Infection, Genetics and Evolution 36 (2015) 108–115

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

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

Evolutionary characterization of the emerging porcine epidemic diarrhea virus worldwide and 2014 epidemic in Taiwan

Ming-Hua Sung [a,b], Ming-Chung Deng [c], Yi-Hsuan Chung [b], Yu-Liang Huang [c], Chia-Yi Chang [c], Yu-Ching Lan [d], Hsin-Lin Chou [d], Day-Yu Chao [a,*]

[a] Graduate Institute of Microbiology and Public Health, College of Veterinary Medicine, National Chung Hsing University, Taichung 40227, Taiwan
[b] Taichung City Animal Protection and Health Inspection Office, Taichung 408, Taiwan
[c] Animal Health Research Institute, Council of Agriculture, New Taipei City 25158, Taiwan
[d] Department of Health Risk Management, School of Public Health, China Medical University, Taichung, Taiwan

ARTICLE INFO

Article history:
Received 15 May 2015
Received in revised form 3 September 2015
Accepted 12 September 2015
Available online 13 September 2015

Keywords:
Porcine epidemic diarrhea virus
Likelihood mapping
Phylogenetic signal
Evolutionary rate
Taiwan
Nsp3
Spike

ABSTRACT

Since 2010, a new variant of PEDV belonging to Genogroup 2 has been transmitting in China and further spreading to the Unites States and other Asian countries including Taiwan. In order to characterize in detail the temporal and geographic relationships among PEDV strains, the present study systematically evaluated the evolutionary patterns and phylogenetic resolution in each gene of the whole PEDV genome in order to determine which regions provided the maximal interpretative power. The result was further applied to identify the origin of PEDV that caused the 2014 epidemic in Taiwan. Thirty-four full genome sequences were downloaded from GenBank and divided into three non-mutually exclusive groups, namely, worldwide, Genogroup 2 and China, to cover different ranges of secular and spatial trends. Each dataset was then divided into different alignments by different genes for likelihood mapping and phylogenetic analysis. Our study suggested that both nsp3 and S genes contained the highest phylogenetic signal with substitution rate and phylogenetic topology similar to those obtained from the complete genome. Furthermore, the proportion of nodes with high posterior support (posterior probability >0.8) was similar between nsp3 and S genes. The nsp3 gene sequences from three clinical samples of swine with PEDV infections were aligned with other strains available from GenBank and the results suggested that the virus responsible for the 2014 PEDV outbreak in Taiwan clustered together with Clade I from the US within Genogroup 2. In conclusion, the current study identified the nsp3 gene as an alternative marker for a rapid and unequivocal classification of the circulating PEDV strains which provides complementary information to the S gene in identifying the emergence of epidemic strain resulting from recombination.

1. Introduction

Porcine epidemic diarrhea virus (PEDV) causes an acute and highly contagious enteric tract infection, characterized by severe villus atrophy and congestion of the thin-walled intestines, which usually lead to high morbidity and mortality, especially in piglets (Song and Park, 2012). The disease was first observed among fattening pigs from United Kingdom in 1971 and the etiological agent was identified in Belgium as a new coronavirus, which was designated as PEDV prototype strain CV777 (Jung and Saif, 2015; Pensaert and de Bouck, 1978). Within the next two decades, PEDV was reported in several other European countries including Hungary, Italy, Germany, France, Switzerland and the Czech Republic, causing only isolated outbreaks in Europe (Hanke et al., 2015; Jung and Saif, 2015). In Asia, PEDV was first identified in 1982 and is now considered an endemic, causing devastating enteric diseases and substantial economic losses to the pork industry in many Asian countries such as China, South Korea, Japan, Thailand, and Taiwan (Lin et al., 2014; Park et al., 2007, 2011; Song and Park, 2012; Temeeyasen et al., 2014; Zhang et al., 2013).

Comparisons of full-length genomes showed that different PEDV strains are more closely related to alphacoronaviruses in a bat than other known alphacoronaviruses, suggesting that interspecies transmission of coronavirus may have occurred decades ago between bats and pigs or through other intermediate hosts (Huang et al., 2013; Tang et al., 2006; Woo et al., 2012). PEDV, belonging to the genus *Alphacoronavirus* within the family *Coronaviridae*, is an enveloped, single-stranded RNA genome with a 5′ cap and a 3′ polyadenylated tail (Jung and Saif, 2015). The size of its genome is approximately 28 Kb with two thirds of the 5′ genome containing two large open reading frames (ORFs), 1a and 1b, that encode two nonstructural polyproteins, pp1a and pp1b, which are involved in genome replication and transcription (Brian and Baric, 2005). The remaining PEDV genome contains ORFs encoding four structural and one nonstructural proteins in the following order: spike (S), ORF3, envelope (E), membrane (M)

* Corresponding author.
E-mail address: dychao@nchu.edu.tw (D.-Y. Chao).

http://dx.doi.org/10.1016/j.meegid.2015.09.011
1567-1348/

and nucleoprotein (N). Similar to S proteins from other coronaviruses (CoVs), the PEDV S protein is a glycoprotein on the viral surface and has a pivotal function in regulating interactions with specific host cell receptor glycoproteins to mediate viral entry (Li, 2015). Thus, the S glycoprotein is often used as an important viral component to understand the genetic relationships of different PEDV strains and the epidemiological status of PEDV in the field (Chen et al., 2013b; Tian et al., 2013). Nevertheless, the high variability of other genes including ORF3, E, M or N protein has been previously utilized for phylogenetic inference (Chen et al., 2013a; Ge et al., 2013; Kubota et al., 1999; Pan et al., 2012; Temeeyasen et al., 2014; Yang et al., 2013). In order to get better phylogenetic resolution, several studies used PEDV full genome sequences for analysis (Chen et al., 2014; Pan et al., 2012; Sun et al., 2015; X. Wang et al., 2013); however, sequencing the full genome is expensive from a laboratory perspective and also limits the rapid characterization of novel PEDV strains. Furthermore, for many computationally intensive analyses, utilizing the full genome is unfeasible. It would be beneficial to use only those genomic regions that contain the highest phylogenetic signal to reduce cost without losing valuable information.

Since 2010, massive PED outbreaks were reported in China characterized by 80 to 100% illness among infected swine herds and a 50 to 90% mortality rate among infected suckling piglets (Zhang et al., 2013). The emerging strains in Asia are distinct from previous endemic PEDV strains characterized by multiple amino acid insertions/deletions or mutations on the S protein (J. Wang et al., 2013). The mutations on the S protein have been speculated to be associated with escape from neutralizing antibodies due to the use of a bivalent vaccine against transmissible gastroenteritis virus (TGEV) and PEDV (Chen et al., 2013a). This novel PEDV strain has continued to spread throughout the United States and Taiwan where no vaccination program has been implemented (Chen et al., 2014; Deng et al., 2014). Therefore, tracing the transmission and evolutionary changes of PEDV is important for future public health intervention. The objectives of the present study are (1) to systematically evaluate the evolutionary patterns and phylogenetic resolution in each gene in order to determine which regions provided the maximum interpretative power to infer temporal and geographic relationships among PEDV strains by using 34 complete genomic sequences from worldwide collection of PEDV field strains; and (2) to identify the origin of PEDV that caused the epidemic in Taiwan in 2014.

2. Materials and methods

2.1. Sequence data

We downloaded all full genome sequences (34) available in GenBank as of February 26, 2014 (http://www.ncbi.nlm.nih.gov/) for which the sampling year and country of collection was recorded (Supplementary Table 1). Since Genogroup 1 contained only 8 full-length sequences and no PEDV in the US was found until 2013, dataset was divided into three non-mutually exclusive groups: worldwide (34 sequences), Genogroup 2 (24 sequences) including all the recent emerging strains, and China (20 sequences) including all strains obtained from China, to cover different ranges of secular and spatial trends. Each of those datasets was then divided into different alignments by genes: structural (S, E, M, N) and non-structural (NSP1, NSP2, NSP3, NSP4, 3C-like protease, NSP6, NSP7, NSP8, NSP9, NSP10, RNA-dependent RNA polymerase, helicase, exoribonuclease, uridylate-specific endoribonuclease and putative 2′-O-methyl transferase) based on the sequence prediction from the whole genome of the strain CV777 (GenBank accession no. NC_003436). We created two additional concatenated datasets with length >2000 nucleotides for comparison: (1) NSP1 and NSP2 (denoted as 5′nsp1-2); (2) ORF3, E, M, and N (denoted as 3′OEMN). Sequences were aligned using ClustalW2 available from the European Molecular Biology Laboratory (EMBL, http://www.ebi.ac.uk/Tools/msa/clustalw2/) and manually edited.

Diarrhea among the piglets from the swine farms in Taiwan began in December 2013 and the etiologic agent of this epidemic was confirmed to be PEDV (Lin et al., 2014). The total nucleic acids from three representative clinical samples were extracted as previously described (Deng et al., 2014) and the cDNA was produced using the SuperScript III Reverse Transcription kit with random hexamers (Invitrogen, Carlsbad, CA, USA). The complete sequence of the NSP3 gene in each clinical sample was determined using the primers summarized in Table 1 and are available under GeneBank accession number KR632490-2.

2.2. Comparative analysis

To evaluate the phylogenetic resolution of each gene from each dataset, two bioinformatics approaches were employed: likelihood mapping (LM) and phylogenetic analysis involving the determination of the model of evolution and Bayesian Markov chain Monte-Carlo (MCMC) phylogenetic analyses.

2.2.1. LM analysis

To investigate the phylogenetic signal of each gene from all three datasets, LM was performed using TREE-PUZZLE program by analyzing 10,000 random quartets (Schmidt et al., 2002). LM assesses if a dataset is suitable for phylogenetic reconstruction by analyzing four randomly chosen groups of sequences, termed quartets. Each quartet was evaluated using maximum likelihood and three possible unrooted tree topologies were weighted for each quartet. The posterior weights were then plotted onto a triangular surface. Based on this method, the fully resolved tree topologies were plotted in the three corners, indicating the presence of a tree-like phylogenetic signal, and the unresolved quartets were shown in the central region of the triangle. The phylogenetic noise was computed probabilistically and a star-like signal was shown when more than 30% of the dots fell within the central area representing unresolved phylogenies (Strimmer and von Haeseler, 1997).

2.2.2. Phylogenetic analysis

Before proceeding to the phylogenetic analysis, each gene and each of the four datasets were tested against the best fitting nucleotide substitution model as specified. The best-fit model of nucleotide substitution was selected under the Akaike information criterion (AIC) and Bayesian information criterion (BIC) as implemented in jModelTest (Posada, 2008).

To estimate the genealogy and the evolutionary timescale of NSP3, S and E sequence alignments using the worldwide, Genogroup 2 and China datasets, the Bayesian framework implemented in BEAST software package version 2.1.3 (Bouckaert et al., 2014) under the designated molecular clock model and the nucleotide substitution model predetermined by jModelTest was used. The BEAST program was run with two different rate categories for codon position 1 + 2 and position 3, allowing gamma-distributed rate variation when model selection results were suggested. The published year of isolation was assigned to each sequence (tip) to calibrate the rate estimate and the expected time to coalescence was modeled under constant population sizes. The MCMC process was run for 100 million iterations until convergence

Table 1
Oligonucleotide primers used for amplifications of the PEDV nsp3 gene by PCR and sequencing.

Primer ID	Sequence (5′ to 3′)	Position*	Used to amplify fragment
Nsp3-F2	TCCCACCGATGGTAATAGTG	2646–2665	PCR forward primer
Nsp3-R2	TGAACAGACACAAAAACCAGAAG	8088–8110	PCR reverse primer
Nsp3-F2-1	TTGGGTGATGTGTCGGCTTG	3706–3725	Sequencing primer
Nsp3-R2-1	GCTTCTTACAGAACTTAGAACC	6960–6981	Sequencing primer
Nsp3-F2-2	AGGAAGATGTTCAACAAGTTTC	4700–4721	Sequencing primer
Nsp3-R2-2	ACACTGTAATTAAATTACGTGAC	6176–6197	Sequencing primer

* Positions correspond to the PEDV CV777 strain (GenBank accession no. NC_003436).