

Origin, Evolution, and Genotyping of Emergent Porcine Epidemic Diarrhea Virus Strains in the United States

Yao-Wei Huang,^{a,b} Allan W. Dickerman,^c Pablo Piñeyro,^a Long Li,^d Li Fang,^d Ross Kiehne,^e Tanja Opriessnig,^{f,g} Xiang-Jin Meng^a

Department of Biomedical Sciences and Pathobiology, College of Veterinary Medicine, Virginia Polytechnic Institute and State University, Blacksburg, Virginia, USA^a; College of Animal Sciences, Zhejiang University, Hangzhou, China^b; Virginia Bioinformatics Institute, Virginia Polytechnic Institute and State University, Blacksburg, Virginia, USA^c; Hangzhou Beta Veterinary Diagnostic Laboratory, Hangzhou, China^d; Swine Veterinary Center, P.A., St. Peter, Minnesota, USA^e; Department of Veterinary Diagnostic and Production Animal Medicine, College of Veterinary Medicine, Iowa State University, Ames, Iowa, USA^f; The Roslin Institute and The Royal (Dick) School of Veterinary Studies, University of Edinburgh, Edinburgh, United Kingdom^g

ABSTRACT Coronaviruses are known to infect humans and other animals and cause respiratory and gastrointestinal diseases. Here we report the emergence of porcine epidemic diarrhea virus (PEDV) in the United States and determination of its origin, evolution, and genotypes based on temporal and geographical evidence. Histological lesions in small intestine sections of affected pigs and the complete genomic sequences of three emergent strains of PEDV isolated from outbreaks in Minnesota and Iowa were characterized. Genetic and phylogenetic analyses of the three U.S. strains revealed a close relationship with Chinese PEDV strains and their likely Chinese origin. The U.S. PEDV strains underwent evolutionary divergence, which can be classified into two sublineages. The three emergent U.S. strains are most closely related to a strain isolated in 2012 from Anhui Province in China, which might be the result of multiple recombination events between different genetic lineages or sublineages of PEDV. Molecular clock analysis of the divergent time based on the complete genomic sequences is consistent with the actual time difference, approximately 2 to 3 years, of the PED outbreaks between China (December 2010) and the United States (May 2013). The finding that the emergent U.S. PEDV strains share unique genetic features at the 5'-untranslated region with a bat coronavirus provided further support of the evolutionary origin of PEDV from bats and potential cross-species transmission. The data from this study have important implications for understanding the ongoing PEDV outbreaks in the United States and will guide future efforts to develop effective preventive and control measures against PEDV.

IMPORTANCE The sudden emergence of porcine epidemic diarrhea virus (PEDV), a coronavirus, for the first time in the United States causes significant economic and public health concerns. Since its recognition in May 2013, PEDV has rapidly spread across the United States, resulting in high mortality in piglets in more than 17 States now. The ongoing outbreaks of Middle East respiratory syndrome coronavirus in humans from countries in or near the Arabian Peninsula and the historical deadly nature of the 2002 outbreaks of severe acute respiratory syndrome coronavirus create further anxiety over the emergence of PEDV in the United States due to the lack of scientific information about the origin and evolution of this emerging coronavirus. Here we report the detailed genetic characterization, origin, and evolution of emergent PEDV strains in the United States. The results provide much needed information to devise effective preventive and control strategies against PEDV in the United States.

Received 2 September 2013 Accepted 19 September 2013 Published 15 October 2013

Citation Huang YW, Dickerman AW, Piñeyro P, Li L, Fang L, Kiehne R, Opriessnig T, Meng XJ. 2013. Origin, evolution, and genotyping of emergent porcine epidemic diarrhea virus strains in the United States. *mBio* 4(5):00737-13. doi:10.1128/mBio.00737-13.

Editor Diane Griffin, Johns Hopkins University School of Public Health

Copyright © 2013 Huang et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution-Noncommercial-ShareAlike 3.0 Unported license](https://creativecommons.org/licenses/by-nc-sa/4.0/), which permits unrestricted noncommercial use, distribution, and reproduction in any medium, provided the original author and source are credited.

Address correspondence to Xiang-Jin Meng, xjmeng@vt.edu, and Yao-Wei Huang, yhuang@vt.edu.

Coronaviruses in the subfamily *Coronavirinae* of the family *Coronaviridae* of the order *Nidovirales* are single-stranded, positive-sense RNA viruses with the largest genome that are known to infect humans, other mammals, and birds, usually causing subclinical or respiratory and gastrointestinal diseases (1, 2). Currently, the subfamily *Coronavirinae* is classified into four genera, including *Alphacoronavirus*, *Betacoronavirus*, *Gammacoronavirus*, and *Deltacoronavirus* (1). The first two genera are believed to originate from bats, whereas the last two are believed to have derived from birds (3, 4). The emergences of two highly pathogenic human betacoronaviruses, the severe acute respiratory syndrome coronavirus (SARS-CoV) in 2002 and the Middle East respiratory syndrome coronavirus (MERS-CoV) in 2012, have posed serious

public health concerns over pandemic diseases associated with novel coronaviruses (4–6). While the MERS-CoV currently continues to spread in countries in or near the Arabian Peninsula, most recently, an alphacoronavirus, porcine epidemic diarrhea virus (PEDV) has suddenly emerged in the United States and rapidly spread across the country, resulting in high mortality in infected newborn piglets in more than 17 states in less than 3 months (7).

Porcine epidemic diarrhea (PED) was first recognized as a devastating enteric disease in feeder and fattening pigs, resembling transmissible gastroenteritis (TGE) in pigs in the United Kingdom in 1971. The etiological agent was identified as a coronavirus, PEDV (strain CV777), in Belgium in 1978 (8). The full-length

genomic sequence of the prototype Belgian CV777 strain was determined in 2001 (9), which was more closely related to a *Scotophilus* bat coronavirus (BtCoV) 512/2005 than to other known alphacoronaviruses, such as TGE virus (TGEV) and human coronaviruses 229E and NL63, in phylogeny as well as genome organization (10), suggesting that PEDV and BtCoV/512/2005 had a common evolutionary precursor and that cross-species transmission of coronavirus might have occurred between bats and pigs. Outbreaks of PED have been documented in many European and Asian countries in the past (11). PED has been documented in China since the 1980s; however, variant strains of PEDV associated with large-scale outbreaks of diarrhea with 80 to 100% morbidity and 50 to 90% mortality in suckling piglets have emerged in China since 2010 and pose a serious concern for the swine industry of China (12–14).

Emergence of PEDV in the United States. PEDV has been exotic in the United States until May 2013. Currently PEDV is spreading rapidly in swine farms in the United States, posing significant economic and public health concerns. To determine the origin and evolution of the PEDV outbreaks in the United States, we characterized four PED cases from Minnesota and two cases from Iowa. Clinical signs were characterized by acute vomiting, anorexia, and watery diarrhea, with high mortality in pigs less than 10 days old. For the four dead neonatal piglets from Minnesota, upon microscopic examination, the piglets had signs of emaciation and dehydration. The gross pathological lesions were confined to the small intestine and were characterized by thin translucent intestinal walls that contained moderate amounts of yellow watery feces without macroscopic traces of blood (see Fig. S1A in the supplemental material). No other gross abnormalities were noticed. Histological evaluation revealed regions of small intestines with villus blunting and fusion and minimal lymphoplasmacytic infiltration of the villi of the lamina propria (see Fig. S1B). The gross and histological lesions from the PEDV outbreaks in the United States are similar to those observed in China (14). PEDV RNA was detected in porcine small intestine samples from all four dead piglets from Minnesota by reverse transcription-PCR (RT-PCR) with primers targeting the N gene (data not shown). Additionally, we also tested fecal and small intestinal samples from diseased suckling pigs from two different Iowa farms. The complete genomes of three representative strains of PEDV from the ongoing outbreaks in the United States—one from Minnesota and two from Iowa, designated strains MN and IA1 and IA2, respectively—were amplified by RT-PCR from total RNAs extracted from the fecal or small intestine sample and genetically characterized.

Determination of the full-length genomic sequences of the emergent PEDV strains in the United States. To amplify the complete genomic sequence of the emergent U.S. strains of PEDV, the extreme 5' and 3' termini of the MN strain were first determined by rapid amplification of cDNA ends (RACE). Subsequently, a total of eight overlapping fragments covering the entire PEDV genome for each of the three emergent U.S. PEDV strains were amplified by RT-PCR using PfuUltra II high-fidelity DNA polymerase (Agilent, Santa Clara, CA), with primers based upon the conserved regions among the CV777 and the Chinese PEDV sequences (Fig. 1A; see Table S1 in the supplemental material). The RT-PCR products were individually excised from the agarose gel, purified, and subsequently cloned into a pSC-B-amp/kan vector (Agilent) or a pCR-Blunt vector (Invitrogen). For each ampli-

con, three to five individual clones were sequenced to determine the consensus sequence of any given genomic region. Sequence contigs with the consensus sequence were assembled into the full-length genome for each of the three emergent U.S. PEDV strains using the Lasergene package (DNASar, Inc., Madison, WI).

Nucleotide sequence accession numbers. The complete genome sequences of the three U.S. PEDV strains have been deposited in GenBank under accession no. KF468752 (MN), KF468753 (IA1), and KF468754 (IA2).

Unique genetic features of the emergent PEDV strains in the United States. The three emergent U.S. PEDV genomic sequences have the same size of 28,038 nucleotides (nt), excluding the polyadenosine tail, and share the genome organization with the prototype PEDV CV777 strain characterized by a gene order of 5'-open reading frame 1a/1b (ORF1a/1b)-S-ORF3-E-M-N-3' (Fig. 1B.) These three U.S. PEDV sequences shared 99.8 to 99.9% nucleotide identities. In particular, strains MN and IA2 had only 11 nucleotide differences across the entire genome. To date, there are a total of 23 complete genomic sequences of PEDV available in the GenBank database, and 19 of them were isolated from pigs in China. Multiple sequence alignments among these 23 sequences together with the 3 emergent U.S. PEDV genomes at the nucleotide as well as the amino acid sequence levels of the replicase (ORF1a/1b), S, ORF3, E, M, and N were performed in an attempt to identify any unique nucleotides and amino acids in the 3 emergent U.S. PEDV genomes. We identified a total of 29 nucleotide differences between the United States strains and the consensus sequence (data not shown), which resulted in eight amino acid changes (Fig. 1B). Interestingly, the three emergent U.S. strains all share four additional unique nucleotides (7191A, 10861T, 15594C, and 20343C; two are nonsynonymous changes) (Fig. 1B) with the Chinese AH2012 strain (GenBank accession no. KC210145) and share another four additional unique nucleotides (21869A, 21935T, 23225T, and 25294C [all of them are synonymous]) with the Chinese CH/ZMDZY/11 strain (GenBank accession no. KC196276). Among the eight amino acid changes unique to the U.S. strains plus the two unique to the United States and AH2012, four single amino acid substitutions are located within nonstructural protein 1 (nsp1), nsp13, nsp15, and S protein genes, respectively. The remaining six amino acid substitutions are found in the nsp3 gene, including three in the papain-like proteinase 2 (PLP2) domain and two in the Y domain (Fig. 1B). The identification of these unique nucleotide and amino acid changes can now be used as the genetic markers to differentiate the emergent U.S. strains from the other PEDV strains, especially those from China.

Pairwise comparison of the prototype PEDV CV777 genomic sequence with the other 25 sequences using the mVISTA program (<http://genome.lbl.gov/vista/mvista/submit.shtml>) also revealed that the nsp3 gene had relatively large regions of dissimilarity among different strains (see Fig. S2 in the supplemental material), and the two most dissimilar regions were in the N-terminal domain (NTD) of the S gene and the ORF3 gene (see Fig. S2). The three emergent U.S. strains possess two notable insertions at aa 56 to 59 and 139, respectively, and one deletion between aa 160 and 161, compared to the CV777 strain (Fig. 1C), which has been identified in all of the PEDV strains recently isolated in South Korea and China (13–15).

In addition, like the other newly isolated PEDV strains, the three emergent U.S. strains contain a notable U insertion at nt 48,

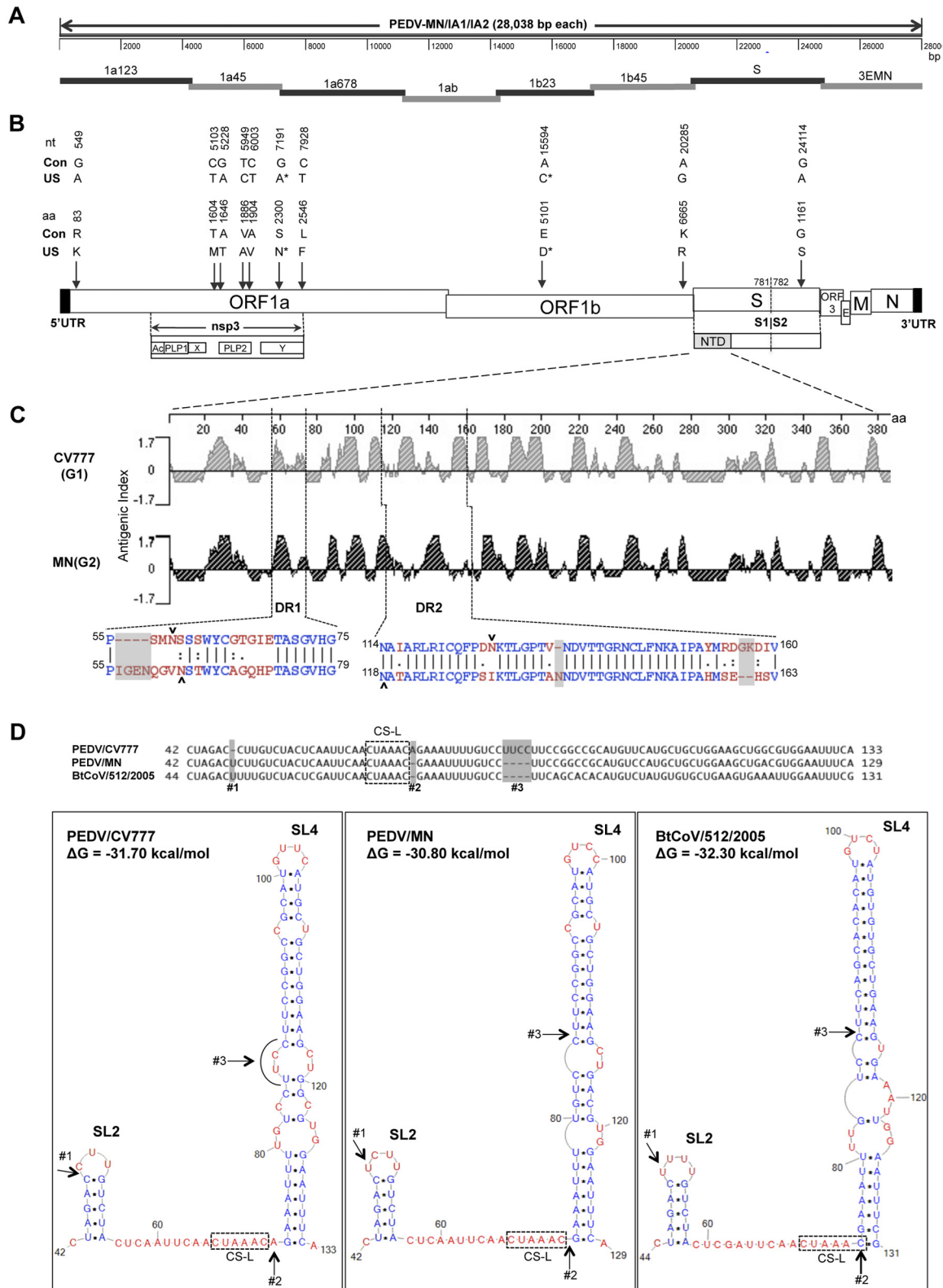


FIG 1 Schematic diagrams of genomic structure, the strategy for genomic cDNA cloning, and molecular characterization of unique features of three emergent U.S. PEDV strains (MN, IA1, and IA2) isolated in Minnesota and Iowa in 2013. (A) Eight overlapping cDNA fragments covering the entire PEDV genome for each of the three U.S. strains, represented by thick lines, were amplified by RT-PCR from total RNAs extracted from fecal or small intestine samples. The names of each fragment are indicated. The location and primer sequences for the RT-PCR and the RACE-PCR used to generate the overlapping RT-PCR products are available in Table S1 in the supplemental material. The numbers on the scale bar indicate distances from the 5' end of the genome. (B) Organization of the PEDV genome and locations of unique amino acid (aa) changes identified in the three U.S. strains. The approximate positions and sizes of genes in the PEDV genome

(Continued)

and two notable nucleotide deletions between nt 72 and 73 (an A deletion) and between nt 84 and 85 (a 4-nt UUCC deletion), respectively, at the 5'-untranslated region (UTR), in comparison with the prototype CV777 strain (Fig. 1D). Only two PEDV strains isolated from the early period, LZC and SM98, had the same sequences to the CV777 strain in this region. The 5'-UTRs of coronaviruses form conserved RNA structural elements that are involved in viral replication, subgenomic RNA (sgRNA) transcription, and translation (16, 17). Unexpectedly, further comparison with the corresponding regions from the other alphacoronaviruses revealed that the bat coronavirus BtCoV/512/2005 isolated in China had the same one insertion and two deletions at the same positions (denoted by no. 1 to 3 in Fig. 1D). These changes do not alter the conserved RNA secondary structures, including stem-loop 2 (SL2) and SL4, which are conserved in all members of *Coronavirinae* (Fig. 1D) (16–18). The U insertion at position 1 is located on the loop of SL2. It has been shown that sequence conservation and the numbers of nucleotides in the predicted SL2 loop play an important role in betacoronavirus replication (17). The A deletion at position 2 immediately follows the core sequences (CUAAAC) of the PEDV leader transcription-regulating sequence (TRS), which may slightly change the efficiency of base pairing of the nascent minus-strand sgRNAs with the positive-strand leader TRS in the discontinuous transcription (19, 20). SL4 functions as a spacer element in mouse hepatitis virus (MHV) to direct sgRNA synthesis (18). A minor change (a 4-nt deletion at position 3) in the stem helix of SL4 might not disturb the spacer role of this structure. However, whether the three changes could collectively alter the efficiency of viral replication and sgRNA synthesis in the emergent U.S. PEDV strains, leading to enhanced pathogenicity, is an important subject for further investigation. The identification of identical unique nucleotide changes of the U.S. PEDV strains and BtCoV/512/2005 is also intriguing. It is possible that the emergent U.S. PEDV and BtCoV/512/2005 strains may have a convergent sequence evolution in this region, although it cannot be ruled out that the newly emergent PEDV strains acquired this segment by recombination from a BtCoV/512/2005-related coronavirus through cross-species transmission. It has never been established whether PEDV represents a genuine swine virus, although it is believed that all the alphacoronaviruses evolutionarily originate from the bat species (3). In fact, an early report had described a PEDV-like virus in minks (21), although further independent verification is needed.

Genotyping of the emergent PEDV strains in the United States. The phylogenetic tree based upon a multiple sequence

alignment of 26 complete PEDV genomes along with the bat coronavirus BtCoV/512/2005 sequence as an outgroup was constructed using the MEGA5.2 program. The PEDV strains fell clearly into two distinct genogroups, designated genogroup 1 (G1) and genogroup 2 (G2) (Fig. 2A). At least three clusters were classified as G1. We designated the first cluster as subgroup 1a, including the prototype CV777 strain and strains LZC and SM98, which had the same SL2 and SL4 sequences at the 5'-UTR. These strains were isolated from the early period. The second cluster, named subgroup 1b, contains five strains—one from South Korea (the DR13 attenuated vaccine strain) and the others from China. (The two strains JS2008 and JS2008/new may have been submitted redundantly.) Although the JS2008 and SD-M strains were claimed to be isolated from infected pigs (SD-M was only passaged on cells for 4 times) (22, 23), we found that both of them had the same characteristic amino acid deletions in the acidic domain of nsp3 (GLPVAPET) and in the C terminus of ORF3 as the other three cell-adapted strains (see Fig. S2 in the supplemental material), suggesting that these two strains were probably isolated from pigs vaccinated with the PEDV attenuated DR13 vaccine or a related vaccine. Therefore, all PEDV strains in subgroup 1b are likely derived from the same source, although some were isolated (SD-M in 2012) or developed as an attenuated vaccine (KC189944 in 2012) after the PED outbreaks in China in late 2010. The 8-aa deletion in nsp3 and the large ORF3 deletion at the C terminus can be used as a genetic signature for this subgroup. The third cluster in G1 consists of the virulent DR13 strain isolated in South Korea and the oldest Chinese PEDV strain, CH/S. We tentatively proposed this cluster as subgroup R, since they may represent recombinants of the other genogroups based upon the phylogenetic relationship (Fig. 2A). Furthermore, these two strains showed a combination VISTA profile compared to the other subgroup strains (see Fig. S2), indicating that this subgroup likely contributed to the diversity of new PEDV strains during the evolution of PEDV by potential recombination events.

The proposed G2 PEDV consists of two subgroups designated 2a and 2b (Fig. 2A). With the exception of the three emergent U.S. strains, all of the other PEDV strains in G2 were isolated in China during 2011 to 2012. The U.S. strains and a Chinese strain, AH2012, share several unique nucleotides, as described above, and are also clustered together in a separate clade within subgroup 2a, suggesting that AH2012 or a closely related strain may be the origin of the PEDV emergence in the United States. Strains MN and IA2 had 99.6% and strain IA1 had 99.5% nucleotide identity with AH2012, respectively. While preparing the manuscript, the

Figure Legend Continued

that correspond to the scale bar are shown in panel A. The putative S1/S2 boundary (amino acid positions) of the S protein is also shown. Nucleotide (nt) and amino acid differences between the 3 emergent U.S. PEDV sequences and the consensus sequences of 23 other known PEDV strains and their positions are depicted. Con, PEDV consensus sequences; US, unique sequences in three U.S. PEDV strains; *, unique amino acids and nucleotides shared by the three U.S. strains and one Chinese AH2012 strain; UTR, untranslated region; nsp, nonstructural protein; Ac, acidic domain; PLP, papain-like proteinase; X, X domain (ADP-ribose-1'-phosphatase [ADRP]); Y, unknown Y domain; NTD, N-terminal domain of the spike gene. (C) Comparison of antigenic index profiles of the NTD (aa 1 to 380) of S protein between the prototype strain CV777 (genogroup 1 [G1]) and U.S. strain MN (genogroup 2 [G2]). The corresponding alignment of amino acid sequences and positions of the two regions containing amino acid deletions/insertions (indicated by dashes and shaded) are shown. DR, region containing deletions. Identical amino acids are marked in blue, whereas mismatches are marked in red. Favorable amino acid mismatches are displayed as colons, whereas neutral mismatches are depicted as periods. The predicted N-linked glycosylation sites in DR1 and DR2 are indicated by arrowheads. (D) Comparisons of the primary sequences and the predicted secondary structures of a 5'-proximal region in the 5'-UTR between Belgian PEDV/CV777 (nt 42 to 133), U.S. PEDV/MN (nt 42 to 129), and a bat coronavirus BtCoV/512/2005 (nt 44 to 131). Three deletions of nucleotides (no. 1 to 3) are indicated by dashes and shaded in the primary sequences and are indicated by arrows in the secondary structures. The core sequences (CUAAAC) of leader transcription-regulating sequences (CS-L) are boxed by dashed lines. Two stem-loops (SL2 and SL4) conserved in all coronaviruses are indicated. (There is no SL3 in alphacoronaviruses.) Nucleotides in the stems are marked in blue, whereas nucleotides in the loops are marked in red.

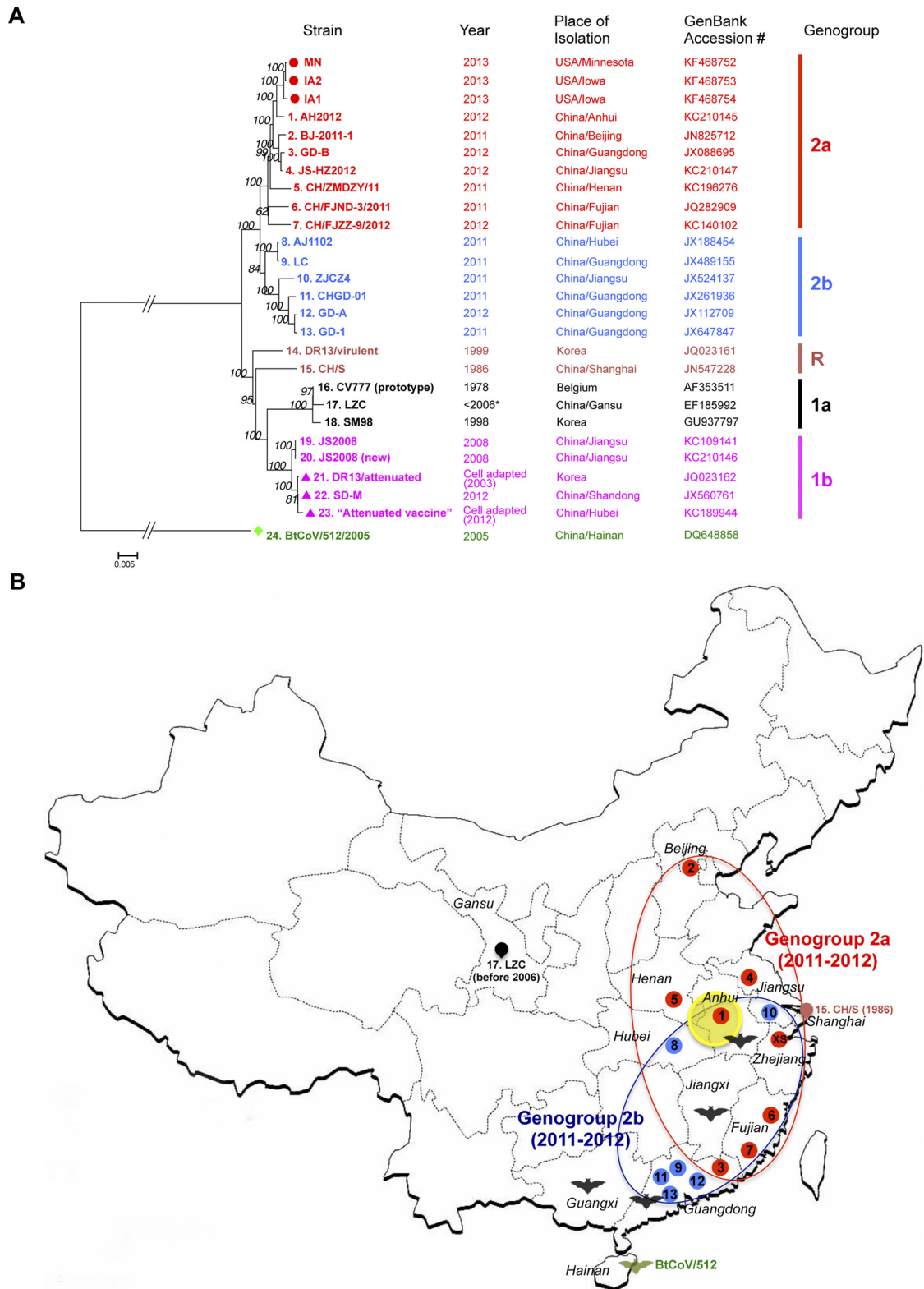


FIG 2 (Continued)

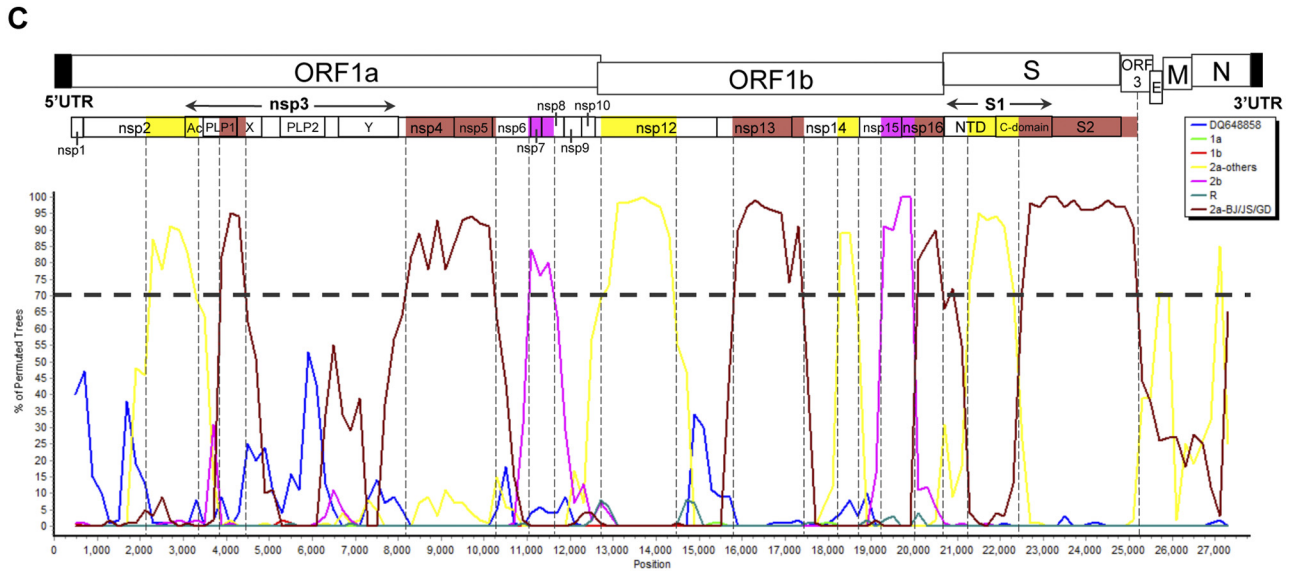


FIG 2 Genotyping and origin of the emergent U.S. PEDV strains based on full-length genomic sequence analyses. (A) Phylogeny-based genotyping of 26 PEDV strains with available complete genomic sequences, including the 3 U.S. PEDV strains. The tree was constructed by the neighbor-joining method, based upon the full-length genomic nucleotide sequences using the bat coronavirus BtCoV/512/2005 sequence as an outgroup. Bootstrap values are indicated for each node from 1,000 resamplings. The names of the strains, years and places of isolation, GenBank accession numbers, and genogroups and subgroups proposed in this study are shown. (An asterisk indicates that the isolation year of LZC is unknown but should be before 2006 according to the GenBank submission date.) Red solid circles, the three U.S. PEDV strains; purple solid triangles, cell-culture-adapted PEDV strains or vaccines; green solid diamonds, bat coronavirus BtCoV/512/2005. (B) Phylogeny-based geographical dissection of genogroup 2 Chinese PEDV strains. The map of China shows all of the provinces where genogroup 2 PEDV strains with the available complete genomic sequences were isolated. The numbers in order and the colors for PEDV strains correspond to those labeled in panel A: genogroup 2a strains are in red, and genogroup 2b strains are in blue. The coverage area for each subgroup (depicted by the red or the blue oval) is deduced based on the distributions of the strains. XS indicates a representative strain (XS2012) isolated in Zhejiang Province of eastern China that belongs to genogroup 2a based on the sequence of the S gene from this study. The yellow shaded circle indicates the hypothetical location of the origin of the U.S. PEDV, where the closely related AH2012 strain was identified. The two early PEDV strains CH/S and LZC and their locations are also shown. Five provinces where bat coronaviruses phylogenetically related to PEDV were isolated (10) are also marked by the “bat” symbols. In particular, the BtCoV/512/2005 strain, isolated in Hainan Province, is marked in green. (C) Bootsacan analysis for possible recombination events of lineage US-AH strains (three U.S. strains plus the AH2012 strain) in subgroup 2a as the query group throughout the genome compared to the other two lineages. 2a-BJ/JS/GD includes strains BJ-2011-1, GD-B and JS-HZ2012 (denoted by the brown line), and “2a-others” includes CH/ZMDZY/11, CH/FJND-3/2011, and CH/FLZZ-9/2012 (denoted by the yellow line) in the same subgroup (based upon the phylogenetic tree constructed in panel A), subgroup 1a (green line), 1b (red line), R (cyan line), 2b (purple line), and BtCoV/512/2005 (accession no. DQ648858 [blue line]). Bootscaffing was conducted by SimPlot (version 3.5.1; window size: 1,000 bp; step, 200 bp), and the cutoff value of bootstrap support for clustering was set to 70. The putative recombinant regions corresponding to functional domains in the PEDV genome are shown at the top.

sequences of three additional U.S. PEDV strains (CO/13, IA2013, and IN2013) were reported online without in-depth analyses (24, 25). These 3 additional strains were reported to have 99.5% identities with the AH2012 strain. Therefore, the U.S. PEDV strains underwent evolutionary divergence and have been further classified into two sublineages: MN-IA2 and IA1-CO/13 (or possible IA1-CO/13-IA2013-IN2013).

The significant amino acid changes at the NTD of the S gene differentiate G1 and G2 (Fig. 1C). Comparison of antigenic index profiles of the NTD between the two genogroups indicated that the second deletion region (DR2) appears to have a higher degree of antigenic change than DR1 and has the distinct N-linked glycosylation site, although DR1 has higher sequence variability (Fig. 1C). Therefore, the attenuated PEDV vaccines based on the historical CV777-derived G1a strains or DR13-derived G1b strains may be antigenically less related to the newly emergent G2 PEDV strains that have antigenic variations in the NTD.

Temporal and geographical evidence for the origin and evolution of the emergent PEDV strains in the United States. Molecular clock analysis was conducted to trace the temporal patterns of PEDV evolution leading to the recent PED outbreaks in the

United States by using the BEAST program under both constant and exponentially growing population size models (for details, see Table S2 in the supplemental material and the methods described above). We focused on two landmarks: the earliest age of intra-United States PEDV divergence and the latest age of putative Asia-to-United States PEDV divergence. These two landmarks should bracket the time at which the virus was transferred from Asia to the United States. The results dated the latest divergence of PEDV from China to United States to approximately 5 to 6 years ago and dated the first divergence of PEDV within the United States to about one and a half years ago. This estimation (between 5 and 6 to ~1.5 years from the present) is consistent with the actual time difference—approximately 2 to 3 years—of the PED outbreaks between China (late 2010) and the United States (late spring 2013).

The possible geographical dissection of the Chinese G2 PEDV strains was plotted to search for a clue regarding the origin of the emergent PEDV strains in the United States (Fig. 2B). It was found that the seven G2a strains were distributed throughout the eastern region of China, from Beijing to Guangdong Province, whereas the G2b strains were mainly located in southeastern China and

have not been identified in northern China. The Chinese AH2012 strain, which is most closely related to the emergent U.S. strains, was isolated in Anhui Province, which is located in the overlapping geographical region where G2a and G2b are circulating (Fig. 2B). The AH2012-like virus was possibly transmitted to the eastern China regions and then transported to the United States. To determine whether the AH2012-related PEDV (designated “lineage US-AH”) could be identified in Zhejiang Province, an adjacent province to the east of Anhui Province, we performed sequencing and sequence analyses of the S genes of PEDV strains isolated since 2012 from different pig farms in Zhejiang Province (represented by the XS2012 strain; GenBank accession no. KF468755), and the results showed that the strains from Zhejiang Province were almost identical, belonging to subgroup 2a but forming a clade distinct from AH2012 and the three U.S. strains (data not shown), suggesting that the AH2012-related PEDV has probably not spread into this area in China.

We also hypothesized that the PEDV in the lineage US-AH may have originated from recombination events between different subgroups of PEDV due to the geographical location of the AH2012 strain and possible cocirculation of different PEDV G2 lineages. Bootscan analysis showed a clearly mosaic structure when the consensus genome of lineage US-AH was used as query against BtCoV/512/2005, subgroups G1a, G1b, and G2b and the other lineages in G2a (Fig. 2C), indicative of multiple recombination events between different PEDV lineages in G2. Interestingly, the BtCoV/512/2005-like bat coronaviruses, which are phylogenetically related to PEDV, were also identified in five provinces in southern and eastern China, including Anhui Province as described previously (Fig. 2C) (10). Taken together, the available sequence and phylogenetic data indicate that the PEDV strains emerging in the United States originated from China.

In summary, we report here the emergence of PEDV in the United States and detailed genetic and phylogenetic analyses of the complete genomic sequences of three emergent U.S. PEDV strains from Minnesota and Iowa. The findings that the emergent U.S. PEDV strains are most closely related to Chinese G2a strains suggest that the emergence of the PEDV in the United States likely originated from China. However, the exact source of the origin is difficult to identify at this point. The finding that the emergent PEDV strains in the United States share unique genetic features with a bat coronavirus further suggests a possible evolutionary origin of PEDV from bat species and potential cross-species transmission. The information presented in this study will guide the current control measures to stop the ongoing spread of PEDV in the United States and also provides important clues for the development of an effective vaccine against the emergent PEDV strains.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at <http://mbio.asm.org/lookup/suppl/doi:10.1128/mBio.00737-13/-/DCSupplemental>.

Figure S1, PDF file, 1.6 MB.

Figure S2, PDF file, 0.8 MB.

Table S1, DOCX file, 0.1 MB.

Table S2, DOCX file, 0.1 MB.

REFERENCES

- de Groot RJ, Baker SC, Baric R, Enjuanes L, Gorbalenya AE, Holmes KV, Perlman S, Poon L, Rottier PJM, Talbot PJ, Woo PCY, Ziebuhr J. 2011. Coronaviridae, p 806–828. *In* King AMQ, Adams MJ, Carstens EB, Lefkowitz EJ (ed), *Virus taxonomy: ninth report of the International Committee on Taxonomy of Viruses*. Elsevier Academic, London, United Kingdom.
- Brian DA, Baric RS. 2005. Coronavirus genome structure and replication. *Curr. Top. Microbiol. Immunol.* 287:1–30.
- Woo PC, Lau SK, Lam CS, Lau CC, Tsang AK, Lau JH, Bai R, Teng JL, Tsang CC, Wang M, Zheng BJ, Chan KH, Yuen KY. 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.
- Graham RL, Baric RS. 2010. Recombination, reservoirs, and the modular spike: mechanisms of coronavirus cross-species transmission. *J. Virol.* 84:3134–3146.
- van Boheemen S, de Graaf M, Lauber C, Bestebroer TM, Raj VS, Zaki AM, Osterhaus AD, Haagmans BL, Gorbalenya AE, Snijder EJ, Fouchier RA. 2012. Genomic characterization of a newly discovered coronavirus associated with acute respiratory distress syndrome in humans. *mBio* 3(6):e00473-12. doi:10.1128/mBio.00473-12.
- Ksiazek TG, Erdman D, Goldsmith CS, Zaki SR, Peret T, Emery S, Tong S, Urbani C, Comer JA, Lim W, Rollin PE, Dowell SF, Ling AE, Humphrey CD, Shieh WJ, Guarner J, Paddock CD, Rota P, Fields B, DeRisi J, Yang JY, Cox N, Hughes JM, LeDuc JW, Bellini WJ, Anderson LJ, SARS Working Group. 2003. A novel coronavirus associated with severe acute respiratory syndrome. *N. Engl. J. Med.* 348:1953–1966.
- Mole B. 2013. Deadly pig virus slips through US borders. *Nature* 499:388. doi:10.1038/499388a.
- Pensaert MB, de Bouck P. 1978. A new coronavirus-like particle associated with diarrhea in swine. *Arch. Virol.* 58:243–247.
- Kocherhans R, Bridgen A, Ackermann M, Tobler K. 2001. Completion of the porcine epidemic diarrhoea coronavirus (PEDV) genome sequence. *Virus Genes* 23:137–144.
- Tang XC, Zhang JX, Zhang SY, Wang P, Fan XH, Li LF, Li G, Dong BQ, Liu W, Cheung CL, Xu KM, Song WJ, Vijaykrishna D, Poon LL, Peiris JS, Smith GJ, Chen H, Guan Y. 2006. Prevalence and genetic diversity of coronaviruses in bats from China. *J. Virol.* 80:7481–7490.
- Song D, Park B. 2012. Porcine epidemic diarrhoea virus: a comprehensive review of molecular epidemiology, diagnosis, and vaccines. *Virus Genes* 44:167–175.
- Bi J, Zeng S, Xiao S, Chen H, Fang L. 2012. Complete genome sequence of porcine epidemic diarrhoea virus strain AJ1102 isolated from a suckling piglet with acute diarrhoea in China. *J. Virol.* 86:10910–10911.
- 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 diarrhoea virus in China. *Virol. J.* 9:195. doi:10.1186/1743-422X-9-195.
- Li W, Li H, Liu Y, Pan Y, Deng F, Song Y, Tang X, He Q. 2012. New variants of porcine epidemic diarrhoea virus, China, 2011. *Emerg. Infect. Dis.* 18:1350–1353.
- Lee DK, Park CK, Kim SH, Lee C. 2010. Heterogeneity in spike protein genes of porcine epidemic diarrhoea viruses isolated in Korea. *Virus Res.* 149:175–182.
- Raman S, Brian DA. 2005. Stem-loop IV in the 5′ untranslated region is a *cis*-acting element in bovine coronavirus defective interfering RNA replication. *J. Virol.* 79:12434–12446.
- Liu P, Li L, Millership JJ, Kang H, Leibowitz JL, Giedroc DP. 2007. A U-turn motif-containing stem-loop in the coronavirus 5′ untranslated region plays a functional role in replication. *RNA* 13:763–780.
- Yang D, Liu P, Giedroc DP, Leibowitz J. 2011. Mouse hepatitis virus stem-loop 4 functions as a spacer element required to drive subgenomic RNA synthesis. *J. Virol.* 85:9199–9209.
- Zúñiga S, Sola I, Alonso S, Enjuanes L. 2004. Sequence motifs involved in the regulation of discontinuous coronavirus subgenomic RNA synthesis. *J. Virol.* 78:980–994.
- Sola I, Moreno JL, Zúñiga S, Alonso S, Enjuanes L. 2005. Role of nucleotides immediately flanking the transcription-regulating sequence core in coronavirus subgenomic mRNA synthesis. *J. Virol.* 79:2506–2516.
- Have P, Moving V, Svansson V, Uttenthal A, Bloch B. 1992. Coronavirus infection in mink (*Mustela vison*). Serological evidence of infection with a coronavirus related to transmissible gastroenteritis virus and porcine epidemic diarrhoea virus. *Vet. Microbiol.* 31:1–10.
- Zhao M, Sun Z, Zhang Y, Wang G, Wang H, Yang F, Tian F, Jiang S.

2012. Complete genome sequence of a Vero cell-adapted isolate of porcine epidemic diarrhea virus in eastern China. *J. Virol.* **86**:13858–13859.
23. Li B, Liu H, He K, Guo R, Ni Y, Du L, Wen L, Zhang X, Yu Z, Zhou J, Mao A, Lv L, Hu Y, Yu Y, Zhu H, Wang X. 2013. Complete genome sequence of a recombinant porcine epidemic diarrhea virus strain from eastern China. *Genome Announc.* **1**(2):e00105-13. doi:[10.1128/genomeA.00105-13](https://doi.org/10.1128/genomeA.00105-13).
24. Marthaler D, Jiang Y, Otterson T, Goyal S, Rossow K, Collins J. 2013. Complete genome sequence of porcine epidemic diarrhea virus strain USA/Colorado/2013 from the United States. *Genome Announc.* **1**(4):e00555-13. doi:[10.1128/genomeA.00555-13](https://doi.org/10.1128/genomeA.00555-13).
25. Stevenson GW, Hoang H, Schwartz KJ, Burrough ER, Sun D, Madson D, Cooper VL, Pillatzki A, Gauger P, Schmitt BJ, Koster LG, Killian ML, Yoon KJ. 2013. Emergence of porcine epidemic diarrhea virus in the United States: clinical signs, lesions, and viral genomic sequences. *J. Vet. Diagn. Invest.* **25**:649–654.