

# The genetic diversity and complete genome analysis of two novel porcine deltacoronavirus isolates in Thailand in 2015

Athip Lorsirigool<sup>1</sup> · Kepalee Saeng-chuto<sup>1</sup> · Adthakorn Madapong<sup>1</sup> ·  
Gun Temeeyasen<sup>1</sup> · Thitima Tripipat<sup>1</sup> · Pavita Kaewprommal<sup>2</sup> ·  
Angkana Tantituvanont<sup>3</sup> · Jittima Piriyaongsa<sup>2</sup> · Dachrit Nilubol<sup>1</sup> 

Received: 17 August 2016 / Accepted: 18 November 2016  
© Springer Science+Business Media New York 2016

**Abstract** Porcine deltacoronavirus (PDCoV) was identified in intestinal samples collected from piglets with diarrhea in Thailand in 2015. Two Thai PDCoV isolates, P23\_15\_TT\_1115 and P24\_15\_NT1\_1215, were isolated and identified. The full-length genome sequences of the P23\_15\_TT\_1115 and P24\_15\_NT1\_1215 isolates were 25,404 and 25,407 nucleotides in length, respectively, which were relatively shorter than that of US and China PDCoV. The phylogenetic analysis based on the full-length genome demonstrated that Thai PDCoV isolates form a new cluster separated from US and China PDCoV but relatively were more closely related to China PDCoV than US isolates. The genetic analyses demonstrated that Thai PDCoVs have 97.0–97.8 and 92.2–94.0% similarities with China PDCoV at nucleotide and amino acid levels, respectively, but share 97.1–97.3 and 92.5–93.0 similarity with US PDCoV at the nucleotide and amino acid levels, respectively. Thai PDCoV

possesses two discontinuous deletions of five amino acids in ORF1a/b region. One additional deletion of one amino acid was identified in P23\_15\_TT\_1115. The variation analyses demonstrated that six regions (nt 1317–1436, 2997–3096, 19,737–19,836, 20,277–20,376, 21,177–21,276, and 22,371–22,416) in ORF1a/b and spike genes exhibit high sequence variation between Thai and other PDCoV. The analyses of amino acid changes suggested that they could potentially be from different lineages.

**Keywords** Porcine deltacoronavirus · Full-length genome · Thailand

## Introduction

Porcine deltacoronavirus (PDCoV), a novel pathogen belonged to the family *Coronaviridae*, genus *Deltacoronavirus* [1], causes an enteric disease in pigs characterized by watery diarrhea similar to porcine epidemic diarrhea (PED) and transmissible gastroenteritis (TGE) [2]. PDCoV is an enveloped, single-stranded, positive-sense RNA virus. The full-length genome of PDCoV is approximately 25 kb in length and comprises the 5'-untranslated region (UTR), open reading frames (ORFs) including ORF1a and ORF1b, spike (S) envelope (E), membrane (M), non-structural protein 6 (Nsp6), nucleoprotein (N) non-structural protein 7 (Nsp7), and the 3'-UTR [1]. ORF1a and ORF1b occupy two-thirds of the genome encoding two overlapping replicase polyproteins. S, E, M, and N genes are located downstream encoding S, E, M and N proteins, respectively. Nsp6 and Nsp7 are located upstream and a section of N gene, respectively. S glycoprotein contains two domains including S1 and S2 playing an important role in binding to specific host cell receptors. E and M proteins are

Edited by Keizo Tomonaga.

**Electronic supplementary material** The online version of this article (doi:10.1007/s11262-016-1413-z) contains supplementary material, which is available to authorized users.

✉ Dachrit Nilubol  
dachrit@gmail.com

<sup>1</sup> Department of Veterinary Microbiology, Faculty of Veterinary Science, Chulalongkorn University, Bangkok 10330, Thailand

<sup>2</sup> Genome Technology Research Unit, National Center for Genetic Engineering and Biotechnology, National Science and Technology Development Agency, Pathumthani 12120, Thailand

<sup>3</sup> Department of Pharmaceutics and Industrial Pharmacy, Faculty of Pharmaceutical Sciences, Chulalongkorn University, Bangkok 10330, Thailand

transmembrane proteins associated with viral envelope formation and release [3]. N protein functions in viral replication and pathogenesis [4].

PDCoV was first detected in Hong Kong as isolates HKU15-44 and HKU15-155 in 2012 [1]. In February 2014, PDCoV was first detected in Ohio, United States, in association with PED cases. Since then, PDCoV has been detected in most pig producing states of the US and Canada [5–7]. The retrospective investigation demonstrated the presence of PDCoV in the US as early as 2013 [8]. Recently, PDCoV was identified for the first time in South Korea and China [9, 10]. At present, two groups of PDCoV based on the origin of discovery have been identified including US-like (G1) and China-like (G2) groups.

The Thai swine industry has experienced diarrhea outbreaks with milder forms of clinical disease compared to PED since 2014. The causative agent was considered to be a variant of PED virus (PEDV). However, PEDV was not detected in intestinal samples from the suspected herds. The role of PDCoV in the outbreak, although suspected, was not investigated at that time. Following the negative detection of PEDV in samples, PDCoV was increasingly suspected when re-breaks of clinical enteric disease similar to PED occurred every two months in some herds, which is too frequent compared to the period of 6-month protection reported earlier [11]. We therefore investigated the presence of PDCoV in intestinal samples collected from pig farms with diarrhea outbreak in Thailand using PCR. PDCoV was then identified in two pig herds from Ratchaburi and Chonburi, provinces in the western and eastern regions of Thailand, respectively. The genetic analyses revealed a novel PDCoV clustered separately from other PDCoVs. The full-length genome of Thai PDCoV isolates were characterized herein compared to previously reported PDCoV.

## Materials and methods

### Farms and sample preparation

Ten intestinal samples (five of each) were collected from two pig farms in Ratchaburi and Chonburi, provinces in the western and eastern regions of Thailand, in November and December 2015, respectively. Both farms have an inventory of 2500 and 4000 sows and are located in a region with a high density of pig farms. Intestinal samples were ground into small pieces and suspended in phosphate-buffered saline (PBS; 0.1 M, pH 7.2). The suspensions were centrifuged at  $10,000\times g$  for 10 min followed by filtering the supernatant through 0.45- $\mu\text{m}$  filters for viral RNA extraction.

### Reverse transcription polymerase chain reaction and sequence determination

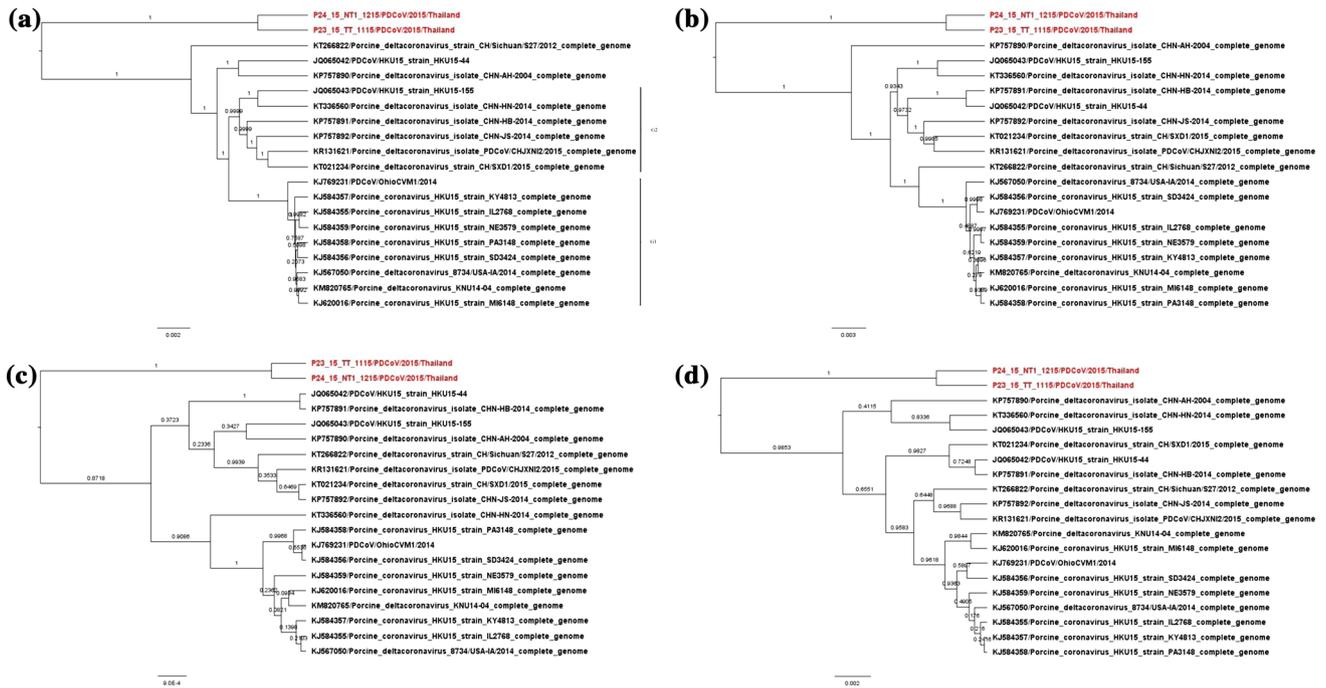
Total viral RNA was extracted from the supernatant using the Nucleospin<sup>®</sup> viral RNA isolation kit (Macherey-Nagel Inc., Duren, Germany) in accordance with the manufacturer's instructions. cDNA was synthesized from the extracted RNA using M-MuLV Reverse Transcriptase (BioLabs Inc., Ipswich, MA, USA). The cDNA was used for PCR amplification and was purified using a Nucleospin Plasmid kit (Macherey-Nagel Inc., Bethlehem, PA, USA). PCR amplification of the cDNA was performed using Platinum<sup>®</sup> *Taq* DNA polymerase High Fidelity (Invitrogen, CA, USA) according to the manufacturer's protocol. To amplify the complete ORF1a/1b, S, E, M, and N genes, 26 primer pairs specific to each gene were designed (Supplement 1). The PCR products were visualized by agarose gel electrophoresis. Positive samples were purified using the Nucleospin Plasmid kit (Macherey-Nagel Inc., Bethlehem, PA, USA) and were sequenced in both directions using an ABI Prism 3730XL sequencer performing at First BASE Laboratory (Selangor, Malaysia). The 5' and 3' terminal regions were determined using a kit for rapid amplification of 5' and 3' cDNA ends (5' and 3'-RACE) (Clontech, Japan).

### Sequence analysis

Nucleotide and amino acid sequence alignments were created using the CLUSTALW program [12]. Phylogenetic analyses based on the full-length genome, S, M, and N genes were separately constructed together with 18 other PDCoV isolate sequences (Supplement 2) using a Bayesian Markov chain Monte Carlo (BMCMC) method implemented in the program BEAST v1.8.3 [13, 14] for substitution trees. A BEAST run was performed based on TN93+G+I (Fig. 1a, b, d) and JC (Fig. 1c) substitution models with a coalescent constant sample size tree prior for each analysis using at least 200 million generations with sampling of every 10,000 generations and the first 10% discarded as burn-in. Tracer v1.6 was used to confirm that post-burn-in trees yielded an effective sample size (ESS) of  $>200$  for all parameters. The resulting tree was viewed and generated in FigTree v1.4.2. The percentages of nucleotide and amino acid sequence identity between isolates were also calculated.

### Sliding window analysis of sequence variation

Genome alignment between two Thai PDCoV isolates and the two other isolates (8734/USA-IA/2014 and CHN-HN-2014) representing US and China PDCoVs, respectively, was performed to determine nucleotide variation sites using the CLUSTALW program [12]. Only protein-encoding sequences were included in the analysis. A sliding



**Fig. 1** Phylogenetic analyses of porcine deltacoronavirus (PDCoV) based on the nucleotide sequences of the full-length genome (a), and S (b), M (c), and N (d) genes were separately performed using a

Bayesian Markov chain Monte Carlo (BMCMC) method. *Red fronts* represent the Thai PDCoV isolates (Color figure online)

window of 100 bp with a step size of 20 bp was used to evaluate sequence diversity for complete alignment. The variation coefficient value, defined as the number of variable points, for each window was calculated according to the method described in Sun et al. [15].

**Calculation of antigenic index and hydrophilicity plots**

Antigenicity and hydrophilicity analyses were performed based on the S protein amino acid sequences of Thai PDCoV isolate (P23\_15\_TT\_1115), 8734/USA-IA/2014, and CHN-HN-2014 isolates. Jameson–Wolf antigenic indexes [16] and Kyte–Doolittle hydrophilicity plots [17] of these sequences were constructed using Protean of the DNASTAR Lasergene software package (DNASTAR, Inc., Madison, WI, USA).

**Results**

**Full-length genome sequences**

Two Thai PDCoV isolates including P23\_15\_TT\_1115 and P24\_15\_NT1\_1215 were identified in samples collected from pig farms in Ratchaburi and Chonburi, respectively. The full-length genome sequences of P23\_15\_TT\_1115 and P24\_15\_NT1\_1215 were characterized and deposited in

GenBank under accession number KU984334 and KX361345, respectively. The full-length genome of Thai PDCoV, P23\_15\_TT\_1115 and P24\_15\_NT1\_1215, had a size of 25,404 and 25,407 nucleotides (nt) in length, respectively, which are relatively shorter in comparison to China and US PDCoV (Table 1). The sequence alignments demonstrated that their genome organization is similar to that of all previously reported PDCoV genomes, which are characterized by the gene order of 5'-ORF1a/1b-S-E-M-Nsp6-N-Nsp7-3' (Table 1). The untranslated regions (UTRs) were present at both ends (5' UTR, nt 536 and 3' UTR, nt 729).

The nucleotide and deduced amino acid sequences of the full-length genome of Thai PDCoV isolates along with that of US and China PEDV isolates were aligned. The relative shorter genome was due to the deletions in ORF1a/b and S genes. Thai PDCoV isolates possess discontinuous deletions (Table 2). The 5'UTR and 3'UTR contain one deletion of 3 and 1 nucleotides, respectively. Two discontinuous deletions of 5 amino acids in the ORF1a/1b gene were identified compared to the US and China PDCoV. P23\_15\_TT\_1115 possesses one deletion of one amino acid in S gene. In contrast, P24\_15\_NT1\_1215 does not possess this deletion.

**Phylogenetic analyses**

The phylogenetic tree based on the full-length genome sequences of PDCoV isolates demonstrated that the PDCoV isolates are clustered mainly into two different

**Table 1** Genome organization of Thai porcine deltacoronavirus (PDCoV) isolates compared to that of US and China PDCoV groups

Gene	P23_15_TT_1115				P24_15_NT1_1215			
	Start	Stop	Size (nt)	Size (aa)	Start	Stop	Size (nt)	Size (aa)
5'UTR	1	536	536	–	1	536	536	–
ORF1a/1b	537	19,322	18,786	6262	537	19,322	18,786	6262
S	19,307	22,783	3477	1159	19,307	22,786	3480	1160
E	22,780	23,028	249	83	22,783	23,031	249	83
M	23,024	23,674	651	217	23,027	23,677	651	217
Nsp6	23,677	23,958	282	94	23,680	23,961	282	94
N	23,982	25,007	1026	342	23,985	25,010	1026	342
Nsp7	24,076	24,675	600	200	24,079	24,678	600	200
3'UTR	24,676	25,404	729	–	24,679	25,407	729	–
Gene	US PDCoV groups				China PDCoV groups			
	Start	Stop	Size (nt)	Size (aa)	Start	Stop	Size (nt)	Size (aa)
5'UTR	1	539	539	–	1	539	539	–
ORF1a/1b	540	19,340	18,801	6267	540	19,340	18,801	6267
S	19,324	22,803	3480	1160	19,324	22,800	3477	1159
E	22,800	23,048	249	83	22,797	23,045	249	83
M	23,044	23,694	651	217	23,041	23,691	651	217
Nsp6	23,697	23,978	282	94	23,694	23,975	282	94
N	24,002	25,027	1026	342	23,999	25,024	1026	342
Nsp7	24,096	24,695	600	200	24,093	24,692	600	200
3'UTR	24,696	25,423	728	–	24,693	25,420	728	–

groups, the US-like (G1) and China-like groups (G2), excluding CH, HKU1515-44, and CHN-AH-2004. However, both Thai PDCoV isolates belong to a new group that cluster separately from the US and China PDCoVs, and the three isolates (Fig. 1).

The full-length genome analyses comparing both Thai PDCoV isolates and PDCoV from G1 and G2 groups demonstrated that both Thai PDCoV have higher genetic similarity with the PDCoV isolate in the G2 than the PDCoV isolates from the G1. Thai PDCoV shares 97.0–97.8 and 92.2–94.0% genetic similarities with the G2 at the nucleotide and amino acid levels, respectively. The genetic similarities of both Thai PDCoV isolates with the isolates in the G1 are 97.1–97.3 and 92.5–93.0% at the nucleotide and amino acid levels, respectively (Table 3).

Phylogenetic trees based on the S, M, and N genes demonstrated a similar clustering pattern to that of the full-length genome tree (Fig. 1). PDCoV isolates are clustered mainly into two different groups, the G1 and G2. Based on the three phylogenetic trees, both Thai isolates were clustered in a novel group separately from US and China PDCoVs. Although China PDCoV detected in 2004 (CHN-AH-2004) was clustered separately from the G1 and G2 based on the phylogenetic tree of the S gene, Thai PDCoV isolates and CHN-AH-2004 were grouped in different

clusters (Fig. 1). The results suggest that Thai PDCoV may have evolved from a different lineage compared to the currently identified PDCoV.

#### Genetic analyses and variation analysis of the complete ORF1a/1b, S, E, M, and N genes of Thai PDCoV compared with PDCoV isolates from other countries

ORF1a/1b gene of the both Thai isolates are 18,786 nt in length and encodes 6262 amino acids. Substitutions are occurred in several positions and 2 discontinuous deletions of 5 amino acids (<sup>401</sup>LK<sup>402</sup> and <sup>758</sup>PVG<sup>760</sup>) were identified compared to PDCoV in G1 and G2 groups. The similarity between Thai isolates and the isolates in the G1 are 97.1–97.4 and 97.6–98.5% at the nucleotide and amino acid levels, respectively (Table 3). The pair-wise nucleotide and amino acid identities between Thai PDCoV isolates and the isolates in the G2 were 97.0–97.9 and 97.6–98.8%, respectively.

The S gene of the P23\_15\_TT\_1115 and P24\_15\_NT1\_1215 isolates are 3477 and 3480 nt in length and encodes 1159 and 1160 amino acids, respectively. The S1 and S2 domains of the P23\_15\_TT\_1115 isolate were located at amino acid positions 68–522 and 531–1147,

**Table 2** Nucleotide (amino acid) deletions and insertions in the Thai porcine deltacoronavirus (PDCoV) isolates compared to that of US and China PDCoV isolates

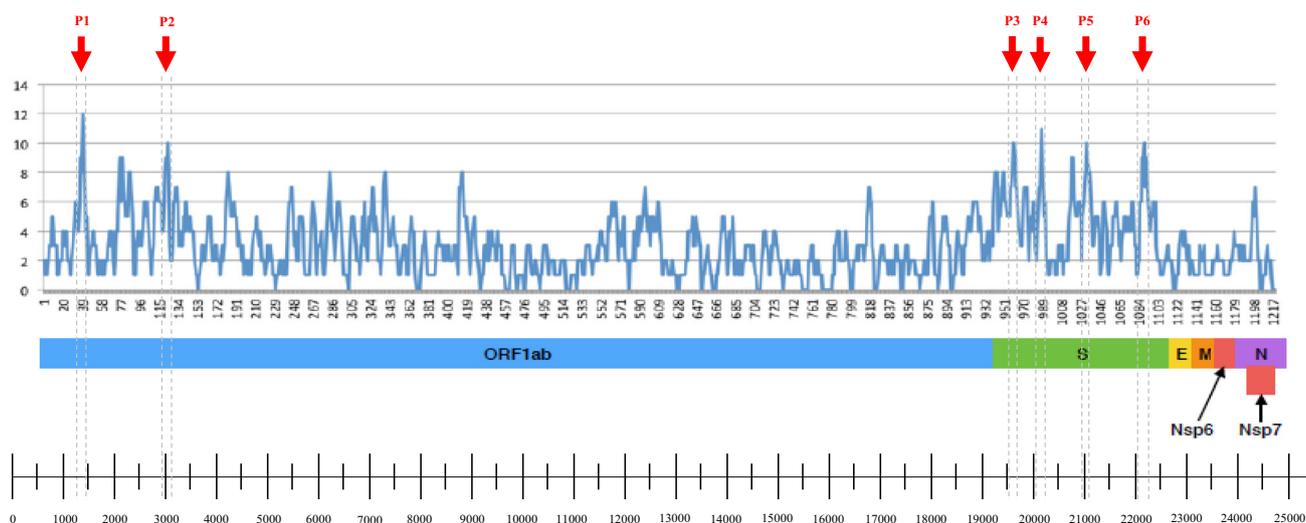
Thai PDCoV isolates	Genes	PDCoV groups			
		US PDCoV		China PDCoV	
		Insertions	Deletions	Insertions	Deletions
P23_15_TT_1115	5'UTR	–	<sup>116</sup> CCT <sup>118</sup>	–	<sup>116</sup> TCT <sup>118</sup>
	ORF1a/1b	–	<sup>1739</sup> TTTGAA <sup>1744</sup> ( <sup>401</sup> LK <sup>402</sup> ) <sup>2815</sup> TCGGCAATG <sup>2823</sup> ( <sup>758</sup> PVG <sup>760</sup> )	–	<sup>1739</sup> TTTGAA <sup>1744</sup> ( <sup>401</sup> LK <sup>402</sup> ) <sup>2815</sup> TCGGCAATG <sup>2823</sup> ( <sup>758</sup> PVG <sup>760</sup> )
	S	–	<sup>51</sup> N	–	–
	E	–	–	–	–
	M	–	–	–	–
	Nsp6	–	–	–	–
	N	–	–	–	–
	Nsp7	–	–	–	–
	3'UTR	<sup>25027</sup> C	–	<sup>25027</sup> C	–
	P24_15_NT_1215	5'UTR	–	<sup>116</sup> CCT <sup>118</sup>	–
ORF1a/1b		–	<sup>1739</sup> TTTGAA <sup>1744</sup> ( <sup>401</sup> LK <sup>402</sup> ) <sup>2815</sup> TCGGCAATG <sup>2823</sup> ( <sup>758</sup> PVG <sup>760</sup> )	–	<sup>1739</sup> TTTGAA <sup>1744</sup> ( <sup>401</sup> LK <sup>402</sup> ) <sup>2815</sup> TCGGCAATG <sup>2823</sup> ( <sup>758</sup> PVG <sup>760</sup> )
S		–	–	<sup>51</sup> N	–
E		–	–	–	–
M		–	–	–	–
Nsp6		–	–	–	–
N		–	–	–	–
Nsp7		–	–	–	–
3'UTR		<sup>25027</sup> C	–	<sup>25027</sup> C	–

**Table 3** Comparison of the nucleotide and amino acid sequence similarities (%) of the five structural genes of the Thai porcine deltacoronavirus (PDCoV) isolates and that of US and China PDCoV isolates

Thai isolates	Genes	PDCoV groups			
		US		China	
		Nucleotide	Amino acid	Nucleotide	Amino acid
P23_15_TT_1115	ORF1a/1b	97.4	98.3–98.5	97.3–97.9	98.3–98.8
	S	96.0–96.2	97.1–97.6	96.0–96.7	95.9–98.1
	E	100.0	100.0	100.0	100.0
	M	97.8–98.1	99.5	98.3–98.6	99.5
	N	97.6–97.9	98.2–99.1	97.9–98.5	98.5–99.1
P24_15_NT1_1215	ORF1a/1b	97.1–97.2	97.6–97.7	97.0–97.6	97.6–98.1
	S	95.9–96.1	96.9–97.5	95.2–96.6	95.2–97.9
	E	100.0	100.0	100.0	100.0
	M	97.8–98.1	99.5	98.3–98.6	99.5
	N	97.2–97.5	98.8–99.1	97.5–98.0	98.8–99.4

respectively. The S1 and S2 domains of the P24\_15\_NT1\_1215 isolate were located at amino acid positions 69–523 and 532–1148, respectively. Several substitutions positions at the amino acid level were identified between the Thai PDCoV isolates and the G1 and G2.

Compared to the G1, the P23\_15\_TT\_1115 isolate has a deletion of 1 (<sup>51</sup>N) amino acid at position 51, similar to isolates in the G2. On the other hand, the P24\_15\_NT1\_1215 isolate has an insertion of 1 (<sup>51</sup>N) amino acid at position 51 compared to G2. The Thai



**Fig. 2** Sliding window analysis of genome sequence variation between the P23\_15\_TT\_1115 isolate and two reference isolates (8734/USA-IA/2014 and CHN-HN-2014 isolates). The graph shows the variation coefficient value calculated from the genome sequence

PDCoV isolates shares 95.2–96.7 and 95.2–98.1% similarity with the G2 at the nucleotide and amino acid levels, respectively. The Thai PDCoV isolates have nucleotide and amino acid similarities of 95.9–96.2 and 96.9–97.6%, respectively, with the G1 (Table 3).

The E gene in both Thai PDCoV isolates has 249 nt and encodes 83 amino acids. No mutations were identified in this gene compared to the G1 and G2 groups. The pair-wise nucleotide and amino acid identities between the Thai PDCoV isolates and both PDCoV groups are 100% (Table 3).

The M gene of the Thai PDCoV isolates are 651 nt in length and encodes 217 amino acids. Compared to the isolates in the G1 and G2, a substitution of 1 (V83A) amino acid at position 83 was identified in the Thai PDCoV isolates. The Thai PDCoV isolates shares 98.3–98.6 and 99.5% similarity at the nucleotide and amino acid levels, respectively, with isolates in the G2. The Thai PDCoV isolates shares 97.8–98.1 and 99.5% similarity with the G1 at the nucleotide and amino acid levels, respectively (Table 3).

The N gene of the Thai PDCoV isolates has a length of 1026 nt and encodes 342 amino acids. Four (A24S, V43A, S163P, and G167C) and three (V43A, S163P, and G167C) substitutions at the amino acid level were identified in the P23\_15\_TT\_1115 and the P24\_15\_NT1\_1215 isolates, respectively, compared to the isolates in both PDCoV groups. The nucleotide and amino acid similarities between the Thai PDCoV isolates and the isolates in the G2 are 97.5–98.5 and 98.5–99.4%, respectively. The Thai PDCoV isolates shares nucleotide and amino acid similarities of 97.2–97.9 and 98.2–99.1%, respectively, with the G1 (Table 3).

alignment (window size = 100 bp, step size = 20 bp). Red arrows represent the positions of high-mutation regions including the ORF1a/1b and S genes. The positions and sizes of the PDCoV genes correspond to the scale bar (Color figure online)

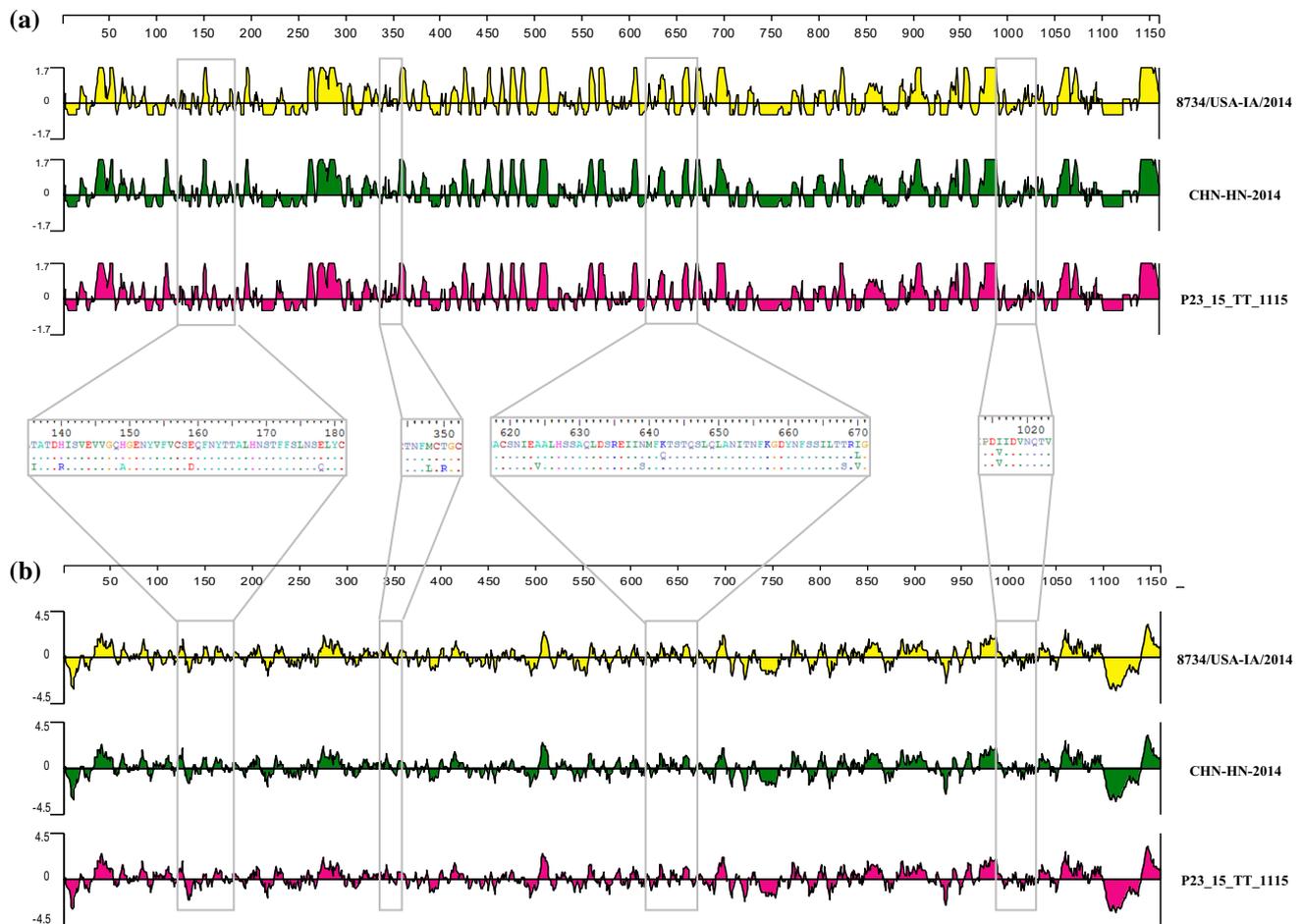
### Sliding window analysis of the full-length genome sequence of the P23\_15\_TT\_1115 isolate compared with PDCoV isolates from other countries

Both Thai PDCoV, P23\_15\_TT\_1115 and P24\_15\_NT1\_1215, share high genetic similarity 99.4 and 98.6% at nucleotide and amino acid levels, respectively. Therefore, to identify the genome regions exhibiting sequence variation, the P23\_15\_TT\_1115 isolate was selected to compare with China PDCoV isolates (CHN-HN-2014) and US PDCoV isolates (8734/USA-IA/2014).

Sliding window analysis identified six regions, named P1, P2, P3, P4, P5, and P6 (Fig. 2), exhibiting high sequence variation among the three isolates (P23\_15\_TT\_1115, CHN-HN-2014, and 8734/USA-IA/2014 isolates) (Fig. 2). P1 and P2 are located on the ORF1a/1b gene at positions nt 1317–1436 and nt 2997–3096, respectively, of the full-length genome. P3 and P4 are located on the S1 domain at positions nt 19,737 to 19,836 and nt 20,277 to 20,376, respectively. P5 and P6 are located on the S2 domain at positions nt 21,177 to 21,276 and nt 22,371 to 22,416, respectively. These results suggest that the ORF1a/1b and S genes are the most variable regions in PDCoV genome.

### Antigenic index and hydrophilicity analyses of the P23\_15\_TT\_1115 isolate compared with PDCoV isolates from other countries

The S gene regions exhibited the highest sequence variation among the P23\_15\_TT\_1115 isolate and the two other PDCoV groups. The antigenic index and the hydrophilicity



**Fig. 3** Antigenic index (a) and hydrophilicity plots (b) based on the amino acid sequences of the divergent region of the S protein fragment (amino acid position 1–1159). The *dashed lines* indicate the

regions exhibiting differences among the P23\_15\_TT\_1115 isolate and two reference isolates (8734/USA-IA/2014 and CHN-HN-2014 isolates) (Color figure online)

plots of the S protein of the P23\_15\_TT\_1115 isolate were compared with those of the China PDCoV isolates (CHN-HN-2014) and US PDCoV isolates (8734/USA-IA/2014) (Fig. 3). The major differences in the antigenic index and hydrophilicity values are located in four regions of the amino acid sequence at positions 144–179, 324–357, 624–657, and 1004–1032. These regions exhibited deletions and substitutions leading to separation between the two groups and the Thai PDCoV isolate (Fig. 3).

## Discussion

Since the identification in Hong Kong in 2012 and the US in 2014, PDCoV has increasingly been detected in other countries, including Canada, South Korea, and China [1, 7, 9, 18]. PDCoV was identified in Thai swine farms in 2015, and the full-length genome of the P23\_15\_TT\_1115 and P24\_15\_NT1\_1215 isolates were characterized herein.

The present study revealed several important findings based on the genetic analyses. Although the genome organization of Thai PDCoV is similar to that of previously reported PDCoVs, the sizes of the full-length genome sequence of the Thai PDCoV isolates were 25,404 and 25,407 nucleotides (nt) in length, which were relatively shorter than that of US and China PDCoV. The genome size of US and China PDCoV were 25,422 and 25,421–25,426 nt in length, respectively [1, 7, 9, 18]. We therefore individually analyzed each gene by comparison with PDCoVs isolated from other countries. The results demonstrated several substitutions based on amino acid level. In addition, discontinuous deletions of nucleotides compared to US and China PDCoVs were observed in the 5'UTR, ORF1a/1b and spike regions. Based on the analysis, the shorter genome size of Thai PDCoV compared to that of US and China PDCoV is due to nucleotide deletions, especially in ORF1a/b and S genes. A deletion at a similar position of the S gene in isolates from China was

reported previously [19]. China isolate contains three nucleotide deletion representing one amino acid deletion at position <sup>51</sup>N [1, 7, 9, 18]. Only one Thai PDCoV also contain amino acid deletion at the similar position. In contrast, both Thai isolates contains deletions in ORF1a/b in which are first identified compared other PDCoV isolates previously reported [1, 7, 9, 18]. Functions and important characteristic of these deletions are still not known and require further investigation.

A phylogenetic tree based on the full-length genome demonstrated that the Thai PDCoV isolates forms a group separated from US and China PDCoV isolates. The results suggest that Thai PDCoV isolates were clustered in a novel group of PDCoV. Previous reports investigating the identification of PDCoV in other countries including US, China and South Korea demonstrated that PDCoV has further evolved into two different groups including US and China PDCoV (Fig. 1a). The Thai PDCoV were clustered separated from all PDCoV. The finding suggested that Thai PDCoV isolates are novel PDCoV and could be evolve from the same ancestor as other PDCoV but different lineage, undergoing evolution for sometimes until complete separation from other PDCoV. Further genetic analyses including molecular clock, molecular epidemiology, and more retrospective investigation in samples collected prior to 2015 are urgently needed to determine the presence of PDCoV in Thailand.

Thai PDCoV isolates are genetically distinct from other PDCoV. We therefore analyzed the genetic difference compared with other PDCoV. The analysis identified six different regions exhibiting high sequence variation among the PDCoV isolates from US, China, and Thailand. The first two hypervariable regions, P1 and P2, are located on the ORF1a/1b gene at positions 1317–1436 and 2997–3096 bp, respectively. These two hypervariable regions are closed to the deletion region in ORF1a/1b. The deletion and insertion could contribute to the high sequence variation. Other functions were still unknown and needed further investigation.

The S gene of the Thai PDCoV isolates, both S1 and S2 domains, exhibit the highest percentage of sequence variation compared to that of the two other PDCoV groups (96.0–96.7 and 95.9–98.1% similarities at the nucleotide and amino acid levels). In addition to substitutions positions at the amino acid level and deletion/insertion of 1 (<sup>51</sup>N) amino acid at position 51 compared to isolates in the China-like and US-like groups, four additional hypervariable regions are located at positions 19,737–19,836, 20,277–20,376, 21,177–21,276, and 22,371–22,416 bp, respectively (Fig. 2). The changes in these four regions could potentially affect the antigenicity of the virus. The functions of S1 and S2 domains of PDCoV are not clear, however, might resemble functions of S protein of PEDV,

which belongs to the same family. These four hypervariable regions in the spike gene of PDCoV could be neutralizing epitope. Four neutralizing epitopes were identified in spike gene of PEDV [20–22].

In conclusion, the genetic analyses based on full-length genome sequences demonstrated that Thai PDCoV isolates form a new PDCoV cluster that is separated from PDCoV isolates from China, South Korea, and the US. Thai PDCoV isolates are new variants, closely related with Chinese PDCoV and possess four discontinuous deletions of seven amino acids in the ORF1a/b and S genes. The origin and source of virus introduction into Thailand are not known. The viruses could have been in this region for some time, similar to China in which the detection of PDCoV dated back to 2004 and continuously evolved until separated into different lineages, or the viruses were introduced from different ancestors or sources. Further retrospective investigations are urgently needed to elucidate source and evolution. In addition, further analysis and molecular epidemiology based on the complete genome sequence and pathogenicity studies of this PDCoV isolate are urgently needed.

**Acknowledgements** This work was supported by the National Research Council of Thailand [Grant Number 5080001], and Agricultural research development agency (Public organization). The funding provided by Special Task Force for Activating Research (STAR), swine viral evolution and vaccine research (SVEVR), Chulalongkorn University.

**Authors' contribution** C.J.S. and D.N. conceived and designed the experiments; C.J.S., G.T. and T.T. performed the experiments; C.J.S. and K.S. analyzed the data; C.J.S., T.T. and A.T. contributed reagents/materials/analysis tools; C.J.S. and D.N. composed the manuscript.

#### Compliance with ethical standards

**Ethical approval** All applicable international, national, and/or institutional guidelines for the care and use of animals were followed.

## References

1. P.C. Woo, S.K. Lau, C.S. Lam, C.C. Lau, A.K. Tsang, J.H. Lau, R. Bai, J.L. Teng, C.C. Tsang, M. Wang, B.J. Zheng, K.H. Chan, K.Y. Yuen, 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 (2012)
2. K. Jung, H. Hu, B. Eyerly, Z. Lu, J. Chepngeno, L.J. Saif, Pathogenicity of 2 porcine deltacoronavirus strains in gnotobiotic pigs. *Emerg. Infect. Dis.* **21**, 650–654 (2015)
3. P.C. Woo, Y. Huang, S.K. Lau, K.Y. Yuen, Coronavirus genomics and bioinformatics analysis. *Viruses* **2**, 1804–1820 (2010)

4. R. McBride, M. van Zyl, B.C. Fielding, The coronavirus nucleocapsid is a multifunctional protein. *Viruses* **6**, 2991–3018 (2014)
5. N. Homwong, M.C. Jarvis, H.C. Lam, A. Diaz, A. Rovira, M. Nelson, D. Marthaler, Characterization and evolution of porcine deltacoronavirus in the United States. *Prev. Vet. Med.* **123**, 168–174 (2016)
6. D. Marthaler, Y. Jiang, J. Collins, K. Rossow, Complete genome sequence of strain SDCV/USA/Illinois121/2014, a porcine deltacoronavirus from the United States. *Genome Announc.* (2014). doi:[10.1128/genomeA.00218-14](https://doi.org/10.1128/genomeA.00218-14)
7. L. Wang, B. Byrum, Y. Zhang, Porcine coronavirus HKU15 detected in 9 US states, 2014. *Emerg. Infect. Dis.* **20**, 1594–1595 (2014)
8. A. Sinha, P. Gauger, J. Zhang, K.J. Yoon, K. Harmon, PCR-based retrospective evaluation of diagnostic samples for emergence of porcine deltacoronavirus in US swine. *Vet. Microbiol.* **179**, 296–298 (2015)
9. S. Lee, C. Lee, Complete genome characterization of Korean porcine deltacoronavirus strain KOR/KNU14-04/2014. *Genome Announc.* (2014). doi:[10.1128/genomeA.01191-14](https://doi.org/10.1128/genomeA.01191-14)
10. D. Song, X. Zhou, Q. Peng, Y. Chen, F. Zhang, T. Huang, T. Zhang, A. Li, D. Huang, Q. Wu, H. He, Y. Tang, Newly emerged porcine deltacoronavirus associated with diarrhoea in swine in China: identification, prevalence and full-length genome sequence analysis. *Trans. Emerg. Dis.* **62**, 575–580 (2015)
11. D. Goede, M.P. Murtaugh, J. Nerem, P. Yeske, K. Rossow, R. Morrison, 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 (2015)
12. J.D. Thompson, D.G. Higgins, T.J. Gibson, 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 (1994)
13. A.J. Drummond, A. Rambaut, BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* **7**, 214 (2007)
14. A.J. Drummond, M.A. Suchard, D. Xie, A. Rambaut, Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* **29**, 1969–1973 (2012)
15. M. Sun, J. Ma, Y. Wang, M. Wang, W. Song, W. Zhang, C. Lu, H. Yao, 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 (2015)
16. B.A. Jameson, H. Wolf, The antigenic index: a novel algorithm for predicting antigenic determinants. *Comput. Appl. Biosci.* **4**, 181–186 (1988)
17. J. Kyte, R.F. Doolittle, A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**, 105–132 (1982)
18. L. Wang, B. Byrum, Y. Zhang, Detection and genetic characterization of deltacoronavirus in pigs, Ohio, USA, 2014. *Emerg. Infect. Dis.* **20**, 1227–1230 (2014)
19. Q. Chen, G. Li, J. Stasko, J.T. Thomas, W.R. Stensland, A.E. Pillatzki, P.C. Gauger, K.J. Schwartz, D. Madson, K.J. Yoon, G.W. Stevenson, E.R. Burrough, K.M. Harmon, R.G. Main, J. Zhang, 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 (2014)
20. D.A. Brian, R.S. Baric, Coronavirus genome structure and replication. *Curr. Top. Microbiol. Immunol.* **287**, 1–30 (2005)
21. D.Q. Liu, J.W. Ge, X.Y. Qiao, Y.P. Jiang, S.M. Liu, Y.J. Li, High-level mucosal and systemic immune responses induced by oral administration with *Lactobacillus*-expressed porcine epidemic diarrhea virus (PEDV) S1 region combined with *Lactobacillus*-expressed N protein. *Appl. Microbiol. Biotechnol.* **93**, 2437–2446 (2012)
22. D.B. Sun, L. Feng, H.Y. Shi, J.F. Chen, S.W. Liu, H.Y. Chen, Y.F. Wang, Spike protein region (aa 636789) of porcine epidemic diarrhea virus is essential for induction of neutralizing antibodies. *Acta Virol.* **51**, 149–156 (2007)