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.

Infection, Genetics and Evolution 44 (2016) 114–121

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Research paper

Full-length genome analysis of two genetically distinct variants of porcine epidemic diarrhea virus in Thailand

Thaniwan Cheun-Arom [a], Gun Temeeyasen [a], Thitima Tripipat [a], Pavita Kaewprommal [b], Jittima Piriyapongsa [b], Suchada Sukrong [c], Wanchai Chongcharoen [d], Angkana Tantituvanont [d], Dachrit Nilubol [a,*]

[a] Department of Veterinary Microbiology, Faculty of Veterinary Science, Bangkok, Thailand
[b] Genome Technology Research Unit, National Center for Genetic Engineering and Biotechnology, National Science and Technology Development Agency, Pathumthani, Thailand
[c] Department of Pharmacognosy and Pharmaceutical Botany, Faculty of Pharmaceutical Sciences, Chulalongkorn University, Bangkok, Thailand
[d] Department of Pharmaceutics and Industrial Pharmacy, Faculty of Pharmaceutical Sciences, Chulalongkorn University, Bangkok 10330, Thailand

ARTICLE INFO

Article history:
Received 29 September 2015
Received in revised form 17 June 2016
Accepted 23 June 2016
Available online 25 June 2016

Keywords:
Porcine epidemic diarrhea virus
Full-length genome
Classical
Pandemic
Antigenic index
Thailand

ABSTRACT

Porcine epidemic diarrhea virus (PEDV) has continued to cause sporadic outbreaks in Thailand since 2007 and a pandemic variant containing an insertion and deletion in the spike gene was responsible for outbreaks. In 2014, there were further outbreaks of the disease occurring within four months of each other. In this study, the full-length genome sequences of two genetically distinct PEDV isolates from the outbreaks were characterized. The two PEDV isolates, CBR1/2014 and EAS1/2014, were 28,039 and 28,033 nucleotides in length and showed 96.2% and 93.6% similarities at nucleotide and amino acid levels respectively. In total, we have observed 1048 nucleotide substitutions throughout the genome. Compared to EAS1/2014, CBR1/2014 has 2 insertions of 4 (^{56}GENQ^{59}) and 1 (^{140}N) amino acid positions 56–59 and 140, and 2 deletions of 2 (^{160}DG^{161}) and 1 (^{1199}Y) amino acid positions 160–161 and 1199. The phylogenetic analysis based on full-length genome of CBR1/2014 isolate has grouped the virus with the pandemic variants. In contrast, EAS1/2014 isolate was grouped with CV777, LZC and SM98, a classical variant. Our findings demonstrated the emergence of EAS1/2014, a classical variant which is novel to Thailand and genetically distinct from the currently circulating endemic variants. This study warrants further investigations into molecular epidemiology and genetic evolution of the PEDV in Thailand.

1. Introduction

Porcine epidemic diarrhea virus (PEDV) is an enveloped, single stranded positive-sense RNA virus belonging to the genus *Alphacoronavirus*, family *Coronaviridae* (Bridgen et al., 1998; Duarte et al., 1993; Kocherhans et al., 2001). PEDV is a causative agent of porcine epidemic diarrhea (PED), a devastating enteric disease characterized by watery diarrhea, dehydration and high mortality in pigs younger than one week of age (Pensaert and De Bouck, 1978). PED was first reported in Belgium and the United Kingdom in 1976–1978 (Chasey and Cartwright, 1978; Pensaert and De Bouck, 1978; Wood, 1977). The disease was subsequently reported in several Asian countries, including Japan, South Korea, China, Vietnam and Thailand, and became endemic, causing significant welfare and economic problems (Li et al., 2012; Park et al., 2007a; Song and Park, 2012; Temeeyasen et al., 2014; Vui et al., 2015). Since 2013, there has been an emergence of a highly virulent PEDV in the North American continent and its subsequent reports from several countries worldwide.

The genome of PEDV is approximately 28 kb in length and comprises seven open reading frames (ORFs): ORF1a and ORF1b, spike (S), ORF3, envelope (E), membrane (M), and nucleoprotein (N) (Kocherhans et al., 2001). Two thirds of the genome is occupied by ORF1a and ORF1b, which encode nonstructural proteins. The other 4 structural proteins, S, E, M, and N, are located downstream of the ORF1a/1b gene. ORF3 is located between S and E and encodes an accessory protein. The S gene encodes S glycoprotein, which can be divided into the S1 and S2 domains. The S glycoprotein plays an important role in binding to specific host cell receptors. The S1 domain of S glycoprotein stimulates the production of neutralizing antibodies by the host (Chang et al., 2002; Cruz et al., 2006, 2008; Duarte and Laude, 1994; Sun et al., 2008; Sun et al., 2007). The ORF3 gene is the only accessory gene and is an important determinant of virulence in PEDV (Park et al., 2008; Park et al., 2007b). Vaccine-derived isolates have a unique deletion of 17 amino acids at positions 82 to 99 (Park et al., 2008; Park et al., 2007b). The other E and M

* Corresponding author at: Department of Veterinary Microbiology, Faculty of Veterinary Science, Chulalongkorn University, Henry Dunant Road, Pathumwan, Bangkok 10330, Thailand.
E-mail address: dachrit@gmail.com (D. Nilubol).

http://dx.doi.org/10.1016/j.meegid.2016.06.046
1567-1348/

genes are associated with viral envelope formation and release (Kocherhans et al., 2001). The N protein is the predominate antigen produced in coronavirus-infected cells and is thus a major viral target (Kocherhans et al., 2001). Two PEDV variants, classic and pandemic, are currently recognized (Sun et al., 2015) which typed as genogroups 1 and 2, respectively (Chen et al., 2014), and can be distinguished by the spike gene. The pandemic variant contains two insertions of 4 (^{56}GENQ^{59}) and 1 (^{140}N) amino acids at positions 55–60 and 140, respectively, and a deletion of 2 amino acids (^{160}DG^{161}) at positions 160–161 (Li et al., 2012).

PED has been observed in Thailand since 2007 (Temeeyasen et al., 2014) and has become endemic, causing sporadic outbreaks in Thai swine herds. The variant of PEDV responsible for the outbreaks and currently circulating in Thailand is a pandemic variant closely related to Chinese strains (Temeeyasen et al., 2014). However, in 2014, several herds in the eastern region of Thailand experienced recurring outbreaks of PED. The investigation revealed two genetically distinct PEDV variants based on complete S gene sequencing. The full-length genome sequences of these variants are reported here.

2. Materials and methods

2.1. Sample preparation and virus isolation

Two genetically distinct variants based on the S gene were isolated from one farm located in Chonburi, a province in the eastern region of Thailand. The investigated farm has a sow inventory of 2000 sows and is located in a high pig density area. The herd is farrow-to-finish; nursery and finishing facilities are located approximately 500 m from the sow herd. The sow herd has 4 breeding and gestating houses and 6 farrowing houses which operate all-in/all-out on a weekly basis. An evaporative cooling system is implemented in all buildings. Strict biosecurity was implemented in the herd. Gilts are introduced from an external source that has no history of PED outbreak.

The investigated herds experienced PED outbreaks twice within 4 months. In each outbreak, 2–3 intestinal samples were collected from 3- to 4-day-old pigs from the farm displaying clinical signs of vomiting and watery diarrhea associated with PED. Intestinal tissue samples of infected pigs were minced into small pieces and suspended in PBS. The suspension was separated by centrifugation at 10,000 $\times g$ for 10 min. The supernatants were filtered through 0.45-μm filters and stored at −80 °C before use. PEDV was isolated from PEDV supernatant samples using a continuous Vero cell line (ATCC, CCL-81).

The PEDV isolate CBR1/2014 (accession number KR610993) was collected during the first outbreak, and isolate EAS1/2014 (accession number KR610991) was collected in the second outbreak.

2.2. Reverse transcription polymerase chain reaction and sequence determination

Viral RNA was extracted from the infected Vero culture supernatant using the Nucleospin® viral RNA isolation kit (Macherey-Nagel Inc., Duren, Germany) in accordance with the manufacturer's instructions. Complementary DNA (cDNA) was synthesized from the extracted RNA using M-MuLV Reverse Transcriptase (BioLabs Inc., Ipswich, MA, USA). Twelve pairs of the oligonucleotide primers were used to amplify the different regions of PED (Vui et al., 2015). PCR amplification was performed using Platinum® *Taq* DNA polymerase High Fidelity (Invitrogen, CA, USA) according to the manufacturer's protocol. The PCR products were purified and cloned into pGEM-T® Easy Vector vectors for the subsequent transformation of *Escherichia coli* using a commercial kit (Promega, Madison, WI, USA). Bacterial colonies were randomly selected from each sample for plasmid purification using a Nucleospin® Plasmid kit (Macherey-Nagel, Duren, Germany) according to the manufacturer's instructions and the 5 selected colonies were grown in LB broth for 24 h. Purified plasmid was sequenced in both directions using an ABI Prism 3730XL sequencer. The 5′ and 3′ terminal regions were determined by using a kit for rapid amplification of 5′ and 3′ cDNA ends (5′ and 3′-RACE) (Clontech, Japan).

2.3. Sequence analysis

The full-length PEDV genome sequencing data were assembled. Nucleotide and deduced amino acid sequence alignments were analyzed using the CLUSTALW program (Thompson et al., 1994). The percentages of homology between the isolates at the nucleotide and amino acid levels were calculated. Phylogenetic analyses of the full-length nucleotide sequences and the complete S and ORF3 genes of the two PEDV isolates in this study along with 35 other PEDV isolate sequences (Supplementary Material 1) were separately performed using a Bayesian Markov chain Monte Carlo (BMCMC) method implemented in the program BEAST v1.7.4 (Drummond et al., 2012). The best-fit substitution model and among-site rate heterogeneity model for each sequence data set were determined using ModelGenerator v0.85 (Keane et al., 2006). A strict molecular clock model with a coalescent constant size tree prior and empirical base frequencies was used for all analyses. For each analysis, BEAST run was performed using 50 million generations with sampling of every 10,000 generations and the first 10% discarded as burn-in. For ORF3 analysis, a chain length of 10 million generations sampled every 1000 generations was applied. Tracer v1.5 (http://beast.bio.ed.ac.uk/Tracer) was used to confirm that the post-burn-in trees yielded an effective sample size (ESS) of >200 for all parameters. The resulting tree was viewed in FigTree v1.4.0 (http://tree.bio.ed.ac.uk/software/figtree/).

2.4. Sliding window analysis of sequence variation

To detect nucleotide variation sites, genome alignment between CBR1/2014 and EAS1/2014 isolates excluding UTR regions was performed using CLUSTALW (Thompson et al., 1994). A sliding window of 100 bp with a step size of 20 bp was used to estimate sequence diversity for the complete alignment. In each window, the percentage of sequence difference was plotted against the mid-point position of the window.

2.5. Calculation of antigenic index and hydrophilicity plots

Antigenicity and hydrophilicity analyses based on amino acid sequences of S protein between the 2 genetically distinct Thai PEDV variants were performed along with 4 other isolates including Virulent DR13, CV777, USA/Colorado and AH2012. These 4 isolates represent PEDV in groups G1 and G2 and subgroups G2-1 and G2-2. The Jameson-Wolf antigenic index (Jameson and Wolf, 1988) and Kyte-Doolittle hydrophilicity plots (Kyte and Doolittle, 1982) of these sequences were constructed using Protean in the DNASTAR Lasergene software package (DNASTAR, Inc., Madison, WI, USA).

2.6. Identifying signatures of natural selection in different gene regions

A genome scan for signals of natural selection was performed to determine the variability of selective pressure among different gene regions. Ratios of non-synonymous to synonymous nucleotide substitutions rates (Ka/Ks) were estimated for the entire alignment between the CBR1/2014 and EAS1/2014 gene sequences using the software package KaKs_Calculator 2.0 (Wang et al., 2010). A sliding window of 90 bp with a step size of 6 bp and the LWL method were used for all analyses.

3. Results

3.1. Virus and full-length genome sequences

Two PEDV variants, designated CBR1/2014 (accession number KR610993) and EAS1/2014 (accession number KR610991), were characterized. The full-length genome sequences of CBR1/2014 and EAS1/2014 are 28,039 and 28,033 nucleotides (nt) in length, respectively. The genome organization of the variants resembles that of all previously sequenced PEDV genomes, with gene order 5′-ORF1a/1b-S-ORF3-E-M-N-3′. The untranslated regions (UTRs) were identified at both ends (5′ UTR, nt 296 and 3′ UTR, nt 334). ORF1a/1b encodes non-structural proteins and was subdivided into 1a and 1b, with sizes of 12,309 nt (nt 297–12,605) and 8037 nt (nt 12,605–20,641), respectively. The S gene of EAS1/2014 is 6 nt shorter than that of CBR1/2014. The sizes of the genes encoding the other 3 structural proteins (E, M, and N) are identical between the two variants and are 231, 681, and 1326 nt, respectively. The accessory protein ORF3 is encoded by genes of similar sizes, 675 nt (CBR1/2014, nt 24,795–25,469; and EAS1/2014, nt 24,789–25,463) in the two variants.

3.2. Phylogenetic analyses

The nucleotide and deduced amino acid sequences of the full-length genome of the two genetically distinct Thai PEDV isolates were aligned, and a phylogenetic tree based on the full-length genome nucleotide sequences of the CBR1/2014 and EAS1/2014 isolates was constructed. The percentages of similarity between the isolates at the nucleotide and amino acid levels are displayed in Table 1. The phylogenetic analysis revealed two major groups, designated G1 and G2. The G2 group was further divided into two subgroups, 2-1 and 2-2 (Fig. 1).

Based on the full-length genome, Thai PEDV isolate CBR1/2014 belongs to G2-2, a sub-group inside G2 (Fig. 1). The isolate shares high genetic similarity with the PEDV variant from China (GD-1, GD-A, GHGD-01, CH/GDGZ, ZJCZ4, LC, and AJ1102) in sub-group G2-2. The similarities with other isolates in the subgroup G2-2 at the nucleotide and amino acid levels range from 98.4% to 98.8% and 97.3% to 98.1%, respectively (Table 1). By contrast, isolate EAS1/2014 is included in group G1 (Fig. 1). The isolate is closely related to the CV777, LZC and SM98 isolates, with 99.3–99.5% and 98.8–99.1% similarities at the nucleotide and amino acid levels. The isolate EAS1/2014 is genetically distinct from the prevalent PEDV variant in Thailand, suggesting that EAS1/2014 is a novel isolate in the country.

3.3. Phylogenetic analyses and sequencing data of complete S and ORF3 genes

The S gene of CBR1/2014 is 4158 nucleotides in length and encodes 1386 amino acids. The S gene of EAS1/2014 is 4152 nucleotides in length and encodes 1384 amino acids. The S gene of EAS1/2014 is 6 nt shorter than that of CBR1/2014. Compared to CBR1/2014, EAS1/2014 has 2 insertions of 4 (^{56}GENQ^{59}) and 1 (^{140}N) amino acids at positions 56–59 and 140 and 2 deletions of 2 (^{160}DG^{161}) and 1 (^{1199}Y) amino acids at positions 160–161 and 1199 (Fig. 2). The insertions and deletions are located in the hypervariable domain in the N-terminus of the S1 region. Phylogenetic analysis based on S gene nucleotide sequences demonstrated the further evolution of PEDV into 2 separate groups, including G1 and G2 (Fig. 1). G1 comprises five strains from China (CH/S, attenuated vaccine, SD-M, JS2008, and LZC), three strains from Korea (attenuated strains DR13, virulent DR13, and SM98), one European strain (CV777), and the EAS1/2014 strain. G2 further evolved into three subgroups (2-1, 2-2, and 2-3). CBR1/2014 belongs to subgroup 2-2, which includes nine Chinese strains isolated in China in 2011–2012. Previously reported Thai isolates from 2008–2013 belong to subgroup 2-3 (Fig. 1).

ORF3, which is located between the S and E genes, is 675 nucleotides in length and encodes a protein of 225 amino acids. Based on phylogenetic analysis of ORF3, all PEDV isolates were divided into two groups (Fig. 1). CBR1/2014 belongs to group 1, which consists of seven strains from China, including ZJCZ4, GD-A, GD-1, CH/GDGZ, CHGD-01, AJ1102, and LC. The attenuated vaccine in group 2 has a unique 17-amino-acid deletion at positions 82–98 (Fig. 3). Group 2 has two subgroups (2-1

Table 1
Comparison of the nucleotide and amino acid sequence identity (%) of full-length of CBR1/2014 and EAS1/2014 with those of PEDV reference strains.

Virus strain	*Group	Subgroup	Countries	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33
1. USA-Iowa/18984	2	2-1	USA		99.9	99.9	99.9	99.8	99.8	99.8	99.5	99.2	99.3	99.2	99.0	99.0	98.9	98.9	98.2	98.2	98.2	98.0	98.4	98.0	98.7	98.7	96.6	96.4	96.6	96.8	97.1	97.2	97.2	97.3	97.7	97.3
2. USA-Indiana	2	2-1	USA	99.9		99.9	99.9	99.8	99.8	99.8	99.5	99.2	99.3	99.2	99.0	99.0	98.9	98.9	98.2	98.2	98.2	97.9	98.4	98.0	98.7	98.7	96.6	96.4	96.5	96.8	97.1	97.2	97.2	97.3	97.7	97.3
3. KUIDL-001	2	2-1	Korea	99.9	99.9		99.9	99.8	99.8	99.8	99.6	99.3	99.3	99.2	99.0	99.0	99.0	98.9	98.2	98.3	98.3	98.0	98.4	98.0	98.7	98.7	96.6	96.4	96.6	96.8	97.1	97.2	97.2	97.3	97.7	97.4
4. KNU-1305	2	2-1	Korea	99.9	99.9	99.9		99.8	99.8	99.8	99.5	99.2	99.3	99.2	99.0	99.0	98.9	98.9	98.2	98.2	98.3	97.9	98.4	98.0	98.7	98.7	96.6	96.4	96.6	96.8	97.1	97.2	97.2	97.3	97.7	97.3
5. USA-Colorado	2	2-1	USA	99.8	99.8	99.8	99.8		99.9	99.9	99.5	99.3	99.4	99.3	99.1	99.0	98.9	98.8	98.2	98.2	98.2	97.9	98.4	98.0	98.7	98.7	96.6	96.4	96.6	96.8	97.1	97.2	97.2	97.3	97.7	97.3
6. IA1	2	2-1	USA	99.7	99.7	99.8	99.8	99.9		99.9	99.4	99.3	99.3	99.2	99.1	99.0	98.9	98.8	98.2	98.2	98.2	97.9	98.3	98.0	98.7	98.7	96.6	96.4	96.5	96.8	97.0	97.1	97.1	97.3	97.7	97.3
7. 13-019349	2	2-1	USA	99.7	99.7	99.8	99.8	99.9	99.8		99.4	99.3	99.3	99.2	99.1	98.9	98.9	98.8	98.2	98.2	98.2	97.9	98.3	98.0	98.7	98.7	96.6	96.4	96.5	96.8	97.0	97.1	97.2	97.3	97.7	97.3
8. AH2012	2	2-1	China	99.3	99.3	99.4	99.4	99.2	99.2	99.2		99.4	99.4	99.3	98.7	99.0	99.0	98.9	98.2	98.2	98.3	98.0	98.4	98.1	98.8	98.8	96.7	96.4	96.6	96.8	97.1	97.2	97.2	97.3	97.8	97.4
9. GD-B	2	2-1	China	98.8	98.8	98.9	98.9	98.9	98.8	98.9	99.0		99.9	99.6	98.7	99.1	99.0	98.9	98.3	98.3	98.3	98.0	98.5	98.1	98.9	98.9	98.9	96.7	96.4	96.6	97.1	97.2	97.2	97.4	97.8	97.4
10. JS-HZ2012	2	2-1	China	99.0	99.0	99.0	99.0	99.0	99.0	99.0	99.1	99.8		99.7	98.8	99.1	99.1	98.9	98.3	98.4	98.4	98.1	98.5	98.1	98.9	98.9	96.7	96.5	96.7	96.9	97.2	97.3	97.3	97.5	97.8	97.4
11. BJ-2011-1	2	2-1	China	98.7	98.7	98.8	98.8	98.8	98.7	98.7	98.7	99.3	99.4		98.7	99.0	98.9	98.8	98.3	98.3	98.4	98.1	98.5	98.1	98.9	98.9	96.8	96.6	96.7	97.0	97.1	97.2	97.3	97.4	97.8	97.3
12. OH851	2	2-1	USA	98.2	98.2	98.3	98.3	98.4	98.3	98.3	97.7	97.7	97.8	97.6		98.8	98.4	98.4	97.6	97.7	97.7	97.4	97.8	97.6	98.1	98.1	96.9	96.6	96.8	97.1	97.4	97.5	97.5	97.6	97.9	97.5
13. CH/ZMDZY/11	2	2-1	China	98.2	98.2	98.3	98.3	98.2	98.2	98.2	98.1	98.1	98.2	98.0	98.1		98.8	98.7	98.1	98.2	98.2	97.9	98.1	98.0	98.7	98.7	96.6	96.4	96.6	96.8	97.1	97.2	97.2	97.2	97.6	97.2
14. CH/FJND-3	2	2-1	China	98.1	98.1	98.1	98.1	98.0	98.0	98.0	98.1	98.0	98.2	97.8	96.9	97.5		98.8	98.0	98.1	98.1	97.8	98.1	97.9	98.7	98.7	96.6	96.4	96.6	96.8	97.2	97.3	97.3	97.4	97.7	97.4
15. CH/FJZZ-9	2	2-1	China	97.9	97.9	98.0	98.0	97.9	97.9	97.9	98.0	97.9	98.0	98.0	97.7	97.0	97.4		98.1	98.1	98.1	97.8	98.2	97.9	98.7	98.7	96.6	96.3	96.5	96.7	97.1	97.2	97.2	97.3	97.5	97.1
16. GD-1	2	2-2	China	97.3	97.3	97.4	97.4	97.3	97.3	97.3	97.3	97.5	97.6	97.6	96.2	97.1	97.0	97.0		99.9	99.6	99.4	99.2	98.8	98.9	98.9	96.5	96.3	96.5	96.7	97.0	97.1	97.1	97.2	97.6	97.1
17. GD-A	2	2-2	China	97.4	97.4	97.4	97.4	97.4	97.3	97.3	97.3	97.5	97.6	97.6	96.3	97.1	97.0	97.0	99.8		99.6	99.4	99.2	98.8	98.9	98.9	96.5	96.3	96.5	96.7	97.0	97.1	97.1	97.2	97.6	97.2
18. CHGD-01	2	2-2	China	97.3	97.3	97.4	97.4	97.3	97.3	97.3	97.3	97.5	97.6	97.6	96.2	97.1	97.0	97.0	99.4	99.4		99.3	99.0	98.8	99.0	99.0	96.6	96.3	96.5	96.7	97.0	97.1	97.1	97.2	97.6	97.2
19. CHGDGZ	2	2-2	China	96.9	96.9	97.0	96.9	96.9	96.8	96.9	96.9	97.1	97.2	97.1	95.8	96.6	96.5	96.5	99.0	99.0	98.8		98.9	98.6	98.6	98.6	96.3	96.1	96.3	96.5	96.8	96.9	96.9	97.0	97.3	96.9
20. ZJCZ4	2	2-2	China	97.8	97.8	97.8	97.8	97.8	97.7	97.8	97.8	98.1	98.2	98.2	96.6	97.1	97.1	97.4	98.4	98.4	98.2	97.8		98.4	98.6	98.6	96.6	96.4	96.6	96.8	97.1	97.2	97.2	97.3	97.6	97.3
21. CBR1	2	2-2	Thailand	96.6	96.6	96.7	96.7	96.6	96.6	96.6	96.7	96.8	96.9	96.9	95.8	96.6	96.4	96.4	98.1	98.1	98.0	97.8	97.3		98.6	98.6	96.4	96.2	96.3	96.5	96.8	96.9	97.0	97.0	97.3	97.0
22. LC	2	2-2	China	97.7	97.7	97.7	97.7	97.7	97.6	97.6	97.7	97.8	97.9	97.9	96.5	97.4	97.4	97.4	98.7	98.7	98.8	98.0	98.0	97.8		99.9	96.7	96.4	96.6	96.8	97.1	97.2	97.3	97.4	97.7	97.3
23. AJ1102	2	2-2	China	97.7	97.7	97.8	97.7	97.7	97.6	97.6	97.7	97.8	97.9	97.9	96.5	97.4	97.5	97.4	98.7	98.7	98.8	98.0	98.0	97.8	99.9		96.6	96.4	96.6	96.8	97.1	97.2	97.3	97.4	97.7	97.3
24. SM98	1	–	Korea	94.6	94.6	94.6	94.6	94.6	95.0	94.6	94.6	94.6	94.7	94.8	94.9	94.5	94.5	94.5	94.3	94.3	94.4	93.9	94.6	94.0	94.6	94.5		99.4	99.5	99.7	97.7	97.8	97.8	97.7	97.4	97.2
25. EAS1	1	–	Thailand	94.2	94.2	94.2	94.2	94.2	94.1	94.2	94.2	94.2	94.3	94.4	94.5	94.1	94.1	94.0	93.9	93.9	94.0	93.5	94.2	93.6	94.1	94.1	99.0		99.3	99.5	97.5	97.6	97.6	97.5	97.2	97.0
26. LZC	1	–	China	94.5	94.5	94.6	94.6	94.5	94.5	94.5	94.5	94.6	94.7	94.8	94.8	94.4	94.5	94.4	94.2	94.3	94.3	93.8	94.6	93.9	94.5	94.5	99.2	98.8		99.7	97.7	97.7	97.8	97.7	97.4	97.1
27. CV777	1	–	Belgium	94.9	94.9	94.9	94.9	94.9	94.8	94.9	94.9	94.9	95.0	95.1	95.2	94.8	94.8	94.8	94.6	94.6	94.6	94.2	94.9	94.2	94.8	94.8	99.6	99.1	99.6		97.9	98.0	98.0	97.9	97.7	97.4
28. Attenuatedvaccine	1	–	China	95.2	95.2	95.2	95.2	95.2	95.1	95.1	95.2	95.2	95.3	95.2	95.6	95.1	95.3	95.2	95.0	95.0	95.0	94.6	95.3	94.6	95.2	95.2	96.1	95.7	96.0	96.4		99.8	99.8	99.6	97.9	97.6
29. SD-M	1	–	China	95.3	95.3	95.4	95.4	95.3	95.3	95.3	95.3	95.4	95.5	95.4	95.8	95.2	95.4	95.3	95.2	95.2	95.2	94.8	95.5	94.8	95.4	95.4	96.2	95.8	96.1	96.5	99.6		99.9	99.7	98.0	97.7
30. Attenuated DR13	1	–	Korea	95.4	95.4	95.5	95.5	95.4	95.4	95.4	95.5	95.5	95.5	95.8	95.3	95.5	95.4	95.2	95.3	95.3	94.8	95.5	94.9	95.4	95.4	96.2	95.8	96.2	96.5	99.6	99.8		99.7	98.0	97.7
31. JS2008	1	–	China	95.7	95.7	95.8	95.7	95.7	95.7	95.7	95.9	95.9	95.8	96.0	95.4	95.8	95.6	95.5	95.5	95.5	95.1	95.9	95.0	95.6	95.7	96.1	95.7	96.0	96.4	99.1	93.3	99.4		98.0	97.8
32. Virulent DR13	1	–	Korea	96.5	96.5	96.6	96.5	96.5	96.5	96.5	96.5	96.7	96.8	96.7	96.5	96.1	96.4	95.9	96.2	96.3	96.3	95.9	96.3	95.6	96.4	96.4	95.7	95.3	95.7	96.0	96.2	96.4	96.5	96.5		98.0
33. CH/S	1	–	China	95.7	95.7	95.7	95.7	95.7	95.6	95.6	95.7	95.8	95.9	95.6	95.8	95.4	95.6	95.1	95.3	95.3	95.3	94.9	95.6	94.8	95.5	95.5	95.2	94.8	95.1	95.5	95.9	96.1	96.1	96.1	96.7	

*1 = classical variant; 2 = pandemic variant

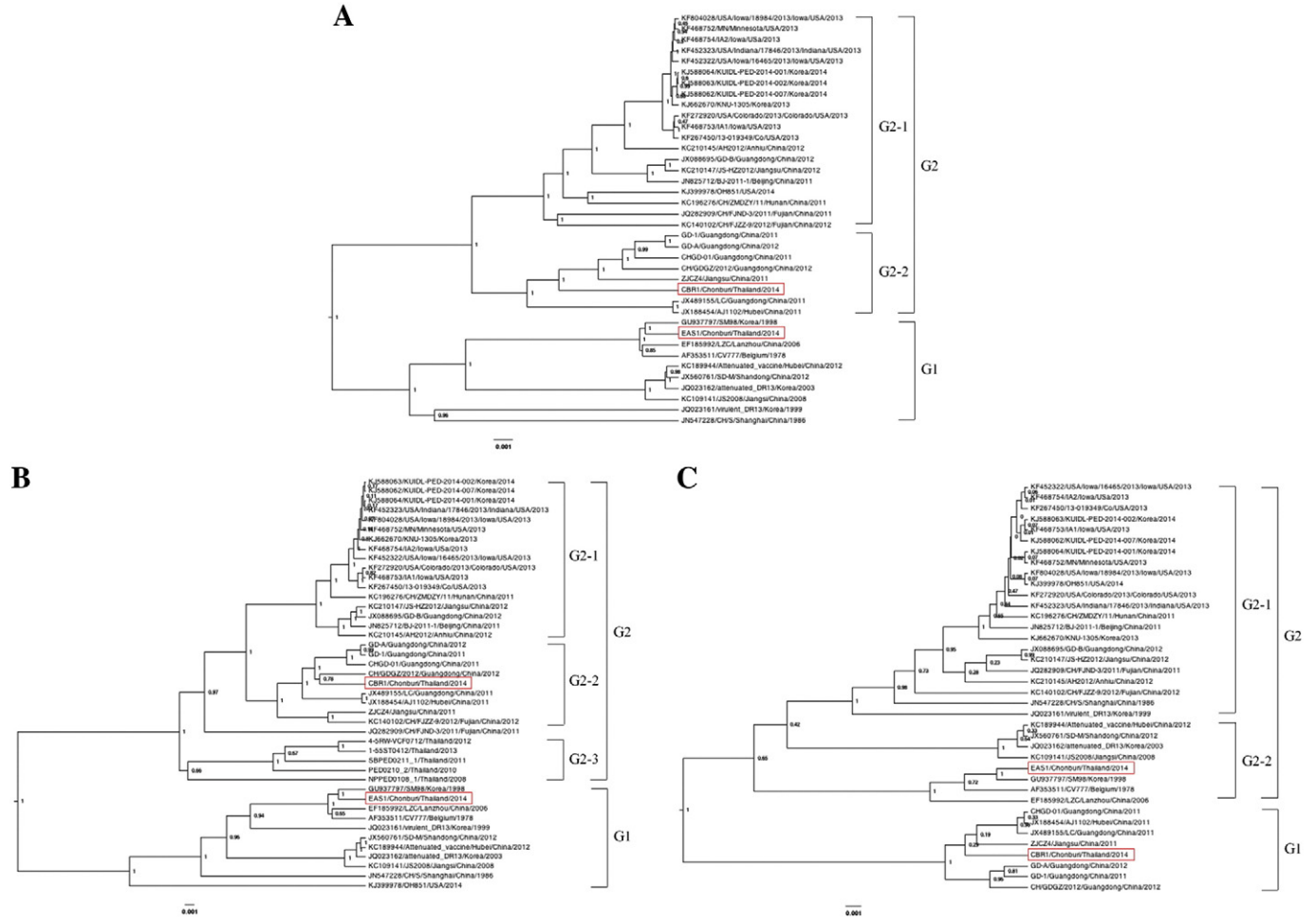

Fig. 1. Phylogenetic analysis of porcine epidemic diarrhea virus (PEDV) based on nucleotide sequences of full-length genome (A), spike (B), and ORF3 (C) were separately performed using a Bayesian Markov chain Monte Carlo (BMCMC) method. Red boxes represent CBR1/2014 and EAS1/2014 isolates, two genetically distinct variant of PEDV isolated in Thailand.

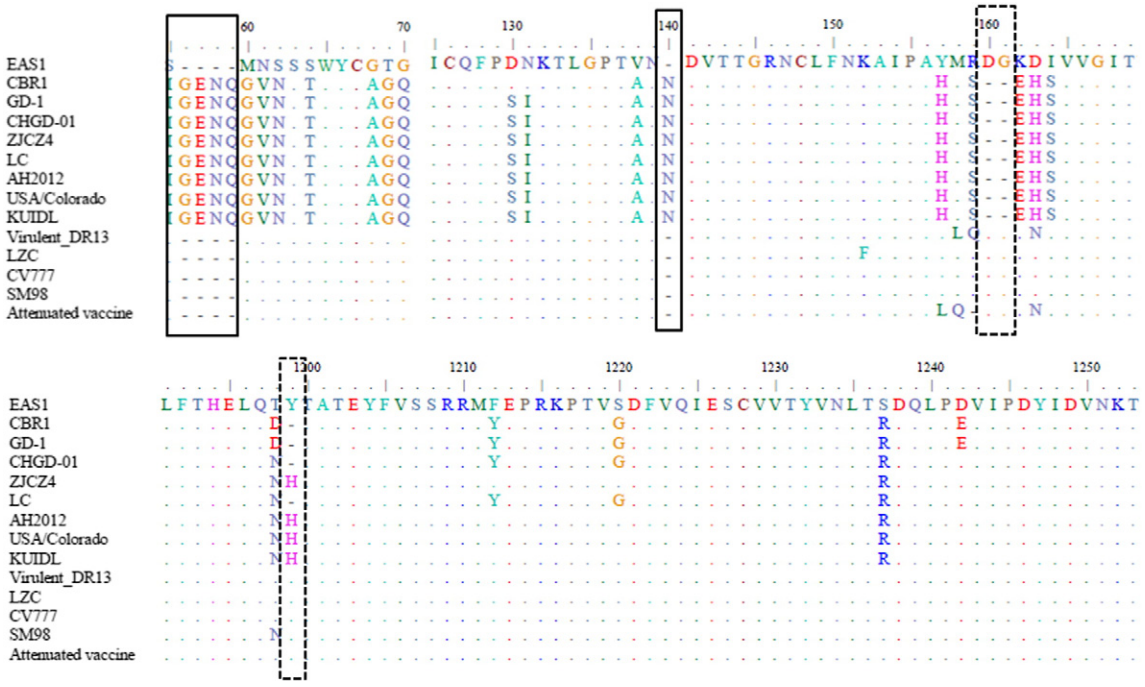

Fig. 2. Alignment of amino acid sequence of S proteins of CBR1/2014, EAS1/2014, and reference strains. The dashes (-) indicate the deleted sequences. Insertions and deletions in PEDV isolates are boxed.

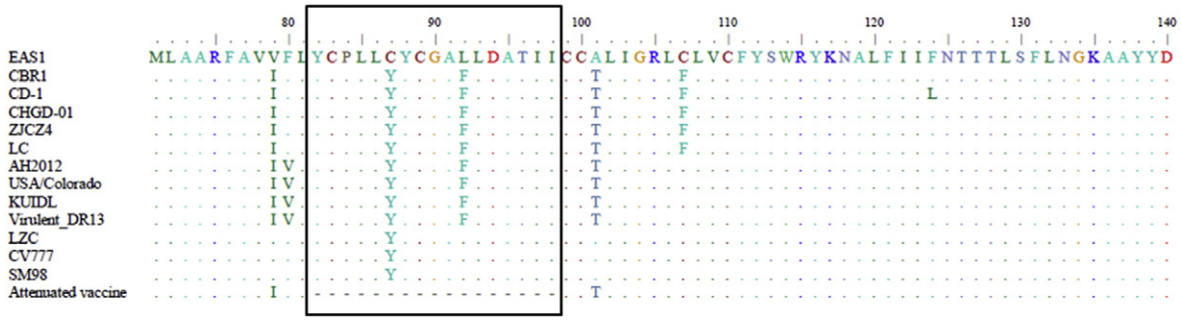

Fig. 3. Alignment of amino acid of the ORF3 proteins of CBR1/2014, EAS1/2014, and reference isolates. The dashes (-) indicate the deleted sequences of attenuated vaccine at position 82–98. Box indicate 17 amino acids deletion of attenuated vaccine compared with field PEDV strains.

and 2-2). EAS1/2014 forms the second subgroup (2-2) with SM98, LZC, and CV777. ORF3 of CBR1/2014 shares 95.4% and 93.3% nucleotide and amino acid identity with EAS1/2014.

3.4. Variation analysis of full-length genome sequences of CBR1/2014 and EAS1/2014

To identify the genome regions displaying sequence variation between the classic and pandemic PEDV variant groups, sequence analysis was performed based on a pairwise genome sequence alignment of representative strains from two groups, CBR1/2014 and EAS1/2014. Overall, a total of 1048 nucleotide mutation positions were located in the alignment. Of these, 1022 positions were substitutions, and 26 positions were insertions/deletions.

Sliding window analysis indicated three regions exhibiting high sequence variation between the two variants (Fig. 4). These regions included parts of three genes, the S gene, ORF3, and N gene, the top three genes with the greatest sequence difference at the nucleotide level (7.67%, 4.59%, and 5.20%, respectively). The total number of substitution positions in the S gene, ORF3, and N gene were 319, 31, and 69, which resulted in 129, 15, and 25 amino acid substitutions, respectively, including 9 insertions/deletions in the S gene. The region exhibiting the highest percentage of sequence difference between the two isolates (Fig. 4) and thus potentially discriminating the pandemic variants from the classic variants was located approximately in the 120- to 260-bp region of the S gene and involved 11 amino acid changes and 4 insertions/deletions.

3.5. Antigenic index and hydrophilicity analyses of the CBR1/2014 and EAS1/2014 isolates

The S gene regions exhibited high sequence variation between the two genetically distinct variants of PEDV, CBR1/2014 and EAS1/2014. To investigate the differences in antigenicity among the PEDV strains, the antigenic index and hydrophilicity plots of the S protein (at amino acid positions 1 to 300 and 500 to 800) of CBR1/2014 and EAS1/2014 were compared with those of classic (virulent DR13 and CV777) and pandemic variant reference strains (AH2012 and USA/Colorado) (Fig. 5). The major differences in the antigenic index and hydrophilicity values were located at five regions of the amino acid sequence: positions 55 to 74, 120 to 179, 557 to 604, 705 to 725, and 769 to 785. These regions exhibited insertions and deletions leading to separation between the classic group and pandemic variant group (Fig. 5).

3.6. Identifying signatures of natural selection in different gene regions of the CBR1/2014 and EAS1/2014

Molecular-level positive selection can be detected in protein-coding genes by pairwise comparisons of the ratios of non-synonymous to synonymous nucleotide substitutions rates (Ka/Ks). A ratio of Ka/Ks > 1 or <1 indicates positive and purifying selection, respectively, and a value near 1 indicates neutral evolution. The signatures of natural selection in different gene regions of CBR1/2014 and EAS1/2014 isolates were detected. The results are displayed in Fig. 6. The analyses revealed that ORF1a was positively selected, with a Ka/Ks ratio of 1.4 and 2.2 at sliding windows of 397–399 (nt 2371–2478) and 501–502 (nt 2995–3090),

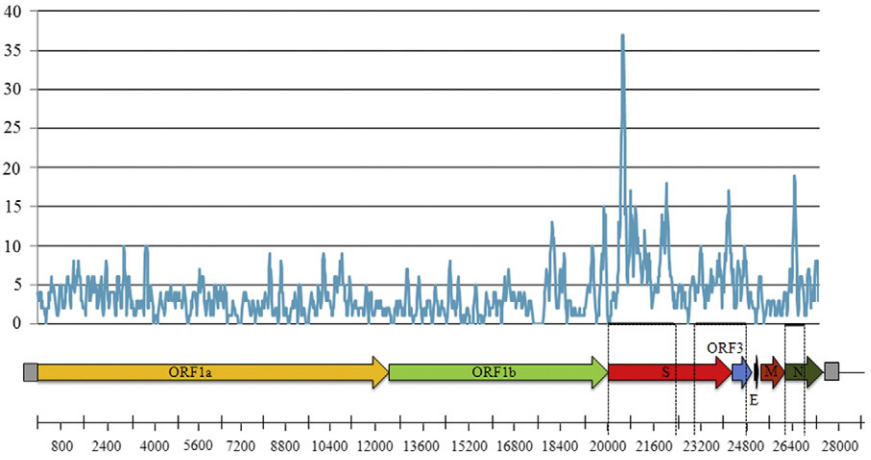

Fig. 4. Sliding window analysis of genome sequence variation between CBR1/2014 and EAS1/2014 isolates. The graph shows the percentage of sequence difference (window size = 100 bp, step size = 20 bp) in the alignment. Dashed and thick black line represents the positions of high mutation regions including S, ORF3, and N genes. The positions and sizes of PEDV genes correspond to the scale bar.

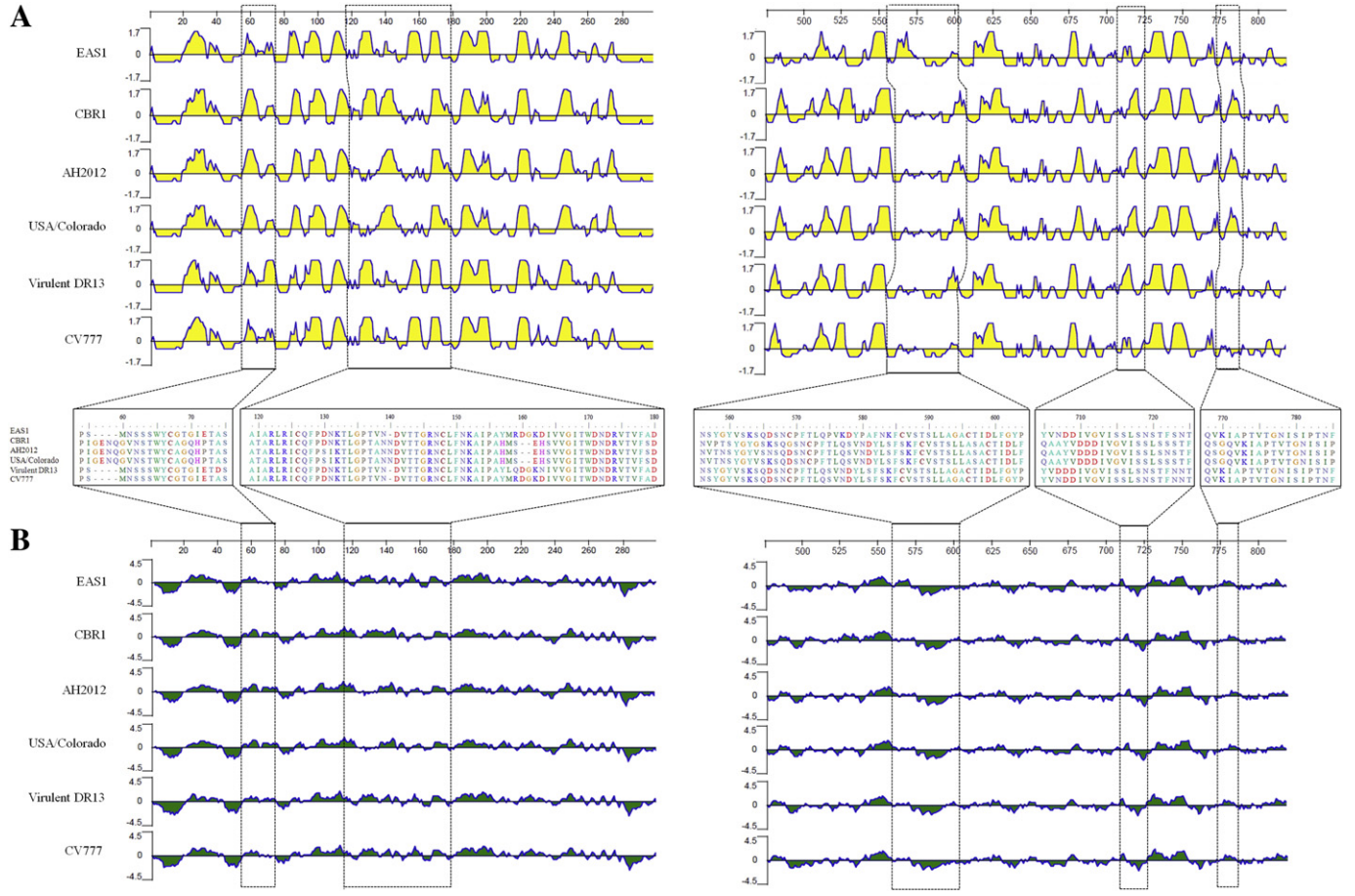

Fig. 5. Antigenic index (A) and hydrophilicity plots (B) based on amino acid sequences of the divergent region of S protein fragments (amino acid position 1 to 300 and 500 to 800). The dash lines indicate the regions showing the differences among EAS1/2014, CBR1/2014, and reference strains, as classical and pandemic variant strains.

respectively (Fig. 6A). The ORF1b gene at various positions was strongly positively selected with Ka/Ks ratios of 4.5, 3.1, and 2.1 at sliding windows of 110–112 (nt 649–750), 113 (nt 667–756), and 114 (nt 673–762), respectively (Fig. 6B). The full-length coding region of the S gene indicated that this gene was nearly neutrally selected, with a Ka/Ks ratio of 0.9–1.1 (Fig. 6C). The ORF3 gene was neutrally selected (Ka/Ks = 1.1) at sliding window 27 (nt 151–240) (Fig. 6D). Other E and M genes were indicated as under purifying selection with a Ka/Ks ration of <1 (Fig. 6E and F, respectively). The N gene was positively selected with a Ka/Ks ratio of 1.4 at the sliding window 191–193 (nt 1135–1236).

4. Discussion

The full-length genomes of two genetically distinct variants of PEDV designated CBR1/2014 and EAS1/2014 were characterized. The two PEDV variants were isolated from a herd that experienced two episodes of PED outbreak 4 months apart in 2014. The isolates responsible for the first and second PED outbreaks were CBR1/2014 and EAS1/2014, respectively. To determine the genetic relationships of the 2 PEDV variants, phylogenetic analyses of the whole genome, and spike and ORF genes were separately constructed, and the results demonstrated that the two PEDV variants were groups in a different cluster in all 3 trees constructed. A previous report demonstrated that PEDV isolated in Thailand in 2007–2013 was grouped in cluster G2, the pandemic group (Temeeyasen et al., 2014). CBR1/2014 clustered in cluster G2, suggesting that the isolate is an endemic Thai PEDV variant. By contrast, EAS1/2014, the PEDV isolate detected in the second outbreak, clustered in G1, the classical variant group, and is more genetically related to LZC, SM98 and CV777. These findings suggest that EAS1/2014 is a novel isolate recently isolated in Thailand and may have been externally introduced into the herd to cause the second outbreak, rather than the product of continuing evolution of existing PEDV in the herd. The antigenic distinction and route of introduction remain to be further elucidated.

Protection against PEDV has been reported to last for at least 6 months (Goede et al., 2015), but in the studied herd, the second PEDV outbreak occurred 4 months after the first outbreak. The short time interval between the first and second outbreaks may indicate a low level of protective immunity against PEDV in the herd and/or the introduction of antigenically distinct PEDV isolates. We therefore explored the antigenic differences between the CBR1/2014 and EAS1/2014 isolates, particularly the S and ORF3 genes, which encode neutralizing epitopes and attenuation markers, respectively. The results of sliding window analysis demonstrated that the spike and ORF3 genes exhibit the highest sequence difference between both isolates, particularly the first 700 amino acids of the spike gene.

The genetic difference between these two variants in Thailand was located in the spike gene regions. CBR1/2014 contains 2 amino acid insertions, ^{56}GENQ^{59} and ^{140}N, and a 1 amino acid deletion at ^{160}DG^{161}. However, EAS1/2014, a novel isolate, contains no insertions or deletions at these sites. In addition, differences between these two variants were identified at amino acid position 504–643 and 769–776, which encode neutralizing epitopes (Chang et al., 2002; Cruz et al., 2008; Sun et al., 2008). Therefore, differences in the neutralizing epitope region may explain the shorter duration of protection compared to the 6-month homologous protection reported previously (Goede et al., 2015). These 2 variants may provide partial protection.

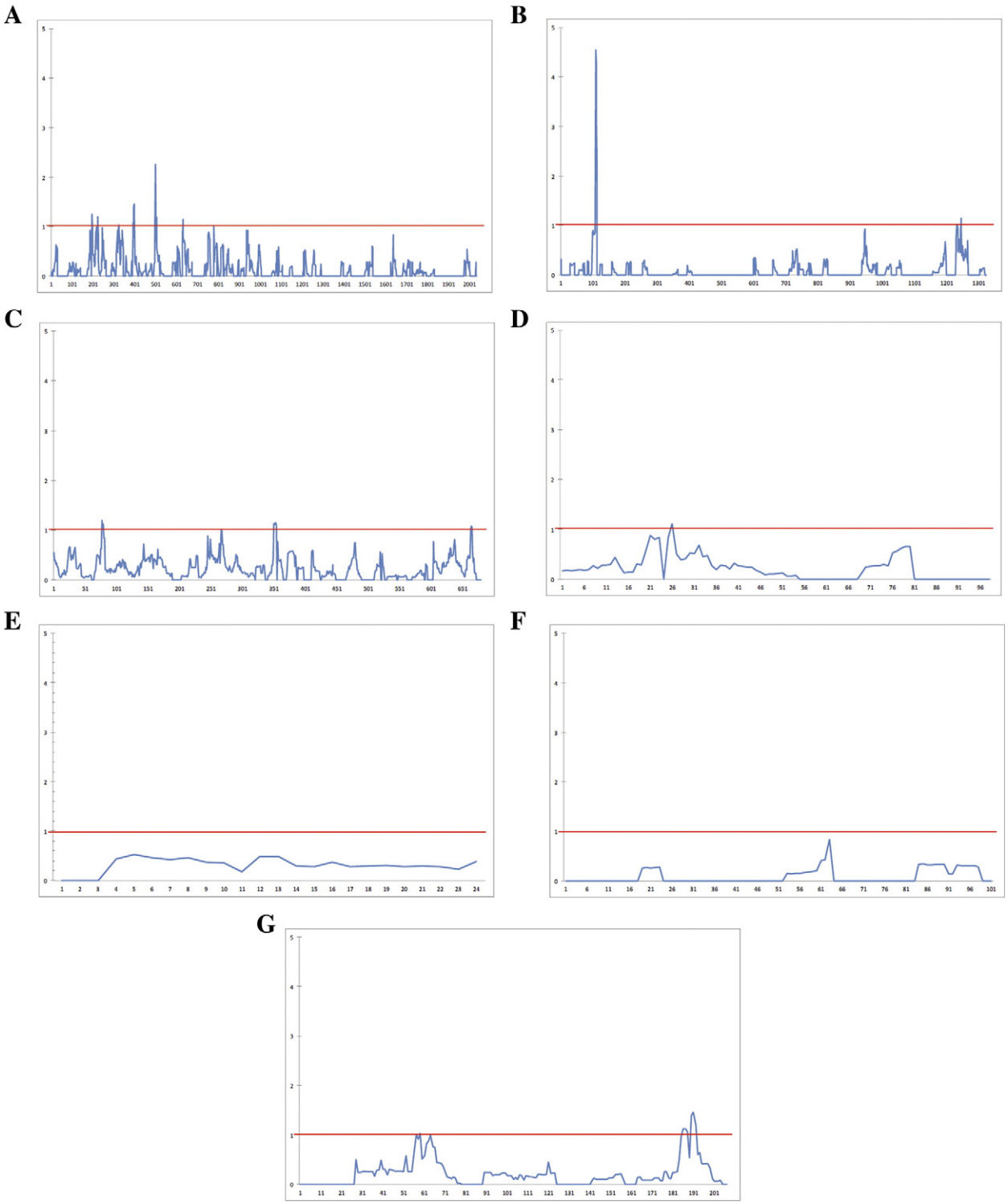

Fig. 6. Identifying signatures of natural selection in different gene regions including ORF1a (A), ORF1b (B), S (C), ORF3 (D), E (E), M (F), and N (G). Ratios of non-synonymous to synonymous nucleotide substitutions rates (Ka/Ks) were estimated for the entire alignment between CBR1/2014 and EAS1/2014 gene sequences. A ratio of Ka/Ks > 1 and <1 indicates positive and purify selection, and a near 1 ratio indicates neutral evolution. A sliding window of 90 bp with a step size of 6 bp and the LWL method were used for all analyses.

In addition to the S gene, the ORF3 gene was further analyzed. ORF3 is an important determinant of virulence in PEDV (Park et al., 2008; Park et al., 2007b) and can be used to differentiate between attenuated vaccine and field isolates. Vaccine-derived isolates have a unique deletion of 17 amino acids at positions 82–99 (Park et al., 2008; Park et al., 2007b). However, the novel isolate EAS1/2014 does not contain the 17-amino-acid deletion in the ORF3 gene that serves as a marker of the attenuated vaccine strain (Park et al., 2008). These findings suggest that in addition to the speculated introduction of the classical variant via a modified live vaccine, there could be other sources of the classical variant, including pig trafficking and the importation of new genetic stock.

To determine if EAS1/2014 is an exotic introduction or the product of continuous evolution from the endemic variant, we identified signatures of natural selection in different gene regions involving in the evolutionary process of PEDV. The ORF1 and N genes were positively selected, suggesting that both genes evolve in the population. In ORF1,

the regions between nt 2995–3090 are the hypervariable region between the nsp2 and nsp3 genes and are reported to be the putative region for a recombination event (Huang et al., 2013). The positive selection in the N gene is consistent with a recent finding suggesting that the N gene also contains a hypervariable region. Interestingly, the S and ORF3 genes were neutrally selected even though both genes have hypervariable regions. This neutral evolution indicates that the evolution was due to mutation and antigenic drift, not genetic selection. The neutral selection in the S and ORF3 genes suggests that the virus was externally introduced rather than continuously evolved in the population, resulting in the emergence of a novel viral isolate.

The classic isolates are trypsin dependent. The propagation of isolates in this groups, for example CV777, requires the presence of trypsin in Vero cells (Wicht et al., 2014). However, DR-13, a cell culture-adapted variant of PEDV, is trypsin independent. The isolate has glycine at amino acid position 890, which is the determinant for trypsin independence. The EAS1/2014 isolate reported in the study is trypsin dependent. Our attempt to propagate the isolate in the absence of trypsin was unsuccessful. In addition, the sequencing data demonstrated that the EAS1/2014 isolate has an arginine at position 890 (R^{890}) of S protein, further suggesting that the isolate is trypsin dependent.

In conclusion, this study characterized two genetically distinct variants of PEDV from the recurring outbreak in current herds. One variant corresponded to the pandemic variant endemic in Thailand since 2007, whereas the other was a classic variant similar to the isolate used for vaccine manufacturing. However, the route of introduction of the classic variant of PEDV into Thailand remains unclear. Further investigation of the genetic diversity and molecular epidemiology of these two variants is needed.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.meegid.2016.06.046.

Acknowledgements

We are grateful to the Rachadapisaeksompote Fund Post-Doctoral Support and funding provided by Special Task Force for Activating Research (STAR), swine viral evolution and vaccine research (SVEVR), Chulalongkorn University. This study was supported by the National Research Council of Thailand (2557-82) and the Ratchadaphiseksomphot Endowment Fund 2013 of Chulalongkorn University (CU-56-527-HR).

References

Bridgen, A., Kocherhans, R., Tobler, K., Carvajal, A., Ackermann, M., 1998. Further analysis of the genome of porcine epidemic diarrhoea virus. Adv. Exp. Med. Biol. 440, 781–786.

Chang, S.H., Bae, J.L., Kang, T.J., Kim, J., Chung, G.H., Lim, C.W., Laude, H., Yang, M.S., Jang, Y.S., 2002. Identification of the epitope region capable of inducing neutralizing antibodies against the porcine epidemic diarrhea virus. Mol. Cell 14, 295–299.

Chasey, D., Cartwright, S., 1978. Virus-like particles associated with porcine epidemic diarrhoea. Res. Vet. Sci. 25, 255.

Chen, Q., Li, G., Stasko, J., Thomas, J.T., Stensland, W.R., Pillatzki, A.E., Gauger, P.C., Schwartz, K.J., Madson, D., Yoon, K.J., Stevenson, G.W., Burrough, E.R., Harmon, K.M., Main, R.G., 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.

Cruz, D.J., Kim, C.J., Shin, H.J., 2006. Phage-displayed peptides having antigenic similarities with porcine epidemic diarrhea virus (PEDV) neutralizing epitopes. Virology 354, 28–34.

Cruz, D.J., Kim, C.J., Shin, H.J., 2008. The GPRLQPY motif located at the carboxy-terminal of the spike protein induces antibodies that neutralize porcine epidemic diarrhea virus. Virus Res. 132, 192–196.

Drummond, A.J., Suchard, M.A., Xie, D., Rambaut, A., 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol. Biol. Evol. 29, 1969–1973.

Duarte, M., Laude, H., 1994. Sequence of the spike protein of the porcine epidemic diarrhoea virus. J. Gen. Virol. 75 (Pt 5), 1195–1200.

Duarte, M., Gelfi, J., Lambert, P., Rasschaert, D., Laude, H., 1993. Genome organization of porcine epidemic diarrhoea virus. Adv. Exp. Med. Biol. 342, 55–60.

Goede, D., Murtaugh, M.P., Nerem, J., Yeske, P., Rossow, K., Morrison, R., 2015. Previous infection of sows with a "mild" strain of porcine epidemic diarrhea virus confers protection against infection with a "severe" strain. Vet. Microbiol. 176, 161–164.

Huang, Y.W., Dickerman, A.W., Pineyro, P., Li, L., Fang, L., Kiehne, R., Opriessnig, T., Meng, X.J., 2013. Origin, evolution, and genotyping of emergent porcine epidemic diarrhea virus strains in the United States. mBio 4, e00737-13.

Jameson, B.A., Wolf, H., 1988. The antigenic index: a novel algorithm for predicting antigenic determinants. Comput. Appl. Biosci. 4, 181–186.

Keane, T.M., Creevey, C.J., Pentony, M.M., Naughton, T.J., McLnerney, J.O., 2006. Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. BMC Evol. Biol. 6, 29.

Kocherhans, R., Bridgen, A., Ackermann, M., Tobler, K., 2001. Completion of the porcine epidemic diarrhoea coronavirus (PEDV) genome sequence. Virus Genes 23, 137–144.

Kyte, J., Doolittle, R.F., 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. 157, 105–132.

Li, W., Li, H., Liu, Y., Pan, Y., Deng, F., Song, Y., Tang, X., He, Q., 2012. New variants of porcine epidemic diarrhea virus, China, 2011. Emerg. Infect. Dis. 18, 1350–1353.

Park, S.J., Moon, H.J., Yang, J.S., Lee, C.S., Song, D.S., Kang, B.K., Park, B.K., 2007a. Sequence analysis of the partial spike glycoprotein gene of porcine epidemic diarrhea viruses isolated in Korea. Virus Genes 35, 321–332.

Park, S.J., Song, D.S., Ha, G.W., Park, B.K., 2007b. Cloning and further sequence analysis of the spike gene of attenuated porcine epidemic diarrhea virus DR13. Virus Genes 35, 55–64.

Park, S.J., Moon, H.J., Luo, Y., Kim, H.K., Kim, E.M., Yang, J.S., Song, D.S., Kang, B.K., Lee, C.S., Park, B.K., 2008. Cloning and further sequence analysis of the ORF3 gene of wild- and attenuated-type porcine epidemic diarrhea viruses. Virus Genes 36, 95–104.

Pensaert, M., De Bouck, P., 1978. A new coronavirus-like particle associated with diarrhea in swine. Arch. Virol. 58, 243–247.

Song, D., Park, B., 2012. Porcine epidemic diarrhoea virus: a comprehensive review of molecular epidemiology, diagnosis, and vaccines. Virus Genes 44, 167–175.

Sun, D.B., Feng, L., Shi, H.Y., Chen, J.F., Liu, S.W., Chen, H.Y., Wang, Y.F., 2007. Spike protein region (aa 636789) of porcine epidemic diarrhea virus is essential for induction of neutralizing antibodies. Acta Virol. 51, 149–156.

Sun, D., Feng, L., Shi, H., Chen, J., Cui, X., Chen, H., Liu, S., Tong, Y., Wang, Y., Tong, G., 2008. Identification of two novel B cell epitopes on porcine epidemic diarrhea virus spike protein. Vet. Microbiol. 131, 73–81.

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.

Temeeyasen, G., Srijangwad, A., Tripipat, T., Tipsombatboon, P., Piriyapongsa, 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.

Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673–4680.

Vui, D.T., Thanh, T.L., Tung, N., Srijangwad, A., Tripipat, T., Chuanasa, T., Nilubol, D., 2015. Complete genome characterization of porcine epidemic diarrhea virus in Vietnam. Arch. Virol. 160, 1931–1938.

Wang, D., Zhang, Y., Zhang, Z., Zhu, J., Yu, J., 2010. KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. Genomics Proteomics Bioinformatics 8, 77–80.

Wicht, O., Li, W., Willems, L., Meuleman, T.J., Wubbolts, R.W., van Kuppeveld, F.M., Rottier, P.J.M., Bosch, B.J., 2014. Proteolytic activation of the porcine epidemic diarrhea coronavirus spike fusion protein by trypsin in cell culture. J. Virol. 88, 7952–7961.

Wood, E.N., 1977. An apparently new syndrome of porcine epidemic diarrhoea. Vet. Rec. 100, 243–244.