Phylogeography and evolutionary dynamics analysis of porcine delta-coronavirus with host expansion to humans

Gen Li[1], Shao-Lun Zhai[2], Xia Zhou[1], Tianbao Chen[1], Jiawei Niu[1], YS Xie[1], Guangbin Si[1], Feng Cong[3], Ruiai Chen[1], and Dongsheng He[1]

[1]South China Agricultural University College of Veterinary Medicine
[2]Institute of Animal Health Guangdong Academy of Agricultural Sciences Scientific Observation and Experiment Station of Veterinary Drugs and Diagnostic Techniques of Guangdong Province Ministry of Agriculture of Rural Affairs and Key Laboratory of Animal Disease Prevention of Guangdong Province Guangzhou 510640 China
[3]Guangdong Laboratory Animals Monitoring Institute and Guangdong Provincial Key Laboratory of Laboratory Animals Guangzhou 510633 China

September 25, 2021

Abstract

From 2003 onwards, three pandemics have been caused by coronaviruses: severe acute respiratory syndrome coronavirus (SARS-CoV); middle east respiratory syndrome coronavirus (MERS-CoV); and, most recently, SARS-CoV-2. Notably, all three were transmitted from animals to humans. This would suggest that animals are potential sources of epidemics for humans. The emerging porcine delta-coronavirus was reported to infect children. This is a red flag that marks the ability of PDCoV to break barriers of cross-species transmission to humans. Therefore, we conducted molecular genetic analysis of global clade PDCoV to characterize spatio-temporal patterns of viral diffusion and genetic diversity. PDCoV was classified into three major lineages, according to distribution and phylogenetic analysis of PDCoV. It can be determined that PDCoV originated in Asia—most likely in Southeast Asia—through inference of migration rate and transmission routes. We also selected six special spike amino acid sequences to align and analyze to find seven significant mutation sites. The accumulation of these mutations may enhance dynamic movements, accelerating spike protein membrane fusion events and transmission. Altogether, our study offers a novel insight into the diversification, evolution, and interspecies transmission and origin of PDCoV and emphasizes the need to study the zoonotic potential of the PDCoV and comprehensive surveillance and enhanced biosecurity precautions for PDCoV.

1 Introduction

Coronaviruses (CoVs) are positive-stranded and single-stranded RNA viruses with an envelope, which belong to members of the order *Nidomiformes* , family*coronaviridae* and have the largest genomes, from 25kb to 32kb, among known RNA viruses (Weiss & Navas-Martin, 2005). CoVs were the first to isolate in 1937 from the chicken and had the ability of cross-species transmission. It can infect birds and mammals, including humans, and mainly cause respiratory and digestive tract symptoms (Weiss & Navas-Martin, 2005). Since the outbreak of SARS in 2003, new coronaviruses continue to be discovered. MERS-CoV has been reported in recent years. SARS-CoV-2 is still sweeping the globe now (Gates, 2020; Lancet, 2013; Li, Niu, Fan, Chen, & He, 2021; Stadler et al., 2003). As of 13 August 2021, there have been 205,338,159 confirmed cases of COVID-19, including 4,333,094 deaths, reported to the WHO (https://covid19.who.int/). Current research suggests that SARS-CoV-2 may have originated from bats or pangolins (Xiao et al., 2020; Zhou et al., 2020).

The emerging porcine delta-coronavirus was first identified in swine specimens collected from 2011 to 2012 in Hong Kong (Woo et al., 2012). In 2014, a severe diarrhea outbreak occurred on a US pig farm, suspected

of PEDV infection. However, PDCoV was detected in the feces of pigs instead of PEDV. Subsequently, PDCoV was detected in Korea, Canada, India, Thailand, and mainland China (Ajayi et al., 2018; Hu et al., 2015; Lee et al., 2016; Lorsirigool et al., 2017), posing a significant threat to the swine industry.

In 2017, Woo found PDCoV may be able to cause respiratory infections in pigs and that in addition to fecal-oral transmission, the virus could possibly spread through the respiratory route (Woo et al., 2017). In 2019, specific pathogen-free (SPF) chickens were inoculated with PDCoV and showed mild diarrhea symptoms and low fecal viral RNA shedding (Liang et al., 2019). From May 2014 to December 2015, Lednicky collected a total of 369 samples from children presenting to the school clinic with acute undifferentiated febrile illness in Ayiti and found three of them were infected with coronavirus strains, which clustered with PDCoV (Lednicky et al., 2021). Therefore, PDCoV represented significant potential in cross-species transmission and spread in new host populations, implying their potential threat to human health.

To dig into the origin, evolution and transmission of PDCoV, phylogenic analysis was used to analyze full genome and S gene sequences of PDCoV to examine source population, time of origin, evolutionary rate, amino acid sites and potential dissemination routes. The data in the study will help to formulate countermeasures to deal with the possible future risk of zoonotic transmission of PDCoV to humans.

2 Materials and methods

2.1 Sample collection and virus isolation

A total of 314 rectal swabs were collected from eight intensive pig farm, located in Guangdong Province, China between 2017 and 2019. PDCoV isolation was performed using LLC porcine kidney (LLC-PK) cell line. The optimal cell culture conditions to isolate and propagate PDCoV in LLC-PK required Dulbecco's modified eagle medium (DMEM, Gibco, USA), supplemented with 10% (v/v) fetal bovine serum (FBS, Gibco), and 10 μg/mL of trypsin (Chen et al., 2016; Hu et al., 2015; Jung, Miyazaki, Hu, & Saif, 2018). The result of isolation was confirmed by an indirect immunofluorescence assay (ELISA) and reverse transcription-polymerase chain reaction (RT-PCR) (Supplementary Tables S1).

2.2 Virus extraction and complete gene sequencing

Viral RNA was extracted from 200 μL of sample using TRIzol reagent (TaKaRa Biotechnology, Dalian, China) according to the manufacturer's instructions, and nucleic acids were finally eluted in 30 μL of the RNase free H_2O. Viral genomic RNA was immediately used for PCR or stored at -80 °C until use. In total, 15 pairs were designed for PCR reactions to sequence the PDCoV genomes (Supplementary Tables S2). Positive plasmids of PDCoV were extracted using a Plasmid Mini Kit I (200) (Omega Bio-tek, Norcross, GA, USA). Then the positive plasmids of PDCoV were sequenced by Sangon Biotech (Shanghai, China) Co., Ltd.

2.3 Sequence analysis and model test

We accessed all available PDCoV genomes sequences up to June 2021 from NCBI GenBank (https://www.ncbi.nlm.nih.gov/). A total of 130 PDCoV genomes (Supplementary Tables S3)—two of which were sequenced by our laboratory—and 130 spike genes sequences, were retrieved by PhyloSuite (version 1.2.2) to attain no identical sequences (Chen et al., 2016; Hu et al., 2015; Jung, Miyazaki, Hu, & Saif, 2018). These PDCoV genomes sequences were input to MAFFT (multiple sequence alignment based on fast Fourier transform) to align (Katoh, Kuma, Toh, & Miyata, 2005; Katoh, Misawa, Kuma, & Miyata, 2002). Then aligned sequences were input to Gblocks to determine the conserved domains and edited manually (Talavera & Castresana, 2007). The best fit nucleotide substitution models were tested by ModelFinder (version 1.6.8) (Kalyaanamoorthy, Minh, Wong, von Haeseler, & Jermiin, 2017).

2.4 Maximum-likelihood and Bayesian Inference analysis of the PDCoV

The best fit general time reversible substitution model with an empirical base frequencies and relaxing gamma distributed rate heterogeneity (GTR+F+R5) with 1000 bootstraps for Maximum-likelihood (ML) trees of PDCoV genomes was constructed by IQ-RTEE (version 1.6.8) according to corrected Akaike information criterion (AICc) (Minh et al., 2020).

The best fit general time reversible substitution model with a proportion of invariant sites, empirical base frequencies and gamma distributed rate heterogeneity (GTR+F+I+G4) with 2000000 generations for Bayesian Inference (BI) trees of PDCoV genomes was input to MrBayes (version 3.2.6) according to corrected Akaike Information Criterion (AICc) (Huelsenbeck & Ronquist, 2001; Ronquist et al., 2012; Ronquist & Huelsenbeck, 2003).

The final results were illustrated in the Interactive Tree of Life (iTOL, https://itol.embl.de/) (Letunic & Bork, 2007; 2021).

2.5 Phylogenetic analysis of spike gene

Bayesian evolutionary analysis by sampling trees (BEAST) package was used to estimate the time to the most recent common ancestor (tMRCA) and evolutionary rates. A total of 130 spike genes sequences were selected from 130 PDCoV genomes and input to IQ-TREE to reconstruct the ML tree. Then we used TempEst (version 1.5.3) to ensure that these sequences had sufficient temporal symbol (Rambaut, Lam, Max Carvalho, & Pybus, 2016). According to the ModelFinder (version 1.6.8) result for BEAST (version 1.10.4), the Bayesian Markov chain Monte Carlo (MCMC) method implemented in the BEAST (BEAST, version 1.10.4) package and BEAGLE was used (Ayres et al., 2012; Suchard & Rambaut, 2009), with the GTR+G nucleotide substitution model, a strict molecular clock model and the Coalescent Bayesian Skyline tree prior (Drummond, Rambaut, Shapiro, & Pybus, 2005; Ferreira & Suchard, 2008). The chain length was 500000000 with a sampling frequency every 50000 steps. Tracer software (version 1.7) was used to test that all parameter estimates yielded effective sampling size >200 and burning up the period of 10% of the total chain length (Rambaut, Drummond, Xie, Baele, & Suchard, 2018). The final Bayesian maximum clade credibility (MCC) tree was generated by TreeAnnotator (version 2.6.3) (http://beast.community/treeannotator) and illustrated in Figtree (version 1.4.3) (http://tree.bio.ed.ac.uk/software/figtree/). Then, to estimate the population dynamics of PDCoV, Bayesian Skyline Plot was used to reconstruct the population history. The results were shown by analysis of Bayesian Skyline Reconstruction in Tracer (version 1.7) (Barido-Sottani et al., 2018).

2.6 Phylogeographic inference

To estimate phylogenies and divergence times and explore phylogeography, the same preprocessing as 2.5 on a total of 130 spike genes sequences were input to BEAST (version 1.10.4) package, with the GTR+G nucleotide substitution model, an uncorrelated relaxed molecular clock model and the BSSVS (Bayesian Stochastic Search Variable Selection) traits model (Barido-Sottani et al., 2018)(Li & Drummond, 2012; Su et al., 2015), according to the results from path sampling and stepping-stone sampling (Supplementary Tables S4)(Lartillot & Philippe, 2006). The chain length was 500000000 with a sampling frequency every 50000 steps. Tracer software (version 1.7) was used to test that all parameter estimates yielded effective sampling size >200 and burning the period of 10% of the total chain length. The trees were generated by TreeAnnotator (version 2.6.3) (http://beast.community/treeannotator) and illustrated in Figtree (version 1.4.3) (http://tree.bio.ed.ac.uk/software/figtree/). The phylogeography results were analyzed by Spatial Phylogenetic Reconstruction of Evolutionary Dynamics (SPREAD3, version 0.9.7) (Bielejec, Rambaut, Suchard, & Lemey, 2011).

2.7 Analysis of spike protein

According to the features of PDCoV from GenBank, Illustrator for Biological Sequences (IBS, version 1.0.3) was used to visualize protein-encoding segments in PDCoV (Liu et al., 2015). Amino acid sequences of spike protein were input to SWISS-MODEL to find proper models and illustrated by Visual Molecular Dynamics (VMD, version 1.9.4) (http://www.ks.uiuc.edu/) (Waterhouse et al., 2018). BioAider (version 1.334) was used to analyze mutation of these protein sequences (Zhou, Qiu, Pu, Huang, & Ge, 2020).

3 Results

3.1 Detection and isolation of PDCoV

There was a total of 42 positive samples for PDCoV in 314 samples, which were suspected of PDCoV, with a positive rate of 13.4%. Two PDCoV completed genome were obtained in this study, whose accessions were MH715491 and MT263013 in GenBank. Moreover, a strain of PDCoV (MT263013), which could be stably passaged on the LLC-PK cell line, was successfully isolated and confirmed by an indirect immunofluorescence assay (ELISA) and reverse transcription-polymerase chain reaction (RT-PCR) (Fig. 1).

3.2 Distribution and phylogenetic analysis of PDCoV

In 2012, PDCoV was first found in Hong Kong, China, and broke out in the United States in 2014. At present, PDCoV has been reported in more than 11 countries all over the world. In China, pig farms in 14 provinces were infected with PDCoV (Fig. 2). It has spread all over the world, showing a global trend. The analysis results from the ML tree and BI tree for PDCoV completed genome were almost identical (Fig. 3). The results indicated that the full genomes would be classified into three major lineages, which named the Southeast Asia (SEA) lineage, including Thailand, Vietnam and Laos, America lineage, including the USA, Peru, Japan, and South Korea (JSK), and China (CHN) lineage. We also observed that MW685622 and MW685624 in Ayiti were highly similar to KY065120 in Tianjin, China (99.8%) and MW685623 in Ayiti were highly similar to KR150443 in Arkansas, USA (98.9%).

Bayesian Skyline Plot analysis revealed that the estimated effective population size went up from 15 to 45 between 1989 and 2010, after a brief fluctuation, and the effective population floated in a range of around 70 up to 2019 (Fig. 4a).

According to the root-to-tip regression from TempEst (version 1.5.3), the analysis of temporal structure revealed aspects of the clock-like structure of spike gene (n = 130, correlation coefficient = 0.56, R^2 = 0.32), which indicated the sufficiently strong temporal signal to estimate time-calibrated phylogenies using molecular clock models (Supplementary Figs S1). Similar to the full genome, spike genes were also classified into three major lineages by the analysis of maximum clade credibility (MCC) trees (Fig. 4b). Moreover, our reconstruction confirmed that the virus spread from the CHN lineage, however, it was interesting that SEA lineage was the origin according the BSSVS analysis. The MCC tree indicated that the probability for PDCoV originating from the CHN lineage (49.05%) and SEA lineage (48.45%) is similar. BSSVS analysis demonstrated PDCoV spread from Southeast Asia to the USA, Ayiti, Peru and China with high BF value and posterior probability (Fig. 5a, Supplementary Tables S5) and spread from China with low BF value and posterior probability. Combining these two points, it can be determined that PDCoV originated in Asia, more likely in Southeast Asia, as consistent with the results of Xiao's phylogenetic analysis (Ye et al., 2020).

The phylogeographic inference indicated that PDCoV might have originated around June of 1989 (June 1982–January 1996, 95% highest posterior density).

3.3 The spread of PDCoV

The worldwide spatial dispersal networks of PDCoV were reconstructed. We selected the transmission routes with BF values exceeding 3 and posterior probability exceeding 0.5 to analyze (Ye et al., 2020). There were six discrete sampling locations and nine significant transmission routes (Fig. 6). SEA and JSK were the major output of PDCoV. SEA was linked with four locations, including Ayiti (BF = 68.10, migration rate = 0.924), JSK (BF = 9.48, migration rate = 0.938), the USA (BF = 6.97, migration rate = 0.936), and China (BF = 38425, migration rate = 1.345). JSK had connections to four locations, including Peru (BF = 54.13, migration rate = 0.913), USA (BF = 4.62, migration rate = 0.927), SEA (BF = 9.93, migration rate = 0.958) and Ayiti (BF = 70.21, migration rate = 0.933). In addition, there was a transmission routes from the USA to JSK (BF = 7.52, migration rate = 3.553) (Fig. 5b, Supplementary Tables S5). China and Southeast Asia are adjacent to each other, the distance is about 2000km and the communication is relatively close. Therefore, there was the highest BF value and high migration rate. Remarkably, a special case got our attention. We observed that a strong signal of viral dissemination from the USA to JSK, even though the two places are approximately 11200 km apart, suggesting that the PDCoV in the JSK may spread from the USA, as consistent with the results of MCC tree. In Ayiti, pigs were reintroduced from North American populations, mainly the USA and in small part, Canada (Alexander, 1992) and there was a certain link

between Ayiti and JSK. Infections in Ayiti were likely to be associated with the importation of pigs from the USA.

3.4 Protein structure analysis

Two samples named LC216914 and LC216915 in GenBank were collected from pigs' nasopharyngeal, suggesting PDCoV may be able to cause respiratory infections in pigs (Woo et al., 2017). A sample named MK248485 in GenBank was collected from chickens, suggesting PDCoV can infect chickens (Boley et al., 2020). Three samples named MW685622, MW685623, and MW685624 in GenBank were collected from children's blood, suggesting PDCoV has the potential to infect humans (Lednicky et al., 2021). Comparing the six special sequences with all the sequences, it was found that there were similar changes in seven amino acid sites. The structure of the S protein from residues 52 to 1017 was shown in Fig. 7, because this region of beginning and end are hydrophobic and can adversely affect protein solubility (Lednicky et al., 2021).

Residue 38 was mutated from P to L, which may affect the secondary structure of the protein, because proline is a subamino acid, which cannot form intra-chain hydrogen bonds and is prone to β-turn angle formation. Residue 40 was mutated from R to S, which reduces the space resistance and enhances hydrophilicity. There was a N-glycosylation site between residue 41 and 44 so residue deletion at site 45 may affect glycosylation (Lednicky et al., 2021). The mutation between A and V at residue 137 and 551 may eliminates specific Van der Waals contact, potentially enhancing protein flexibility and dynamic movement of S1. Moreover, this change may represent a common mechanism that enhances dynamic movements, accelerating virus membrane fusion events and transmission (Lednicky et al., 2021; Thompson et al., 2021). Mutation at residue 670 altered the spatial site resistance. The phosphorylation of the protein is mainly carried out on tyrosine, serine, and threonine residues in the peptide chain. Residue 689 was mutated from S to A, with a phosphorylation site losing (Fig. 8 and Table. 1).

4 Discussion

The outbreak of COVID-19 has raised great concern all over the world and poses serious threat to global public health. Therefore, Covs gained increasing attention. When researchers examined blood samples from Ayiti children in early 2021, three samples were found to be infected with PDCoV (Lednicky et al., 2021). Of note, PDCoV has repeatedly crossed the host barrier to chicken (Boley et al., 2020; Liang et al., 2019). This phenomenon exhibits the same cross-species transmission as SARS-COV-2, which suggested that the emerging porcine delta-coronavirus may be a potential threat to human health. Here, we adopted a phylogeographic approach to examine the time of origin, evolutionary rate, diversification patterns and potential transmission routes of PDCoV and selected six special spike amino acid sequences to align and analyze.

PDCoV is an emerging coronavirus causing severe enteritis, diarrhea, and vomiting in piglets. PDCoV was first discovered in Hong Kong, China in 2012, but began to spread in the United States in 2014. Now many countries in the world have reported that pigs were infected with PDCoV. We attained 42 positive samples for PDCoV from eight intensive pig farms, located in Guangdong Province, China and sequenced two PDCoV completed genomes. Moreover, a strain of PDCoV, which could be stably passaged on the LLC-PK cell line was successfully isolated. According to phylogenetic analysis of PDCoV, our findings indicated that PDCoV would be classified into three major lineages: China lineage; USA lineage; and Southeast Asia lineage. There were three PDCoV in Ayiti that can infect children, two of which were highly similar to KY065120 in Tianjin, China (99.8%), and the rest were highly similar to KR150443 in Arkansas, USA (98.9%). However, after suffering from African swine fever in Ayiti, pigs were reintroduced mainly from American populations. The worldwide spatial dispersal networks of PDCoV also indicated the US lineage was one of the origins to Ayiti. Therefore, PDCoV infected in Ayiti was most likely from America (Alexander, 1992; Lednicky et al., 2021).

The MCC tree showed that PDCoV originated in Asia, between China and southeast Asia, and more likely in Southeast Asia in 1989. In addition, according to Bayesian Skyline Plot analysis, effective population size of PDCoV tended to level off and stabilize at around 70. We observed the association between geographic distance and migration rate to clarify space dynamics. However, we observed a strong signal of viral dissem-

ination (migration rate = 3.55) from US to Japan and South Korea, with a long distance (about 11200 km). This showed there was no certain connection between geographic distance and migration rate. Moreover, the scatter plot of distance and migration rate also did not display a certain trend (Supplementary Tables S6, Fig S2). It may be due to the swine trade between countries and artificial or feed contamination that affects the rate of natural transmission.

To analyse the effect of amino acid mutation on cross-species transmission, we found six PDCoV sequences that can spread across species, and align the amino acids of their spike protein. By analyzing the amino acid sequence of spike protein, seven significant mutation sites were defined. The accumulation of these mutations may enhance dynamic movements, accelerating virus membrane fusion events and transmission.

Frankly, this study has several limitations. First, sampling bias, no same number of samples in different regions, may have affected the Bayesian phylogenic reconstruction and inference of the transmission networks (Alexander, 1992; Lednicky et al., 2021). Second, the information from sequences contains the most regions infected with PDCoV but does not cover all regions, because there are some places infected with PDCoV, but no sequence uploaded in GenBank.

In summary, our findings offer novel insights into the diversification, evolution, and interspecies transmission and origin of PDCoV. We also briefly analyzed why humans are infected with PDCoV. Increasing evidence strongly implicates that CoVs have a high capacity for cross-species transmission. Given that pigs are in frequent contact with humans and wild animals, the lower interspecies hurdles in pigs would make them a potential mixing vessel for CoVs (Ye et al., 2020). Therefore, comprehensive surveillance and enhanced biosecurity precautions for PDCoV should be taken to prevent further pandemic such as COVID-19.

Declaration of Competing Interest

The authors state that they have no competing interests

Acknowledgments

This work was supported by Key Laboratory of Zoonosis Prevention and Control of Guangdong Province, the Guangdong Province Pig Industrial System Innovation Team (Grant Number 2018LM1103).

Ethical Approval

Ethical statement is not applicable in this study. The sampling and data publication were approved by the owners of animal. Animals were humanly treated during samples collected.

References:

Ajayi, T., Dara, R., Misener, M., Pasma, T., Moser, L., & Poljak, Z. (2018). Herd-level prevalence and incidence of porcine epidemic diarrhoea virus (PEDV) and porcine delta-coronavirus (PDCoV) in swine herds in Ontario, Canada. *Transboundary and Emerging Diseases,*65(5)**,** 1197-1207.doi:10.1111/tbed.12858

Alexander, F. C. (1992). Experiences with African swine fever in Haiti.*Annals of the New York Academy of Sciences,* 653**,** 251-256

Ayres, D. L., Darling, A., Zwickl, D. J., Beerli, P., Holder, M. T., Lewis, P. O.,... Suchard, M. A. (2012) BEAGLE: an application programming interface and high-performance computing library for statistical phylogenetics..

Barido-Sottani, J., Bošková, V., Plessis, L. D., Kühnert, D., Magnus, C., Mitov, V.,... Stadler, T. (2018). Taming the BEAST-A Community Teaching Material Resource for BEAST 2. *Systematic Biology,* 67(1)**,**170-174.doi:10.1093/sysbio/syx060

Bielejec, F., Rambaut, A., Suchard, M. A., & Lemey, P. (2011). SPREAD: spatial phylogenetic reconstruction of evolutionary dynamics.*Bioinformatics (Oxford, England),* 27(20)**,**2910-2912.doi:10.1093/bioinformatics/btr481

Boley, P. A., Alhamo, M. A., Lossie, G., Yadav, K. K., Vasquez-Lee, M., Saif, L. J.,... Kenney, S. P. (2020). Porcine Delta-coronavirus Infection and Transmission in Poultry, United States(1). *Emerging Infectious Diseases,* 26(2)**,** 255-265.doi:10.3201/eid2602.190346

Chen, J. F., Wang, X. B., Jiao, H. X., Shi, H. Y., Zhang, X., Liu, J. B.,... Li, F. (2016). Isolation and identification of the first porcine delta-coronavirus strain in China. *Chinese Journal of Preventive Veterinary Medicine*

Drummond, A. J., Rambaut, A., Shapiro, B., & Pybus, O. G. (2005). Bayesian coalescent inference of past population dynamics from molecular sequences. *Molecular Biology and Evolution,* 22(5)**,**1185-1192

Ferreira, M. A. R., & Suchard, M. A. (2008). Bayesian analysis of elapsed times in continuous-time Markov chains. *Canadian journal of statistics,* 36(3)**,** 355-368.doi:10.1002/cjs.5550360302

Gates, B. (2020). Responding to Covid-19 - A Once-in-a-Century Pandemic? *The New England journal of medicine,* 382(18)**,**1677-1679.doi:10.1056/NEJMp2003762

Hofstad, M. S., & Yoder, H. W. J. (1966) Avian infectious bronchitis–virus distribution in tissues of chicks.

Hu, H., Jung, K., Vlasova, A. N., Chepngeno, J., Lu, Z., Wang, Q.,... Saif, L. J. (2015). Isolation and characterization of porcine delta-coronavirus from pigs with diarrhea in the United States.*Journal of Clinical Microbiology,* 53(5)**,**1537-1548.doi:10.1128/JCM.00031-15

Huelsenbeck, J. P., & Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics (Oxford, England),*17(8)**,** 754-755

Jia, Y., Cao, J., & Wei, Z. (2021). Bioinformatics Analysis of Spike Proteins of Porcine Enteric Coronaviruses. *Biomed Research International,* 2021**,** 6689471.doi:10.1155/2021/6689471

Jung, K., Miyazaki, A., Hu, H., & Saif, L. J. (2018). Susceptibility of porcine IPEC-J2 intestinal epithelial cells to infection with porcine delta-coronavirus (PDCoV) and serum cytokine responses of gnotobiotic pigs to acute infection with IPEC-J2 cell culture-passaged PDCoV. *Veterinary Microbiology,* 221**,**49-58.doi:10.1016/j.vetmic.2018.05.019

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., & Jermiin, L. S. (2017) ModelFinder: fast model selection for accurate phylogenetic estimate

Katoh, K., Kuma, K., Toh, H., & Miyata, T. (2005) MAFFT version 5: improvement in accuracy of multiple sequence alignment.

Katoh, K., Misawa, K., Kuma, K., & Miyata, T. (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform.

Lancet, T. (2013) MERS-CoV: a global challenge.

Lartillot, N., & Philippe, H. (2006). Computing Bayes factors using thermodynamic integration. *Systematic Biology,* 55(2)**,** 195-207

Lednicky, J. A., Tagliamonte, M. S., White, S. K., Elbadry, M. A., Alam, M. M., Stephenson, C. J.,... Morris, J. G. (2021) Emergence of porcine delta-coronavirus pathogenic infections among children in Haiti through independent zoonoses and convergent evolution..

Lee, J. H., Chung, H. C., Nguyen, V. G., Moon, H. J., Kim, H. K., Park, S. J.,... Park, B. K. (2016). Detection and Phylogenetic Analysis of Porcine Delta-coronavirus in Korean Swine Farms, 2015. *Transboundary and Emerging Diseases,* 63(3)**,**248-252.doi:10.1111/tbed.12490

Lemey, P., Rambaut, A., Drummond, A. J., & Suchard, M. A. (2009). Bayesian phylogeography finds its roots. *Plos Computational Biology,* 5(9)**,** e1000520.doi:10.1371/journal.pcbi.1000520

Letunic, I., & Bork, P. (2007). Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and. *Bioinformatics (Oxford, England),* 23(1)**,** 127-128

Letunic, I., & Bork, P. (2021). Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and. *Nucleic Acids Research,*49(W1)**,** W293-W296.doi:10.1093/nar/gkab301

Li, G., Niu, J., Fan, X., Chen, T., & He, D. (2021) The chilly climate may increase the chance of infecting COVID-19.

Li, W. L. S., & Drummond, A. J. (2012). Model averaging and Bayes factor calculation of relaxed molecular clocks in Bayesian phylogenetics.*Molecular Biology and Evolution,* 29(2)**,**751-761.doi:10.1093/molbev/msr232

Liang, Q., Zhang, H., Li, B., Ding, Q., Wang, Y., Gao, W.,... Hu, H. (2019). Susceptibility of Chickens to Porcine Delta-coronavirus Infection. *Viruses,* 11(6).doi:10.3390/v11060573

Liu, W., Xie, Y., Ma, J., Luo, X., Nie, P., Zuo, Z.,... Ren, J. (2015). IBS: an illustrator for the presentation and visualization of biological sequences. *Bioinformatics (Oxford, England),* 31(20)**,**3359-3361.doi:10.1093/bioinformatics/btv362

Lorsirigool, A., Saeng-Chuto, K., Madapong, A., Temeeyasen, G., Tripipat, T., Kaewprommal, P.,... Nilubol, D. (2017). The genetic diversity and complete genome analysis of two novel porcine delta-coronavirus isolates in Thailand in 2015. *Virus Genes,*53(2)**,** 240-248.doi:10.1007/s11262-016-1413-z

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A.,... Lanfear, R. (2020). IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era.*Molecular Biology and Evolution,* 37(5)**,**1530-1534.doi:10.1093/molbev/msaa015

Müller, N. F., Rasmussen, D., & Stadler, T. (2018). MASCOT: parameter and state inference under the marginal structured coalescent.*Bioinformatics (Oxford, England),* 34(22)**,**3843-3848.doi:10.1093/bioinformatics/bty406

RAGGI, L. G., & LEE, G. G. (1965). LACK OF CORRELATION BETWEEN INFECTIVITY, SEROLOGIC RESPONSE AND CHALLENGE RESULTS IN IMMUNIZATION WITH AN AVIAN INFECTIOUS BRONCHITIS VACCINE. *Journal of immunology (Baltimore, Md. : 1950),* 94**,** 538-543

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., & Suchard, M. A. (2018). Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7.*Systematic Biology,* 67(5)**,**901-904.doi:10.1093/sysbio/syy032

Rambaut, A., Lam, T. T., Max Carvalho, L., & Pybus, O. G. (2016). Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus evolution,* 2(1)**,** w7

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S.,... Huelsenbeck, J. P. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space.*Systematic Biology,* 61(3)**,**539-542.doi:10.1093/sysbio/sys029

Ronquist, F., & Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics (Oxford, England),* 19(12)**,** 1572-1574

Shang, J., Zheng, Y., Yang, Y., Liu, C., Geng, Q., Tai, W.,... Li, F. (2018). Cryo-Electron Microscopy Structure of Porcine Delta-coronavirus Spike Protein in the Prefusion State. *Journal of Virology,*92(4)**,** e1517-e1556.doi:10.1128/JVI.01556-17

Stadler, K., Masignani, V., Eickmann, M., Becker, S., Abrignani, S., Klenk, H.,... Rappuoli, R. (2003). SARS–beginning to understand a new virus.*Nature reviews. Microbiology,* 1(3)**,** 209-218

Su, Y. C. F., Bahl, J., Joseph, U., Butt, K. M., Peck, H. A., Koay, E. S. C.,... Smith, G. J. D. (2015). Phylodynamics of H1N1/2009 influenza reveals the transition from host adaptation to immune-driven selection.*Nature Communications,* 6**,** 7952.doi:10.1038/ncomms8952

Suchard, M. A., & Rambaut, A. (2009). Many-core algorithms for statistical phylogenetics. *Bioinformatics (Oxford, England),*25(11)**,** 1370-1376.doi:10.1093/bioinformatics/btp244

Talavera, G., & Castresana, J. (2007). Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Systematic Biology,* 56(4)**,** 564-577

Thompson, C. N., Hughes, S., Ngai, S., Baumgartner, J., Wang, J. C., McGibbon, E.,... Fine, A. D. (2021). Rapid Emergence and Epidemiologic Characteristics of the SARS-CoV-2 B.1.526 Variant - New York City, New York, January 1-April 5, 2021. *MMWR. Morbidity and mortality weekly report,* 70(19)**,** 712-716.doi:10.15585/mmwr.mm7019e1

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R.,... Schwede, T. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Research,*46(W1)**,** W296-W303.doi:10.1093/nar/gky427

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(4)**,** 635-664

Winterfield, R. W., Fadly, A. M., & Hoerr, F. J. (1976). Immunity to infectious bronchitis virus from spray vaccination with derivatives of a Holland strain. *Avian Diseases,* 20(1)**,** 42-48

Woo, P. C. Y., Lau, S. K. P., Lam, C. S. F., Lau, C. C. Y., Tsang, A. K. L., Lau, J. H. N.,... Yuen, K. (2012). Discovery of seven novel Mammalian and avian coronaviruses in the genus delta-coronavirus supports bat coronaviruses as the gene source of alphacoronavirus and betacoronavirus and avian coronaviruses as the gene source of gammacoronavirus and delta-coronavirus. *Journal of Virology,*86(7)**,** 3995-4008.doi:10.1128/JVI.06540-11

Woo, P. C., Lau, S. K., Tsang, C., Lau, C. C., Wong, P., Chow, F. W.,... Yuen, K. (2017). Coronavirus HKU15 in respiratory tract of pigs and first discovery of coronavirus quasispecies in 5'-untranslated region.*Emerging Microbes & Infections,* 6(6)**,**e53.doi:10.1038/emi.2017.37

Xiao, K., Zhai, J., Feng, Y., Zhou, N., Zhang, X., Zou, J.,... Shen, Y. (2020) Isolation of SARS-CoV-2-related coronavirus from Malayan pangolins..

Ye, X., Chen, Y., Zhu, X., Guo, J., Da, X., Hou, Z.,... Xiao, S. (2020). Cross-Species Transmission of Deltacoronavirus and the Origin of Porcine. *Evolutionary Applications,* 13(9)**,**2246-2253.doi:10.1111/eva.12997

Zhang, D., Gao, F., Jakovlić, I., Zou, H., Zhang, J., Li, W. X.,... Wang, G. T. (2020) PhyloSuite: An integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies..

Zhang, J., Chen, Y., Shan, N., Wang, X., Lin, S., Ma, K.,... Qi, W. (2020). Genetic diversity, phylogeography, and evolutionary dynamics of highly pathogenic. *Virus evolution,* 6(2)**,**a79.doi:10.1093/ve/veaa079

Zhou, P., Yang, X., Wang, X., Hu, B., Zhang, L., Zhang, W.,... Shi, Z. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature,* 579(7798)**,**270-273.doi:10.1038/s41586-020-2012-7

Zhou, Z., Qiu, Y., Pu, Y., Huang, X., & Ge, X. (2020). BioAider: An efficient tool for viral genome analysis and its application in tracing SARS-CoV-2 transmission. *Sustainable Cities and Society,* 63**,**102466.doi:10.1016/j.scs.2020.102466

Fig.1 Isolation and identification of PDCoV

(a) The samples were tested by RT-PCR to determine that the samples were infected with PDCoV and no other common virus were present. (b) Isolation and passage of PDCoV in LLC-PK cells. Magnification: 200×. (c) Indirect immunofluorescence assay on LLC-PK cells infected with PDCoV. (d) Indirect immunofluorescence assay on the intestine of PDCoV-infected pigs. The primary antibody is a monoclonal antibody against N protein.

Fig. 2 Geographical distribution of PDCoV around the world and in China

(a) Different countries are denoted by different colors; (b) Different provinces in China are denoted by different colors.

Fig. 3 Maximum-likelihood tree and bayesian inference tree

(a) ML tree of completed PDCoV gene. (b) BI tree of completed PDCoV gene. Different colors represent different regions. Japan and South Korea, JSK. Southeast Asia, SEA. USA, the Unite States America.

Fig. 4 Demographic history of PDCoV spike gene

(a) Demographic history was inferred by bayesian skyline analysis. The median and 95% HPD intervals were plotted. (b) A MCC tree of the S gene sampled in nine different countries. PDCoV were classified into three major lineages. The pie charts indicate the probability of chosen nodes. Except for the four chosen nodes shown, the probability of the rest of the nodes is greater than 90%, so they are not shown.

Fig. 5 BF supports and migration rate for each transmission routes

(a) Sufficient signal transmission routes were shown. (b) Migration rates were indicated for each support transition routes. (c) The migration changes of PDCoV in each region were exhibited.

Fig. 6 Spatial diffusion of PDCoV

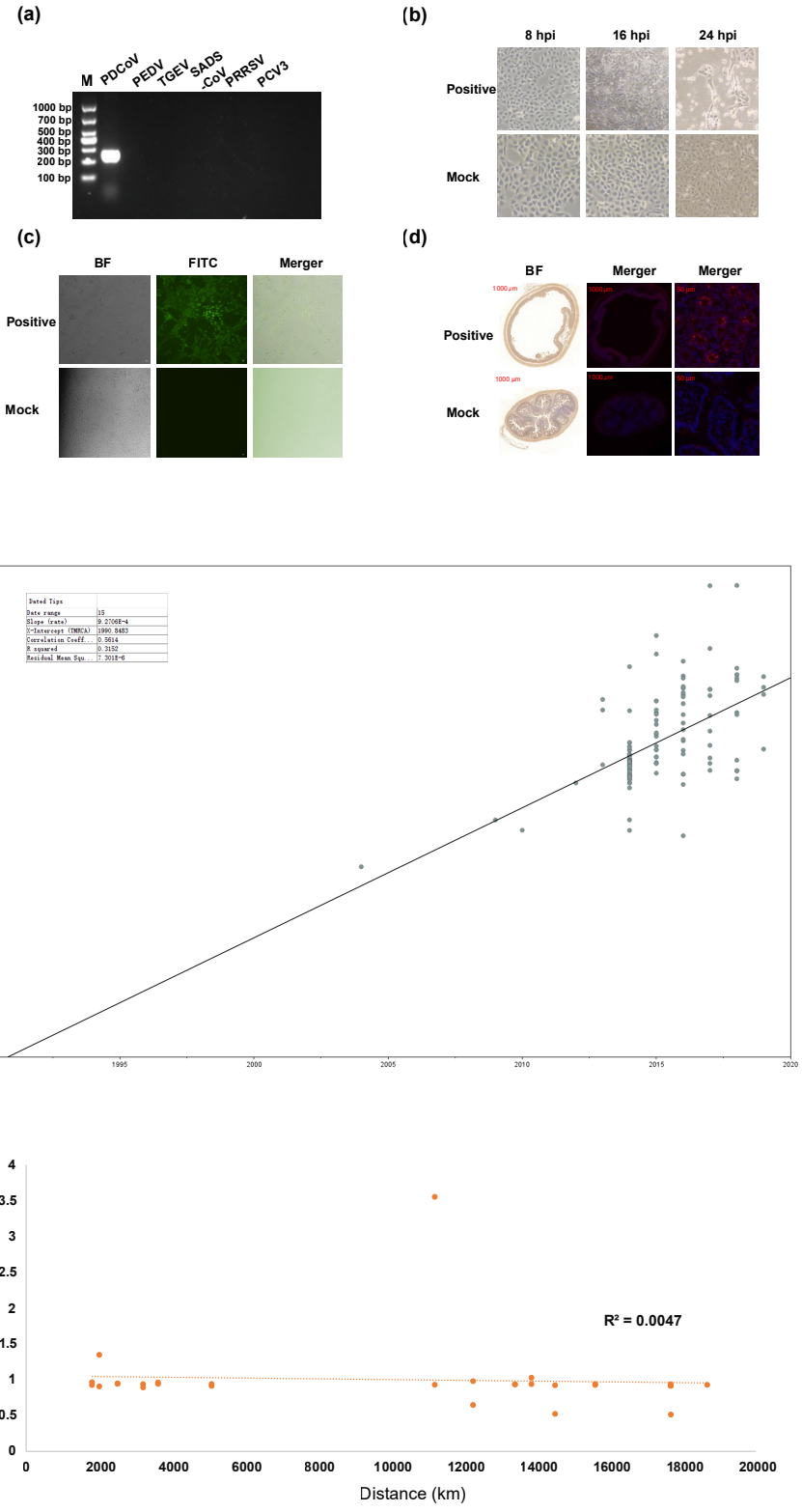
The spatial diffusion of PDCoV was determined by Bayesian phylogeography inference of spike gene sequences. Curves show the among-country virus lineage transitions statistically supported with BF >3 and posterior probability >0.5 for PDCoV. Widths and colors of the curve represent migration rate and BF support, respectively.

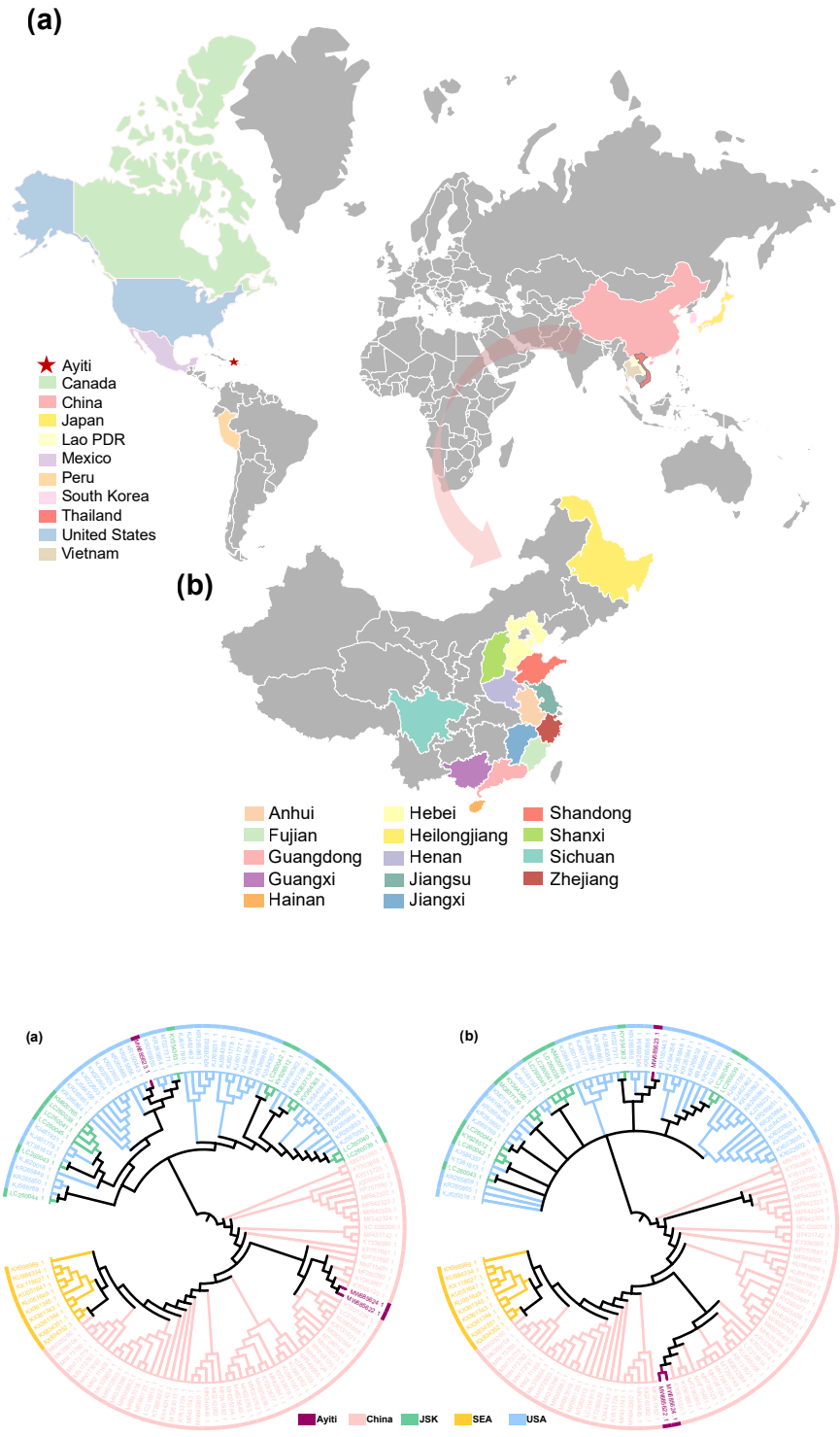
Fig. 7 Visualization of PDCoV structure

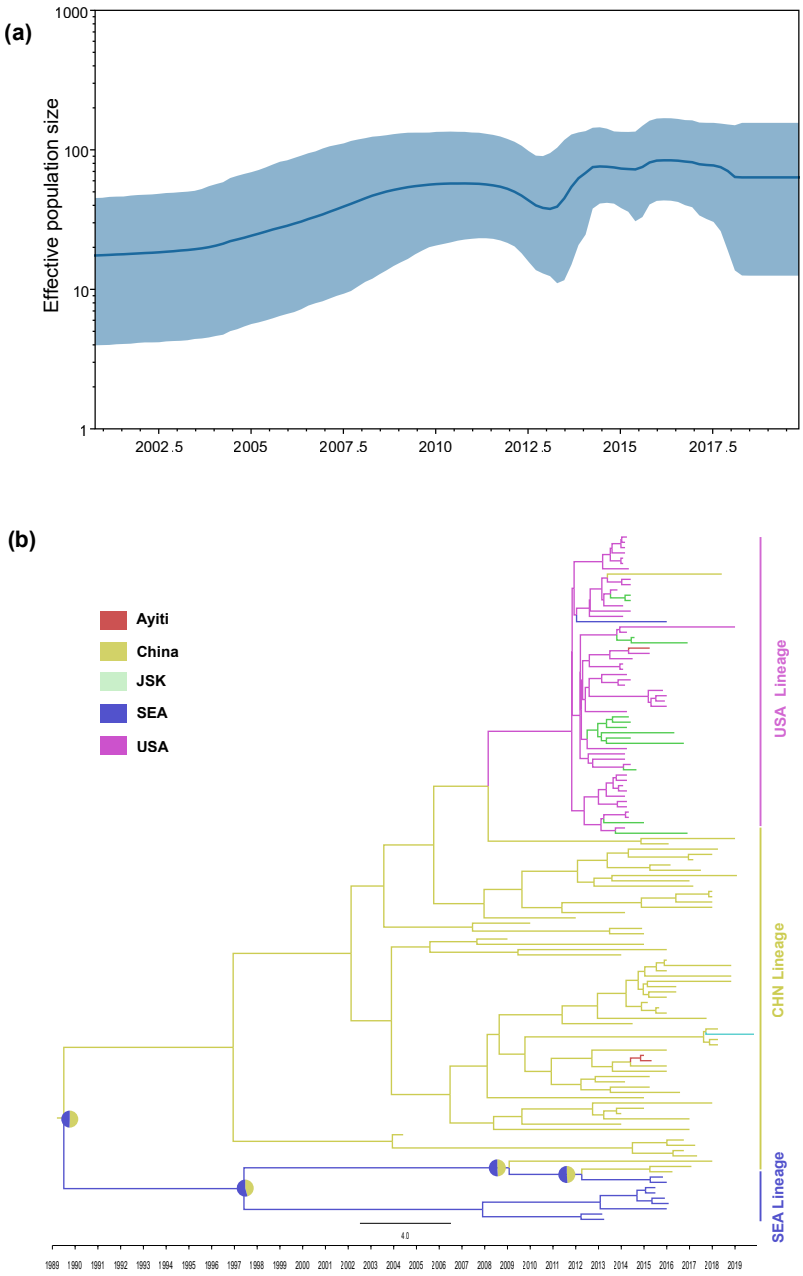
(a) Overall structure of PDCoV contains two large ORF (opening reading frame), four structural proteins and tree accessory proteins. (b) Structure of spike protein is classified into three parts, S1-NTD (N-terminal domain), S1-CTD (C-terminal domain) and S2(Shang et al., 2018).

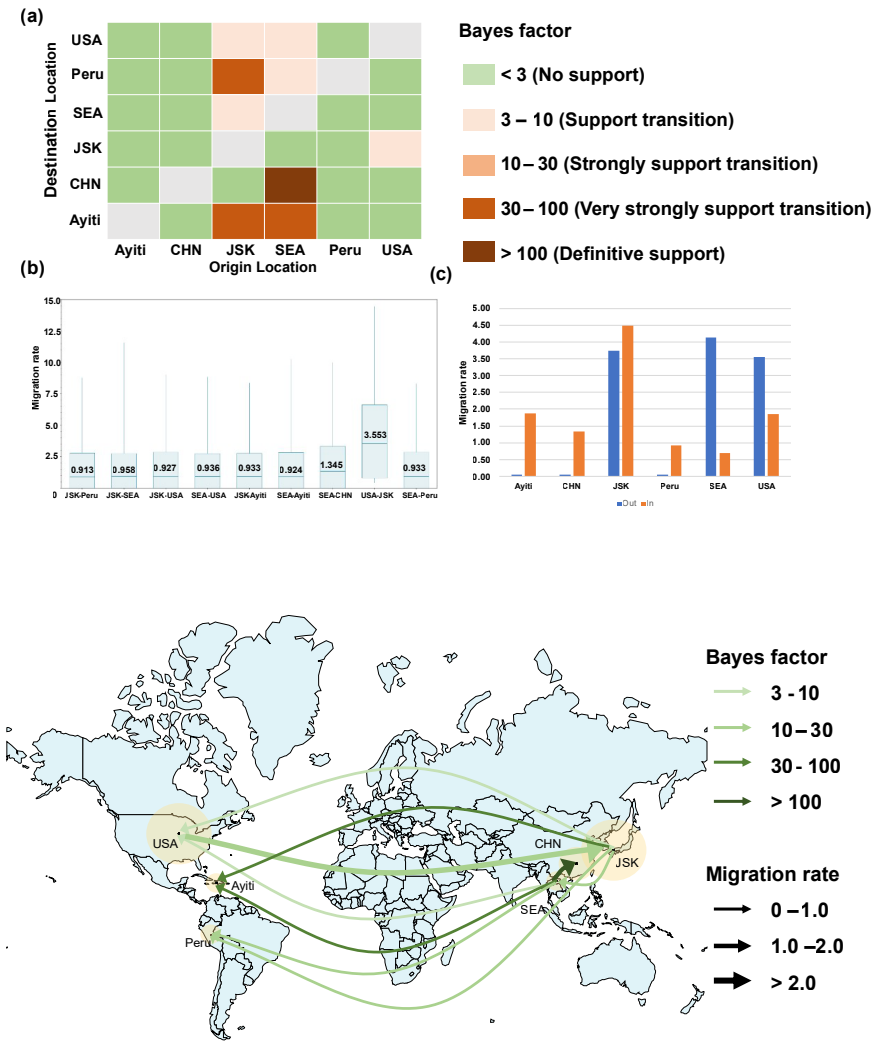
Fig. 8 Three-dimensional view of S protein

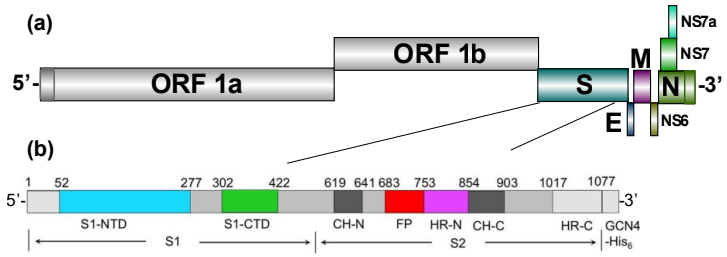
The five selected amino acids were marked in red and the locations indicated by arrows. The other mutations (position 38, 40 and 50) were not resolved in the structure.

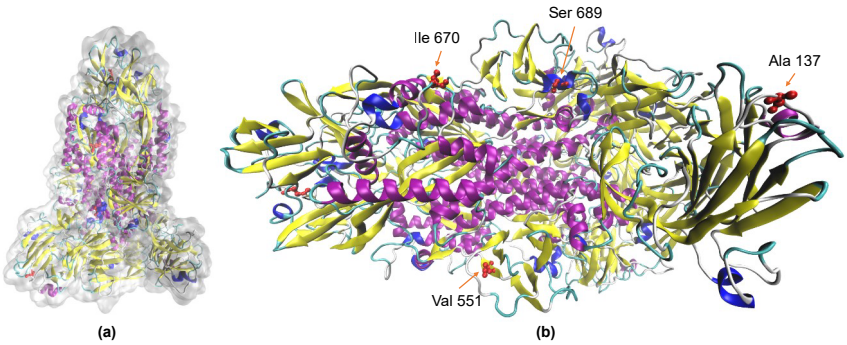












Hosted file

Table. 1.docx available at https://authorea.com/users/437035/articles/538954-phylogeography-and-evolutionary-dynamics-analysis-of-porcine-delta-coronavirus-with-host-expansion-to-humans