan from 2013 to 2014. *Infect Genet Evol* 36:363–368
- Swofford D (2003) PAUP\*: phylogenetic analysis using parsimony, version 4.0b10
- Takahashi K, Okada K, Ohshima K (1983) An outbreak of swine diarrhoea of a new-type associated with coronavirus-like particles in Japan *Nihon juigaku zasshi*. *Jpn J Vet Sci* 45:829
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL\_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 25:4876–4882
- Toan NT, Puranaveja S, Thanawongnuwech R (2011) Genetic characterization of porcine epidemic diarrhoea virus (PEDV) isolates from southern Vietnam during 2009–2010 outbreaks. *Thai J Vet Med* 41:55
- Vlasova AN et al (2014) Distinct characteristics and complex evolution of PEDV strains, North America, May 2013–February 2014. *Emerg Infect Dis* 20:1620–1628
- Yamamoto T, Suzuki T, Ohashi S, Miyazaki A, Tsutsui T (2016) Genomic motifs as a novel indicator of the relationship between strains isolated from the epidemic of porcine epidemic diarrhoea in 2013–2014. *PLoS ONE* 11:e0147994
- Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24:1586–1591
- Yoon SH, Park W, King DP, Kim H (2011) Phylogenomics and molecular evolution of foot-and-mouth disease virus. *Mol Cells* 31:413–421
- Yoon SH, Kim H, Kim J, Lee H-K, Park B, Kim H (2013) Complete genome sequences of porcine reproductive and respiratory syndrome viruses: perspectives on their temporal and spatial dynamics. *Mol Biol Rep* 40:6843–6853
- Zeder MA (2008) Domestication and early agriculture in the Mediterranean Basin: Origins, diffusion, and impact. *Proc Natl Acad Sci USA* 105:11597–11604