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.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Research paper

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

Xinliang Fu¹,², Bo Fang¹,², Yixing Liu¹,², Mengkai Cai¹,², Junming Jun¹,², Jun Ma²,³, Dexin Bu¹,², Lifang Wang¹,², Pei Zhou²,³, Heng Wang²,³,⁎, Guihong Zhang¹,²,⁎

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

ARTICLE INFO

Keywords:
Coronavirus
Diarrhea
PEAV
Origin
Evolutionary analysis

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 possibly 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 $1.93 \times 10^{-4}$ 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.

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 Bouch, 1978; Zhang et al., 2017). Porcine hemorrhagic 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, 2007; Wang et al., 2014b; Woo et al., 2005). 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 pigs. 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, pH7.4) and subsequently centrifuged at 10,000 × g for 10 min 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′-TTTTGGTTCTCACGGGCTGTT-3′; reverse: 5′-CAAACGTACGCCGTGCAACT-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 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 (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).

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 1000 bootstrap replicates in MEGA 5.0 (Tamura et al., 2011). Consideration of 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., 2018; Woo et al., 2012). Specifically, analyses were performed under the GTR + I + F 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 (TMRA) 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′ (Fig. 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 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).

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′ (Fig. 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 BrRF-AlphacCov/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↓SAIEDLLF), were found at positions 546 and 673 in the S protein of PEAV, respectively (Fig. 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 (endordomain), which contains conserved cysteine residues (Fig. 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.

### 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<sup>a</sup> | Genome Features | Pairwise amino acid identity (%) |
|--------------------------|-----------------|---------------------------------|
|                          | Size (bases)    | G + C content (%)   | ORF1ab | RdRp | S   | E   | M   |
| Alpha-CoV group A        |                |                    |        |      |     |     |     |
| TGEV                     | 28,614          | 37.58              | 55.7   | 75.6 | 25.2 | 27.6 | 52.4 |
| FIPV                     | 29,355          | 38.14              | 55.5   | 75.5 | 25.5 | 27.6 | 52.4 |
| PRCV                     | 27,550          | 37.46              | 55.7   | 75.5 | 24.0 | 27.6 | 54.6 |
| Alpha-CoV group B        |                |                    |        |      |     |     |     |
| HCoV-229E                | 27,317          | 38.26              | 60.9   | 80.8 | 25.1 | 51.3 | 56.9 |
| HCoV-NL63                | 27,553          | 34.46              | 60.0   | 78.9 | 25.5 | 49.3 | 58.4 |
| PEDV                     | 28,033          | 42.02              | 60.1   | 78.0 | 25.2 | 47.3 | 64.6 |
| Bat-CoV HKU2             | 27,165          | 39.28              | 98.3   | 99.1 | 85.2 | 97.3 | 91.6 |
| BtRF-CoV YN2012          | 26,975          | 37.80              | 94.5   | 98.9 | 78.6 | 96.0 | 96.9 |
| PEAV-GD-01               | 27,155          | 39.34              | 99.5   | 99.4 | 98.1 | 97.3 | 98.2 |
| PEAV-GD504               | 27,155          | 39.41              | NA<sup>b</sup> | NA | NA | NA | NA |
| PEAV-GD-CH               | 27,155          | 39.41              | NA<sup>b</sup> | NA | NA | NA | NA |
| Beta-CoV group A         |                |                    |        |      |     |     |     |
| HCoV-HKU1                | 29,926          | 32.06              | 36.1   | 56.6 | 26.9 | 25.0 | 35.1 |
| HCoV-OC43                | 30,746          | 36.65              | 35.9   | 57.6 | 27.7 | 24.0 | 35.6 |
| MHV                      | 3,616           | 41.78              | 36.5   | 56.4 | 26.6 | 25.0 | 37.4 |
| PHEV                     | 30,480          | 32.25              | 35.6   | 57.4 | 27.2 | 25.3 | 37.3 |
| Beta-CoV group B         |                |                    |        |      |     |     |     |
| SARS-CoV                 | 29,751          | 40.76              | 37.8   | 59.7 | 25.8 | 25.3 | 32.1 |
| Beta-CoV group C         |                |                    |        |      |     |     |     |
| Bat-CoV HKU5             | 30,482          | 43.19              | 38.2   | 59.0 | 26.4 | 22.7 | 33.2 |
| Beta-CoV group D         |                |                    |        |      |     |     |     |
| Bat-CoV HKU9             | 29,114          | 41.05              | 36.5   | 58.0 | 26.0 | 18.4 | 34.4 |
| Gamma-CoV                |                |                    |        |      |     |     |     |
| IBV                      | 27,679          | 37.93              | 36.3   | 59.3 | 21.1 | 17.1 | 21.5 |
| Delta-CoV                | 25,404          | 43.28              | 32.3   | 50.2 | 23.2 | 18.3 | 22.0 |
| PDCoV                    |                |                    |        |      |     |     |     |
| 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 diarrheea 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. |
respectively (Figs. 2 and 3). Obviously, all PEAV strains cluster with bat-CoVs based on the nucleotide sequences of the whole genome and relationship and the potential recombination of PEAV with other coronaviruses (Lau et al., 2010; Vijgen 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 investigation. The same result can also be observed from the phylogenetic tree that was constructed based on the amino acid sequences of ORF1ab, RdRp, S, M and N proteins (Fig. 3). These results are consistent with identical amino acid analysis during 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; Adlhoch et al., 2012), but whether wild boars plays an important role during the interspecies transmission of bat-HKU2 needs to further investigate.

### 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 (Figs. 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 (Fig. 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 (Fig. 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-CoVs but are separate from all four known subgroups of beta-CoVs; we defined this distinct lineage as the beta-like group (Fig. 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 (Fig. 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 (Fig. 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 (Fig. 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 SARS-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 MRCA 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 MRCA for all coronaviruses was estimated at 3914 BC (95% HPD, 8637 BC to 45 BC), approximately 6000 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 $1.93 \times 10^{-4}$ (95% HPD, $1.27 \times 10^{-4}$ to $3.57 \times 10^{-4}$) 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; Adlhoch et al., 2012), 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, 2007; Wang et al., 2014b; Woo et al., 2005, 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 pigs, 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.
(caption on next page)
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 (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 coronaviruses (Lu et al., 2015). In addition, the arginine (position 12) at the backbone of alpha-CoV and the S gene from an unrecognized beta-CoV (Fu et al., 2018). Based on the associated circovirus from bat to swine (Fu et al., 2018). Based on the interspecies transmission of these bird-CoVs to other bird species and generated gamma-CoV and delta-CoV. In- terceptors 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 supposed to be generated from the interspecies transmission of bat-associated circovirus from bat to swine (Fu et al., 2018). 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 systematically 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.
(caption on next page)
Fig. 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 1000 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.

Fig. 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.

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 Xiniang 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.
Amino, 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. J. Clin. Microbiol. 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, 1262–1267.
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 Evol. Biol. 7, 214.
Firth, C., Charleston, M.A., Duffy, S., Shapiro, B., Holmes, E.C., 2009. Insights into the
