Virus Research 149 (2010) 175–182

Contents lists available at ScienceDirect

Virus Research

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

Heterogeneity in spike protein genes of porcine epidemic diarrhea viruses isolated in Korea

Dong-Kyu Lee[a,1], Choi-Kyu Park[b,1], Seong-Hee Kim[b], Changhee Lee[a,*]

[a] Department of Microbiology, College of Natural Sciences, Kyungpook National University, Daegu 702-701, Republic of Korea
[b] Animal Disease Diagnostic Center, National Veterinary Research and Quarantine Services, Anyang 430-824, Republic of Korea

ARTICLE INFO

Article history:
Received 19 November 2009
Received in revised form 16 January 2010
Accepted 26 January 2010
Available online 2 February 2010

Keywords:
PEDV
Spike proteins
Korean isolates
Phylogenetic analysis

ABSTRACT

Porcine epidemic diarrhea virus (PEDV) has plagued the domestic swine industry in Korea causing significant economic impacts on pig production nationwide. In the present study, we determined the complete nucleotide sequences of the spike (S) glycoprotein genes of seven Korean PEDV isolates. The entire S genes of all isolates were found to be nine nucleotides longer in length than other PEDV reference strains. This size difference was due to the combined presence of notable 15 bp insertion and 6 bp deletion within the N-terminal region of the S1 domain of the Korean isolates. In addition, the largest number of amino acid variations was accumulated in the S1 N-terminal region, leading to the presence of hypervariability in the isolates. Sequence comparisons at the peptide level of the S proteins revealed that all seven Korean isolates shared diverse similarities ranging from a 93.6% to 99.6% identity with each other but exhibited a 92.2% to 93.7% identity with other reference strains. Collectively, the sequence analysis data indicate the diversity of the PEDV isolates currently prevalent in Korea that represents a heterogeneous group. Phylogenetic analyses showed two separate clusters, in which all Korean field isolates were grouped together in the second cluster (group 2). The results indicate that prevailing isolates in Korea are phylogenetically more closely related to each other rather than other reference strains. Interestingly, the tree topology based on the nucleotide sequences representing the S1 domain or the S1 N-terminal region most nearly resembled the full S gene-based phylogenetic tree. Therefore, our data implicates a potential usefulness of the partial S protein gene including the N-terminal region in unveiling genetic relatedness of PEDV isolates.

1. Introduction

Porcine epidemic diarrhea (PED) is a devastating swine disease that is characterized by acute enteritis and lethal watery diarrhea followed by dehydration leading to death with a high mortality rate in piglets (Debouck and Pensaert, 1980; Pensaert and Yeo, 2006; Pijpers et al., 1993). The disease was first recognized in England in 1971 (Oldham, 1972), and since then, outbreaks have been reported in Europe and Asia (Chen et al., 2008; Pensaert et al., 1981; Puranaveja et al., 2009; Takahashi et al., 1983). The causative agent of this disease, the PED virus (PEDV), was first described in 1978 (Pensaert and Debouck, 1978). Thereafter, a suitable cell culture system for PEDV propagation was developed and successfully used for virus isolation (Hofmann and Wyler, 1988). In Korea, PEDV was first recognized in 1992 (Kweon et al., 1993); however, a retrospective study indicated that the virus had already existed as early as 1987 (Park and Lee, 1997). Since the emergence, a periodic vaccination strategy has been applied nationwide to control the disease on pig farms but PEDV outbreaks have continuously occurred causing tremendous economical losses to the Korean swine industry.

PEDV, a member of the family *Coronaviridae*, is an enveloped virus possessing a single-stranded positive-sense approximately 28 kb RNA genome with a 5′ cap and a 3′ polyadenylated tail (Pensaert and Debouck, 1978; Pensaert and Yeo, 2006). The PEDV genome is composed of the 5′ untranslated region (UTR), at least 7 open reading frames (ORF1a, ORF1b, and ORF2–6), and the 3′ UTR (Kocherhans et al., 2001). The two large ORF1a and 1b cover the 5′ two-third of the genome and encode the non-structural replicase genes. The remaining ORFs in the 3′ terminal region code for four major structural proteins, the 150–220 kDa glycosylated spike (S) protein, 20–30 kDa membrane (M) protein, 7 kDa envelop (E) protein, and 58 kDa nucleocapsid (N) protein (Duarte and Laude, 1994; Pensaert and Yeo, 2006).

The S protein of PEDV is a type I membrane glycoprotein composed of 1383 amino acids (aa), which contains a signal peptide (1–24 aa), a large extracellular region, a single transmembrane domain (1334–1356 aa), and a short cytoplasmic tail. In addition,

* Corresponding author. Tel.: +82 53 950 7365; fax: +82 53 955 5522.
E-mail address: changhee@knu.ac.kr (C. Lee).
[1] These authors contributed equally to this work.

0168-1702/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.virusres.2010.01.015

the S protein can be divided into S1 (1–735 aa) and S2 (736–1383 aa) domains based on its homology with S proteins of other coronaviruses (Jackwood et al., 2001; Sturman and Holmes, 1984). As with other coronavirus S proteins, the PEDV S glycoprotein is known to play pivotal roles in interacting with the cellular receptor to mediate viral entry and inducing neutralizing antibodies in the natural host (Bosch et al., 2003; Chang et al., 2002; Godet et al., 1994). Thus, the S glycoprotein would be a primary target for the development of effective vaccines against PEDV. It is also the major envelope glycoprotein found in the virion, which serves as an important viral component in the study to understand genetic relationships among PEDV isolates and the epidemiological status of PEDV in the field (Park et al., 2007b; Puranaveja et al., 2009; Spaan et al., 1988). In the present study, therefore, we sought to determine the full-length sequences of the S glycoproteins of PEDV field isolates in order to investigate genetic relatedness and to understand the diversity and prevalence of PEDV isolates in Korea.

2. Materials and methods

2.1. Sample collection

Small intestine or stool specimens were taken from piglets showing acute enteritis and watery diarrhea on seven different swine farms in the Gyeongbuk province in 2008–2009. Intestinal samples were prepared into 10% suspensions through homogenization with phosphate-buffered saline (PBS). Fecal samples were diluted with PBS to be 10% suspensions. The suspensions were then vortexed and centrifuged for 10 min at 1700 × *g* (Hanil Centrifuge FLETA 5, South Korea). The clarified supernatants were initially subjected to RT-PCR with PEDV N gene-specific primers N-Fwd and N-Rev (Table 1), and nucleotide sequencing in order to confirm PEDV infection. The PEDV-positive supernatants were stored at −80 °C as the field isolate until use.

2.2. RT-PCR, DNA cloning and sequence analysis

In order to determine the full-length S glycoprotein gene sequences of the Korean PEDV isolates, primers were first selected based on published known sequences of reference PEDV strains (Table 1). Two overlapping cDNA fragments spanning the entire S gene were RT-PCR amplified by using gene-specific primer sets. Briefly, viral RNA was extracted from the PBS-diluted or homogenized samples with the RNeasy Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. Reverse transcription (RT) was performed by using 1 μg of viral RNA, specific reverse primers, and SuperScript III First-Strand cDNA Synthesis Kit (Invitrogen, Carlsbad, CA) according to the manual provided by the manufacturer. PCR was conducted to amplify each cDNA fragment from the RT product by using TaKaRa Ex Taq DNA polymerase (TaKaRa, Japan) according to the manufacturer's protocol under the following conditions; denaturation at 94 °C for 5 min, 30 cycles of denaturation at 94 °C for 30 s, annealing at 58 °C for 30 s and extension at 72 °C for 2 min, followed by a final extension at 72 °C for 10 min. The individual cDNA amplicons were gel-purified, cloned into pGEM-T easy (Promega, Madison, WI), and sequenced in both directions using two commercial vector-specific T7 and SP6 primers and the S gene-specific primers. General procedures for DNA manipulation and cloning were performed according to standard procedures (Sambrook and Russell, 2001). The complete S glycoprotein gene sequences of the Korean PEDV isolates were deposited to the GenBank database under their accession numbers (Table 2).

2.3. Multiple alignments and phylogenetic analyses

The sequences of 17 fully and 8 partially sequenced S genes of PEDV isolates were independently used in sequence alignments and phylogenetic analyses (Table 2). The multiple-sequencing alignments were generated with ClustalX 2.0 program (Thompson et al., 1997), and percentage nucleotide sequence divergences were further assessed by using the same software application. Phylogenetic trees were constructed from the aligned nucleotide sequences by using the neighbor-joining method and subsequently subjected to bootstrap analysis with 1000 replicates in order to determine percentage reliability values on each internal node of the tree (Saitou and Nei, 1987). All tree figures were produced by the Mega 4.0 software (Kumar et al., 2004). A similarity plot was generated by Simplot ver. 3.5.1 (Lole et al., 1999) to compare the full S genes of the PEDV isolates with those of the published reference strains.

3. Results

3.1. Nucleotide and amino acid sequence analyses

The full-length nucleotide sequences of S glycoprotein genes of PEDV isolates were determined to investigate their genetic characterization. The sequence data found that the S genes of seven Korean PEDV isolates (KNU-0801, KNU-0802, and KNU-0901–KNU-0905) consisted of 4161 nucleotides (nt) in length, which encoded 1386 amino acid residues. The complete nucleotide sequences of the Korean PEDV isolates were further compared to those of reference PEDV strains described in Table 2. The full-length S genes of the Korean PEDV isolates were determined to be 9 nt (3 amino acids) longer in size than that of other PEDV strains. This consequence was due to the presence of the number of inserted or deleted nucleotides that was accumulated in the N-terminus of the S1 domain (Fig. 1). Remarkably, the first notable insertion existed at amino acid positions 56–59 of the S protein for all seven Korean PEDV isolates followed by the presence of one additional inserted aspartic acid residue at position 140. Interestingly, the unique deletion was also located between amino acids 156 and 157 of the PEDV S protein, which is commonly present in reference PEDV strains. The S proteins of seven Korean PEDV isolates contained 25–27 potential *N*-linked glycosylation sites (identified by the sequence NXS/T, where X is any amino acid except proline). Deletion or modification of *N*-glycosylation sites was found predominantly in the N-terminus of the S1 domain. Furthermore, the largest number of amino acid differences was also clustered in the N-terminal region of the S1 domain. In addition, the genetic distances between the query strain CV777 and the Korean field isolates were plotted against the nucleotide position. Likewise, the N-terminal region of the S1 domain (1–800 bp) showed the lowest similarity with the PEDV reference strain (Fig. 2). Taken together, our data indicate that the S1 domain including the peptide signal sequences of the Korean field isolates contain regions that exhibited sequence hypervariablity.

The similarity between the S protein genes was measured, and the sequence homology results are described in Table 3. The comparison of nucleotide sequences showed variable similarities among the seven Korean PEDV isolates, ranging from 94.8% to 99.7% identity. The sequence analysis was further extended to the deduced amino acid level and exhibited 93.6–99.6% identity with each other. In addition, all isolates shared diverse identities with the first Korean field isolate Chinju99, showing 93.8–97.7% homology at the nucleotide sequence level and accordingly, the percent identity ranged between 91.8% and 96.8%. However,