



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

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



Evolutionary characterization of the emerging porcine epidemic diarrhea virus worldwide and 2014 epidemic in Taiwan



Ming-Hua Sung^{a,b}, Ming-Chung Deng^c, Yi-Hsuan Chung^b, Yu-Liang Huang^c, Chia-Yi Chang^c, Yu-Ching Lan^d, Hsin-Lin Chou^d, Day-Yu Chao^{a,*}

^a Graduate Institute of Microbiology and Public Health, College of Veterinary Medicine, National Chung Hsing University, Taichung 40227, Taiwan

^b Taichung City Animal Protection and Health Inspection Office, Taichung 408, Taiwan

^c Animal Health Research Institute, Council of Agriculture, New Taipei City 25158, Taiwan

^d Department of Health Risk Management, School of Public Health, China Medical University, Taichung, Taiwan

ARTICLE INFO

Article history:

Received 15 May 2015

Received in revised form 3 September 2015

Accepted 12 September 2015

Available online 13 September 2015

Keywords:

Porcine epidemic diarrhea virus

Likelihood mapping

Phylogenetic signal

Evolutionary rate

Taiwan

Nsp3

Spike

ABSTRACT

Since 2010, a new variant of PEDV belonging to Genogroup 2 has been transmitting in China and further spreading to the United States and other Asian countries including Taiwan. In order to characterize in detail the temporal and geographic relationships among PEDV strains, the present study systematically evaluated the evolutionary patterns and phylogenetic resolution in each gene of the whole PEDV genome in order to determine which regions provided the maximal interpretative power. The result was further applied to identify the origin of PEDV that caused the 2014 epidemic in Taiwan. Thirty-four full genome sequences were downloaded from GenBank and divided into three non-mutually exclusive groups, namely, worldwide, Genogroup 2 and China, to cover different ranges of secular and spatial trends. Each dataset was then divided into different alignments by different genes for likelihood mapping and phylogenetic analysis. Our study suggested that both nsp3 and S genes contained the highest phylogenetic signal with substitution rate and phylogenetic topology similar to those obtained from the complete genome. Furthermore, the proportion of nodes with high posterior support (posterior probability > 0.8) was similar between nsp3 and S genes. The nsp3 gene sequences from three clinical samples of swine with PEDV infections were aligned with other strains available from GenBank and the results suggested that the virus responsible for the 2014 PEDV outbreak in Taiwan clustered together with Clade I from the US within Genogroup 2. In conclusion, the current study identified the nsp3 gene as an alternative marker for a rapid and unequivocal classification of the circulating PEDV strains which provides complementary information to the S gene in identifying the emergence of epidemic strain resulting from recombination.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Porcine epidemic diarrhea virus (PEDV) causes an acute and highly contagious enteric tract infection, characterized by severe villus atrophy and congestion of the thin-walled intestines, which usually lead to high morbidity and mortality, especially in piglets (Song and Park, 2012). The disease was first observed among fattening pigs from United Kingdom in 1971 and the etiological agent was identified in Belgium as a new coronavirus, which was designated as PEDV prototype strain CV777 (Jung and Saif, 2015; Pensaert and de Bouck, 1978). Within the next two decades, PEDV was reported in several other European countries including Hungary, Italy, Germany, France, Switzerland and the Czech Republic, causing only isolated outbreaks in Europe (Hanke et al., 2015; Jung and Saif, 2015). In Asia, PEDV was first identified in 1982 and is now considered an endemic, causing devastating enteric diseases and

substantial economic losses to the pork industry in many Asian countries such as China, South Korea, Japan, Thailand, and Taiwan (Lin et al., 2014; Park et al., 2007, 2011; Song and Park, 2012; Temeeyasen et al., 2014; Zhang et al., 2013).

Comparisons of full-length genomes showed that different PEDV strains are more closely related to alphacoronaviruses in a bat than other known alphacoronaviruses, suggesting that interspecies transmission of coronavirus may have occurred decades ago between bats and pigs or through other intermediate hosts (Huang et al., 2013; Tang et al., 2006; Woo et al., 2012). PEDV, belonging to the genus *Alphacoronavirus* within the family *Coronaviridae*, is an enveloped, single-stranded RNA genome with a 5' cap and a 3' polyadenylated tail (Jung and Saif, 2015). The size of its genome is approximately 28 Kb with two thirds of the 5' genome containing two large open reading frames (ORFs), 1a and 1b, that encode two nonstructural polyproteins, pp1a and pp1b, which are involved in genome replication and transcription (Brian and Baric, 2005). The remaining PEDV genome contains ORFs encoding four structural and one nonstructural proteins in the following order: spike (S), ORF3, envelope (E), membrane (M)

* Corresponding author.

E-mail address: dychao@nchu.edu.tw (D.-Y. Chao).

and nucleoprotein (N). Similar to S proteins from other coronaviruses (CoVs), the PEDV S protein is a glycoprotein on the viral surface and has a pivotal function in regulating interactions with specific host cell receptor glycoproteins to mediate viral entry (Li, 2015). Thus, the S glycoprotein is often used as an important viral component to understand the genetic relationships of different PEDV strains and the epidemiological status of PEDV in the field (Chen et al., 2013b; Tian et al., 2013). Nevertheless, the high variability of other genes including ORF3, E, M or N protein has been previously utilized for phylogenetic inference (Chen et al., 2013a; Ge et al., 2013; Kubota et al., 1999; Pan et al., 2012; Temeeyasen et al., 2014; Yang et al., 2013). In order to get better phylogenetic resolution, several studies used PEDV full genome sequences for analysis (Chen et al., 2014; Pan et al., 2012; Sun et al., 2015; X. Wang et al., 2013); however, sequencing the full genome is expensive from a laboratory perspective and also limits the rapid characterization of novel PEDV strains. Furthermore, for many computationally intensive analyses, utilizing the full genome is unfeasible. It would be beneficial to use only those genomic regions that contain the highest phylogenetic signal to reduce cost without losing valuable information.

Since 2010, massive PED outbreaks were reported in China characterized by 80 to 100% illness among infected swine herds and a 50 to 90% mortality rate among infected suckling piglets (Zhang et al., 2013). The emerging strains in Asia are distinct from previous endemic PEDV strains characterized by multiple amino acid insertions/deletions or mutations on the S protein (J. Wang et al., 2013). The mutations on the S protein have been speculated to be associated with escape from neutralizing antibodies due to the use of a bivalent vaccine against transmissible gastroenteritis virus (TGEV) and PEDV (Chen et al., 2013a). This novel PEDV strain has continued to spread throughout the United States and Taiwan where no vaccination program has been implemented (Chen et al., 2014; Deng et al., 2014). Therefore, tracing the transmission and evolutionary changes of PEDV is important for future public health intervention. The objectives of the present study are (1) to systematically evaluate the evolutionary patterns and phylogenetic resolution in each gene in order to determine which regions provided the maximum interpretative power to infer temporal and geographic relationships among PEDV strains by using 34 complete genomic sequences from worldwide collection of PEDV field strains; and (2) to identify the origin of PEDV that caused the epidemic in Taiwan in 2014.

2. Materials and methods

2.1. Sequence data

We downloaded all full genome sequences (34) available in GenBank as of February 26, 2014 (<http://www.ncbi.nlm.nih.gov/>) for which the sampling year and country of collection was recorded (Supplementary Table 1). Since Genogroup 1 contained only 8 full-length sequences and no PEDV in the US was found until 2013, dataset was divided into three non-mutually exclusive groups: worldwide (34 sequences), Genogroup 2 (24 sequences) including all the recent emerging strains, and China (20 sequences) including all strains obtained from China, to cover different ranges of secular and spatial trends. Each of those datasets was then divided into different alignments by genes: structural (S, E, M, N) and non-structural (NSP1, NSP2, NSP3, NSP4, 3C-like protease, NSP6, NSP7, NSP8, NSP9, NSP10, RNA-dependent RNA polymerase, helicase, exoribonuclease, uridylylate-specific endoribonuclease and putative 2'-O-methyl transferase) based on the sequence prediction from the whole genome of the strain CV777 (GenBank accession no. NC_003436). We created two additional concatenated datasets with length >2000 nucleotides for comparison: (1) NSP1 and NSP2 (denoted as 5'nsp1-2); (2) ORF3, E, M, and N (denoted as 3'OEMN). Sequences were aligned using ClustalW2 available from the European Molecular Biology Laboratory (EMBL, <http://www.ebi.ac.uk/Tools/msa/clustalw2/>) and manually edited.

Diarrhea among the piglets from the swine farms in Taiwan began in December 2013 and the etiologic agent of this epidemic was confirmed to be PEDV (Lin et al., 2014). The total nucleic acids from three representative clinical samples were extracted as previously described (Deng et al., 2014) and the cDNA was produced using the SuperScript III Reverse Transcription kit with random hexamers (Invitrogen, Carlsbad, CA, USA). The complete sequence of the NSP3 gene in each clinical sample was determined using the primers summarized in Table 1 and are available under GeneBank accession number KR632490-2.

2.2. Comparative analysis

To evaluate the phylogenetic resolution of each gene from each dataset, two bioinformatics approaches were employed: likelihood mapping (LM) and phylogenetic analysis involving the determination of the model of evolution and Bayesian Markov chain Monte-Carlo (MCMC) phylogenetic analyses.

2.2.1. LM analysis

To investigate the phylogenetic signal of each gene from all three datasets, LM was performed using TREE-PUZZLE program by analyzing 10,000 random quartets (Schmidt et al., 2002). LM assesses if a dataset is suitable for phylogenetic reconstruction by analyzing four randomly chosen groups of sequences, termed quartets. Each quartet was evaluated using maximum likelihood and three possible unrooted tree topologies were weighted for each quartet. The posterior weights were then plotted onto a triangular surface. Based on this method, the fully resolved tree topologies were plotted in the three corners, indicating the presence of a tree-like phylogenetic signal, and the unresolved quartets were shown in the central region of the triangle. The phylogenetic noise was computed probabilistically and a star-like signal was shown when more than 30% of the dots fell within the central area representing unresolved phylogenies (Strimmer and von Haeseler, 1997).

2.2.2. Phylogenetic analysis

Before proceeding to the phylogenetic analysis, each gene and each of the four datasets were tested against the best fitting nucleotide substitution model as specified. The best-fit model of nucleotide substitution was selected under the Akaike information criterion (AIC) and Bayesian information criterion (BIC) as implemented in jModelTest (Posada, 2008).

To estimate the genealogy and the evolutionary timescale of NSP3, S and E sequence alignments using the worldwide, Genogroup 2 and China datasets, the Bayesian framework implemented in BEAST software package version 2.1.3 (Bouckaert et al., 2014) under the designated molecular clock model and the nucleotide substitution model predetermined by jModelTest was used. The BEAST program was run with two different rate categories for codon position 1 + 2 and position 3, allowing gamma-distributed rate variation when model selection results were suggested. The published year of isolation was assigned to each sequence (tip) to calibrate the rate estimate and the expected time to coalescence was modeled under constant population sizes. The MCMC process was run for 100 million iterations until convergence

Table 1

Oligonucleotide primers used for amplifications of the PEDV nsp3 gene by PCR and sequencing.

Primer ID	Sequence (5' to 3')	Position*	Used to amplify fragment
Nsp3-F2	TCCCACCGATGGTAAATAGTG	2646–2665	PCR forward primer
Nsp3-R2	TGAACAGACACAAAAACGAGAAG	8088–8110	PCR reverse primer
Nsp3-F2-1	TTGGGTGATGTGTCGGCTTG	3706–3725	Sequencing primer
Nsp3-R2-1	GCTTCTTACAGAAGTCTAGAAC	6960–6981	Sequencing primer
Nsp3-F2-2	AGGAAGATGTTCAACAAGTTTC	4700–4721	Sequencing primer
Nsp3-R2-2	ACACTGTAATTAATTACGTGAC	6176–6197	Sequencing primer

* Positions correspond to the PEDV CV777 strain (GenBank accession no. NC_003436).

with sampling every 10,000th generation (Drummond et al., 2012). The results were visualized using the Tracer program v.1.6 (<http://beast.bio.ed.ac.uk/Tracer>), and convergence of the Markov chain was assessed by calculating the effective sampling size (ESS) for each parameter. All the MCMC runs were repeated at least three times to confirm the convergence. All ESS values >200 indicated sufficient sampling. For each dataset, the maximum clade credibility (MCC) tree, which is the tree with the largest product of posterior clade probabilities, was selected from the posterior tree distribution after 10% burn-in using the program TreeAnnotator version 2.1.2 and displayed by FigTree v.1.4.2.

2.2.3. Evaluation of the molecular clock hypothesis

Molecular clock analyses were performed based on the NSP3, S and E genes from all three worldwide, Genogroup 2 and China datasets. The strict molecular clock assumes the same evolutionary rates along all branches in the tree, while the relaxed molecular clock allows different rates along different tree branches. These two models were compared by calculating the Bayes Factor (BF), which is the ratio of the marginal likelihoods (marginal with respect to the prior) between the two models being compared (Suchard et al., 2001; Vexler et al., 2013). The approximate marginal likelihood for each coalescent model was calculated through 1000 bootstrap samplings using the harmonic mean of the sampled likelihoods and the natural log difference of marginal likelihood between any two models is the \log_e of the Bayes Factor, $\log_e(\text{BF})$. Evidence against the null model (i.e., the one with lower marginal likelihood) is indicated by $2 \log_e(\text{BF})$. It is moderately rejected when the value is greater than 3 and strongly rejected when it is greater than 10. The calculations were performed with BEAST version 2.1.3 and Tracer v.1.5.

2.3. Statistical analysis

The proportion of constant sites between different datasets was analyzed by the nonparametric Mann–Whitney U test. The linear relationship between nucleotide length and phylogenetic noise was examined using Pearson correlation test.

3. Results

3.1. Phylogenetic signal using LM

LM analysis was used to calculate the phylogenetic noise for each gene in each dataset and it has been shown that datasets with less than 30% noise are usually reliable for phylogenetic inference based on previous simulation studies (Strimmer and von Haeseler, 1997). In our study, the phylogenetic noise ranged from 55.9% (nsp7) to 3% (nsp3) in the worldwide dataset (Table 2) and only four individual genes (nsp1, 7, 9 and 10) contained noise greater than 30%. Similar results were found when only the China dataset was used, among which four genes (nsp 1, 7, 10 and E) contained noise greater than 30%. However, when the Genogroup 2 dataset was examined, 10 genes contained noise greater than 30%, including those genes detected in the worldwide and China datasets (Supplementary Tables 2 and 3). Overall, in the worldwide dataset, the triangle of LM analysis showed that nsp3 and spike genes contained only 3.2% and 4.1% of the probabilities in the center region representing the star-like topologies (unresolved phylogenies), respectively. Similar results were also found in the Genogroup 2 and China datasets.

Next, we further examined the constant sites and alpha values among different genes since the sequence numbers were equivalent among the three datasets. The results suggested that Genogroup 2 dataset consistently contained higher constant sites than the worldwide and China datasets, particularly in 3C-like protease, nsp 6, 8, uridylylate-specific endoribonuclease, M and N genes, with statistical significance ($p < 0.05$). Furthermore, a broad range of alpha values of the gamma-distribution was observed in all three datasets. Although different

Table 2
Determination of phylogenetic signal/noise by likelihood mapping analysis.

Dataset	Gene	Length	% Noise	Constant sites (%) [#]	Alpha
Worldwide	Nsp1	330	32.4	90	0.18
Worldwide	Nsp2	2355	9.8*	89.1	0.13
Worldwide	Nsp3	4863	3.2*	91.9	0.03
Worldwide	Nsp4	1443	13.5*	92.5	0.34
Worldwide	3C-like protease (nsp5)	906	24.2*	94	0.02
Worldwide	Nsp6	840	17.4*	91.9	0.38
Worldwide	Nsp7	249	55.9	94	0.52
Worldwide	Nsp8	585	15.1*	93.2	0.4
Worldwide	Nsp9	324	30.1	94.1	0.03
Worldwide	Nsp10	405	47.4	96.8	0.02
Worldwide	RNA-dependent RNA polymerase (nsp12)	2781	10*	93.5	0.02
Worldwide	Helicase (nsp13)	1557	12.9*	93.8	0.02
Worldwide	Exoribonuclease (nsp14)	1785	6.1*	93.1	0.02
Worldwide	Uridylylate-specific endoribonuclease (nsp15)	1017	16.9*	93.6	0.03
Worldwide	Putative 2'-O-methyl transferase (nsp16)	903	20.8*	94.6	0.03
Worldwide	S	4180	4.1*	85.7	0.1
Worldwide	ORF3	675	16.7*	77.9	0.09
Worldwide	E	231	26.6*	80.5	3.13
Worldwide	M	681	27.9*	91.1	0.02
Worldwide	N	1326	16.6*	87	0.16
Worldwide	5'(nsp1–2)	2685	7*	89.2	0.1
Worldwide	3'OEMN	2911	8.7*	85.4	0.02

* Strong phylogenetic signal (<30% noise).

[#] Average of constant sites among all genes is 90.6%, which is statistically significantly lower than that from Genogroup 2 dataset.

gamma-distributions might affect the levels of phylogenetic noise, no correlation was found. However, when the relationship between nucleotide length and phylogenetic signal was plotted, a linear correlation was observed with statistical significance (worldwide: $R^2 = 0.52$, $p < 0.001$; Genogroup 2: $R^2 = 0.50$, $p < 0.001$; China: $R^2 = 0.51$, $p < 0.001$) (Fig. 1). The results were further supported by the concatenated alignment of multiple genes. Both 5'nsp1–2 and 3'OEMN reduced the noise to 10%, which was significantly lower compared to the noises obtained from individual gene analysis (Table 2). In summary, the LM analysis results suggested that while length is significantly correlated with phylogenetic signal, the nsp3 and S genes contained the greatest phylogenetic signal, and had similar constant sites but different gamma-distributions under different temporal and spatial contents. We therefore focused on these two genes for further phylogenetic analysis.

3.2. Evolutionary model and rate estimation

The rate matrix, alpha value in the gamma distribution and the proportion of the invariable site were calculated for the complete, nsp3, S and E genes in all three datasets (Supplementary Table 4). In the worldwide dataset, the estimated rate matrix was similar and the same evolutionary model (GTR + I + G) was selected for all three different gene sets except for the E gene. However, when the genogroup reduced or geographically reduced datasets were used, different evolutionary models for each gene were obtained. The small sample size of sequences in this study does not permit for further investigation.

The coalescent-based Bayesian MCMC approach was further used to estimate the evolutionary rate of the nsp3, S and E genes from all three worldwide, Genogroup 2 and China datasets. The strict (SC) and relaxed molecular clock (RC) hypotheses were compared by using the BEAST program that incorporated the year of sample collection. Analyses were conducted using a constant coalescent prior and the specified substitution models based on the results were shown in Supplementary Table 4. For all three datasets, the SC hypothesis was strongly rejected (Bayes Factors > 70) when compared to RC model for the nsp3 and S

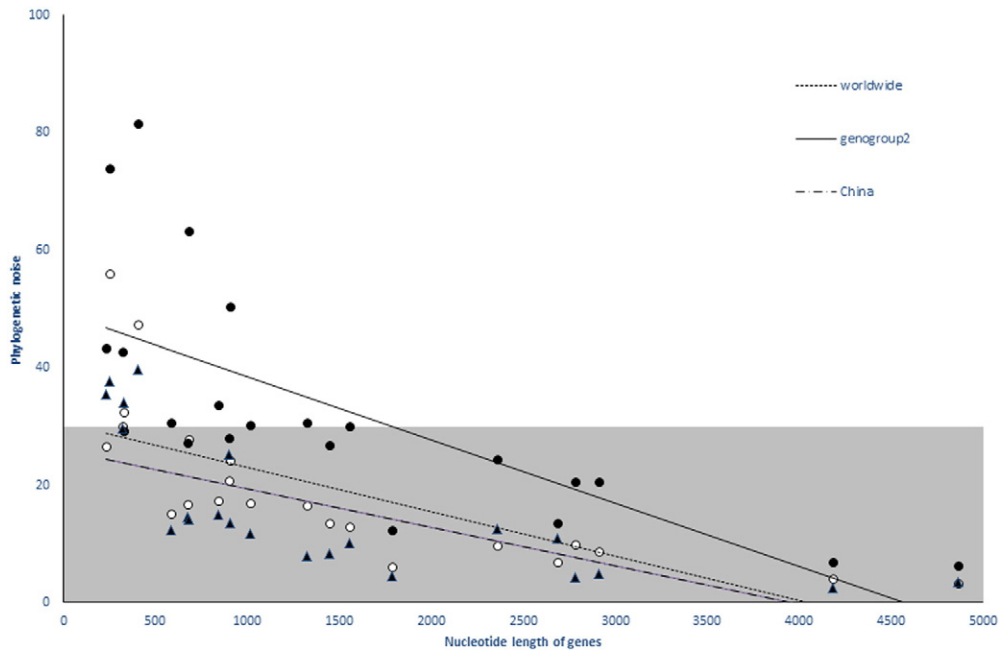


Fig. 1. Correlation between nucleotide length and phylogenetic noise. Length of each gene tested for phylogenetic noise is plotted on the x-axis, and the phylogenetic noise as measured using TREE-PUZZLE program is plotted on the y-axis. Open circles indicate measurements from the worldwide dataset, filled circles indicate measurements from the Genogroup 2 dataset, and filled triangles indicate measurements from the China dataset. A linear regression line was plotted against each dataset (dashed = worldwide, solid = Genogroup 2, intermittent = China) with statistical significance (worldwide: $R^2 = 0.82$, $p < 0.001$; Genogroup 2: $R^2 = 0.68$, $p < 0.001$; China: $R^2 = 0.82$, $p < 0.001$) by Pearson correlation test. The gray filled region at the bottom of the graph denoted phylogenetic noise <30%.

genes, which was consistent with that obtained using complete genome sequences, suggesting rate heterogeneity among branches (Table 3). For the worldwide dataset, although the 95% high posterior density intervals (95% HPDs) were overlapping, the mean evolutionary rate estimated under the RC assumption was slightly faster for S (2.22×10^{-2}) than for nsp3 (1.63×10^{-2}) gene or complete genome (1.37×10^{-2}). Similar findings were also found in the Genogroup 2 and China datasets. On the contrary, the E gene didn't show rate heterogeneity among branches

since neither the RC model nor the SC model favored the result based on the Bayes Factor calculation. Furthermore, the E gene showed the highest evolutionary rate based on SC model estimation from all three different datasets.

Although different evolutionary rates were estimated for all three different genes, the TMRCA was similar for all three genes. For worldwide dataset, about 36 to 38 years before present were estimated from the RC model when complete genomes, nsp3 and S genes were

Table 3

Bayesian estimation of the molecular clock, evolutionary rate, and time to most recent common ancestor (TMRCA).

Dataset	Gene	Clock model	Bayes factor	Evolutionary rate (substitutions/site/year)	Lower HPD	Upper HPD	TMRCA (year)	Lower HPD	Upper HPD	
Worldwide (N = 34)	Complete	Strict		1.31E-02	1.14E-02	1.48E-02	119.8	102.2	134.5	
		Relaxed	494.9	1.37E-02	9.77E-03	1.80E-02	38.4	35	45.8	
	NSP3	Strict		2.16E-02	1.81E-02	2.50E-02	60.4	53.5	67.6	
		Relaxed	146.2	1.63E-02	1.09E-02	2.19E-02	37	35	41.2	
	S	Strict		2.24E-02	1.89E-02	2.63E-02	91.5	77.7	106.9	
		Relaxed	288.1	2.28E-02	1.50E-02	3.10E-02	36.1	35	38.7	
	E	Strict		5.81E-02	4.30E-02	7.42E-02	35.7	35	37	
		Relaxed	39.9	2.90E-02	1.77E-02	4.14E-02	36	35	38.3	
	Genogroup2 (N = 24)	Complete	Strict		3.93E-02	3.14E-02	4.78E-02	8.8	7.5	10.1
			Relaxed	357	4.59E-02	2.66E-02	6.71E-02	3.2	2.3	4.5
NSP3		Strict		5.81E-02	4.42E-02	7.38E-02	4.6	3.8	5.3	
		Relaxed	108.9	4.50E-02	2.56E-02	6.51E-02	2.9	2.2	3.8	
S		Strict		6.07E-02	4.63E-02	7.78E-02	6.3	5.2	7.5	
		Relaxed	103.8	5.67E-02	3.24E-02	8.46E-02	3.2	2.3	4.4	
E		Strict		1.29E-01	7.89E-02	1.85E-01	2.8	2.2	3.4	
		Relaxed	15.3	8.27E-02	4.01E-02	1.32E-01	2.6	2.1	3.2	
China (N = 20)		Complete	Strict		3.35E-02	2.95E-02	3.78E-02	29.6	27.8	31.3
			Relaxed	464	1.37E-02	9.83E-03	1.80E-02	26.9	26	29
	NSP3	Strict		3.50E-02	2.82E-02	4.22E-02	26.6	26	27.8	
		Relaxed	97.3	2.31E-03	1.95E-03	2.65E-03	26.2	26	26.9	
	S	Strict		5.15E-02	4.01E-02	6.35E-02	26.5	26	27.5	
		Relaxed	159.3	3.05E-02	1.79E-02	4.43E-02	27.2	26	29.8	
	E	Strict		6.90E-02	4.97E-02	9.03E-02	26.5	26	27.5	
		Relaxed	40.1	2.63E-02	1.40E-02	4.00E-02	26.8	26	28.9	

used. Similar results were obtained when the E gene was used. For Genogroup 2 dataset, the TMRCA was estimated to be about 2 years before present, which was consistent among all genes. For the China dataset, about 26 years before present was estimated, which was also consistent among all genes.

3.3. Genealogy among different genes

In order to determine how the topology and support for the genealogies differed between genes with a strong phylogenetic signal (nsp3 and S genes) and weak signal (E gene), we examined both the topology and posterior probability of the MCC tree for the nsp3, S and E genes in all three different datasets under the relaxed molecular clock and the constant size coalescent prior. The example of the MCC trees based on three different genes from the worldwide dataset is shown in Fig. 2. Each gene is comprised of the same number of taxa (n = 34), and each tree is a bifurcating rooted tree, so the number of internal nodes (including the root) will be the same. To compare the degree of posterior support for the MCC tree inferred from each gene, we calculated the sum of the posterior support for all internal nodes of the MCC tree, which represented the estimate of the total probability of the given tree topology. The results suggested that the nsp3 and S genes showed a similar sum of the posterior probability in all three datasets and the E gene had the lowest results (Table 4). In order to compare the particular aspects of the topologies of each genealogy, clades with relatively high posterior support (posterior probability > 0.8) were counted and the proportion of nodes with high support was compared among three different genes from all three datasets. The result with the highest support was obtained from nsp3 and S genes from all three datasets, despite that a slightly lower proportion (0.52) of high support for nsp3 gene from the Genogroup 2 dataset was noticed than that of S gene (0.61) (Table 4). Similar to the low sum of posterior probability of the

Table 4

Sum of the posterior probabilities of all internal nodes for all three different genes from the worldwide, Genogroup 2 and China databases.

Dataset	Gene	Sum of posterior probabilities	Proportion of nodes with posterior probability >0.8
Worldwide (N = 34)	NSP3	27.48	0.76(25/33)
	S	26.79	0.73(24/33)
	E	17.85	0.45(15/33)
Genogroup2 (N = 24)	NSP3	17.5	0.52(12/23)
	S	17.32	0.61(14/23)
	E	6.75	0.13(3/23)
China (N = 20)	NSP3	17.82	0.84(16/19)
	S	18.55	0.84(16/19)
	E	13.48	0.63(12/19)

E gene, the proportion of high support was lowest for the E gene among all three datasets.

3.4. Phylogenetic analysis

Since the nsp3 and S genes showed similar evolutionary rate and phylogenetic topology, the nsp3 genes from three clinical samples of swine with PEDV infection were obtained and aligned with 34 nsp3 genes available in GenBank. The results suggested that the viruses from the 2014 PEDV epidemic in Taiwan were highly clustered with the viruses from the US within Genogroup 2 (Fig. 3A). While preparing the manuscript, comprehensive PEDV full-genome sequences collected from the US were obtained and the phylogenetic tree was further constructed based on the nsp3 gene including 74 and 3 PEDV sequences from the US and Taiwan, respectively. The results suggested that the virus responsible for the 2014 PEDV outbreak in Taiwan clustered closely together with Clade I from the US (Fig. 3B).

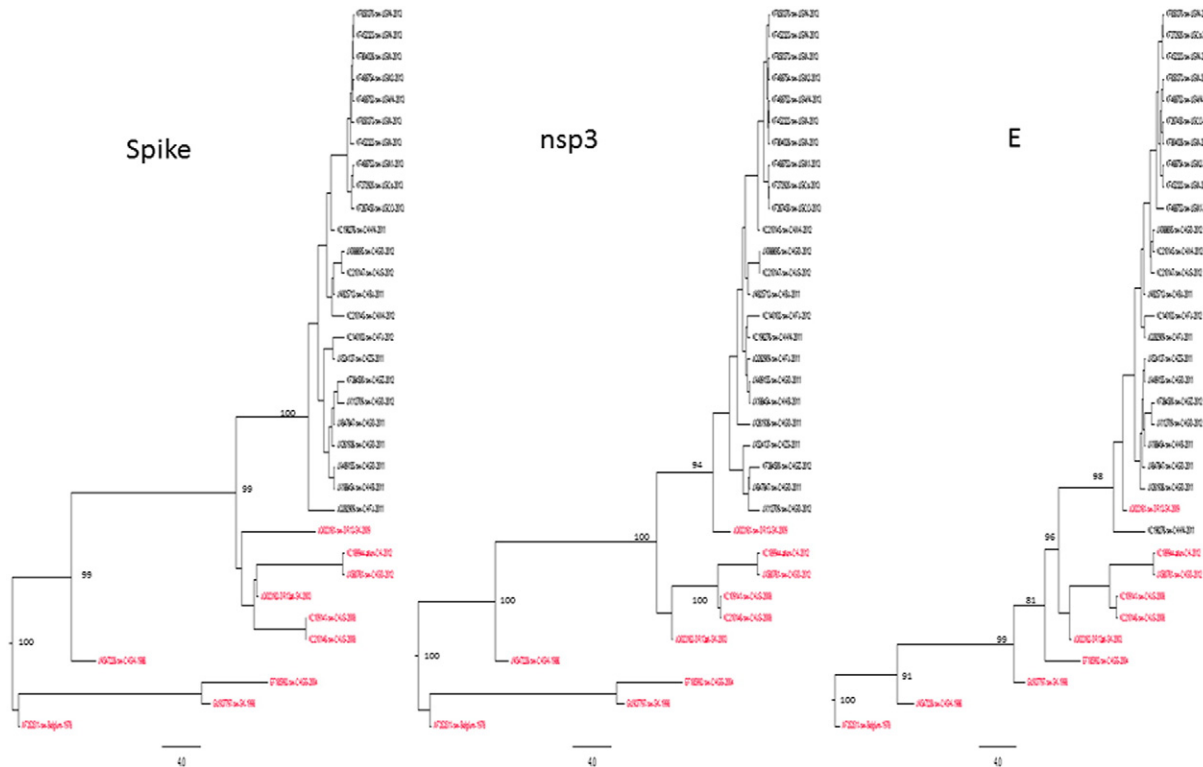


Fig. 2. Bayesian maximum clade credibility phylogenetic trees based on the spike (S, left), nsp3 (middle), and E (right) genes for the worldwide dataset according to the indicated substitution models and molecular clock suggested by the ModelTest and Bayes factors. Branch lengths are scaled in years. Branch with posterior support >0.8 are labeled inside the node. Sequences colored in red clustered within Genogroup 1 and black within Genogroup 2 based on the full-genome phylogenetic analysis (Huang et al., 2013). Each strain is labeled based on the accession number on GeneBank followed by the host, country and year, which could refer to supplementary Table S1 for details.

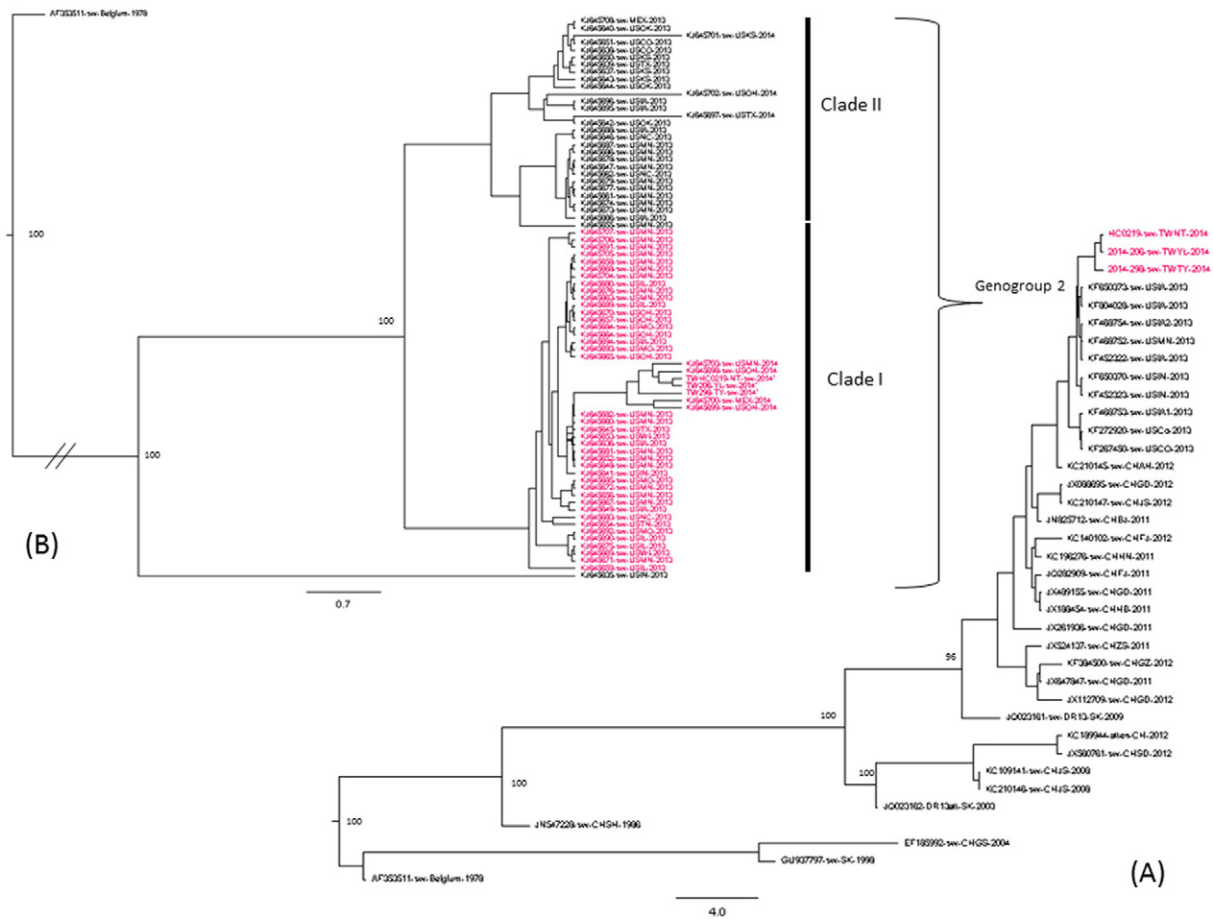


Fig. 3. Bayesian maximum clade credibility phylogenetic trees with three PEDV strains collected during the 2014 Taiwan epidemic based on the nsp3 for the worldwide dataset (A) and North America dataset (B) according to the substitution model of GTR + I + G and GTR + I suggested by jModeltest, respectively. Branch with posterior support >0.8 are labeled inside the node. Sequences within each clade are colored consistently across genes. Three local Taiwan PEDV strains are marked with asterisk. North American PEDV strains are labeled based on the accession number on GeneBank (KJ645635–708) followed by the host, country and year.

4. Discussion

Phylogeographic inference of the routes and rates of the dissemination are important to understand in order to plan effective intervention, particularly for new and potentially much more deadly infectious diseases. Until now, only a few studies that used full-genome sequences were able to define a meaningful geographic structure in Asia (Huang et al., 2013; Sun et al., 2015). However, because of the relatively high cost of sequencing the full genome and the enormous computational time required for long sequences, comprehensive phylogenetic analyses have been unfeasible. Although several investigators have reported that the S gene is appropriate to study the genetic relatedness of PEDV (Chen et al., 2014; Lee et al., 2010), a comprehensive analysis has not been performed to determine which genes are the most appropriate for such phylogenetic studies. In the current study, the phylogenetic noises have been evaluated by likelihood mapping method and validated by the phylogenetic method which compares the substitution model, evolution rate, tree topology and node support proportion. Our study suggested that both nsp3 and S genes are suitable for phylogenetic tree construction and phylogenetic analysis.

RNA synthesis during coronavirus infection is carried out by a replicase–transcriptase composed of 16 nonstructural protein (nsp) subunits. Among these, nsp3 is the largest and the first to be inserted into the endoplasmic reticulum (Hagemeijer et al., 2014). Nsp3 comprises multiple structural domains including two papain-like proteases (PLPs), which are predicted to be conserved among all CoVs. Other than polyprotein processing of the cleavage between nsp1–nsp2 and

nsp2–nsp3, PLPs also possess a deubiquitinase activity, which acts to counter host innate immunity (Mielech et al., 2014). Also, the N-terminal region of the nsp3 is highly conserved, containing a ubiquitin-like globular fold followed by a flexible, extended acidic-domain rich in glutamic acid and catalytically active ADP-ribose-100-phosphatase (ADRP, app-100-pase) domain, which are thought to play a role during the synthesis of viral subgenomic RNAs (Báez-Santos et al., 2015; Oostra et al., 2008). Interestingly, comparative genome and proteome analyses of two bovine CoV (BCoV) isolates showed a predominant clustering of mutations within the nsp3 multi-domain (Chouljenko et al., 2001). Despite that the detailed function of nsp3 during PEDV infection is currently unknown, the multifunctionality of the nsp3, the frequency of point mutations observed in nsp3 domains, and the involvement of nsp3 in structural arrangements of the replicase complex together with double-membrane vesicles may render pleiotropic effects, not only in pathogenicity, but also on future emerging coronaviruses (Snijder et al., 2003). A recent study identified four hypervariable regions suitable for phylogenetic study from whole genome analysis including regions covering the nsp3 and S genes (Sun et al., 2015). Our study confirmed the results and further suggested that differential evolutionary patterns were responsible for both genes under different genotypes or spatial transmission patterns.

The 3' end of PEDV genome containing four structural proteins (S, E, M and N) and one nonstructural protein (ORF3) plays an important role during RNA binding after phosphorylation, virus assembly, immunogenic activity and virulence (de Haan et al., 1998). The E protein, a small transmembrane protein of 76 to 109 amino acids in length, plays a

pivotal role in the assembly of virions by inducing membrane curvature or aid in membrane scission (Fischer et al., 1998; Khattari et al., 2006). In addition, the E protein has ion channel activity and interacts with host proteins (Alvarez et al., 2010; Pervushin et al., 2009; Teoh et al., 2010; Wilson et al., 2004). In the case of severe acute respiratory syndrome (SARS)-CoV and other coronaviruses, the deletion of the E protein reduces its growth and virus production in vitro and in vivo, resulting in an attenuated virus (DeDiego et al., 2007; Dediego et al., 2008; Netland et al., 2010; Ortego et al., 2007). Although little is known about the function of E protein in PEDV, the E gene has been previously used for a phylogenetic study (Park et al., 2013). Therefore, the 3' end of PEDV genome was the most commonly used genomic region for phylogenetic studies of PEDV (Chen et al., 2013a; Ge et al., 2013; Kubota et al., 1999; Pan et al., 2012; Temeeyasen et al., 2014; Yang et al., 2013). Our study found that a 26.6% noise from the E gene was sufficient to affect the statistical support for the phylogeny and topology of the genealogies despite previous simulation studies suggesting that less than 30% noise is usually reliable for phylogenetic inference (Strimmer and von Haeseler, 1997). Additionally, the E gene showed the highest evolutionary rate based on SC model estimation from all three different datasets.

Lack of geographic structure based on phylogenetic analysis using only structural genes has been previously suggested in other viruses (Gray et al., 2010). Previous studies also found that the E, M and N genes are relatively conserved, and various PEDV strains may cluster together and show monophylogeny without reflecting the true genetic differences exhibited at the whole-genome level (Chen et al., 2014; Vlasova et al., 2014). Our study strengthens the importance of evaluating the evolutionary characteristics among different genes before performing the phylogenetic analysis.

Divergence of PEDV could be driven by genetic recombination, as suggested in other coronaviruses (Graham and Baric, 2010; Huang et al., 2013). Multiple recombination events related to the emergence of PEDV strain with epidemic potential has been suggested and three recombination sites were proposed to be located close to 3' end of ORF 1a including partial of nsp3 gene and partial S-ORF3-E-M-partial N region (Tian et al., 2014; Vlasova et al., 2014). It was also hypothesized that the emergent US PEDV strains are possibly the descendent of two major lineages derived from the ZMDZY and AH2012 lineage through recombination (Vlasova et al., 2014). Our phylogenetic analysis also showed that strains FJND-3 (accession number JQ282909) and CH/S (accession number JN547228) isolated from China during 2011 and 1986, respectively, clustered with different lineages, depending on the nsp3 or S genes used (Fig. 2). Therefore, the phylogenetic analysis relying only on the S gene might ignore the important recombination event with epidemiological significance. The results in our study suggested that other than the S gene, the target nsp3 gene of the non-structural proteins in the 5' PEDV genome region would complement the molecular characterization and identification of novel and emerging PEDV strains.

Since 2010, a new variant of PEDV belonging to Genogroup 2 has been transmitting in China, further spreading to the United States and other Asian countries including Taiwan. Since the nsp3 and S genes showed similar evolutionary rate and phylogenetic topology, the nsp3 genes from three clinical samples of swine with PEDV infection in Taiwan were analyzed here. The phylogenetic analysis suggested that viruses from the 2014 PEDV epidemic in Taiwan were closely related to those from US, which was consistent with the results published previously using the S gene (Lin et al., 2014). Our study further demonstrated that the PEDV responsible for the 2014 epidemic in Taiwan clustered together with Clade I of northern US viruses (Fig. 3B). The relatively short length and abundance in the database of either nsp3 or S gene allow for landscape phylodynamic analysis, which incorporates phylogenetics and geographic information system (GIS) frameworks, to provide a better interpretation of the causes and consequences of epidemics in Taiwan.

In conclusion, we have shown that the S and nsp3 genes contained the lowest phylogenetic noise, which are both suitable for phylogenetic analysis and molecular characterization. Although differential evolutionary patterns were responsible for the evolution of both genes under different geographic or genogroup-reduced databases, the constructed phylogenetic trees utilizing the S and nsp3 genes could complement each other to identify the emergence of epidemic strain resulting from recombination. The robustness of the results between the S and nsp3 genes leads us to recommend the nsp3 genomic region as an alternative marker for a rapid and unequivocal classification of the circulating PEDV strains and in providing complementary information to the understanding of PEDV epidemiology.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.meegid.2015.09.011>.

References

- Alvarez, E., DeDiego, M., Nieto-Torres, J., Jiménez-Guardeño, J., Marcos-Villar, L., Enjuanes, L., 2010. The envelope protein of severe acute respiratory syndrome coronavirus interacts with the non-structural protein 3 and is ubiquitinated. *Virology* 402, 281–291.
- Báez-Santos, Y., St John, S., Mesecar, A., 2015. The SARS-coronavirus papain-like protease: structure, function and inhibition by designed antiviral compounds. *Antivir. Res.* 115.
- Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C., Xie, D., Suchard, M., Rambaut, A., Drummond, A., 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* 10, e1003537.
- Brian, D., Baric, R., 2005. Coronavirus genome structure and replication. *Curr. Top. Microbiol. Immunol.* 287, 1–30.
- Chen, J., Liu, X., Lang, H., Wang, Z., Shi, D., Shi, H., Zhang, X., Feng, L., 2013a. Genetic variation of nucleocapsid genes of porcine epidemic diarrhea virus field strains in China. *Arch. Virol.* 158, 1397–1401.
- Chen, J., Liu, X., Shi, D., Shi, H., Zhang, X., Li, C., Chi, Y., Feng, L., 2013b. Detection and molecular diversity of spike gene of porcine epidemic diarrhea virus in China. *Viruses* 5, 2601–2613.
- Chen, Q., Li, G., Stasko, J., Thomas, J., Stensland, W., Pillatzki, A., Gauger, P., Schwartz, K., Madson, D., Yoon, K., Stevenson, G., Burrough, E., Harmon, K., Main, R., Zhang, J., 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.
- Chouljenko, V.N., Lin, X.Q., Storz, J., Kousoulas, K.G., Gorbalenya, A.E., 2001. Comparison of genomic and predicted amino acid sequences of respiratory and enteric bovine coronaviruses isolated from the same animal with fatal shipping pneumonia. *J. Gen. Virol.* 82, 2927–2933.
- de Haan, C., Kuo, L., Masters, P., Vennema, H., Rottier, P., 1998. Coronavirus particle assembly, primary structure requirements of the membrane protein. *J. Virol.* 72, 6838–6850.
- DeDiego, M., Alvarez, E., Almazan, F., Rejas, M., Lamirande, E., Roberts, A., Shieh, W., Zaki, S., Subbarao, K., Enjuanes, L., 2007. A severe acute respiratory syndrome coronavirus that lacks the E gene is attenuated in vitro and in vivo. *J. Virol.* 81, 1701–1713.
- Dediego, M., Pewe, L., Alvarez, E., Rejas, M., Perlman, S., Enjuanes, L., 2008. Pathogenicity of severe acute respiratory coronavirus deletion mutants in hACE-2 transgenic mice. *Virology* 376, 379–389.
- Deng, M., Chang, C., Huang, T., Kuo, S., Tsai, H., Chang, C., Huang, Y., 2014. The outbreak of porcine epidemic diarrhea in Taiwan. *Taiwan Vet. J.* 40, 115–121.
- Drummond, A., Suchard, M., Xie, D., Rambaut, A., 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* 29, 1969–1973.
- Fischer, F., Stegen, C., Masters, P., Samsonoff, W., 1998. Analysis of constructed E gene mutants of mouse hepatitis virus confirms a pivotal role for E protein in coronavirus assembly. *J. Virol.* 72, 7885–7894.
- Ge, F., Yang, D., Ju, H., Wang, J., Liu, J., Liu, P., Zhou, J., 2013. Epidemiological survey of porcine epidemic diarrhea virus in swine farms in Shanghai, China. *Arch. Virol.* 158, 2227–2231.
- Graham, R., Baric, R., 2010. Recombination, reservoirs, and the modular spike: mechanisms of coronavirus cross-species transmission. *J. Virol.* 84, 3134–3146.
- Gray, R.R., Veras, N.M., Santos, L.A., Salemi, M., 2010. Evolutionary characterization of the West Nile Virus complete genome. *Mol. Phylogenet. Evol.* 56(1), 195–200 (Jul).
- Hagemeijer, M., Monastyrska, I., Griffith, J., van der Sluijs, P., Voortman, J., van Bergen en Henegouwen, P., Vonk, A., Rottier, P., Reggiori, F., de Haan, C., 2014. Membrane rearrangements mediated by coronavirus nonstructural proteins 3 and 4. *Virology* 458–459.
- Hanke, D., Jenckel, M., Petrov, A., Ritzmann, M., Stadler, J., Akimkin, V., Blome, S., Pohlmann, A., Schirmer, H., Beer, M., Höper, D., 2015. Comparison of porcine epidemic diarrhea viruses from Germany and the United States, 2014. *Emerg. Infect. Dis.* 21, 493–496.
- Huang, Y., Dickerman, A., Piñeyro, P., Li, L., Fang, L., Kiehne, R., Opriessnig, T., Meng, X., 2013. Origin, evolution, and genotyping of emergent porcine epidemic diarrhea virus strains in the United States. *MBio* 4 (e00737-00713).

- Jung, K., Saif, L., 2015. Porcine epidemic diarrhea virus infection: etiology, epidemiology, pathogenesis and immunoprophylaxis. *Vet. J.* S1090-0233 (00078-00077).
- Khattari, Z., Brotons, G., Akkawi, M., Arbely, E., Arkin, I., Salditt, T., 2006. SARS coronavirus E protein in phospholipid bilayers: an x-ray study. *Biophys. J.* 90, 2038–2050.
- Kubota, S., Sasaki, O., Amimoto, K., Okada, N., Kitazima, T., Yasuhara, H., 1999. Detection of porcine epidemic diarrhea virus using polymerase chain reaction and comparison of the nucleocapsid protein genes among strains of the virus. *J. Vet. Med. Sci.* 61, 827–830.
- Lee, D., Park, C., Kim, S., Lee, C., 2010. Heterogeneity in spike protein genes of porcine epidemic diarrhea viruses isolated in Korea. *Virus Res.* 149.
- Li, F., 2015. Receptor recognition mechanisms of coronaviruses: a decade of structural studies. *J. Virol.* 89, 1954–1964.
- Lin, C., Chung, W., Chang, S., Wen, C., Liu, H., Chien, C., Chiou, M., 2014. US-like strain of porcine epidemic diarrhea virus outbreaks in Taiwan, 2013–2014. *J. Vet. Med. Sci.* 76, 1297–1299.
- Mielech, A., Chen, Y., Mesecar, A., Baker, S., 2014. Nidovirus papainlike proteases: multi-functional enzymes with protease, deubiquitinating and deISGylating activities. *Virus Res.* 194, 184–190.
- Netland, J., DeDiego, M., Zhao, J., Fett, C., Alvarez, E., Nieto-Torres, J., Enjuanes, L., Perlman, S., 2010. Immunization with an attenuated severe acute respiratory syndrome coronavirus deleted in E protein protects against lethal respiratory disease. *Virology* 399, 120–128.
- Oostra, M., Hagemeijer, M., van Gent, M., Bekker, C., Te Lintelo, E., Rottier, P., De Haan, C., 2008. Topology and membrane anchoring of the coronavirus replication complex: not all hydrophobic domains of nsp3 and nsp6 are membrane spanning. *J. Virol.* 82, 12392–12405.
- Ortego, J., Ceriani, J., Patino, C., Plana, J., Enjuanes, L., 2007. Absence of E protein arrests transmissible gastroenteritis coronavirus maturation in the secretory pathway. *Virology* 368, 296–308.
- Pan, Y., Tian, X., Li, W., Zhou, Q., Wang, D., Bi, Y., Chen, F., Song, Y., 2012. Isolation and characterization of a variant porcine epidemic diarrhea virus in China. *Viol.* 9, 195.
- Park, S.J., Moon, H.J., Yang, J.S., Lee, C.S., Song, D.S., Kang, B.K., Park, B.K., 2007. Sequence analysis of the partial spike glycoprotein gene of porcine epidemic diarrhea viruses isolated in Korea. *Virus Genes* 35, 321–332.
- Park, S.J., Kim, H.K., Song, D.S., Moon, H.J., Park, B.K., 2011. Molecular characterization and phylogenetic analysis of porcine epidemic diarrhea virus (PEDV) field isolates in Korea. *Arch. Virol.* 156, 577–585.
- Park, S., Song, D., Park, B., 2013. Molecular epidemiology and phylogenetic analysis of porcine epidemic diarrhea virus (PEDV) field isolates in Korea. *Arch. Virol.* 158, 1533–1541.
- Pensaert, M., de Bouck, P., 1978. A new coronavirus-like particle associated with diarrhea in swine. *Arch. Virol.* 58, 243–247.
- Pervushin, K., Tan, E., Parthasarathy, K., Lin, X., Jiang, F., Yu, D., Vararattanavech, A., Soong, T., Liu, D., Torres, J., 2009. Structure and inhibition of the SARS coronavirus envelope protein ion channel. *PLoS Pathog.* 5, e1000511.
- Posada, D., 2008. jModelTest: phylogenetic model averaging. *Mol. Biol. Evol.* 25, 1253–1256.
- Schmidt, H., Strimmer, K., Vingron, M., von Haeseler, A., 2002. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* 18, 502–504.
- Snijder, E.J., Bredenbeek, P.J., Dobbe, J.C., Thiel, V., Ziebuhr, J., Poon, L.L., Guan, Y., Rozanov, M., Spaan, W.J., Gorbalenya, A.E., 2003. Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage. *J. Mol. Biol.* 331, 991–1004.
- Song, D., Park, B., 2012. Porcine epidemic diarrhoea virus: a comprehensive review of molecular epidemiology, diagnosis, and vaccines. *Virus Genes* 44, 167–175.
- Strimmer, K., von Haeseler, A., 1997. Likelihood-mapping: a simple method to visualize phylogenetic content of a sequence alignment. *Proc. Natl. Acad. Sci. U. S. A.* 94, 6815–6819.
- Suchard, M., Weiss, R., Sinsheimer, J., 2001. Bayesian selection of continuous-time Markov chain evolutionary models. *Mol. Biol. Evol.* 18, 1001–1013.
- 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.
- Tang, X., Zhang, J., Zhang, S., Wang, P., Fan, X., Li, L., Li, G., Dong, B., Liu, W., Cheung, C., Xu, K., Song, W., Vijaykrishna, D., Poon, L., Peiris, J., Smith, G., Chen, H., Guan, Y., 2006. Prevalence and genetic diversity of coronaviruses in bats from China. *J. Virol.* 80, 7481–7490.
- Temeeyasen, G., Srijangwad, A., Tripipat, T., Tipsombatboon, P., Piriyaopongsa, J., Phoolcharoen, W., Chuanasa, T., Tantituvanont, A., Nilubol, D., 2014. Genetic diversity of ORF3 and spike genes of porcine epidemic diarrhea virus in Thailand. *Infect. Genet. Evol.* 21, 205–213.
- Teoh, K., Siu, Y., Chan, W., Schlüter, M., Liu, C., Peiris, J., Bruzzone, R., Margolis, B., Nal, B., 2010. The SARS coronavirus E protein interacts with PALS1 and alters tight junction formation and epithelial morphogenesis. *Mol. Biol. Cell* 21, 3838–3852.
- Tian, Y., Yu, Z., Cheng, K., Liu, Y., Huang, J., Xin, Y., Li, Y., Fan, S., Wang, T., Huang, G., Feng, N., Yang, Z., Yang, S., Gao, Y., Xia, X., 2013. Molecular characterization and phylogenetic analysis of new variants of the porcine epidemic diarrhea virus in Gansu, China in 2012. *Viruses* 5, 1991–2004.
- Tian, P., Jin, Y., Xing, G., Qv, L., Huang, Y., Zhou, J., 2014. Evidence of recombinant strains of porcine epidemic diarrhea virus, United States, 2013. *Emerg. Infect. Dis.* 20, 1735–1738.
- Vexler, A., Deng, W., Wilding, G., 2013. Nonparametric Bayes factors based on empirical likelihood ratios. *J. Stat. Plan Infer.* 143, 611–620.
- Vlasova, A., Marthaler, D., Wang, Q., Culhane, M., Rossow, K., Rovira, A., Collins, J., Saif, L., 2014. Distinct characteristics and complex evolution of PEDV strains, North America, May 2013–February 2014. *Emerg. Infect. Dis.* 20, 1620–1628.
- Wang, J., Zhao, P., Guo, L., Liu, Y., Du, Y., Ren, S., Li, J., Zhang, Y., Fan, Y., Huang, B., Liu, S., Wu, J., 2013a. Porcine epidemic diarrhea virus variants with high pathogenicity, China. *Emerg. Infect. Dis.* 19, 2048–2049.
- Wang, X., Niu, B., Yan, H., Gao, D., Yang, X., Chen, L., Chang, H., Zhao, J., Wang, C., 2013b. Genetic properties of endemic Chinese porcine epidemic diarrhea virus strains isolated since 2010. *Arch. Virol.* 158, 2487–2494.
- Wilson, L., McKinlay, C., Gage, P., Ewart, G., 2004. SARS coronavirus E protein forms cation-selective ion channels. *Virology* 330, 322–331.
- Woo, P., Lau, S., Lam, C., Lau, C., Tsang, A., Lau, J., Bai, R., Teng, J., Tsang, C., Wang, M., Zheng, B., Chan, K., Yuen, K., 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. *J. Virol.* 86, 3995–4008.
- Yang, X., Huo, J., Chen, L., Zheng, F., Chang, H., Zhao, J., Wang, X., Wang, C., 2013. Genetic variation analysis of reemerging porcine epidemic diarrhea virus prevailing in central China from 2010 to 2011. *Virus Genes* 46, 337–344.
- Zhang, Q., Hu, R., Tang, X., Wu, C., He, Q., Zhao, Z., Chen, H., Wu, B., 2013. Occurrence and investigation of enteric viral infections in pigs with diarrhea in China. *Arch. Virol.* 158, 1631–1636.