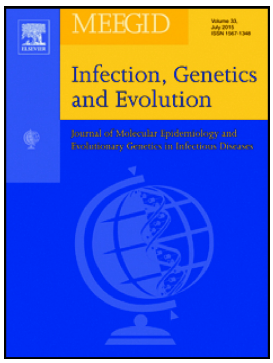


Newly emerged porcine enteric alphacoronavirus in southern China: Identification, origin and evolutionary history analysis

Fu Xinliang, Fang Bo, Liu Yixing, Cai Mengkai, Huang Junming, Ma Jun, Bu Dexin, Wang Lifang, Zhou Pei, Wang Heng, Zhang Guihong

PII: S1567-1348(18)30210-7
DOI: doi:10.1016/j.meegid.2018.04.031
Reference: MEEGID 3497

To appear in: *Infection, Genetics and Evolution*

Received date: 11 January 2018
Revised date: 2 April 2018
Accepted date: 24 April 2018

Please cite this article as: Fu Xinliang, Fang Bo, Liu Yixing, Cai Mengkai, Huang Junming, Ma Jun, Bu Dexin, Wang Lifang, Zhou Pei, Wang Heng, Zhang Guihong , Newly emerged porcine enteric alphacoronavirus in southern China: Identification, origin and evolutionary history analysis. The address for the corresponding author was captured as affiliation for all authors. Please check if appropriate. Meegid(2018), doi:10.1016/j.meegid.2018.04.031

Newly Emerged Porcine Enteric Alphacoronavirus in Southern China: Identification, Origin and Evolutionary History Analysis

Fu Xinliang[1,2], Fang Bo[1,2], Liu Yixing[1,2], Cai Mengkai[1,2], Huang Junming[1,2], Ma Jun[1,3], Bu Dexin[1,2], Wang Lifang[1,3], Zhou Pei[1,3], Wang Heng[1,3,*] wangheng2009@scau.edu.cn, Zhang Guihong[1,2,*] guihongzh@scau.edu.cn

[1]College of Veterinary Medicine, South China Agricultural University, Guangzhou, China

[2]Key Laboratory of Zoonosis Prevention and Control of Guangdong Province, Guangzhou, China

[3]Key Laboratory of Comprehensive Prevention and Control for Severe Clinical Animal Diseases of Guangdong Province, Guangzhou, China

[*]Corresponding authors.

Abstract: Coronaviruses have a wide host range and can cause a variety of diseases with varying severity in different animals. Several enteric coronaviruses have been identified that are associated with diarrhea in swine and that have caused substantial economic losses. In this study, a newly emerged porcine enteric alphacoronavirus (PEAV), PEAV-GD-CH/2017, was identified from suckling piglets with diarrhea in southern China, and a full-length genome sequence of PEAV was obtained for systematic analysis. The novel PEAV sequence was most identical to that of bat-HKU2, and the differences between them were comprehensively compared, especially the uniform features of the S protein, which was shown to have a close relationship with betacoronaviruses and to perhaps represent unrecognized betacoronaviruses. In addition, Bayesian analysis was conducted to address the origin of PEAV, and the divergence time between PEAV and bat-HKU2 was estimated at 1926, which indicates that PEAV is not newly emerged and may have circulated in swine herds for several decades since the interspecies transmission of this coronavirus from bat to swine. The evolutionary rate of coronaviruses was estimated to be 1.93×10^{-4} substitutions per site per year for the RdRp gene in our analysis. For the origin of PEAV, we suspect that it is the result of the interspecies transmission of bat-HKU2 from bat to swine. Our results provide valuable information about the uniform features, origin and evolution of the novel PEAV, which will facilitate further

investigations of this newly emerged pathogen.

Keywords: coronavirus; diarrhea; PEAV; origin; evolutionary analysis

1. Introduction

Coronaviruses (CoVs) are enveloped viruses with a single-stranded, positive-sense RNA genome, they belong to the family *Coronaviridae*, and they are found in a wide variety of animals in which they can cause respiratory, hepatic, enteric and neurological diseases of varying severity (Weiss and Navas-Martin, 2005; Woo et al., 2006). CoVs are separated into four distinct genera based on genotypic and serological characterization: alpha-CoV, beta-CoV, gamma-CoV and delta-CoV (Su et al., 2016). To date, several enteric CoVs that are attributed to diarrhea in swine have been identified and have caused substantial economic losses. Transmissible gastroenteritis virus (TGEV) and porcine epidemic diarrhea virus (PEDV) belong to alpha-CoV, and both of them cause life-threatening acute enteric disease in suckling piglets (Pensaert and de Bouck, 1978; Zhang et al., 2017). Porcine hemagglutinating encephalomyelitis virus (PHEV) is a beta-CoV that primarily affects pigs under 3 weeks of age (Pensaert and Callebaut, 1974; Rho et al., 2011). Porcine deltacoronavirus (PDCoV) is a newly identified enteric coronavirus in swine and belongs to delta-CoV (Wang et al., 2014a). The outbreak of severe acute respiratory syndrome (SARS) and the identification of SARS-CoV-like viruses from wild animals in China have boosted interest in the discovery of novel CoVs in both humans and animals. For example, human coronaviruses NL63 and HKU1 were discovered in 2004 and 2005, respectively, and MERS-CoV emerged in 2012 (Fouchier et al., 2004; Woo et al., 2005; Zaki et al., 2012). For animal CoVs, SARS-CoV-like viruses and bat-CoV-HKU2 were discovered in horseshoe bats; novel delta-CoVs, in birds and swine; and additional novel CoVs, in bats and other animals (Chu et al., 2008; Dong et al., 2007; Lau et al., 2005; Lau et al., 2007; Wang et al., 2014b; Woo et al., 2012). Recently, a novel bat-HKU2-like coronavirus that can cause diarrhea in suckling piglets was discovered in swine by two research groups in China (Gong et al., 2017; Pan et al., 2017). This novel enteric coronavirus shares high nucleotide identities (approximately 95%) with the reported

bat-HKU2 strains at the full genome level and is tentatively named porcine enteric alphacoronavirus (PEAV) (Gong et al., 2017).

In this retrospective study, we report the identification of this newly emerged PEAV from a pig farm in Guangdong Province, China, which outbreaks of severe diarrhea in suckling piglets in March 2017. We analyzed and described the genome characteristic of this novel PEAV systematically and the phylogenetic relationship of this virus with other groups of CoVs. Bayesian analysis was also conducted to address the origin and evolutionary history of PEAV, and our results indicate that PEAV emerged approximately 91 years ago and may have circulated in swine herds for several decades.

2. Materials and methods

2.1 Sample collection and disease diagnosis

In March 2017, an acute diarrheal outbreak of newborn-piglet diarrhea occurred in a commercial pig farm in Guangdong Province, China. The clinical manifestations included vomiting, acute watery diarrhea and dehydration in ill suckling piglets. Small intestinal and fecal samples were collected from ill pigs and submitted to the Animal Disease Detection Diagnosis Center of Southern China Agricultural University for pathogen detection. The small intestinal samples were homogenized with phosphate-buffered saline (PBS; 0.1 M, pH 7.4) and subsequently centrifuged at 10,000×g for 10 minutes at 4°C. The fecal samples were resuspended with PBS and centrifuged as described above. Both supernatants were collected for RNA extraction using a TaKaRa MiniBEST Universal RNA Extraction Kit (TaKaRa, Dalian, China), and first-strand cDNA was synthesized using a PrimeScript™ 1st Strand cDNA Synthesis Kit (TaKaRa, Dalian, China) following the manufacturer's instructions. PCR was used for the detection of common enteric viral pathogens as previously described, including PEDV, TGEV, PDCoV and porcine group A rotaviruses (RVAs) (Amimo et al., 2013; Kim et al., 2000; Liu and Wang, 2016; Song et al., 2015). However, all samples were negative for PEDV, TGEV, PDCoV and RVAs. Subsequently, we suspected PEAV infection and conducted a retrospective study of these samples after the report of PEAV in Guangdong (Gong et al., 2017).

2.2 PEAV detection and complete genome sequencing

A pair of primers (forward: 5'-TTTTGGTTCTTACGGGCTGTT-3'; reverse: 5'-CAAACTGTACGCTGGTCAACT-3') based on RNA-dependent RNA polymerase (RdRp) gene of a known bat-HKU2 strain (EF203065) was designed for PEAV detection. After PEAV was detected, 18 pairs of primers were designed based on the bat-HKU2 genome to amplify the full genome (these primer sequences are available on request), and the PCR-amplified products were analyzed by electrophoresis on 1.5% agarose gels and purified using a MiniBEST DNA Extraction Kit (TaKaRa, Dalian, China). The purified PCR product was cloned into the pMD18-T (TaKaRa, Dalian, China) vector for sequencing. Sequences of fragments were assembled using the DNAStar program to produce the final viral genome sequence and used for further analysis.

2.3 Genome analysis and phylogenetic analysis

The complete genome sequence of PEAV and the deduced amino acid sequences of the open reading frames (ORFs) were compared to those of other known CoVs as previously reported (Woo et al., 2012). Multiple sequence alignments were performed by MAFFT, and a phylogenetic tree based on the full-length genome nucleotide sequences of PEAV and of other representative CoVs was constructed using the neighbor-joining method with 1,000 bootstrap replicates in MEGA 5.0 (Tamura et al., 2011). Consideration the extensive divergence between the nucleotide sequences of different coronavirus genera, phylogenetic trees for the ORF1ab, RdRp, S, M, and N proteins were also constructed based on the corresponding amino acid sequences. Bootscan analysis was also performed to detect if a potential recombination event occurred for PEAV using Simplot 3.5.1 with the genome sequence of PEAV as the query. Prediction of transmembrane domains was performed using TMHMM (http://www.cbs.dtu.dk/services/TMHMM/).

2.4 Evolutionary dynamics and estimation of the divergence time of PEAV

The Bayesian Markov chain Monte Carlo (MCMC) method was used to infer the divergence time of PEAV with other members of CoVs in BEAST 1.8.3 as described previously (Drummond and Rambaut, 2007; Fu et al., 2017; Woo et al., 2012). Specifically, analyses

were performed under the GTR+I+Γ nucleotide substitution model for the RdRp gene (2781 bp) and using an unrelaxed lognormal distribution molecular clock with a constant size model. The MCMC algorithm was run for a 100 million step chain and sampled every 10,000 states, and 10% of the chain was removed as burn-in. The maximum clade credibility (MCC) tree was inferred by the Tree Annotator program included in the BEAST package. The mean time of the most recent common ancestor (TMRCA) and the highest posterior density (HPD) regions at 95% were calculated in Tracer 1.6, and posterior probability values provided an assessment of the degree of support for the key node of the tree. The nucleotide substitution rate (per site per year) for coronaviruses was also estimated in this analysis.

3. Results

3.1 Diagnosis and detection of PEAV

All samples were negative for RT-PCR detection of common enteric viruses, including PEDV, TGEV, PDCoV and RVAs. Subsequently, a newly emerged PEAV that can cause diarrhea in suckling piglets was reported in Guangdong, China (Gong et al., 2017); we suspected PEAV infection and conducted a retrospective study of these samples. Considering the high nucleotide identities (approximately 95%) of PEAV with reported bat-HKU2 strains (Gong et al., 2017), we designed a pair of primers based on RNA-dependent RNA polymerase (RdRp) gene of a known bat-HKU2 strain for PEAV detection. To our surprise, an expected 750 bp fragment was amplified from all samples, and the PCR products were further sequenced. The sequences of the PCR products were subjected to BLAST searches in the GenBank database, showed the highest identity to bat-HKU2 strains (approximately 97%), and corresponded to nucleotide positions 12,837-13,570 in the bat-HKU2 genome. The full-length genome of PEAV was finally obtained by segment amplification and named PEAV-GD-CH/2017 (MG742313).

3.2 Genome and S protein feature analysis

The genomic structure of PEAV is organized with the same gene order as that of bat-HKU2, namely, 5'-ORF1a/1b (ORF1ab)-S-ORF3-E-M-N-NS7a-3' (Figure 1), and the genome sequence length of PEAV-GD-CH/2017 is 27,155 nt, excluding the poly (A) tail, which is

similar to previous reports (Gong et al., 2017; Pan et al., 2017). The G+C content of PEAV ranges from 39.34% to 39.41% (Table 1), and the genome nucleotide identities of PEAV-GD-CH/2017 with PEAV-GDS04 (MF167434) and PEAV-GD-01(MF370205) are 99.7% and 99.8%, respectively. All three known PEAV strains are most identical to bat-HKU2 and BtRF-AlphaCoV/YN2012, with approximately 95.0% and 87.5% nucleotide identities, respectively. In addition, comparison of the genomic features of PEAV and of other coronaviruses and the amino acid identities between the predicted ORF1ab, RdRp, S, E, M and N proteins of PEAV and the corresponding proteins of other coronaviruses are summarized in Table 1. Notably, most of these PEAV proteins share higher identities to alpha-CoVs (group B) than the other three groups of coronaviruses, except the S protein, which shares only approximately 25% amino acid identity to that of alpha-CoVs (Table 1). The putative transcription regulatory sequence (TRS) motif, 5'-AACUAAA-3', precedes each ORF of PEAV (Table 2) and has the same TRS sequence as bat-HKU2 and HCoV-NL63 (Lau et al., 2007; Pyrc et al., 2004). The coding potential and putative TRS sequence for each ORF of PEAV are summarized in Table 2. Similar to bat-HKU2, one ORF was observed between the S and E genes, which encodes a putative 229-amino acid nonstructural protein, NS3 (Lau et al., 2007). The NS3 protein of PEAV shares 94% amino acid identity to that of bat-HKU2 but only 42% and 35% identities to those of HCoV-NL63 and PEDV, respectively.

The S protein is the main determinant during coronavirus infection, as it possesses both receptor-binding and fusion functions; it is also the crucial determinant of tissue tropism and host range (Millet and Whittaker, 2015). However, the S protein of PEAV is very unique, similar to that of bat-HKU2; because the amino acid identities to the S proteins of all known coronaviruses are lower than 28%, we systematically analyzed the S protein of PEAV and compared it with those of other coronaviruses. The S protein of PEAV contains 1130 amino acid residues, and the insertion of two amino acid residues (serine and isoleucine) at positions 12 and 13 was observed compared to that of bat-HKU2. Two putative cleavage sites, S1/S2 (VRR ↓ MTFE) and S2' (ESR ↓ SAI**EDLLF**), were found at positions 546 and 673 in the S

protein of PEAV, respectively (Figure 1). Interestingly, the arginine at cleavage site S2' is conserved in the S proteins of almost all four genera of coronaviruses, and this cleavage site have a remarkably conserved motif, E-D-L-L-F; in contrast, the arginine (position 545) at cleavage sites S1/S2 is conserved in S proteins from several beta-CoVs (Table S1). The PEAV S protein is predicted to have a transmembrane domain from positions 1069 to 1091, followed by a short cytoplasmic tail (endodomain), which contains conserved cysteine residues (Figure 1). Pairwise comparison of the amino acid sequences of S proteins of PEAV and bat-HKU2 revealed more mutations at the S1 subunit (122 mutations) than the S2 subunit (26 mutations), particularly in the NTD (amino-terminal domain), which may be related to tissue tropism and host range changes and may result in interspecies transmission from bat to swine.

3.3 Phylogenetic analysis and recombination analysis

Phylogenetic analysis was conducted to address the evolutionary relationship and the potential recombination of PEAV with other coronaviruses based on the nucleotide sequences of the whole genome and the amino acid sequences of ORF1ab, RdRp, S, M and N proteins, respectively (Figures 2 and 3). Obviously, all PEAV strains cluster with bat-HKU2 and BtRF-AlphaCoV/YN2012 and form a distinct lineage (defined as HKU2-like, not shown in the tree) closely related to other alpha-CoVs that belong to group 1b based on the whole genome level (Figure 2). The same result can also be observed from the phylogenetic tree that was constructed based on the amino acid sequences of ORF1ab, RdRp, M and N proteins (Figure 3). However, the evolutionary relationship of PEAV exhibited a uniform feature when phylogenetic analysis was conducted based on the S protein. All PEAV strains cluster with bat-HKU2 and BtRF-AlphaCoV/YN2012 along with a newly identified rat-CoV, LRNV. These strains form a distinct lineage and cluster with beta-CoV but are separate from all four known subgroups of beta-CoVs; we defined this distinct lineage as the beta-like group (Figure 3). These results are consistent with identical amino acid analysis and with those of a previous report (Pan et al., 2017). We also conducted recombination analysis to evaluate if recombination has occurred in the PEAV genome, especially in the S gene, but no significant single recombination event was observed when the genome sequence of PEAV was used as

the query (Figure S1). Additionally, recombination was not observed in bat-HKU2 and LRNV genomes in previous studies (Lau et al., 2007; Wang et al., 2015). Noteworthy, another large difference between PEAV and bat-HKU2 is in N protein, it shows distant phylogenetic relationship comparing with analysis of ORF1ab, RdRp, E and M protein (Figure 3), which is consistent with analysis of amino acid identity (Table 1). 22 amino acid mutations were found during pairwise comparison of the amino acid sequences of N proteins of PEAV and bat-HKU2, and most mutations located in carboxyl terminal, however, N protein is highly conserved among different PEAV strains.

3.4 Origin and the divergence time of PEAV

Because the RdRp gene is the most conserved gene between all coronaviruses, the RdRp gene was used for Bayesian analysis to address the divergence time and evolutionary history of PEAV in this study. The MCC tree constructed based on the RdRp gene has a topology similar to that of the phylogenetic tree that was constructed based on the whole genome and the RdRp protein, with high posterior probability values supporting each key node, and the mean TMRCA was estimated with 95% HPD values (Figure 4). Based on our analysis, the mean TMRCA of bovine-CoV and HCoV-OC43 was estimated at 1914 (95% HPD, 1841 to 1981), and the mean TMRCA of human and civet SARSr-CoV was estimated at 2001 (95% HPD, 1998 to 2003). In addition, the divergence time between HKU15 and PDCoV was estimated at 1986 (95% HPD, 1970 to 1994). All of these results are highly consistent with those of previous studies (Lau et al., 2010; Vijgen et al., 2005; Woo et al., 2017) and indicate that our Bayesian analysis is unbiased. The mean TMRCA of PEAV and bat-HKU2 was estimated at 1926 (95% HPD, 1864 to 1984), approximately 91 years ago, which indicates that PEAV is not newly emerged and may have circulated in swine herds for several decades since its interspecies transmission from bat to swine. PEAV clusters with bat-HKU2; these coronaviruses have a common ancestor with another bat-CoV, BtRf-AlphaCoV/YN2012, and the divergence time between them was estimated at 1783 (95% HPD, 1620 to 1943). All of these bat-HKU2-like coronaviruses are closely related to HCoV-229E and HCoV-NL63 and emerged at approximately 277 (95% HPD, 931 BC to 1434). In addition, the TMRCA for

alpha-CoV, beta-CoV and gamma-CoV were also estimated in our analysis at approximately 827 BC (95% HPD, 2626 BC to 1042), 1419 BC (95% HPD, 3561 BC to 867) and 977 BC (95% HPD, 3313 BC to 1090), respectively. In addition, the TMRCA for all coronaviruses was estimated at 3914 BC (95% HPD, 8637 BC to 45 BC), approximately 6,000 years ago, which indicates that coronaviruses have had a very long evolutionary history since their emergence. The mean evolutionary rate of CoVs was estimated to be 1.93×10^{-4} (95% HPD, 1.27×10^{-4} to 3.57×10^{-4}) nucleotide substitutions per site per year for the RdRp gene based on Bayesian analysis, which is consistent with the results of a previous report (Woo et al., 2012).

For the origin of PEAV, we conjecture that the interspecies transmission of bat-HKU2 from bat to swine occurred approximately 90 years ago. As wild boars have been reported as reservoirs for various pathogens, and can transmit these pathogens into domestic swine, such as porcine circovirus type 2 (PCV2), classical swine fever virus (CSFV) and Hepatitis E virus (HEV) (Adlhoch et al., 2009; Firth et al., 2009; Goller et al., 2016), but whether wild boars plays an important role during the interspecies transmission of bat-HKU2 needs to further investigate.

4. Discussion

Coronaviruses are important pathogens that have a wide host range and cause different kinds of diseases in a variety of animals; many novel coronaviruses have been identified in both humans and animals since the outbreak of SARS in 2003 (Fouchier et al., 2004; Lau et al., 2005; Lau et al., 2007; Wang et al., 2014b; Woo et al., 2005; Woo et al., 2012; Zaki et al., 2012). Several enteric coronaviruses that can cause diarrhea in swine have been identified and have circulated in swine herds for a long time; PEDV, TGEV and PHEV are examples of these viruses (Pensaert and de Bouck, 1978; Rho et al., 2011; Zhang et al., 2017). In particular, large-scale outbreaks of PEDV in China and the USA, with high rates of illness and death in suckling piglets, caused substantial economic losses in late 2010 and 2013, respectively (Huang et al., 2013; Wang et al., 2013).

A newly enteric coronavirus, PDCoV, was identified in the USA in 2014, and this coronavirus caused clinical signs in swine similar to those of PEDV (Wang et al., 2014a). In this study, a

novel PEAV (PEAV-GD-CH/2017) strain was identified from suckling piglets with diarrhea, and this strain shares high identities with the other two PEAV strains that were previously reported (Gong et al., 2017; Pan et al., 2017). These novel PEAVs are most identical to bat-HKU2, with 95% nucleotide identity, and have the same genome organization and TRS motif for each ORF. The greatest difference between PEAV and bat-HKU2 is their S proteins, which share 85% amino acid identity each other, a value much lower compared with those of other proteins (Table 1). This difference is caused by amino acid mutations in the S protein, particularly in the NTD in the S1 subunit, which has been proven to be the key factor determining issue tropism and the host range of coronaviruses (Lu et al., 2015). In addition to its low amino acid identity with the S protein of HKU2-like coronavirus, it shares low amino acid identity (lower than 28%) with S proteins of all known coronaviruses. Thus, clarifying the origin of the S proteins of PEAV and HKU2-like coronavirus is important for determining the origin and evolutionary history of these coronaviruses. A previous study showed that the extreme NTD in the S1 subunit of PEAV is structurally similar to that of NL63, while the rest of the S1 subunit is structurally similar to that of MHV (Pan et al., 2017). In addition, a short peptide in the S protein of bat-HKU2 was found to be homologous to a corresponding peptide within the receptor-binding motif (RBM) in the S1 subunit of SARS-CoV (Lau et al., 2007). We also analyzed the arginine (position 545) at cleavage sites S1/S2 of PEAV and found that it is conserved in several beta-CoVs in this study. Moreover, the phylogenetic tree based on the S protein presents a uniform evolutionary relationship; these bat-HKU2-like coronaviruses cluster with a newly identified rat-CoV, LRNV, which represents a novel species of coronaviruses. All of these strains form a distinct lineage and cluster with beta-CoVs but are separate from all four known subgroups of beta-CoVs (Figure 3), which may indicate that these strains are part of a novel subgroup of beta-CoVs. These results suggest that PEAV and HKU2-like coronaviruses may have some relations with beta-CoVs and most likely resulted from recombination with the backbone of alpha-CoV and the S gene from an unrecognized beta-CoV. Another large difference between PEAV and bat-HKU2 is N protein, which share about 93.9% amino acid identity, as N protein is a multifunctional

protein for coronaviruses, which is involving in virus replication, budding and pathogenesis et al. (McBride et al., 2014). While the role of these mutations in N protein between PEAV and bat-HKU2 should further investigate.

The origin and emergence time of a newly emerged pathogen are important issues to answer to determine the evolutionary history and plan methods of prevention for these new pathogens. For example, previous SARS research reported that the interspecies transfer of SARS-like coronaviruses from bats to the amplifying host (e.g., civet) occurred in 1998 and that interspecies transfer from civet to humans occurred in 2002 (Chinese, 2004; Hon et al., 2008; Lau et al., 2010; Song et al., 2005). These results provide insight into the origin and evolutionary history of SARS coronavirus. The origin and divergence time of other coronaviruses have also been estimated previously; the divergence time of bovine-CoV and HCoV-OC43 could be dated back to the end of the 19th century to the beginning of the 20th century and was estimated at 1910 (Vijgen et al., 2006). The TMRCA for all PDCoV strains was reported at 1991, approximately 24 years before PDCoV was identified (Woo et al., 2017). In this study, we also addressed the emergence time and evolutionary history of PEAV and of the other coronaviruses based on the RdRp gene by Bayesian analysis. In anticipation, the mean divergence time of bovine-CoV and HCoV-OC43 was estimated at 1914, and the mean TMRCA of human and civet SARSr-CoV was estimated at 2001 in our analysis (Figure 4). These results are highly consistent with those of a previous report discussed above and further indicate that our analysis is unbiased. The emergence time of PEAV was estimated at 1926 (95% HPD, 1864 to 1984) based on our analysis, which indicates that PEAV is not newly emerged and may have circulated in swine herds for several decades since interspecies transmission from bat to swine occurred. In addition, these HKU2-like coronaviruses have a common ancestor with HCoV-NL63 and HCoV-229E, and the divergence time was estimated at 277, which indicates that these HKU2-like coronaviruses have a long evolutionary history. The mean TMRCA for alpha-CoV, beta-CoV, and gamma-CoV, as well as those of all coronaviruses estimated in this study, were later compared with those of a previous report (Woo et al., 2012). Nevertheless, the mean TMRCA coincides with the regions with 95%

HPD to each other. The evolutionary rate of different coronaviruses was estimated previously; 4.3×10^{-4} substitutions per site per year was estimated for HCoV-OC43 (Vijgen et al., 2005), and the mean evolutionary rate for group 1b coronaviruses was estimated to be 3×10^{-4} substitutions per site per year (Pyrc et al., 2006). In addition, the evolutionary rate for all coronaviruses was estimated to be 1.3×10^{-4} substitutions per site per year (Woo et al., 2012), which is estimated to be 1.93×10^{-4} (95% HPD, 1.27×10^{-4} to 3.57×10^{-4}) nucleotide substitutions per site per year for the RdRp gene in this study, and all of these results are comparable to each other.

Bats and birds are supposed to be the reservoir hosts for coronaviruses; in particular, bats are the reservoir hosts and gene pools of alpha-CoVs and beta-CoVs, while birds are the reservoir hosts and gene pools of gamma-CoVs and delta-CoVs (Woo et al., 2012). However, whether the first coronaviruses occurred in bats or birds is still unknown. To date, the generally acknowledged evolutionary model for coronaviruses is as follows: the ancestor of bat-CoV was transmitted to another species of bat and generated alpha-CoV and beta-CoV. Interspecies transmission of these bat-CoVs to other bat species and other mammals then occurred, and these coronaviruses are circulating in these hosts. Similarly, the ancestor of bird-CoV was transmitted to another species of birds and generated gamma-CoV and delta-CoV. Interspecies transmission of these bird-CoVs to other bird species and accidentally to some mammalian species (e.g., pig and whale) then occurred (Woo et al., 2012). Bat is also supposed to be the origin of other swine pathogens, such as porcine circovirus type 3 (PCV3), which was suspected to be generated from the interspecies transmission of bat-associated circovirus from bat to swine (Fu et al., 2017). Based on the evolutionary relationship and molecular features of PEAV and bat-HKU2-CoV, as well as the important role of bat in the ecology of coronaviruses, we conjecture that the origin of PEAV is the result of the interspecies transmission of bat-HKU2-CoV from bat to swine approximately 90 years ago.

In summary, the novel PEAV was identified from suckling piglets with diarrhea in southern China, and the full-length genome of PEAV-GD-CH/2017 was obtained in this study. The genome and S protein features of PEAV was systematic analyzed, as well as the evolutionary

relationship of PEAV with other coronaviruses, which indicated PEAV may recombination with unrecognized beta-CoV. PEAV emerged approximately 90 years ago and origin from the interspecies transmission of bat-HKU2 from bat to swine, and wild boars may plays an important role in this process. Thus, epidemiological investigations of PEAV should be further conducted in both swine and wild boars to better understand and clarify the origin and evolutionary history of PEAV. Importantly, considering this infectious coronavirus and its serious clinical implications for suckling piglets (Pan et al., 2017), the development of an effective vaccine for PEAV is urgently needed for the prevention of this disease.

Acknowledgements:

This work was supported by the National Key Research and Development Program (2016YFD0500707), the Modern Agro-Industry Technology Research System (CARS-35) and the Science and Technology Project of Guangdong Province (2015B020203006). Fu Xinliang received financial support from the China Scholarship Council (CSC, 201708440274).

Conflict of interest:

The authors declared no potential conflicts of interest with respect to the research, authorship and publication of this article.

References:

Adlhoch, C., Wolf, A., Meisel, H., Kaiser, M., Ellerbrok, H., Pauli, G., 2009. High HEV presence in four different wild boar populations in East and West Germany. Vet Microbiol 139, 270-278.

Amimo, J.O., Vlasova, A.N., Saif, L.J., 2013. Detection and genetic diversity of porcine group A rotaviruses in historic (2004) and recent (2011 and 2012) swine fecal samples in Ohio: predominance of the G9P[13] genotype in nursing piglets. Journal of clinical microbiology 51, 1142-1151.

Chinese, S.M.E.C., 2004. Molecular evolution of the SARS coronavirus during the course of the SARS epidemic in China. Science 303, 1666-1669.

Chu, D.K., Peiris, J.S., Chen, H., Guan, Y., Poon, L.L., 2008. Genomic characterizations of bat coronaviruses (1A, 1B and HKU8) and evidence for co-infections in Miniopterus bats. J Gen Virol 89, 1282-1287.

Dong, B.Q., Liu, W., Fan, X.H., Vijaykrishna, D., Tang, X.C., Gao, F., Li, L.F., Li, G.J., Zhang, J.X., Yang, L.Q., Poon, L.L., Zhang, S.Y., Peiris, J.S., Smith, G.J., Chen, H., Guan, Y., 2007. Detection of a novel and highly divergent coronavirus from asian leopard cats and Chinese ferret badgers in Southern China. J Virol 81, 6920-6926.

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

Firth, C., Charleston, M.A., Duffy, S., Shapiro, B., Holmes, E.C., 2009. Insights into the evolutionary history of an emerging livestock pathogen: porcine circovirus 2. J Virol 83, 12813-12821.

Fouchier, R.A., Hartwig, N.G., Bestebroer, T.M., Niemeyer, B., de Jong, J.C., Simon, J.H., Osterhaus, A.D., 2004. A previously undescribed coronavirus associated with respiratory disease in humans. Proceedings of the National Academy of Sciences of the United States of America 101, 6212-6216.

Fu, X., Fang, B., Ma, J., Liu, Y., Bu, D., Zhou, P., Wang, H., Jia, K., Zhang, G., 2017. Insights into the epidemic characteristics and evolutionary history of the novel porcine circovirus type 3 in southern China. Transboundary and emerging diseases.

Goller, K.V., Gabriel, C., Dimna, M.L., Le Potier, M.F., Rossi, S., Staubach, C., Merboth, M., Beer, M., Blome, S., 2016. Evolution and molecular epidemiology of classical swine fever virus during a multi-annual outbreak amongst European wild boar. J Gen Virol 97, 639-645.

Gong, L., Li, J., Zhou, Q., Xu, Z., Chen, L., Zhang, Y., Xue, C., Wen, Z., Cao, Y., 2017. A New Bat-HKU2-like Coronavirus in Swine, China, 2017. Emerg Infect Dis 23.

Hon, C.C., Lam, T.Y., Shi, Z.L., Drummond, A.J., Yip, C.W., Zeng, F., Lam, P.Y., Leung, F.C., 2008. Evidence of the recombinant origin of a bat severe acute respiratory syndrome (SARS)-like coronavirus and its implications on the direct ancestor of SARS coronavirus. J Virol 82, 1819-1826.

Huang, Y.W., Dickerman, A.W., Pineyro, P., Li, L., Fang, L., Kiehne, R., Opriessnig, T., Meng, X.J., 2013. Origin, evolution, and genotyping of emergent porcine epidemic diarrhea virus strains in the United States. Mbio 4, e00737-00713.

Kim, L., Chang, K.O., Sestak, K., Parwani, A., Saif, L.J., 2000. Development of a reverse transcription-nested polymerase chain reaction assay for differential diagnosis of transmissible gastroenteritis virus and porcine respiratory coronavirus from feces and nasal swabs of infected pigs. Journal of veterinary diagnostic investigation : official publication of the American Association of Veterinary Laboratory Diagnosticians, Inc 12, 385-388.

Lau, S.K., Li, K.S., Huang, Y., Shek, C.T., Tse, H., Wang, M., Choi, G.K., Xu, H., Lam, C.S., Guo, R.,

Chan, K.H., Zheng, B.J., Woo, P.C., Yuen, K.Y., 2010. Ecoepidemiology and complete genome comparison of different strains of severe acute respiratory syndrome-related Rhinolophus bat coronavirus in China reveal bats as a reservoir for acute, self-limiting infection that allows recombination events. J Virol 84, 2808-2819.

Lau, S.K., Woo, P.C., Li, K.S., Huang, Y., Tsoi, H.W., Wong, B.H., Wong, S.S., Leung, S.Y., Chan, K.H., Yuen, K.Y., 2005. Severe acute respiratory syndrome coronavirus-like virus in Chinese horseshoe bats. Proceedings of the National Academy of Sciences of the United States of America 102, 14040-14045.

Lau, S.K., Woo, P.C., Li, K.S., Huang, Y., Wang, M., Lam, C.S., Xu, H., Guo, R., Chan, K.H., Zheng, B.J., Yuen, K.Y., 2007. Complete genome sequence of bat coronavirus HKU2 from Chinese horseshoe bats revealed a much smaller spike gene with a different evolutionary lineage from the rest of the genome. Virology 367, 428-439.

Liu, X., Wang, Q., 2016. Reverse transcription-PCR assays for the differentiation of various US porcine epidemic diarrhea virus strains. Journal of virological methods 234, 137-141.

Lu, G., Wang, Q., Gao, G.F., 2015. Bat-to-human: spike features determining 'host jump' of coronaviruses SARS-CoV, MERS-CoV, and beyond. Trends in microbiology 23, 468-478.

McBride, R., van Zyl, M., Fielding, B.C., 2014. The coronavirus nucleocapsid is a multifunctional protein. Viruses 6, 2991-3018.

Millet, J.K., Whittaker, G.R., 2015. Host cell proteases: Critical determinants of coronavirus tropism and pathogenesis. Virus research 202, 120-134.

Pan, Y., Tian, X., Qin, P., Wang, B., Zhao, P., Yang, Y.L., Wang, L., Wang, D., Song, Y., Zhang, X., Huang, Y.W., 2017. Discovery of a novel swine enteric alphacoronavirus (SeACoV) in southern China. Vet Microbiol 211, 15-21.

Pensaert, M.B., Callebaut, P.E., 1974. Characteristics of a coronavirus causing vomition and wasting in pigs. Archiv fur die gesamte Virusforschung 44, 35-50.

Pensaert, M.B., de Bouck, P., 1978. A new coronavirus-like particle associated with diarrhea in swine. Archives of virology 58, 243-247.

Pyrc, K., Dijkman, R., Deng, L., Jebbink, M.F., Ross, H.A., Berkhout, B., van der Hoek, L., 2006. Mosaic structure of human coronavirus NL63, one thousand years of evolution. Journal of molecular biology 364, 964-973.

Pyrc, K., Jebbink, M.F., Berkhout, B., van der Hoek, L., 2004. Genome structure and transcriptional regulation of human coronavirus NL63. Virol J 1, 7.

Rho, S., Moon, H.J., Park, S.J., Kim, H.K., Keum, H.O., Han, J.Y., Van Nguyen, G., Park, B.K., 2011. Detection and genetic analysis of porcine hemagglutinating encephalomyelitis virus in

South Korea. Virus genes 42, 90-96.

Song, D., Zhou, X., Peng, Q., Chen, Y., Zhang, F., Huang, T., Zhang, T., Li, A., Huang, D., Wu, Q., He, H., Tang, Y., 2015. Newly Emerged Porcine Deltacoronavirus Associated With Diarrhoea in Swine in China: Identification, Prevalence and Full-Length Genome Sequence Analysis. Transboundary and emerging diseases 62, 575-580.

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. Proceedings of the National Academy of Sciences of the United States of America 102, 2430-2435.

Su, S., Wong, G., Shi, W., Liu, J., Lai, A.C., Zhou, J., Liu, W., Bi, Y., Gao, G.F., 2016. Epidemiology, Genetic Recombination, and Pathogenesis of Coronaviruses. Trends in microbiology 24, 490-502.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Molecular biology and evolution 28, 2731-2739.

Vijgen, L., Keyaerts, E., Lemey, P., Maes, P., Van Reeth, K., Nauwynck, H., Pensaert, M., Van Ranst, M., 2006. Evolutionary history of the closely related group 2 coronaviruses: porcine hemagglutinating encephalomyelitis virus, bovine coronavirus, and human coronavirus OC43. J Virol 80, 7270-7274.

Vijgen, L., Keyaerts, E., Moes, E., Thoelen, I., Wollants, E., Lemey, P., Vandamme, A.M., Van Ranst, M., 2005. Complete genomic sequence of human coronavirus OC43: molecular clock analysis suggests a relatively recent zoonotic coronavirus transmission event. J Virol 79, 1595-1604.

Wang, J., Zhao, P., Guo, L., Liu, Y., Du, Y., Ren, S., Li, J., Zhang, Y., Fan, Y., Huang, B., Liu, S., Wu, J., 2013. Porcine epidemic diarrhea virus variants with high pathogenicity, China. Emerg Infect Dis 19, 2048-2049.

Wang, L., Byrum, B., Zhang, Y., 2014a. Detection and genetic characterization of deltacoronavirus in pigs, Ohio, USA, 2014. Emerg Infect Dis 20, 1227-1230.

Wang, L., Byrum, B., Zhang, Y., 2014b. Porcine coronavirus HKU15 detected in 9 US states,

2014. Emerg Infect Dis 20, 1594-1595.

Wang, W., Lin, X.D., Guo, W.P., Zhou, R.H., Wang, M.R., Wang, C.Q., Ge, S., Mei, S.H., Li, M.H., Shi, M., Holmes, E.C., Zhang, Y.Z., 2015. Discovery, diversity and evolution of novel coronaviruses sampled from rodents in China. Virology 474, 19-27.

Weiss, S.R., Navas-Martin, S., 2005. Coronavirus pathogenesis and the emerging pathogen severe acute respiratory syndrome coronavirus. Microbiology and molecular biology reviews : MMBR 69, 635-664.

Woo, P.C., Lau, S.K., Chu, C.M., Chan, K.H., Tsoi, H.W., Huang, Y., Wong, B.H., Poon, R.W., Cai, J.J., Luk, W.K., Poon, L.L., Wong, S.S., Guan, Y., Peiris, J.S., Yuen, K.Y., 2005. Characterization and complete genome sequence of a novel coronavirus, coronavirus HKU1, from patients with pneumonia. J Virol 79, 884-895.

Woo, P.C., Lau, S.K., Lam, C.S., Lau, C.C., Tsang, A.K., Lau, J.H., Bai, R., Teng, J.L., Tsang, C.C., Wang, M., Zheng, B.J., Chan, K.H., Yuen, K.Y., 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.

Woo, P.C., Lau, S.K., Li, K.S., Poon, R.W., Wong, B.H., Tsoi, H.W., Yip, B.C., Huang, Y., Chan, K.H., Yuen, K.Y., 2006. Molecular diversity of coronaviruses in bats. Virology 351, 180-187.

Woo, P.C., Lau, S.K., Tsang, C.C., Lau, C.C., Wong, P.C., Chow, F.W., Fong, J.Y., Yuen, K.Y., 2017. Coronavirus HKU15 in respiratory tract of pigs and first discovery of coronavirus quasispecies in 5'-untranslated region. Emerging microbes & infections 6, e53.

Zaki, A.M., van Boheemen, S., Bestebroer, T.M., Osterhaus, A.D., Fouchier, R.A., 2012. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. The New England journal of medicine 367, 1814-1820.

Zhang, X., Zhu, Y., Zhu, X., Shi, H., Chen, J., Shi, D., Yuan, J., Cao, L., Liu, J., Dong, H., Jing, Z., Zhang, J., Wang, X., Feng, L., 2017. Identification of a natural recombinant transmissible gastroenteritis virus between Purdue and Miller clusters in China. Emerging microbes & infections 6, e74.

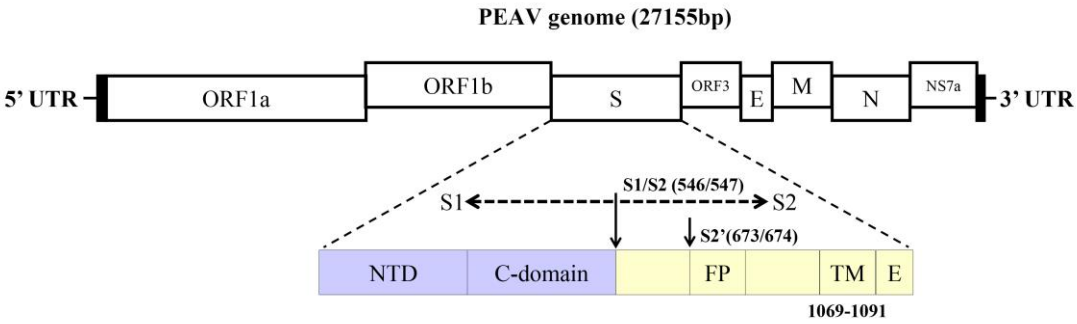
Figure legends:

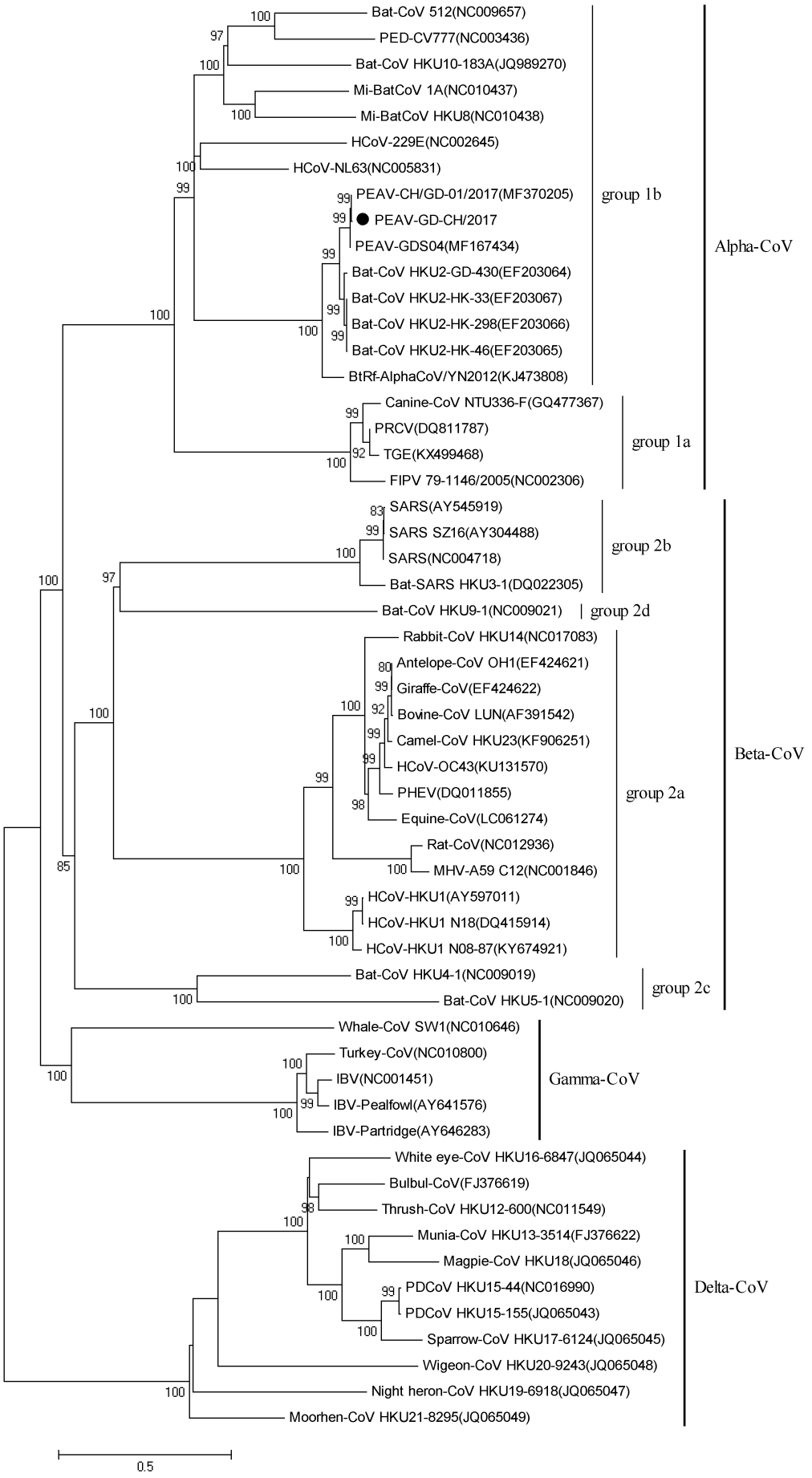
Figure 1: Diagram of the structural organization of the PEAV genome. The putative cleavage sites S1/S2 and S2' in the S protein are shown by arrows, and the numbers indicate the amino acid positions in the S protein of PEAV. The S protein is composed of two subunits: the S1 receptor-binding subunit, and the S2 fusion subunit. NTD: N-terminal domain of S1; C-domain: C-terminal domain of S1; FP: putative fusion peptide; TM: transmembrane domain; E: endodomain. Not drawn to scale.

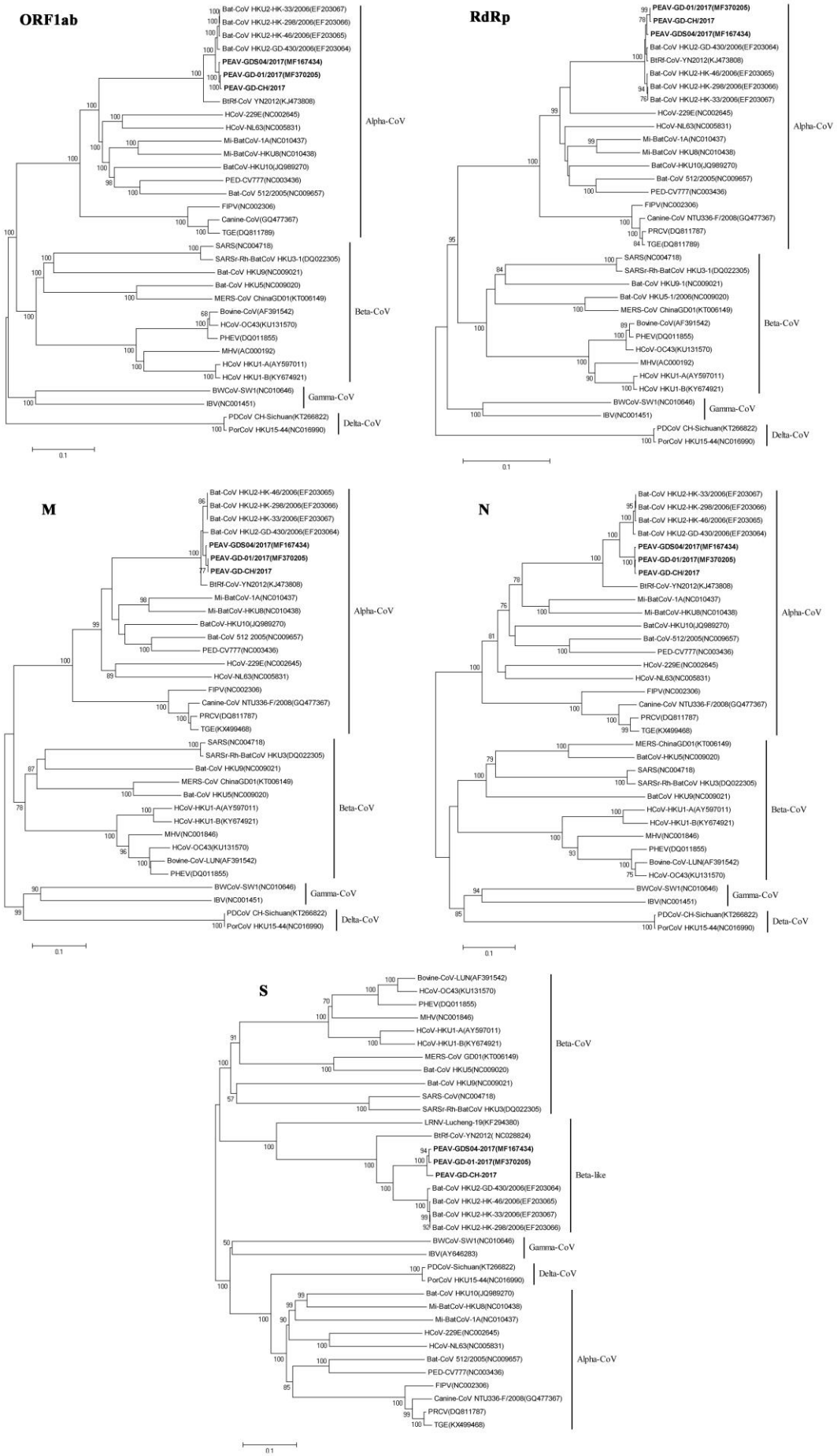
Figure 2: Phylogenetic analysis of PEAV with other four genera of coronaviruses based on full-length genome sequences. The tree was constructed by the neighbor-joining method with 1,000 bootstrap replicates in MEGA 5.0 after multiple sequence alignments by MAFFT. Alpha-CoV and beta-CoV subgroups are shown in the tree, and the PEAV strain (PEAV-GD-CH/2017) identified in this study is indicated with a solid black circle.

Figure 3: Phylogenetic analysis of the ORF1ab, RdRp, M, N and S proteins of PEAV based on the amino acid sequences of these proteins. These trees were constructed using the neighbor-joining method with 1,000 bootstrap replicates in MEGA 5.0. The amino acid lengths of the ORF1ab, RdRp, M, N and S proteins used in this analysis are 6262 aa, 927 aa, 229 aa, 342 aa and 1130 aa, respectively. The PEAV strains are shown in bold in these trees.

Figure 4: Bayesian maximum clade credibility (MCC) phylogenetic tree was constructed in BEAST 1.8.3 using the Markov chain Monte Carlo (MCMC) method based on the RdRp gene (2781 bp). The mean TMRCA (time of the most recent common ancestor) was estimated for each key node with 95% HPD (highest posterior density) and is shown in brackets. High posterior probability values are shown for each key node and provide an assessment of the degree of support for the node on the tree. BC dates are identified with a suffix, while AD dates are not.







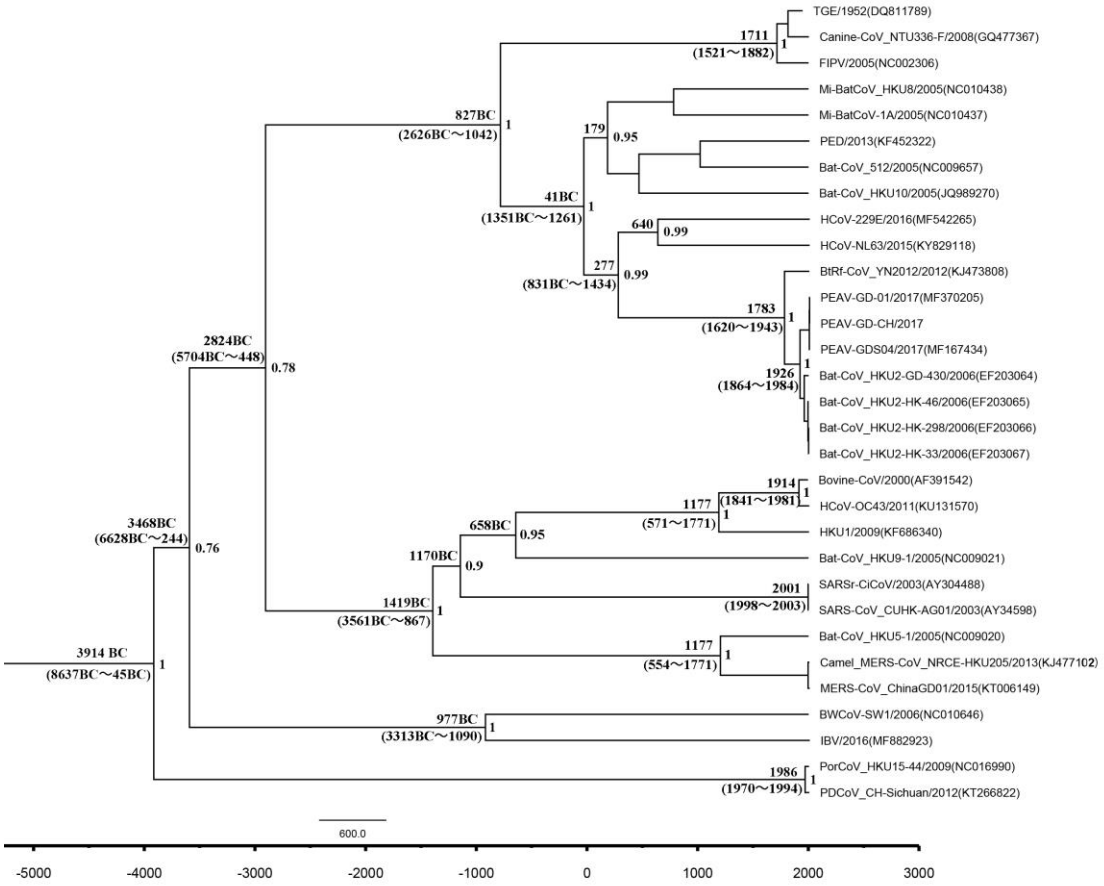


Table 1. Comparison of the genomic features of PEAV and other coronaviruses and amino acid identities between the predicted ORF1ab, RdRp, S, E, M and N proteins of PEAV and the corresponding proteins of other coronaviruses

Coronaviruses[a]	Genome size (bases)	Features G+C content (%)	Pairwise amino acid identity (%)					
			ORF1ab	RdRp	S	E	M	N
Alpha-CoV group A								
TGEV	28,614	37.58	55.7	75.6	25.2	27.6	52.4	41.7
FIPV	29,355	38.14	55.5	75.5	25.5	27.6	52.4	42.7
PRCV	27,550	37.46	55.7	75.5	24.0	27.6	54.6	41.5
Alpha-CoV group B								
HCoV-229E	27,317	38.26	60.9	80.8	25.1	51.3	56.9	46.7
HCoV-NL63	27,553	34.46	60.0	78.9	25.5	49.3	58.4	49.7
PEDV	28,033	42.02	60.1	78.0	25.2	47.3	64.6	47.1
Bat-CoV HKU2	27,165	39.28	98.3	99.1	85.2	97.3	96.1	93.9
BtRF-CoV YN2012	26,975	37.80	94.5	98.9	78.6	96.0	96.9	88.0
PEAV-GD-01	27,155	39.43	99.7	99.9	98.4	98.7	98.7	99.7
PEAV-GDS04	27,154	39.34	99.5	99.4	98.1	97.3	98.2	99.5
PEAV-GD-CH	27,155	39.41	NA[b]	NA	NA	NA	NA	NA
Beta-CoV group A								
HCoV-HKU1	29,926	32.06	36.1	56.6	26.9	25.0	35.1	26.8
HCoV-OC43	30,746	36.65	35.9	57.6	27.7	24.0	35.6	28.6
MHV	3,1616	41.78	36.5	56.4	26.6	25.0	37.4	29.9
PHEV	30,480	37.25	35.6	57.4	27.2	25.3	37.3	27.3
Beta-CoV group B								
SARS-CoV	29,751	40.76	37.8	59.7	25.8	25.3	32.1	25.2
Beta-CoV group C								
Bat-CoV HKU5	30,482	43.19	38.2	59.0	26.4	22.7	33.2	29.5
Beta-CoV group D								
Bat-CoV HKU9	29,114	41.05	36.5	58.0	26.0	18.4	34.4	23.1
Gamma-CoV								
IBV	27,679	37.93	36.3	59.3	21.1	17.1	21.5	25.6
Delta-CoV								
PDCoV	25,404	43.28	32.3	50.2	23.2	18.3	22.0	19.6

[a] TGEV, porcine transmissible gastroenteritis virus; FIPV, feline infectious peritonitis virus; PRCV, porcine respiratory coronavirus; HCoV-229E, human coronavirus 229E; HCoV-NL63, human coronavirus NL63; PEDV, porcine epidemic diarrhea virus; PEAV, porcine enteric alphacoronavirus; HCoV-HKU1, human coronavirus HKU1; HCoV-OC43, human coronavirus OC43; MHV, murine hepatitis virus; PHEV, porcine hemagglutinating encephalomyelitis virus; SARS-CoV, severe acute respiratory syndrome coronavirus; IBV,

infectious bronchitis virus; PDCoV, porcine deltacoronavirus.
[b] NA, data not available for analysis.

Table 2. Coding potential and putative transcription regulatory sequences (TRSs) of PEAV

Coronaviruses	ORFs	Start-end (nucleotide position)	No. of nucleotides	No. of amino acids	Putative TRS	
					Nucleotide position in the genome	TRS
PEAV	1ab	297-20,482 (shift at 12,434)	20,186	6,728	69	AACUAAAC(220[a])AUG
	S	20,479-23,871	3,393	1,130	20,473	AACUAAAUG
	NS3	23,871-24,560	690	229	23,826	AACUAAAC(37)AUG
	E	24,541-24,768	228	75	24,532	AACUAAAC(1)AUG
	M	24,777-25,463	687	228	24,768	AACUAAAC(1)AUG
	N	25,475-26,602	1,128	375	25,463	AACUAAAC(4)AUG
	NS7a	26,614-26,913	300	99	26,606	AACUAAACAUG

[a] Number means the number of nucleotides from the TRS to AUG.

Highlights

- Identify and sequence a PEAV strain from suckling piglets with diarrhea.
- The S protein of PEAV may recombination from unrecognized beta-CoV.
- The novel PEAV was emerged approximately at 1926 based on Bayesian analysis.
- PEAV origin from the interspecies transmission of bat-HKU2 from bat to swine.