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.

Preventive Veterinary Medicine 123 (2016) 175–184

Contents lists available at ScienceDirect

Preventive Veterinary Medicine

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

Genomic and evolutionary inferences between American and global strains of porcine epidemic diarrhea virus

Matthew C. Jarvis[a], Ham Ching Lam[a], Yan Zhang[b], Leyi Wang[b], Richard A. Hesse[c], Ben M. Hause[c], Anastasia Vlasova[d], Qiuhong Wang[d], Jianqiang Zhang[e], Martha I. Nelson[f], Michael P. Murtaugh[g], Douglas Marthaler[a,*]

[a] Department of Veterinary Population Medicine, College of Veterinary Medicine, University of Minnesota, St. Paul, MN, USA
[b] Animal Disease Diagnostic Laboratory, Ohio Department of Agriculture, Reynoldsburg, OH, USA
[c] Kansas State Veterinary Diagnostic Laboratory and Department of Diagnostic Medicine and Pathobiology, Kansas State University, Manhattan, KS, USA
[d] Food Animal Health Research Program, Department of Veterinary Preventive Medicine, Ohio Agricultural Research and Development Center, College of Food Agriculture and Environmental Sciences, The Ohio State University, Wooster, OH, USA
[e] Department of Veterinary Diagnostic and Production Animal Medicine, College of Veterinary Medicine, Iowa State University, Ames, IA, USA
[f] Division of International Epidemiology and Population Studies, Fogarty International Center, National Institutes of Health, Bethesda, MD, USA
[g] Department of Veterinary and Biomedical Sciences, College of Veterinary Medicine, University of Minnesota, St. Paul, MN, USA

ARTICLE INFO

Article history:
Received 2 April 2015
Received in revised form
29 September 2015
Accepted 22 October 2015

Keywords:
Porcine epidemic diarrhea virus
Molecular analysis
Bayesian analysis
Complete genome

ABSTRACT

Porcine epidemic diarrhea virus (PEDV) has caused severe economic losses both recently in the United States (US) and historically throughout Europe and Asia. Traditionally, analysis of the spike gene has been used to determine phylogenetic relationships between PEDV strains. We determined the complete genomes of 93 PEDV field samples from US swine and analyzed the data in conjunction with complete genome sequences available from GenBank ($n = 126$) to determine the most variable genomic areas. Our results indicate high levels of variation within the ORF1 and spike regions while the C-terminal domains of structural genes were highly conserved. Analysis of the Receptor Binding Domains in the spike gene revealed a limited number of amino acid substitutions in US strains compared to Asian strains. Phylogenetic analysis of the complete genome sequence data revealed high rates of recombination, resulting in differing evolutionary patterns in phylogenies inferred for the spike region versus whole genomes. These finding suggest that significant genetic events outside of the spike region have contributed to the evolution of PEDV.

1. Introduction

Porcine epidemic diarrhea virus (PEDV) causes diarrhea, vomiting, and dehydration, leading to high mortality (up to 100%) in suckling piglets. PEDV was first discovered in the United Kingdom in 1971, and later was found in Belgium, Hungary, France, Italy, and the Czech Republic (Chasey and Cartwright, 1978; Fan et al., 2012). In 1986, PEDV was first reported in China, and proceeded to spread throughout Asia (Cui, 1990; Song and Park, 2012). In late 2010, a "variant" PEDV strain with increased pathogenesis compared to the classical strains was reported in China, and is regarded as the first of the pandemic strains (Sun et al., 2012). Highly pathogenic PEDV was first detected in the United States (US) in April 2013, and rapidly spread to 31 states, as well as Mexico and Canada (Oka et al., 2014; Ojkic et al., 2015). In December 2013, a second PEDV strain OH851, initially called "US variant" (later renamed S-INDEL), was detected in the US, which shared the same insertions and deletions in the S1 region as the classical PEDV strain, CV777 (Wang et al., 2014; EFSA AHAW Panel, 2014). The S-INDEL strains were reported with lower virulence in the field and were later identified in June 2013 upon a retroactive study (Vlasova et al., 2014). To date, the highly virulent PEDV has been reported in Colombia, Dominican Republic, Japan, Peru, Germany, Portugal, South Korea, and Ukraine (EFSA AHAW Panel, 2014; Lee and Lee, 2014; Hanke et al., 2015; Murakami et al., 2015; Mesquita et al., 2015).

* Corresponding author at: Veterinary Population Medicine, College of Veterinary Medicine, University of Minnesota, 1333 Gortner Ave, Saint Paul, MN 55108, USA.
E-mail address: marth027@umn.edu (D. Marthaler).

http://dx.doi.org/10.1016/j.prevetmed.2015.10.020

PEDV is a single-stranded, positive sense RNA virus belonging to the family *Coronaviridae*, genus *Alphacoronavirus*. The PEDV genome is approximately 28 kb in length and roughly two-thirds of the genome consists of open reading frame (ORF) 1, which encodes 16 non-structural proteins (nsps) (Lai et al., 2013). These nsps play important roles in viral replication, post-translational processing, and immune evasion (Lai et al., 2013). The virus produces various structural proteins, including spike, membrane, and nucleocapsid (Lai et al., 2013). The spike protein is crucial to cell attachment and infection, and the envelope is an integral membrane protein, aiding in membrane fusion while the nucleocapsid protein is necessary for genomic packaging (Hagemeijer and de Haan, 2015). In addition, the PEDV genome includes ORF3, located between the spike and membrane genes, that encodes an ion channel protein possibly associated with PEDV pathogenesis (Park et al., 2008; Wang et al., 2012).

Researchers have explored various regions of the coronavirus (CoV) genome to link specific areas with virulence and host cell attachment. For example, the spike gene codes for a viral attachment protein that can be divided into the S1 (1-789 aa) and S2 (790-1383 aa) regions (Song and Park, 2012). Comparative analysis of transmissible gastroenteritis virus (TGEV), porcine respiratory coronavirus (PRCV), and murine hepatitis virus (MHV) revealed two main antigenic sites in the S1 region: the N-terminal domain (NTD) and the C-terminal receptor binding domain (RBD) (Li et al., 2007). While both domains can influence virus infectivity, such as in TGEV, one domain tends to be central to a CoV's tropism: the NTD is important for MHV tropism, and the RBD is central to PEDV infectivity and virulence (Reguera et al., 2012). The NTD can bind to various sialic acids on the host cell surface (Reguera et al., 2012). The RBD contains residues that bind to the porcine aminopeptidase-N (pAPN), the host receptor utilized by TGEV and PEDV (Delmas et al., 1992).

Since the last large-scale North American PEDV outbreak ended in the spring of 2014, the complete genomes of 93 PEDV strains from the US were sequenced and analyzed to further understand the origin and phylogenetic relationships among the American and global PEDV strains. In-depth nucleotide and amino acid analysis was conducted to identify genes of high diversity. Bayesian analysis was performed to understand the evolution of PEDV and the emergence of different clades within US strains. In addition, the RBD was modeled to visualize the differences between American and Asian strains to better understand how changes in the RBD might affect vaccine efficacy and development.

2. Materials and methods

2.1. Sample acquisition

Samples were routinely submitted to the University of Minnesota Veterinary Diagnostic Laboratory (UMVDL) for pathogen detection. Between January 2014 and December 2014 samples were screened for PEDV by real time RT-PCR (Vlasova et al., 2014). Samples for complete genome sequencing were selected based on the criteria of a high viral concentration from the RT-PCR results and geographical diversity within the US. A total of 83 samples, including fecal ($n = 38$), intestinal homogenate ($n = 21$), fecal swab ($n = 10$), oral fluid ($n = 5$), feedback ($n = 4$), and environmental ($n = 5$) samples were selected for complete genome sequencing using next generation sequencing (NGS) techniques as previously described (GenBank numbers KR265759–KR265834, KR265840–KR265846) (Marthaler et al., 2013; Marthaler et al., 2014). Whole genomic PEDV sequences obtained using NGS techniques were also generously supplied from Iowa State University ($n = 7$, GenBank numbers KM975735–KM975741) and the Ohio Department of Agriculture ($n = 3$, GenBank numbers KP641661–KP641663), using previously described methods (Wang et al., 2014; Chen et al., 2014).

2.2. Nucleotide and amino acid analysis

Using the complete PEDV genome sequences from this study ($n = 93$) and the available PEDV sequences from GenBank ($n = 126$), two nucleotide alignments were created and analyzed to determine the phylogenetic relationships between American and global PEDV sequences: the concatenation of all ORFs (ORF1, S, ORF3, envelope, membrane, and nucleocapsid), and a S1 alignment. Vaccine and cell-passaged strains were excluded from the analysis (Table S1). Nucleotide and amino acid entropy analyses were performed using the MATLAB software (MATLAB v8.0 and Statistics Toolbox v8.1, The MathWorks, Inc., Natick, MA, USA). Threshold values were determined using previously published methods (Shannon, 1948; Litwin and Jores, 1992).

2.3. Recombination and bayesian analysis (BEAST)

Recombination analysis was performed using the Recombination Detection Program (RDP) v4, which uses multiple detection algorithms, including the RDP Method, GENECOV, and MAXCHI, to check for the presence of recombinant sequences in the sequence dataset (Martin et al., 2015). Window size was set to 100 bp. Breakpoints, the presence of major/minor donor sequences, and confidence intervals were used to determine regions that required excision from the alignment, or if entire sequences needed to be removed from the analysis due to multiple recombination events within the sequence. Recombinant sequences were removed only prior to the Bayesian analysis, but remained in the alignments for all entropy analysis and molecular modeling.

Bayesian Markov Chain Monte Carlo (MCMC) approach using BEAST v1.8.1, with a relaxed molecular clock and Bayesian skyline population (BSP) prior, with a general-time reversible nucleotide substitution and gamma distributed among-site rate variation was used to infer time-scaled phylogeny (Drummond et al., 2002, 2005, 2006, 2012Drummond and Rambaut, 2007; Minin et al., 2008; Drummond and Suchard, 2010). The MCMC chain was run for 800 million generations, with sub-sampling every 80,000 iterations. A Maximum clade credibility (MCC) tree was created by discarding the initial 10% of the chains and summarized in TreeAnnotator (v.1.8.0). Key nodes were identified using FigTree (v.1.4.2) to determine time to most recent common ancestor (tMRCA).

2.4. Molecular modeling

The putative pAPN receptor-binding residues were analyzed to determine residue trends between classical and pandemic strains (Reguera et al., 2012). The C-terminal RBD within the S1 region of the spike gene was modeled using the open-source modeling server SWISS-MODEL provided by the Swiss Institute of Bioinformatics (Biasini et al., 2014). Predicted tertiary structure of the PEDV pAPN RBD was modeled using PRCV as a template since a PEDV template was not available. Spike monomer and trimer models were developed using a theoretical SARS-CoV model as a template (Bernini et al., 2004). Illustrations were created using the open-source Java-based molecular viewer Jmol (Herraez, 2006) and the Python-based molecular viewer PyMOL (The PyMOL Molecular Graphics System, Version 1.7.4 Schrödinger, LLC.).

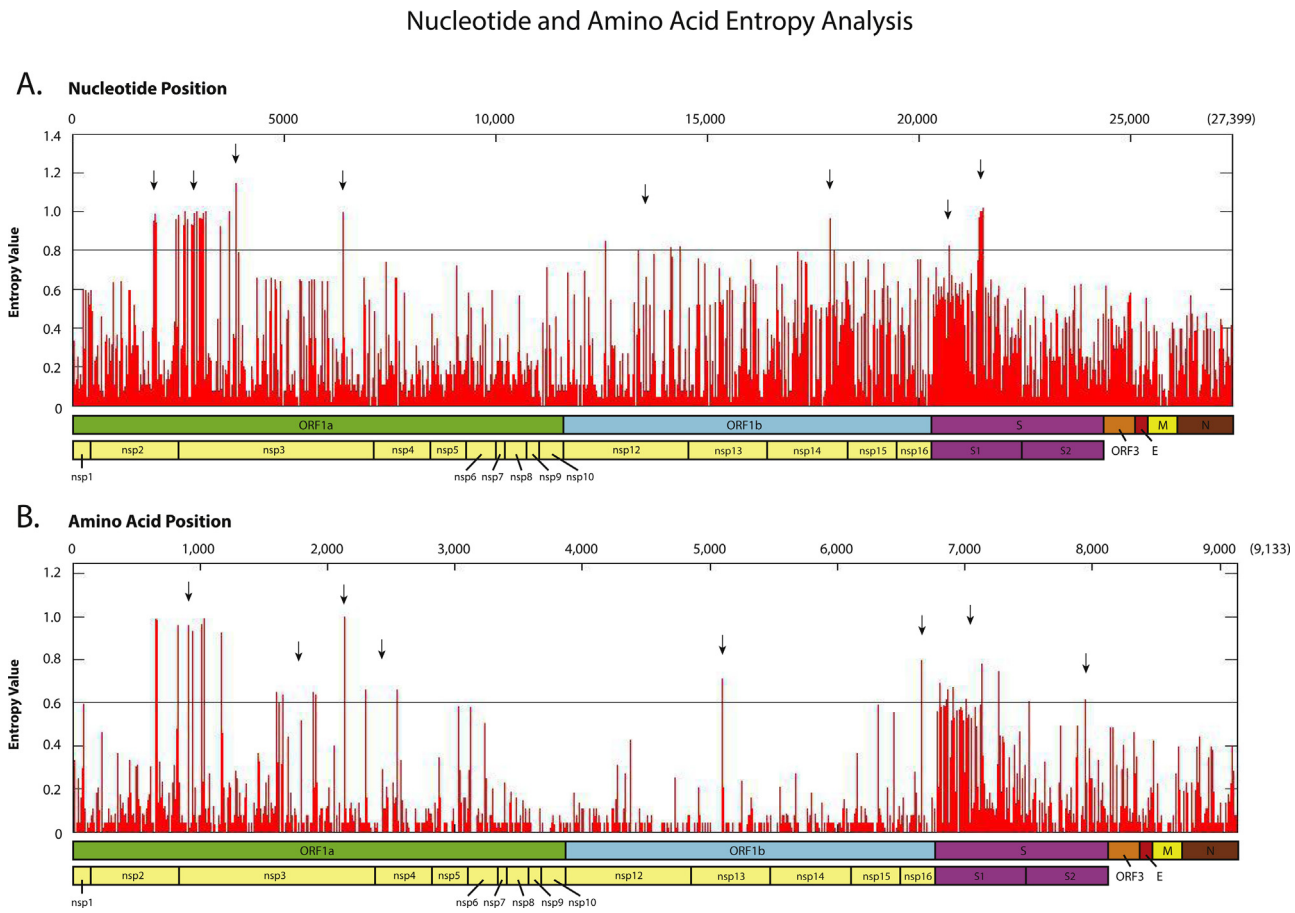

Fig. 1. Nucleotide and amino acid entropy analysis.
The *x*-axis represents position while the *y*-axis represents entropy level. The PEDV non-structural and structural proteins are annotated below each plot. (A) Entropy levels for nucleotide sequences. Black line represents the variance threshold of 0.8. (B) Entropy levels for amino acid sequences. Black line represents the variance threshold of 0.6. Arrows indicate general areas expressing high levels of variance.

Table 1
Regions in the PEDV genome with high variance.

	ORF1a and ORF 1b (20,337 nucleotides and 6779 amino acids)							Spike (S) (4161 nt & 1387 aa)	
	nsp2	nsp3	nsp4	nsp12	nsp13	nsp14	nsp16	S1	S2
Nucelotide positions	1932, 1951, 1976, 2467, 2517, 2652, 2658, 2664, 2667, 2679	2720, 2809, 2856, 2883, 2940, 3003, 3021, 3025, 3078, 3105, 3110, 3171, 3516, 3708, 3876, 6405	None	12603, 14,133, 14,361	None	17901	None	20724, 21,427, 21,447, 21,498, 21,516	None
Amino acid positions	651, 659, 823	907, 937, 1009, 1037, 1172, 1604, 1646, 1886, 1904, 2135, 2300	2546	None	5099	none	6663	6806, 6850, 6861, 6866, 6909, 6910, 7014, 7133, 7266, 7502	7941

3. Results

3.1. Nucleotide and amino acid analysis

The PEDV nucleotide sequences ranged from 27,529 to 28,061 bases in length. Two PEDV genomes from our study had insertions or deletions. Ohio249 had a 3-nt insertion between positions 22,039 and 22,041 in the spike gene while Minnesota309 with a 4-nt deletion from positions 27,768 to 27,771 in the 3′ UTR compared to the original US strain, USA/Colorado/2013. Entropy analysis was conducted with 219 whole nucleotide and amino acid sequences, containing the concatenated ORFs excluding 5′ and 3′ UTRs. Entropy values greater than 0.8 and 0.6 were considered highly variable for the nucleotide and amino acid alignments, respectively, based on the level of diversity in the dataset and previously determined entropy values (Litwin and Jores, 1992). Within the nucleotide alignment, 15 of the 20 PEDV regions lacked positions with entropy levels above 0.8 (nsp1, nsp5-nsp10, nsp13, nsp15, nsp16, S2, ORF3, envelope, membrane, and nucleocapsid) while 5 regions had entropy levels above 0.8 (nsp2, nsp3, nsp12,

nsp14, and S1) (Fig. 1). The nsp2 and 3 were the most divergent regions containing 10 and 16 diverse nucleotide positions, respectively (Table 1). Interestingly, the nsp12 gene contained 4 diverse nucleotide positions, which were absent in the amino acid sequence (Fig. 1A). Inversely, high amino acid diversity was observed in the nsp4, nsp13, nsp16, and S2 genes, which were absent in the nucleotide alignment (Fig. 1B). Higher entropy levels were present in the nsp2, nsp3, and S1 regions in both the nucleotide and amino acid alignments. Overall, the ORF1a entropy levels were higher compared to ORF1b in the amino acid analysis. Of the structural genes, the S gene had the highest entropy levels compared to the envelope, membrane, and nucleocapsid genes.

3.2. Recombination and bayesian analysis (BEAST)

Recombination was detected in 7 main areas of the concatenated full genome, including the nsp2, nsp3, nsp14–16, S1 domain, and nucleocapsid gene (Fig. 2A). In these areas, recombination was present in the majority of the sequences, so the entire region was excised from the alignment prior to Bayesian analysis. In addition, 35 sequences (23 from Asia, 12 from the Americas) were omitted from the Bayesian analysis due to evidence of widespread recombination throughout the genome (Table S1). For example, the pandemic sequence Minnesota211 contained a recombinant region with the characteristic S-INDEL deletions and insertions in the S1 domain, indicating a recombinant event occurred between an S-INDEL strain and a non S-INDEL pathogenic strain in the US (Fig. 2B).

A maximum clade credibility (MCC) phylogeny was inferred for both the concatenated genomic sequences excluding the recombinant regions (12,999 nt) and the spike S1 gene (2142 nt). The analysis was run independently twice until convergence was reached, with high agreement between the two runs. In the concatenated alignment tree, the classical and pandemic Asian strains were positioned as basal to the US strains, consistent with an Asian origin for the US outbreak (Fig. 3). Importantly, the concatenated alignment tree suggests that the US epidemic may have resulted from two independent PEDV introductions into the US, including minor and major clade of viruses. The minor clade contained the American and European S-INDELs, and a small subclade of non S-INDEL sequences from the Ohio, including Ohio249/2014, PC21A/2013, and OH15962/2013. The major clade of US PEDV strains was supported by high posterior probability (100%) and appears to have diverged further into two highly supported sublineages (99% and 100% posterior probability). The phylogeny is consistent with multiple incursions of the major clade of US PEDV viruses into Mexico, Canada, and South Korea. The minor clade includes sequences from late 2013 to early 2014 that are localized to the midwestern and eastern US regions. The estimated tMRCA of the minor clade of US PEDV strains is July 2009–2011, and the estimated tMRCA of the major clade of US PEDV strains is September 2010–August 2012. The estimated evolutionary rate for the complete genome (excluding recombinant regions and sequences) is 6.2×10^{-4} substitutions/site/year ($4.8 \times 10^{-4} - 7.6 \times 10^{-4}$, 95% highest posterior density (HPD)). The rate estimate for the US strains is slightly higher, but not significantly: 5.5×10^{-3} substitutions/site/year ($4.6 \times 10^{-3} - 6.5 \times 10^{-3}$, 95% HPD).

The spike tree illustrates the evolutionary relationship between the classical strains and the S-INDELs, which suggests a classical origin for the S-INDEL genotype (Fig. 4). The pathogenic strains form a highly diverged major clade (Fig. 4B), which braches into 2 large American clades. In addition, the Bayesian analysis of the spike gene might suggest 2 separate introductions of PEDV into the US. The evolutionary rate for the S1 gene is 1.5×10^{-3} substitutions/site/year ($1.1 \times 10^{-3} - 1.9 \times 10^{-3}$, 95% HPD).

3.3. Examination and modeling of RBD residues

Considering the high entropy levels in the spike gene and the evolutionary rate determined from the S1 Bayesian spike tree, the RBD within the S1 was further examined. The S-INDEL and classical PEDV strains shared similar amino acid substitutions, specifically in the NTD of the S1 region (Fig. S1). Furthermore, the pandemic PEDV strains from China had an increased number of substitutions within the S1 domain when compared to the American strains due to the longer circulation time in China. Compared to the attenuated vaccine strain DR13, 29 of 185 (16%) American strains and 19 of 34 (56%) Asian strains had at least one amino acid substitution in the pAPN RBD (Table 2). The majority of the American strains ($n = 156$) did not represent any amino acid differences in the pAPN RBD. In this region, 8 positions in the American strains had amino acid differences compared to the vaccine strain DR13 (Fig. 5A). The most common substitutions were in the fourth region of the pAPN RBD at positions E594D ($n = 10$) and G598D ($n = 9$), which were substituted with aspartic acid. More substitutions ($n = 13$) occurred in the pAPN RBD of the Asian strains (Table 2), with the most common substitutions at position H515L ($n = 5$).

The RBD regions were three-dimensionally modeled to illustrate the 8 and 13 amino acid positions at which substitutions occurred in the North American and Asian strains, respectively. The modeling of the spike protein suggests that the pAPN RBD residues cluster around the inner pore created by the trimer molecule, while the NTD is oriented around the outer surface of the S1 domain (Fig. 5B through 5F).

4. Discussion

Genomic analysis depends critically on complete sequence data to conduct accurate research on phylogeny, evolution, and gene regulation. In the past, it was more economically and time effective to sequence smaller pieces of a genome and develop evolutionary conclusions from these relatively small genomic pieces. However, without full genomic sequences, it is impossible to compare variations within a genome to determine selective pressures on specific genes or regions. Because of NGS technology, tools like site-specific entropy analysis can be used to examine variability throughout the genome of many PEDV sequences. Sun and collaborators reported four regions of diversity within the PEDV genome, including V1 in the nsp2 and nsp3, V2 in the S1, V3 in the S2 and ORF3, and V4 in the nucleocapsid (Sun et al., 2015). While our results support the nucleotide variance in the nsp2, nsp3, and spike genes, high levels of diversity were not present in the S2, ORF3, and the nucleocapsid. This could be due to the omission of PEDV isolates in our analysis, as well as the comparatively large number of new US sequences in our dataset. Our variance results may more accurately represent variance within American PEDV strains, and underrepresent variance within Chinese strains.

The diversity in the S1 region is comprehensible since it is under strong immunological pressure while the S2 region is more conserved throughout CoVs (Aydin et al., 2014). The functions of nsp2 and nsp3 remain relatively ambiguous. Despite being involved in viral growth and propagation, nsp2 is dispensable for viral replication because CoV strains can replicate in absence of the nsp2 (Graham et al., 2006). The function of nsp3 may be related to innate immune evasion since it encodes proteases that facilitate proteasome degradation, changes in intracellular destination, signaling, protein interactivity, and host type I interferon (IFN) antagonist activities (Xing et al., 2013). Due to the multifaceted nature of nsp3, other nsp regions could produce proteins with novel effects not yet understood that mediate the virulence of CoV species. Acquired nucleotide differences throughout the nsp2 and nsp3 regions could

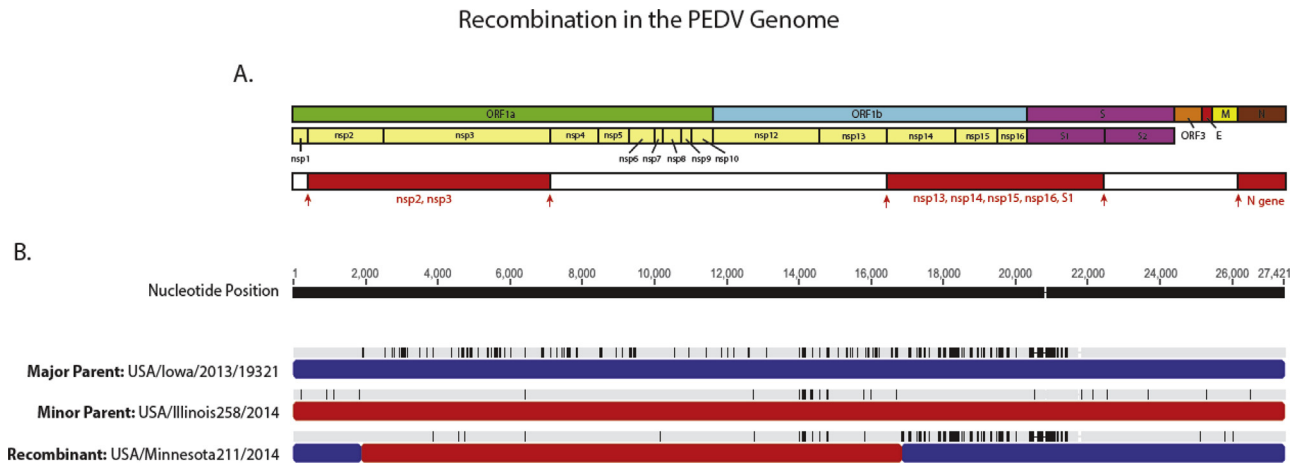

Fig. 2. Areas of recombination in the PEDV genome.
Recombinant regions within the PEDV genome. (A) The PEDV genome with excised recombinant areas highlighted in red. The red arrows designate excision of the recombinant sequence. (B) Recombination in strain Minnesota211, with major and minor parent sequences provided. The blue bars represent sequence originating from the major parent while red bars represent sequence originating from the minor (donor) parent.(For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 2
Amino acid substitutions in the pAPN Receptor-Binding Residues.

	Strain Name	513 to 521	524 to 526	549 to 554	599 to 603
North Amer-i-can Strains	USA/Illinois260/2014	G514D			
	USA/NorthCarolina91/2013	G514D			
	USA/Missouri373/2014	G514D			
	USA/Oklahoma35/2013	H515R			
	USA/Texas128/2014	H515R			
	USA/SouthDakota371/2014	H515R			
	USA/Oklahoma32/2013	G517D			G603V
	USA/Tennesse56/2013	N519D			
	USA/NorthCarolina/49469/2013		T525I		
	USA/SouthDakota336/2014			S551I	
	USA/Ohio10123/2014			S551I	
	USA/Minnesota79/2013				E599D
	USA/Texas31/2013				E599D
	USA/Kansas29/2013				E599D
	USA/Texas424/2014				E599D
	USA/Illinois197/2014				E599D
	USA/Oklahoma471/2014				E599D
	USA/Nebraska288/2014				E599D
	USA/Ohio9097/2014				E599D
	USA/Ohio8593/2014				E599D
	USA/Missouri177/2014				E599D, G603D
	USA/NorthCarolina/35140/2013				G603D
	USA/Kansas46/2013				G603D
	USA/Colorado47/2013				G603D
	USA/Missouri337/2014				G603D
	USA/Illinois307/2014				G603D
	USA/Illinois308/2014				G603D
	USA/Oklahoma1/2014				G603D
	USA/Oklahoma477/2014				G603D
Asian Strains	LZW/2012	G514D			
	BJ-2011-1/2011	H515L			
	PEDV-7C/2011	H515R			
	GDB/2012	H515R			
	CHYJ130330/2013	H515Q	T526I		
	GDZQ/2014	H515L	D524E		G603D
	FJND-3/2011	H515P			S602G
	CH/S/1986	H515S, I521V			G603S
	ZMDZY/11/2011	H515R, I521T			
	CHM2013/2013	H515L, G517S, I521V			E599A
	CV777/1978	H515L, G517S, I521V			E599A
	LZC/2006	H515L, G517S, I521V			E599A
	VIE/JFP1013_1/2013	H516R		T549P, N550T	
	VIE/VAP1113_1/2013			T549P, N550T	
	VIE/KCHY-310113/2013			T549P, N550T	
	PEDV-WS/2014				E599D
	JX-1/2013				E599D, F600Y
	JX-2/2013				E599D, F600Y
	PEDV-CHZ/2013				G603V

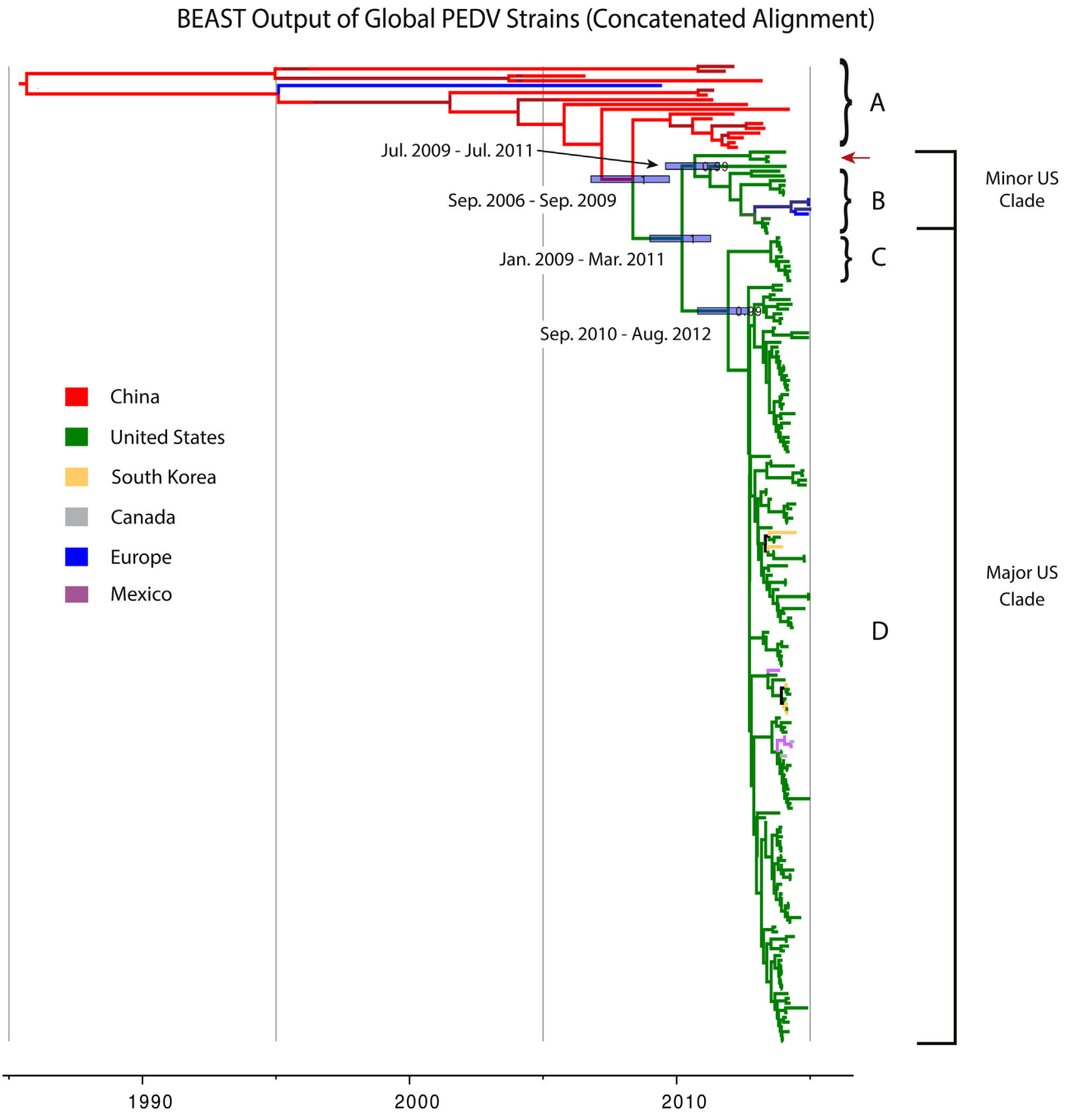

Fig. 3. Concatenated genome bayesian phylogenetic tree.
A time-scaled MCC phylogenetic tree of the full PEDV genome, excluding recombinant regions and sequences. The posterior probabilities are labeled at the major nodes. Branches are colored by country of origin. The red arrow denotes a clade of non-INDEL strains that group with US and European S-INDELs. Note the classical CV777 strain is not included in the figure. The historical strains are represented in clade A. The minor US clade includes the US and European S-INDELs (clade b) and the non-US INDEL (represented by the red arrow). The Major US clade includes a smaller subclade (clade C) separate out from the rest of the US major clade (clade D).(For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

contribute to the evasion of host immunity. Thus, future research should focus on the functionality and importance of all PEDV genes to further understand CoV pathogenesis.

Recombination plays a pivotal role in the evolution of CoVs by creating new strains with altered virulence. The Minnesota211 strain originated from a recombination even between an S-INDEL and a US pandemic strain, which has been associated with altered pathogenesis. While recombination may occur more often during an epidemic, recombination events occurred in most of the Asian strains. Recombination events can affect the phylogenetic analysis because different regions of the genome may have different evolutionary histories (Spade et al., 2015). Our recombination analysis resulted in a significant portion of the complete genome being removed prior to more detailed phylogenetic analysis. At this time, the BEAST program cannot accommodate genetic data that includes recombined regions.

Our analysis supported an Asian origin for the US outbreak while the inference is biased by the lack of background sequences from other regions. Although over 100 genomes were from the US, interpretation of the evolutionary and spatial his-

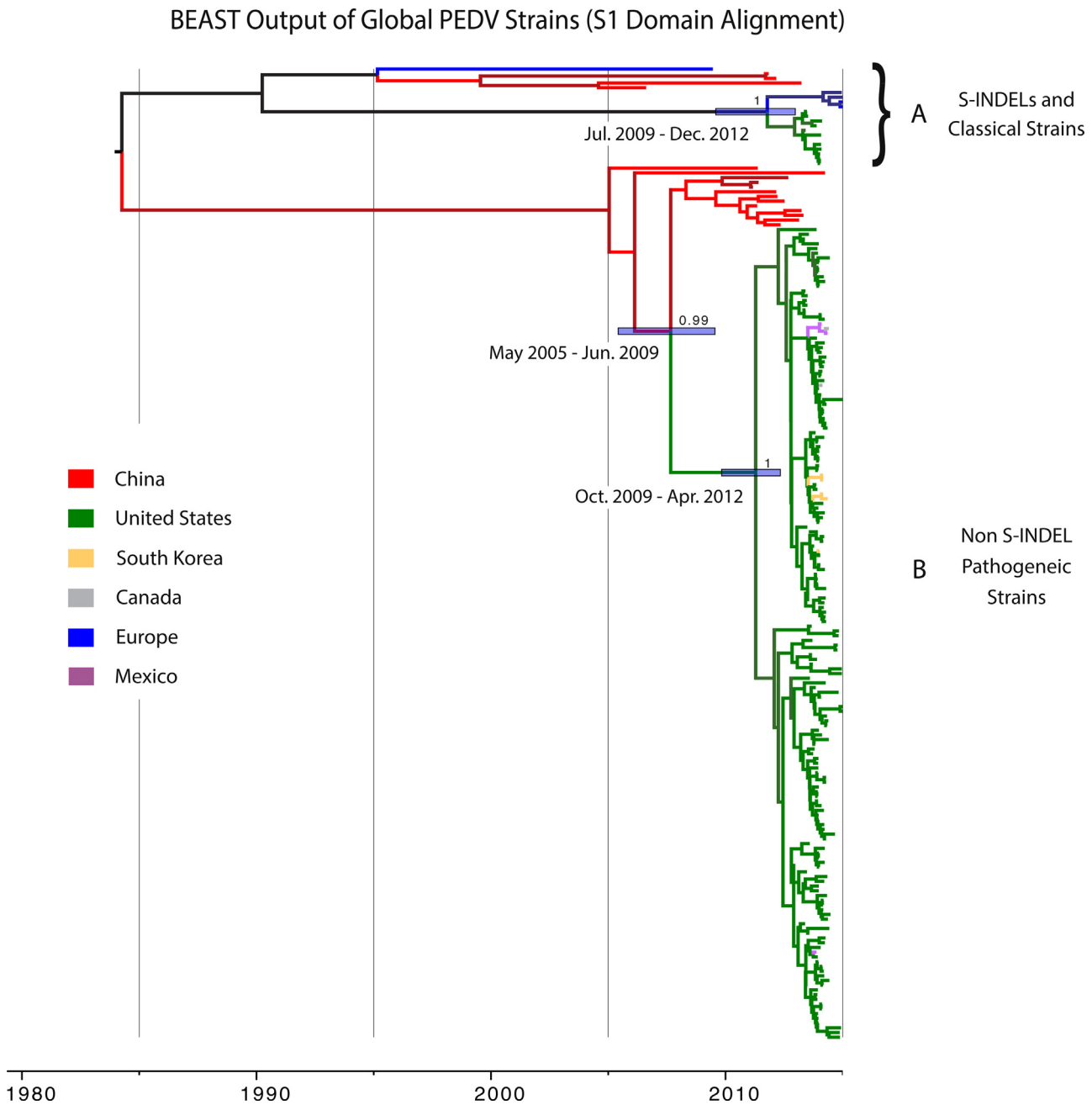

Fig. 4. S1 domain bayesian phylogenetic tree.
A time-scaled MCC phylogenetic tree of the PEDV S1 domain. Branches are shaded by country of origin. Clade A consists of S-INDELs and classical strains, including sequences from America, Europe, and China. Clade B includes global non S-INDEL pathogenic sequences.

tory of this data is limited by the lack of genomic PEDV data from other regions, including Europe and Asia. US strains had a higher evolutionary rate compared to the global strains, but the Bayesian Skyline plot did not show any significant increase in evolutionary rate, possibly due to the lack of temporal sampling in the US dataset. The evolutionary rate for the spike gene was higher compared to the rest of the genome, reflecting greater selective pressure. The overall evolutionary rate of PEDV (6.2×10^{-4} substitutions/site/year) is similar to that of TGEV and wild animal CoVs (6.08×10^{-3} substitutions/site/year, $1.3 \times 10^{-5} - 1.4 \times 10^{-2}$ 95% HPD), but lower than that of SARS-CoV (2.3×10^{-3} nucleotide substitutions/site/year), except during the time of the US PEDV 2013 epidemic (5.5×10^{-3} substitutions/site/year) (Song et al., 2005; Vijaykrishna et al., 2007).

Surprisingly, 3 pandemic strains were positioned within the S-INDEL clade. Possibly, a pandemic and S-INDEL strain were introduced into the Americas, and a recombination event occurred in the NTD that removed the characteristic insertions and deletions of an S-INDEL strain, as indicated in the Minnesota211 sequence. The relationship between the US minor clade and the recent PEDV strains from Europe is less clear. While these viral populations are closely related (posterior probability of 100%), the direction of transmission is unclear at this time. Additional sequences from Europe might help to resolve the origins of these recent European PEDV cases.

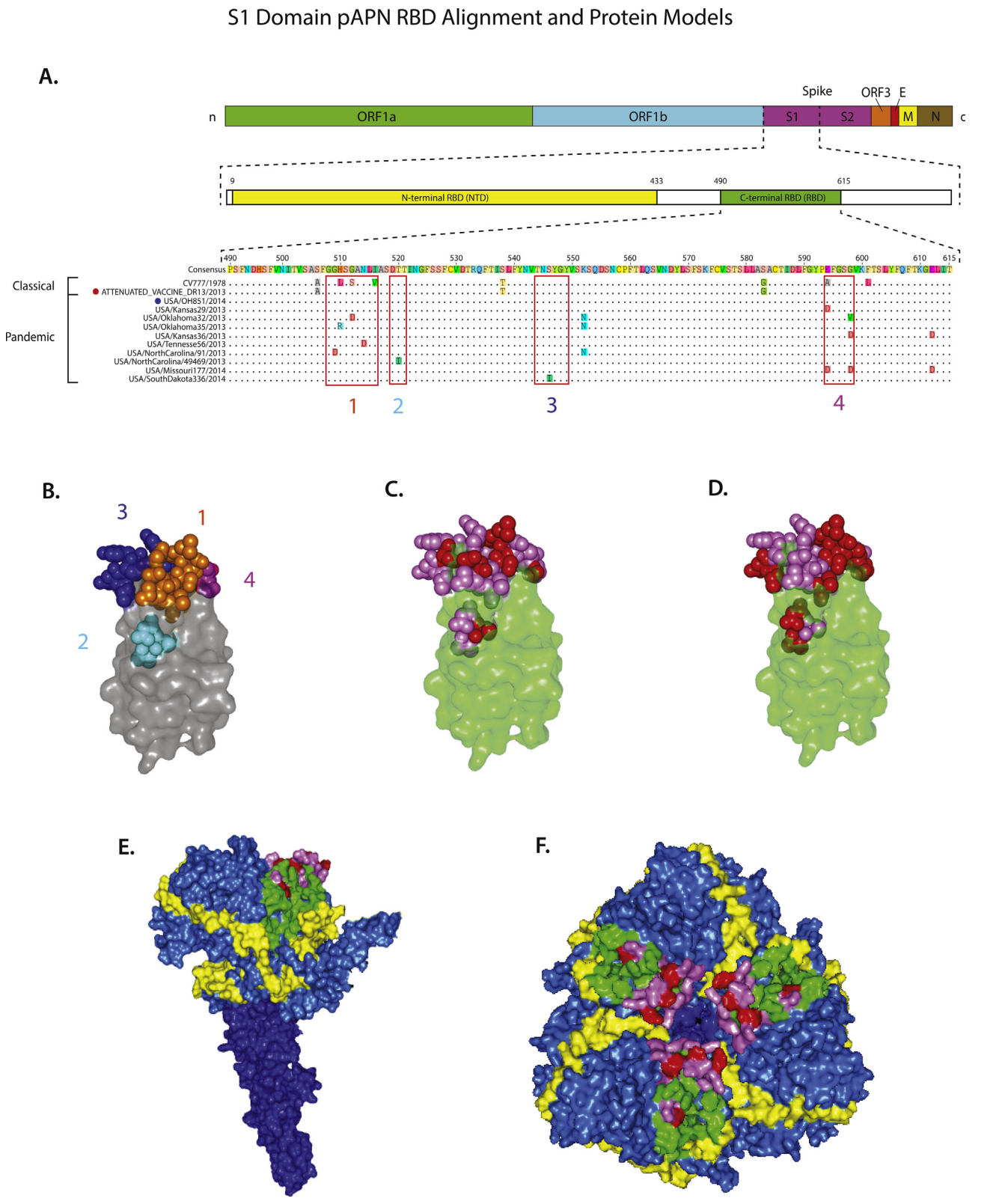

Fig. 5. S1 domain pAPN RBD alignment and protein models.
Amino acid alignment and protein model visualizations of the pAPN RBD. (A) An alignment of the putative C-terminal receptor-binding domain (RBD) representing diverse residues within the American strains. Identical residues are represented as dots while residues that differ from the consensus are colored. Red boxes outline the four specific porcine aminopeptidase-N (pAPN) receptor binding regions. The attenuated vaccine strain is marked by a red dot while a S-INDEL strain is marked by a blue dot. (B) A tertiary model view of the C-terminal RBD, with the four regions of the pAPN RBD (labeled 1 through 4) are highlighted in orange, cyan, blue and magenta, respectively. (C) The pAPN RBD with the residue changes in North American strains compared to the vaccine strain DR13 highlighted in red. (D) The pAPN RBD with the residue changes in Asian strains compared to the vaccine strain DR13 highlighted in red. (E) A monomer model of the PEDV spike protein, with the C-terminal RBD represented in green, dark blue represents the S2 region, light blue represents the S1 region, and yellow represents the N-terminal RBD. (F) A theoretical tertiary structure model of the PEDV spike protein. Blue represents the S1 region, with the specific N- and C-terminal RBDs highlighted in yellow and green, respectively. The pAPN-RBD is shown in violet.(For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Examining the spike gene can reveal interesting conclusions about the pAPN RBD. While the NTD spans a larger region of the S1 domain, it has not been directly linked to PEDV tropism and functionality, as in TGEV and MHV (Reguera et al., 2012). The Korean attenuated vaccine strain and the US strains share similar residues in the pAPN RBD, with numerous differences compared to the older classical strains, suggesting this vaccine may protect against the American strains. However, developing a consistently and longitudinally efficacious vaccine may prove challenging, considering the high evolutionary rate of the S1 region. Failures in the development of an efficacious vaccine have been reported, further supporting the difficulty in generating vaccines for PEDV (Li et al., 2014). Despite some uncertainties in vaccine efficacy, a recent study demonstrated that prior exposure of sows to the S-INDEL strain provided a level of protective immunity when their piglets were challenged with the more virulent original US PEDV strain, which is probably due to conservation within the C-terminal region of the viral genome (Goede et al., 2015).

5. Conclusion

While the exact functionality of all the genes of PEDV and other CoVs is unknown, adding the complete genomes of diverse strains to the global database promotes better understanding of evolutionary and phylogenetic relationships. Multiple regions within the genome are variable, and recombination is common between PEDV strains. Despite excising a large portion of the genome prior to analysis, the Bayesian trees illustrate two distinct entries of PEDV into the US and characterize the evolution of PEDV compared to other CoVs. Modeling of the pAPN RBD region has revealed that Asian strains have increasing diversity compared to previously developed vaccines, and the variability in both the American and Asian strains needs to be considered for future vaccine development. As the US swine industry recovers from the PEDV epidemic of 2013–2014, research is maturing to understand the regions of diversity, evolution, and the RBD of PEDV to prevent future outbreaks and foster vaccine development.

Acknowledgements

This study was supported partially by the Rapid Agricultural Response Fund, established by the Minnesota legislature and administered by the University of Minnesota Agricultural Experiment Station, and by Boehringer Ingelheim Vetmedica, Inc.

The authors thank the faculty and personal at the UMVDL for their technical services.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.prevetmed.2015.10.020.

References

Aydin, H., Al-Khooly, D., Lee, J.E., 2014. Influence of hydrophobic and electrostatic residues on SARS-coronavirus S2 protein stability: insights into mechanisms of general viral fusion and inhibitor design. Protein Sci. 23, 603–617.

Bernini, A., Spiga, O., Ciutti, A., Chiellini, S., Bracci, L., Yan, X., Zheng, B., Huang, J., He, M.L., Song, H.D., Hao, P., Zhao, G., Niccolai, N., 2004. Prediction of quaternary assembly of SARS coronavirus peplomer. Biochem. Biophys. Res. Commun. 325, 1210–1214.

Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Cassarino, T.G., Bertoni, M., Bordoli, L., Schwede, T., 2014. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Res. 42, 252–258.

Chasey, D., Cartwright, S.F., 1978. Virus-like particles associated with porcine epidemic diarrhoea. Res. Vet. Sci. 25, 255–256.

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.

Cui, X., 1990. Studies on the detection of porcine epidemic diarrhea virus by immunofluorescent techniques. Chin. J. Prev. Vet. Med. 5, 20–24.

Delmas, B., Gelfi, J., L'Haridon, R., Vogel, L.K., Sjostrom, H., Noren, O., Laude, H., 1992. Aminopeptidase N is a major receptor for the entero-pathogenic coronavirus TGEV. Nature 357, 417–420.

Drummond, A.J., Ho, S.Y., Phillips, M.J., Rambaut, A., 2006. Relaxed phylogenetics and dating with confidence. PLoS Biol. 4, e88.

Drummond, A.J., Nicholls, G.K., Rodrigo, A.G., Solomon, W., 2002. Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. Genetics 161, 1307–1320.

Drummond, A.J., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. 7, 214.

Drummond, A.J., Rambaut, A., Shapiro, B., Pybus, O.G., 2005. Bayesian coalescent inference of past population dynamics from molecular sequences. Mol. Biol. Evol. 22, 1185–1192.

Drummond, A.J., Suchard, M.A., 2010. Bayesian random local clocks, or one rate to rule them all. BMC Biol. 8, 8–114 114-7007.

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.

EFSA AHAW Panel, 2014. (EFSA Panel on Animal Health and Welfare). Scientific opinion on porcine epidemic diarrhoea and emerginf pig deltacoronavirus. EFSA J. 12, 3877-68.

Fan, H., Zhang, J., Ye, Y., Tong, T., Xie, K., Liao, M., 2012. Complete genome sequence of a novel porcine epidemic diarrhea virus in south China. J. Virol. 86, 10248–10249.

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.

Graham, R.L., Sims, A.C., Baric, R.S., Denison, M.R., 2006. The nsp2 proteins of mouse hepatitis virus and SARS coronavirus are dispensable for viral replication. Adv. Exp. Med. Biol. 581, 67–72.

Hanke, D., Jenckel, M., Petrov, A., Ritzmann, M., Stadler, J., Akimkin, V., Blome, S., Pohlmann, A., Schirrmeier, H., Beer, M., Hoper, D., 2015. Comparison of porcine epidemic diarrhea viruses from Germany and the United States, 2014. Emerg. Infect. Dis. 21, 493–496.

Herraez, A., 2006. Biomolecules in the computer: Jmol to the rescue. Biochem. Mol. Biol. Educ. 34, 255–261.

Lai, M.M.C., Perlman, S., Anderson, L.J., 2013. Nidovirales: coronoviridae and ateriviridae. In: Knipe, D.M., Howley, P.M. (Eds.), Fields Virology. , 6th edition. Lippincott Williams & Wilkins, Philadelphia, PA, USA, pp. 825–841.

Lee, S., Lee, C., 2014. Outbreak-related porcine epidemic diarrhea virus strains similar to US strains, South Korea, 2013. Emerg. Infect. Dis. 20, 1223–1226.

Li, B.X., Ge, J.W., Li, Y.J., 2007. Porcine aminopeptidase N is a functional receptor for the PEDV coronavirus. Virology 365, 166–172.

Li, R., Qiao, S., Yang, Y., Su, Y., Zhao, P., Zhou, E., Zhang, G., 2014. Phylogenetic analysis of porcine epidemic diarrhea virus (PEDV) field strains in central China based on the ORF3 gene and the main neutralization epitopes. Arch. Virol. 159, 1057–1065.

Litwin, S., Jores, R., 1992. In theoretical and experimental insights into immunology.

Marthaler, D., Jiang, Y., Collins, J., Rossow, K., 2014. Complete genome sequence of strain SDCV/USA/Illinois121/2014, a porcine deltacoronavirus from the United States. Genome Announc. 2, 10.1128/genomeA. 00218-14.

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, 10.1128/genomeA. 00555-13.

Martin, D., Murrell, B., Golden, M., Khoosal, A., Muhire, B., 2015. RDP4: detection and analysis of recombination patterns in virus genomes. Virus Evol. 1, 1–5.

Mesquita, J.R., Hakze-van der Honing, R., Almeida, A., Lourenco, M., van der Poel, W.H., Nascimento, M.S., 2015. Outbreak of porcine epidemic diarrhea virus in Portugal, 2015. Transbound Emerg. Dis.

Minin, V.N., Bloomquist, E.W., Suchard, M.A., 2008. Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. Mol. Biol. Evol. 25, 1459–1471.

Murakami, S., Miyazaki, A., Takahashi, O., Hashizume, W., Hase, Y., Ohashi, S., Suzuki, T., 2015. Complete genome sequence of the porcine epidemic diarrhea virus variant tottori2/JPN/2014. Genome Announc. 3, 10.1128/genomeA. 00877-15.

Ojkic, D., Hazlett, M., Fairles, J., Marom, A., Slavic, D., Maxie, G., Alexandersen, S., Pasick, J., Alsop, J., Burlatschenko, S., 2015. The first case of porcine epidemic diarrhea in Canada. Can. Vet. J. 56, 149–152.

Oka, T., Saif, L.J., Marthaler, D., Esseili, M.A., Meulia, T., Lin, C.M., Vlasova, A.N., Jung, K., Zhang, Y., Wang, Q., 2014. Cell culture isolation and sequence analysis of genetically diverse US porcine epidemic diarrhea virus strains including a novel strain with a large deletion in the spike gene. Vet. Microbiol. 173, 258–269.

Reguera, J., Santiago, C., Mudgal, G., Ordono, D., Enjuanes, L., Casasnovas, J.M., 2012. Structural bases of coronavirus attachment to host aminopeptidase N and its inhibition by neutralizing antibodies. PLoS Pathog. 8, e1002859.

C.E. Shannon The mathematical theory of communication The Bell system Technical Journal 27 1948 623–656 379-423.

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

Song, H.D., Tu, C.C., Zhang, G.W., Wang, S.Y., Zheng, K., Lei, L.C., Chen, Q.X., Gao, Y.W., Zhou, H.Q., Xiang, H., Zheng, H.J., Chern, S.W., Cheng, F., Pan, C.M., Xuan, H., Chen, S.J., Luo, H.M., Zhou, D.H., Liu, Y.F., He, J.F., Qin, P.Z., Li, L.H., Ren, Y.Q., Liang, W.J., Yu, Y.D., Anderson, L., Wang, M., Xu, R.H., Wu, X.W., Zheng, H.Y., Chen, J.D., Liang, G., Gao, Y., Liao, M., Fang, L., Jiang, L.Y., Li, H., Chen, F., Di, B., He, L.J., Lin, J.Y., Tong, S., Kong, X., Du, L., Hao, P., Tang, H., Bernini, A., Yu, X.J., Spiga, O., Guo, Z.M., Pan, H.Y., He, W.Z., Manuguerra, J.C., Fontanet, A., Danchin, A., Niccolai, N., Li, Y.X., Wu, C.I., Zhao, G.P., 2005. Cross-host evolution of severe acute respiratory syndrome coronavirus in palm civet and human. Proc. Natl. Acad. Sci. U. S. A. 102, 2430–2435.

Spade, D.A., Herbei, R., Kubatko, L.S., 2015. Geometric ergodicity of a hybrid sampler for Bayesian inference of phylogenetic branch lengths. Math. Biosci. 268, 9–21.

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.

Sun, R.Q., Cai, R.J., Chen, Y.Q., Liang, P.S., Chen, D.K., Song, C.X., 2012. Outbreak of porcine epidemic diarrhea in suckling piglets, China. Emerg. Infect. Dis. 18, 161–163.

Vijaykrishna, D., Smith, G.J., Zhang, J.X., Peiris, J.S., Chen, H., Guan, Y., 2007. Evolutionary insights into the ecology of coronaviruses. J. Virol. 81, 4012–4020.

Vlasova, A.N., Marthaler, D., Wang, Q., Culhane, M.R., Rossow, K.D., Rovira, A., Collins, J., Saif, L.J., 2014. Distinct characteristics and complex evolution of PEDV strains, North America, May 2013-February 2014. Emerg. Infect. Dis. 20, 1620–1628.

Wang, L., Byrum, B., Zhang, Y., 2014. New variant of porcine epidemic diarrhea virus, United States, 2014. Emerg. Infect. Dis. 20, 917–919.

Xing, Y., Chen, J., Tu, J., Zhang, B., Chen, X., Shi, H., Baker, S.C., Feng, L., Chen, Z., 2013. The papain-like protease of porcine epidemic diarrhea virus negatively regulates type I interferon pathway by acting as a viral deubiquitinase. J. Gen. Virol. 94, 1554–1567.