# Isolation and evolutionary analyses of porcine epidemic diarrhea virus in Asia

**Wan Liang** [1, 2], **Danna Zhou** [1], **Chao Geng** [1, 3], **Keli Yang** [1], **Zhengying Duan** [1], **Rui Guo** [1], **Wei Liu** [1], **Fangyan Yuan** [1], **Zewen Liu** [1], **Ting Gao** [1], **Ling Zhao** [2], **Dongwan Yoo** [4], **Yongxiang Tian** [Corresp. 1]

[1] Key Laboratory of Prevention and Control Agents for Animal Bacteriosis (Ministry of Agriculture), Institute of Animal Husbandry and Veterinary Science, Hubei Academy of Agricultural Sciences, Wuhan, China

[2] Key Laboratory of Preventive Veterinary Medicine of Hubei Province, College of Veterinary Medicine, Huazhong Agricultural University, Wuhan, China

[3] College of Animal Sciences, Yangtze University, Jinzhou, China

[4] Department of Pathobiology, College of Veterinary Medicine, University of Illinois at Urbana-Champaign, Champaign, United States

Corresponding Author: Yongxiang Tian
Email address: tyxanbit@163.com

Porcine epidemic diarrhea virus (PEDV) is a leading cause of diarrhea in pigs worldwide. Virus isolation and genetic evolutionary analysis allow investigations into the prevalence of epidemic strains and provide data for the clinical diagnosis and vaccine development. In this study, we investigated the genetic characteristics of PEDV circulation in Asia through virus isolation and comparative genomics analysis. A PEDV strain designated HB2018 was isolated from a pig in a farm experiencing a diarrhea outbreak. The complete genome sequence of HB2018 was 28,138 bp in length. Phylogenetic analysis of HB2018 and 207 PEDVs in Asia showed that most PEDV strains circulating in Asia after 2010 belong to genotype GII, particularly GII-a. The PEDV vaccine strain CV777 belonged to GI, and thus, unmatched genotypes between CV777 and GII-a variants might partially explain incomplete protection by the CV777-derived vaccine against PEDV variants in China. In addition, we found the S protein of variant strains contained numerous mutations compared to the S protein of CV777, and these mutations occurred in the N-terminal domain of the S protein. These mutations may influence the antigenicity, pathogenicity, and neutralization properties of the variant strains.

1  **Isolation and evolutionary analyses of porcine epidemic diarrhea virus in Asia**

2  Wan Liang [1,2], Danna Zhou [1], Chao Geng [1,3], Keli Yang [1], Zhengying Duan [1], Rui Guo [1], Wei Liu

3  [1], Fangyan Yuan [1], Zewen Liu [1], Ting Gao [1], Ling Zhao [2], Dongwan Yoo [4] and Yongxiang Tian

4  [1,*]

6  [1]Key Laboratory of Prevention and Control Agents for Animal Bacteriosis (Ministry of

7  Agriculture), Institute of Animal Husbandry and Veterinary, Hubei Academy of Agricultural

8  Sciences, Wuhan, China; liangwan521521@163.com (WL); zdn_66@126.com (DZ);

9  a18163131821@163.com (CG); keliy6@126.com (KY); zy001d@sina.com (ZD);

10  hlguorui@163.com (RG); liuwei85@126.com (WLiu); hbxms@126.com (FY);

11  liuzwen2004@sina.com (ZL); gt20170719@163.com (TG); tyxanbit@163.com (YT)

12  [2]Key Laboratory of Preventive Veterinary Medicine of Hubei Province, College of Veterinary

13  Medicine, Huazhong Agricultural University, Wuhan, China; liangwan521521@163.com (WL);

14  lingzhao@mail.hzau.edu.cn (LZ).

15  [3]College of Animal Sciences, Yangtze University, Jinzhou, China; a18163131821@163.com

16  (CG)

17  [4]Department of Pathobiology, College of Veterinary Medicine, University of Illinois at Urbana-

18  Champaign, Champaign, United States; dyoo@illinois.edu (DY)

20  * Corresponding author. *E-mail* address: tyxanbit@163.com (Y. Tian)

21    **ABSTRACT**

22    Porcine epidemic diarrhea virus (PEDV) is a leading cause of diarrhea in pigs worldwide. Virus

23    isolation and genetic evolutionary analysis allow investigations into the prevalence of epidemic

24    strains and provide data for the clinical diagnosis and vaccine development. In this study, we

25    investigated the genetic characteristics of PEDV circulation in Asia through virus isolation and

26    comparative genomics analysis. A PEDV strain designated HB2018 was isolated from a pig in a

27    farm experiencing a diarrhea outbreak. The complete genome sequence of HB2018 was 28,138

28    bp in length. Phylogenetic analysis of HB2018 and 207 PEDVs in Asia showed that most PEDV

29    strains circulating in Asia after 2010 belong to genotype GII, particularly GII-a. The PEDV

30    vaccine strain CV777 belonged to GI, and thus, unmatched genotypes between CV777 and GII-

31    a variants might partially explain incomplete protection by the CV777-derived vaccine against

32    PEDV variants in China. In addition, we found the S protein of variant strains contained

33    numerous mutations compared to the S protein of CV777, and these mutations occurred in the N-

34    terminal domain of the S protein. These mutations may influence the antigenicity, pathogenicity,

35    and neutralization properties of the variant strains.

36    **Keywords:** Porcine epidemic diarrhea virus; virus isolation; genetic evolutionary analysis; amino

37    acid mutations; Asia

## INTRODUCTION

Porcine epidemic diarrhea (PED) is a high contagious and devastating disease resulting in the watery diarrhea in suckling pigs with high mortality and morbidity (Zhang et al. 2019). The causative agent of PED, the porcine epidemic diarrhea virus (PEDV), is an enveloped, single-stranded, positive-sense RNA virus belonging to the genus *Alphacoronavirus* in the family *Coronaviridae* (Woo et al. 2012). PEDV possesses a 28-kb genome which encodes seven proteins including ORF1a, ORF1b, spike (S) glycoprotein, ORF3 hypothetical protein, envelop (E) protein, membrane (M) protein and nucleocapsid protein (Guo et al. 2019). Among these proteins, the S protein plays a key role in interaction between the virus and host cells. S protein consists of 1383-amino acids (Aziz et al.), and amino acid changes in S protein may lead to antigenic variations and affect the virus virulence (Gong et al. 2018; Suzuki et al. 2018). Therefore, this protein is commonly used as an important target for analyzing genetic variations and molecular epidemiology of PEDV (Hsueh et al. 2020).

PED outbreaks have been reported continuously in China since 1973. PED was well controlled since administration of a CV777-derived vaccine (Chen et al. 2019a; Wang et al. 2016a). However, recent outbreaks of PED in China since 2010 was due to the re-emergence of PEDV, and the continuous spread of the virus during the last 10 years has resulted in serious economic losses in the pig industry in Asian countries (Yang et al. 2013). In these outbreaks, inactivated vaccines and attenuated live vaccines, which were derived from CV777, were used to control the disease but neither of them provided effective protection (Sun et al. 2012; Zhou et al. 2012). Moreover, the virus has evolved since 2010 (Guo et al. 2019; Hsu et al. 2018; Sun et al. 2019), and acquisition of whole genome features of PEDV provides a convenient tool for the tracking of PEDV epidemiology (Chen et al. 2019b). In addition, virus isolation and genetic analysis allow

61   investigations on the prevalence of epidemic strains and will provide information for diagnosis

62   and vaccine developments (Li et al. 2018). In this study, we isolated a highly pathogenic PEDV

63   strain HB2018 from a pig in a farm experiencing PED outbreaks in Hubei province, China, and

64   determined its complete genome sequence. By comparing the HB2018 genome sequence with the

65   sequences of 207 PEDV isolates circulating in Asia, which were publicly available in the Genbank

66   data base, this study also aims to elucidate the evolutionary and genetic characteristics of PEDV

67   currently circulation in different regions of Asia.

68   **MATERIALS AND METHODS**

69   **Virus detection and isolation**

70       In 2018, an outbreak of diarrhea occurred in a CV777-vaccinated pig farm (numbers of sows

71   $\geq$ 100) in Hubei Province in China. Many pigs in the farm suffered from severe watery diarrhea,

72   and some of them died. Samples of intestinal tissues were collected from dead pigs and sent to the

73   Veterinary Diagnostic Laboratory of Hubei Academy of Agricultural Sciences in Wuhan, China,

74   for diagnosis. Tissues were immersed with Dulbecco's modified Eagle medium (DMEM; Gibco,

75   Grand Island, NY, USA), and were then homogenized using a QIAGEN TissueLyser II (QIAGEN,

76   Dusseldorf, Nordrhein-Westfalen, Germany). The sample homogenates were then frozen at −80

77   °C and thawed for three times. After that, the supernatants were filtered through a 0.22-μm

78   membrane and were harvested for RNA and virus isolation. Total RNAs were extracted using

79   TRIzol (Thermo, Waltham, MA, USA) and were reverse transcribed to cDNA using a Thermo

80   Scientific First Strand cDNA Synthesis kit (Thermo, Waltham, MA, USA). Viral nucleic acids

81   were detected by RT-PCR assays using the cDNA as templates and the primers specific for PEDV

82   (F: 5'-TTCGGTTCTATTCCCGTTGATG-3', R: 5'-CCCATGAAGCACTTTCTCACTATC-3'),

83  TGEV (transmissible gastroenteritis virus) (F: 5'-TTACAAACTCGCTATCGCATGG-3', R: 5'-

84  TCTTGTCACATCACCTTTACCTGC-3')    and    PoRV   (porcine    rotavirus)   (F:    5'-

85  CCCCGGTATTGAATATACCACAGT-3',    R:    5'-TTTCTGTTGGCCACCCTTTAGT-3'),

86  respectively.

87       Vero cells (Purchased from ATCC, Manassas, VA, USA) were used for virus isolation. In

88  brief, homogenate supernatants and trypsin (5 µg/ml) were inoculated into monolayers of Vero

89  cells, which were then incubated in a 37 °C incubator supplemented with 5% $CO_2$. Cells with

90  obvious cytopathic effects (CPEs) were harvested, thawed, and refrozen multiple times. The

91  harvested virus suspension was then inoculated into newly prepared Vero cells for passages, and

92  the propagation was continuously performed for 20 passages (F20). Virus RNA was extracted

93  every five passage for RT-PCR detection of the virus nucleic acids.

94  **Virus titration and serum neutralization**

95       Virus titers were measured on 96-well plates using 10-fold serial dilutions of culture

96  supernatant in triplicate per dilution to determine the quantity of viruses required to produce CPEs

97  in 50% of cells. After incubating for enough time, no more CPEs appeared, and TCID50 was

98  calculated using the Reed-Muench method (Reed & Muench 1983). The virus titer was also

99  determined by plaque assay using Vero cells and expressed as plaque-forming units (PFU) per mL.

100  The serum neutralization (SN) test was performed in 96-well plates with inactivated serum

101  collected from the guinea pigs infected with the vaccine strain CV777. The virus was diluted in

102  serum-free DMEM to make 200 TCID in a 50 µL volume, and mixed with 50 µL of 2-fold serial

103  dilution serum. The mixture was added to cells cultured in 96-well plates and incubated at 37℃

104  for 1 h. After removing the mixture and thoroughly washing three times with PBS, the cells were

105  incubated at 37 °C with 5% $CO_2$ for 2 days. Neutralization titers were calculated as the reciprocal

106  of the highest dilution of serum that inhibits CPEs.

107  **Genome sequencing and annotation**

108  Genomic RNA was extracted using the TAKARA RNA extraction kit (Takara, Kusatsu,

109  Shiga, Japan) following the manufacture instruction. The quantity and quality of the extracted

110  RNA were measured by using a Nanodrop spectrophotometer (Thermo, Waltham, MA, USA). The

111  RNA was then subjected to reverse transcription for cDNA using a cDNA synthesis kit (Thermo,

112  Waltham, MA, USA). Genome sequencing was performed with a paired-end library constructed

113  by using a NEB-Next® DNA Library Prep Master Mix Set for Illumina (NEB, Ipswich, MA, USA)

114  and subsequently sequenced on an Illumina NextSeq 500 with $2 \times 150$ paired end sequencing

115  chemistry. After filtering, the clean reads were assembled using SPAdes v3.10.1 (Bankevich et al.

116  2012) and assembled sequences were mapped to the reference genome. The prediction of the genes

117  and proteins were conducted with Prokka v1.12 and RAST Serve (http://rast.nmpdr.org) (Aziz et

118  al. 2008). The complete genome sequence as well as its annotations were deposited into NCBI

119  GenBank under the accession number MT166307.

120  **Comparative genomics and bioinformatical analysis**

121  The NCBI data was search for "porcine epidemic diarrhea virus" and a total of 207 complete

122  genome sequences were publicly available for PEDV isolates representing different parts of Asia

123  (See Table S1 in supplementary materials). All of these 207 sequences were downloaded for

124  further analysis. The average nucleotide sequence identity between the genomes of HB2018 and

125  CV777 was calculated by ANI calculator (Goris et al. 2007). Sequence alignments were performed

126  using MAFFT v7.4.02 (Katoh & Standley 2013). Nucleotide sequence similarity and the putative

127 recombination sites was assessed by SimPlot v.3.5.1 (Lole et al. 1999), with a sliding window size

128 of 500 bp, step size of 100 nucleotides, and 1,000 bootstrap replicates, using gap-stripped

129 alignments and the F84 (ML) distance model. Phylogenetic trees based on complete genome

130 sequences were generated by using MEGA X software with 1,000 bootstrapping (Kumar et al.

131 2018).The evolutionary history was inferred by using the Maximum Likelihood method and

132 Tamura-Nei model (Tamura & Nei 1993). Initial tree(s) for the heuristic search were obtained

133 automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances

134 estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the

135 topology with superior log likelihood value. The tree is drawn to scale, with branch lengths

136 measured in the number of substitutions per site. A maximum likelihood tree was also generated

137 using the BEAST 2 package (version 2.6.3) (Bouckaert et al. 2019). Gamma correction for site

138 heterogeneity and the GTR model (Gatto et al. 2007) were selected for the tree generation. Both

139 of the trees were annotated and visualized by using the iTOL v.4 online tool (Interactive Tree of

140 Life, http://itol.embl.de/) (Letunic & Bork 2019). Single nucleotide polymorphisms (SNPs)

141 between two genome sequences were determined by the MAUVE package (version 2.4.0) (Darling

142 et al. 2004), and the coding effect of these SNPs were analyzed using a previously reported local

143 Perl command (Peng et al. 2016). Protein structure was generated using SWISS-MODEL

144 (Waterhouse et al. 2018). Protein N-glycosylation sites were predicted using online software

145 (http://www.cbs.dtu.dk/services/NetNGlyc/). Threshold values of greater than 0.5 and Jury

146 agreement 9/9 were used for the high-specificity N-glycosylation sites determination (Sagesser et

147 al. 1997).

148 **RESULTS**

149 **Isolation of PEDV HB2018 and its genomic characteristics**

150    RT-PCR detection of the viral nucleic acids revealed that the intestinal samples from pigs

151    suffered and died from severe watery diarrhea were positive for PEDV but negative for TGEV and

152    PoRV (Figure S1 in supplementary materials). Through virus isolation and purification using Vero

153    cells and determination of PEDV nucleic acids using RT-PCR, a PEDV strain was finally

154    recovered and designated HB2018. The TCID50/0.1 mL value of HB2018 was $10^{5.3}$. The complete

155    genome sequence of PEDV strain HB2018 was 28,138 bp in length. This 2.8-kb genome contained

156    seven open reading frames (ORFs): ORF1a (nucleotide positions 281 to 12634), ORF1b (positions

157    12,664 to 20,625), S gene (positions 24,782 to 25,456), ORF3 (positions 25,675 to 25,667), E gene

158    (positions 25,437 to 25,667), M gene (positions 25,675 to 26,355), and N gene (positions 26,367

159    to 27,692).

160    Phylogenetic analysis based on the complete genome sequence showed that HB2018 was

161    phylogenetically distinct from the vaccine strain CV777 (Figure 1A). According to the genotyping

162    system based on a full-length genomic sequence analysis (Guo et al. 2019; Wang et al. 2016a),

163    HB2018 and CV777 belonged to two different genotype: HB2018 was assigned as a type GII strain

164    while CV777 was a GI strain (Figure 1A). The average nucleotide identity between the genomes

165    of HB2018 and CV777 (GenBank accession no. AF353511) was 96.06% (Figure S2 in

166    supplementary materials). The ORF1, ORF3, E, M, and N genes of HB2018 as well as their

167    encoding proteins were highly homologous to those of CV777 (nucleotide identity ≥ 95% for

168    genes; amino acid similarity ≥ 95% for proteins) (Figure 1B; Table 1). However, the identity of

169    the S genes and proteins between the two strains was relatively low: the homology for nucleotide

170    and amino acid sequences between HB2018 and CV777 were 93.76% and 93.44%, respectively

171    (Figures 1B & 1C; Table 1). SNP analysis determined a total of 946 SNPs in the genome sequence

172    of HB2018 when compared to the genome sequence of the reference strain CV777. Among these

173  SNPs, 925 SNPs including 262 non-synonymous substitutions and 663 synonymous substitutions

174  were located with the ORF regions, with an overall ratio of nonsynonymous to synonymous

175  substitutions (dN/dS) of 0.39 (Table 2). The dN/dS ratios in each of the ORFs encoded by the

176  HB2018 genome ranged from 0.15 to 0.61, with the S protein had the highest dN/dS ratio (Table

177  2).

178      Compared to the S protein of CV777, the S protein of HB2018 had changes, deletions, and/or

179  insertions of amino acids at multiple sites (Table S2 and Figure S3 in supplementary materials).

180  Notably, most of these mutations occurred in the N-terminal domain (NTD, 19-233aa) of the S

181  protein (Figure 1C; Figure S3 in supplementary materials). Interestingly, some of these mutations

182  were located within the neutralizing epitopes of PEDV [COE (499–638), SS2 (748–755), SS6

183  (764–771) and 2C10 (1368–1374)]. In addition, these mutations led to a structural change at some

184  parts of the HB2018 S protein compared to the CV777 S protein (Figures 1D&1E).

185  **Phylogenetic analysis of Aisan PEDV isolates**

186      To explore the phylogenic relationships of the PEDVs currently circulating in Asia, we

187  generated two maximum likelihood trees based on the whole genome sequences, either by using

188  the MEGA X software with the Tamura-Nei model (Figure 2A) or by using the BEAST 2 package

189  with the GTR model (Figure 2B). Both of the results revealed that the 208 PEDV strains in Asia,

190  representing the 207 genome sequences publicly available in NCBI and the HB2018 sequence was

191  divided into two genogroups: GI (classical) and GII (variant). Interestingly, isolates in China

192  before 2010 and the vaccine strain CV777 were included within the GI genogroup. However, most

193  of the PEDV isolates from China as well as the other Asian countries after 2010 belonged to GII

194  genogroup (Figures 2A&2B). The phylogenetic trees also showed that the two genogroups

195  consisted of several subgroups: the genogroup GI was divided into two subgroups, GI-a and GI-b,

196  while the genogroups GII was divided into three subgroups, GII-a, GII-b, and GII-c (Figures

197  2A&2B). The GI-a and GI-b subgroups included isolates from China before 2010 and several

198  Chinese isolates between 2010 and 2015 (Figures 2A&2B). Most of the GII isolates from China

199  and South Korea and all GII isolates from Japan were included within the GII-a subgroup, while

200  less proportion of the Chinese GII isolates and most of the GII-a isolates from Southeast Asia

201  (Vietnam and Thailand) were included within GII-b subgroup (Figures 2A&2B). Interestingly, the

202  GII-c subgroup only consisted of isolates from China (Figures 2A&2B).

203      By analyzing isolation years and genogroups of PEDVs, the history of PEDV and the

204  evolution in China are speculated. Between 1986 and 2008, only five PEDV strains were

205  sequenced in China, and all of them belonged to G1 (Figures 2A&2B). However, the number of

206  PEDV sequences increased significantly after 2010 (Figure 2C). While several PEDV sequences

207  belonged to genogroup GI after 2010, most sequences from China were GII strains (Figures

208  2A&2B&2C).

209  **Analysis on the S protein**

210      Compared to the S proteins of the Chinese GI-a strains, amino acid changes, deletions, and/or

211  insertions were observed at multiple sites within the S proteins of the Chinese GI-b strains (Table

212  S3 and Txt S1 in supplementary materials). Compared to the S proteins of the Chinese GI strains,

213  the S proteins of the Chinese GII strains commonly had amino acid changes, deletions, and/or

214  insertions at several sites (Table S4 and Txt S1 in supplementary materials). Most these mutations

215  occurred in S-NTD (19-233aa) of the S protein (Figure 3A; Txt S1 in supplementary materials).

216      Compared to S proteins of the Chinese GII-a strains, S proteins of most isolates from Japan,

217  South Korea, and Vietnam did not contain characteristic amino acid mutations, with the exception

218  of S proteins of two Japanese strains (NIG-2/JPN/2014 and KMM-1/JPN/2014) which had a

219  continuous deletion of 194 amino acids at sites 23-216 (Figure 3B; Txt S2 in supplementary

220  materials). In addition, four isolates from Japan (GenBank accession num. LC063844~LC063847)

221  and three isolates from South Korea (KNU-141112-S DEL5, KNU-141112-S DEL5ORF3, KNU-

222  1406-1) had a continuous deletion of 5 amino acids (GENQG) at sites 56-60 in their S proteins

223  compared to the S proteins of the Chinese GII-a strains (Figure 3B; Txt S2 in supplementary

224  materials). In addition to the GII-a strains, several isolates from China, South Korea, Thailand, and

225  Vietnam were GII-b strains (Figure 1). Compared to S proteins of most of the Chinese GII-b

226  strains, S proteins of the GII-b isolates from Thailand and Vietnam had amino acid changes at sites

227  130-131 (SI→DN), 182 (Y→H), 287 (I→M), 324 (N→D), 327 (S→A), 358 (A→T), 367 (I→T),

228  433 (D→G), 558-559 (TN→PT), 1287 (E→K), and 1317 (L→F) (Txt S3 in supplementary

229  materials).

230      The N-glycosylation sites in the S proteins of the Asian strains studied were investigated

231  herein. Bioinformatical analysis revealed that most of the 208 Asian isolates contained 7~9 high-

232  specificity N-glycosylation sites in their S proteins. When combining all high-specificity N-

233  glycosylation sites determined and deleting the duplicates, eleven sites appeared in most isolates,

234  including    57NSTW60,    112NATA115,    127NKTL130,    212NVTS215,    320NDTS323,

235  347NSSD350, 510NITV513, 552NVTN555, 777NISI780, 1245NKTL1248, and 1257NRTG1260

236  (Table 3). Among these sites, 212NVTS215, 777NISI780, and 1245NKTL1248 were conserved

237  in almost all of the 208 Asian strains. However, the N-glycosylation sites at sites 57-60 of some

238  strains were "NSSS" rather than "NSTW". This is because the S proteins of these strains had the

239  59QGVN62 deletion compared to the S proteins of the other strains. Due to the amino acids

240  changes, the N-glycosylation sites at sites 347-350 in some strains also changed from "NSSD" to

241  "NSSN" or "NSTN". The 510NITV513 and 552NVTN555 N-glycosylation sites were missing in

242     S proteins of the isolates from South Korea and Vietnam. Due to the large deletion of amino acids

243     at sites 23-216, the 57NSTW60, 112NATA115, 127NKTL130, and 212NVTS215 N-

244     glycosylation sites were missing in S proteins of two Japanese strains NIG-2/JPN/2014 and KMM-

245     1/JPN/2014, and seven N-glycosylation sites, including 130NDTS133, 157NSSN160,

246     494NVTS497, 549NCTE552, 587NISI590, 1055NKTL1058, and 1067NRTG1070 were retained

247     in these two strains.

248     **Analysis of the ORF3-E-M-N proteins**

249     Unlike the S protein, ORF3 is a conserved protein among the PEDV isolates (Wang et al.

250     2016b). However, several PEDVs were found to have characteristic amino acid mutations in the

251     ORF3 (Figure 4). Compared to ORF3 proteins of many G1 strains and all GII strains, nine GI-a

252     PEDVs (ZJUG12013, 85-7, 85-7-mutant 1, 85-7-mutant 2, 85-7-mutant 3, 85-7-mutant 4, 85-7-

253     mutant 5, 85-7-A40, and 85-7-C40) had a continuous deletion of 70 amino acids at their N-terminal

254     (positions 1-70) of ORF3; two GI-a strains (85-7-mutant 2 and 85-7-mutant 4) had a continuous

255     deletion of 47 amino acids at their C-terminal (positions 178-224) of ORF3; while nine GI-b strains

256     (JS2008, AH-M, SD-M, SQ2014, SC1402, HLJBY, PEDV-SX, JSLS-12015 and JS-22015) had a

257     continuous deletion of 133 amino acids at their C-terminal (positions 92-224) of ORF3 (Figure 4).

258     Compared to ORF3 proteins of many GI and GII strains, ORF3 protein of a GII-a strain

259     (CHSXYL2016) had a continuous deletion of 14 amino acids at positions 211-224; ORF3 protein

260     of another GII-a strain (NW17) had a continuous deletion of 6 amino acids (DLYLAI) at positions

261     168-173; while ORF3 proteins of six GII-b strains (YN15, YN30, YN60, YN90, YN144, YN200)

262     had a continuous deletion of 79 amino acids at their C-terminal (positions 146-224) (Figure 4).

263     Compared to ORF3 proteins of the GII-a isolates, more than half of the GII-b isolates had amino

264     acid changes at positions 25 (L→S), 70 (I→V), 80 (V→F), 107 (C→F), 168 (D→N), and 182

265 (Q→H) (Figure 4). Interestingly, these amino acid changes were also found in the ORF3 proteins

266 of many GII-c isolates after 2016.

267      Sequence comparisons revealed that there were no common INDELs or mutations in E

268 proteins of one subgroup of GII strains compared to E proteins of other subgroups of GII strains

269 (Figure 5A). M proteins of most GII-a strains had a glutamine (Q) at site 13; however, all GII-b

270 strains isolated between 2011 and 2012 had a glutamic acid (E) at the same position in their M

271 proteins, and this amino acid change (Q→E) occurred frequently in M proteins of GII-b since 2013

272 (Figure 5B). A similar phenomenon was also observed in the M proteins of the GII-c strains, as

273 most of the GII-c strains isolated before 2016 had a glutamine (Q) at position 13 in their M

274 proteins, but a Q→E change at position 13 was seen in the M proteins of more frequently in strains

275 isolated after 2016. In addition, amino acid changes at positions 192 (G→S) and 214 (S→A)

276 appeared simultaneously in M proteins of some GII-b and GII-c strains. Similarly, in N proteins,

277 amino acid changes at positions 216 (M→V) and 241 (R→K) appeared simultaneously in many

278 GII strains (Figure 5C).

279 **Discussion**

280      As an infectious virus attracted great intention, the PEDV strains were frequently reported and

281 isolated in Asia. The virus isolation and genetic analysis will provide important information for

282 PEDV research and vaccine developments. In this study, we isolated a GII-a strain HB2018 and

283 determined its genomic characteristics (Figure 1A). Comparative genomic analysis revealed that

284 the ORF1, ORF3, E, M, and N genes of HB2018 as well as their encoding proteins were highly

285 homologous to those of CV777 (Figure 1B; Table 1). However, a number of SNPs were

determined within these ORFs, with the S proteins showed the highest dN/dS ratio (Table 2). Since

287 dN/dS ratio is commonly used as a measure of purifying versus diversifying selection (Rocha et

288 al. 2006), the highest dN/dS ratio suggests S protein is under diversifying selection, and this

289 diversifying selection might be associated with its frequent interaction with host cells.

290     Sequence alignments determined many mutations in the genome sequence of HB2018

291 compared to that of the reference strain CV777. These mutations, especially in the S protein, might

292 be the pathogenic determinants for it, because some deletions and insertions in the S protein may

293 change the antigenicity, pathogenicity and neutralization properties (Chen et al. 2019a; Sagesser

294 et al. 1997; Zhang et al. 2015). The presence of these mutations in the NTD of S protein in HB2018

295 might have an effect on the viral pathogenicity since the S-NTD domain is proposed to be the

296 region relevant to the virulence of PEDV (Hou et al. 2017; Su et al. 2018; Su et al. 2019; Suzuki

297 et al. 2018). In addition, the structural changes led by these mutations in S protein of HB2018

298 might influence the immunogenicity. The distinct phylogenetic relationship between our isolation

299 HB2018 and CV777 might partly explain why vaccination of pigs with CV777 did not provide

300 effective protection against the infection of HB2018 in the vaccinated pig farm (Figures 2A&2B).

301 The analysis based on isolation years and genogroups of PEDVs in Asia might also revealed the

302 vaccine CV777 did not match with the pandemic PEDV isolations. Before 2010, all the strains in

303 China belonged to GI (Figures 2A&2B). During this period, PEDV was well controlled in China

304 due to the use of CV777 which was the GI-based vaccine (Chen et al. 2019a; Yang et al. 2013).

305 The phylogenetic analysis of Asian PEDV isolates showed that most of the PEDV isolates from

306  Asia after 2010 belonged to GII genogroup, while the vaccine CV777 were included within GI

307  genogroup. These findings agree with the results of the other studies (Guo et al. 2019; Wang et al.

308  2016a). The unmatched genotypes between CV777 and PEDV epidemic strains in Asia after 2010

309  could explain why vaccination with CV777 could not stop the outbreak of PED in many Asian

310  countries after 2010 and provide effective protection against the current epidemic strains (Chen et

311  al. 2019a; Puranaveja et al. 2009; Sun et al. 2012; Zhou et al. 2012).

312      With the most reported numbers of PEDV strains, China has more genogroups than other

313  countries. The GII-c subgroup only consisted of isolates from China, these findings are also in

314  agreement with previous studies (Guo et al. 2019; Wang et al. 2016a), suggesting that the

315  genotypes of PEDV strains circulating in China might be more heterogeneous than those of the

316  isolates in other Asian countries. These findings may also explain why PEDV vaccines developed

317  in China contain more than one strains that generally include CV777 and at least one more local

318  GII isolate (http://vdts.ivdc.org.cn:8081/cx/#). The new emerged PEDV in 2010 might accelerate

319  numerous isolations and sequencing of PEDVs (Li et al. 2012; Yang et al. 2013). In this article, it

320  was found that the number of PEDV sequences increased significantly after 2010 and most

321  sequences were GII strains. These results are in good agreement with the findings of the PEDV

322  epidemiological investigations in China (Chen et al. 2019a; Sun et al. 2018). It has been reported

323  that PEDV GII isolates were more virulent than GI isolates (Vlasova et al. 2014). This might in

324  part explain why the traditional vaccines had no to little effect on the control and spread of PEDV

325  in China after 2010. It is noteworthy that PEDV GII strains are also responsible for the recent

326    outbreaks of PED in North America and Europe (Choudhury et al. 2016). These findings suggest

327    the circulation of PEDV GII strains also pose a problem to the global pig industry.

328        S protein is the most variable protein of PEDV, the amino acid changes in this protein may

329    lead to virus variation and affect the virus virulence (Gong et al. 2018; Suzuki et al. 2018). The

330    mutations between were found between GI-a strains and GI-b strains, it is still uncertain whether

331    these mutations between them has a biological significance. While, the mutations occurred in S-

332    NTD of the S protein between GI strains and GII strains might in part explain why do the PEDV

333    GII isolates be more pathogenic than the GI isolates (Vlasova et al. 2014), as S-NTD is proposed

334    to be the region relevant to the virulence of PEDV (Hou et al. 2017; Su et al. 2018; Su et al. 2019;

335    Suzuki et al. 2018). It is worthy note that PEDVs with insertions of amino acids at 167-168 and

336    deletions of amino acids at 55-58 and 144 in their S proteins are called S-INDEL strains (Wang et

337    al. 2014). A previous study has found infection of the S-INDEL strains could induce pro-

338    inflammatory cytokines through the non-canonical NF-κB signaling pathway by activating RIG-

339    I; however, infection of the non-S-INDEL strains suppresses the induction of pro-inflammatory

340    cytokines and type-I interferon production by down-regulation of TLRs and downstream signaling

341    molecules (Temeeyasen et al. 2018). Whether the continuous deletion of 194 amino acids occurred

342    in Japanese strains will affect the virulence of these strains are unknown and warrant further

343    exploration. A previous study however has found that a Japanese strain Tottori2, which had the

344    same deletion, had non-lethal effects in piglets (Masuda et al. 2015). The mutations also were

345    found in C domain of S protein, since the NTD and C-domain both can bind to the host cell receptor

346    and function as the receptor-binding domain, the amino acid changes in their sequences may have

347    important role for the virus (Li 2012). It has been reported the N-linked glycosylation sites on the

348    S protein of some coronaviruses such as SARS-CoV play a critical role in the viral entry (Han et

349    al. 2007). The phylogenetic and N-linked glycosylation sites analysis of S protein may offer

350    reasons for further studies. There was no too many mutations were found in the ORF3-E-M-N

351    proteins, it might be because some of them, such as E protein, do not bear too much immune

352    selective pressure since it has no effect on the host cell growth or cell cycle (Xu et al. 2013).

353    **CONCLUSIONS**

354    In conclusion, through virus isolation and complete genome sequencing, we obtained PEDV

355    HB2018 strain. Using this virus, we investigated the genetic and phylogenetic characteristics of

356    PEDV isolates in China as well as in Asia in this study. Phylogenetic analysis revealed

357    heterogeneous genotypes of PEDVs circulate in Asia, but GII particularly GII-a genotype

358    represents the main epidemic genotype in the continent. Our study also revealed that most of the

359    PEDVs currently prevalent in Asian countries displayed a different genotype as well as a distant

360    relationship from the conventional vaccine strain CV777. This finding might explain why CV777-

361    derived vaccine provided poor protection against PEDV epidemics (variant strains) since 2010. In

362    addition, we also identified many mutations in the S, ORF3, E, M, N proteins of the variant strains

363    (GII) compared to those of the classical strains (Temeeyasen et al.). The presence of these

364    mutations, particularly those determined in the S proteins, may affect the antigenicity,

365    pathogenicity, and neutralization properties of the variant strains.

366    **ADDITIONAL INFORMATION AND DECLARATIONS**

367 **Funding**

368     This work was supported by the Key Laboratory of Prevention and Control Agents for Animal

369 Bacteriosis (Ministry of Agriculture) (grant number: KLPCAAB-YTP-1801), the open funds of

370 the Key Laboratory of Preventive Veterinary Medicine of Hubei Province, and China Postdoctoral

371 Science foundation (grant number: 2019M652609). There was no additional external funding

372 received for this study.

373 **Competing Interests**

374     The authors declare there are no competing interests.

375 **Acknowledgments**

376     The authors sincerely appreciate Dr. Zhong Peng at College of Veterinary Medicine,

377 Huazhong Agricultural University, Wuhan, China for the language reversion and suggestions on

378 the analysis.

379 **Author Contributions**

380     Conceptualization, W.L. and Y.T.; methodology, W.L., D.Z., C.G., K.Y., Z.D., R.G., W.Liu.,

381 T.G., Z.L., L.Z., Y.T.; writing—original draft preparation, W.L., D.Y.; writing—review and

382 editing, W.L., D.Z., Y.T., F.Y., D.Y.; supervision, Y.T.; project administration, W.L., Y.T.;

383 funding acquisition, W.L., Y.T.. All authors have read and agreed to the published version of the

384 manuscript.

385 **Data Availability**

386     The following information was supplied regarding data availability: NCBI GenBank under

387 the accession number of MT166307.

388 **Supplemental Materials**

389 **Figure S1:** RT-PCR detection of the viral nucleic acids from the intestinal samples of the pigs

390 suffered and/or died from severe watery diarrhea.

391 **Figure S2:** Sequence alignments on the whole genomes of HB2018 and CV777.

392 **Figure S3:** Sequence alignments on the S proteins of HB2018 and CV777.

393 **Table S1:** List of the 207 Asian PEDV isolates.

394 **Txt S1:** Alignment of the S protein of GI-a and GI-b strains from China.

395 **Txt S2:** Alignment of the S protein of Asian GII-a strains.

396 **Txt S3:** Alignment of the S protein of Asian GII-b strains.

397 **REFERENCES**

398 Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F,
399     Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD,
400     Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, and Zagnitko O. 2008. The RAST Server: rapid
401     annotations using subsystems technology. *BMC Genomics* 9:75. 10.1186/1471-2164-9-75
402 Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski
403     AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, and Pevzner PA. 2012. SPAdes: a new
404     genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19:455-477.
405     10.1089/cmb.2012.0021
406 Bouckaert R, Vaughan TG, Barido-Sottani J, Duchêne S, Fourment M, Gavryushkina A, Heled J, Jones G, Kühnert D,
407     De Maio N, Matschiner M, Mendes FK, Müller NF, Ogilvie HA, du Plessis L, Popinga A, Rambaut A,
408     Rasmussen D, Siveroni I, Suchard MA, Wu CH, Xie D, Zhang C, Stadler T, and Drummond AJ. 2019. BEAST
409     2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS Comput Biol* 15:e1006650.
410     10.1371/journal.pcbi.1006650
411 Chen P, Wang K, Hou Y, Li H, Li X, Yu L, Jiang Y, Gao F, Tong W, Yu H, Yang Z, Tong G, and Zhou Y. 2019a. Genetic
412     evolution analysis and pathogenicity assessment of porcine epidemic diarrhea virus strains circulating in part
413     of China during 2011-2017. *Infect Genet Evol* 69:153-165. 10.1016/j.meegid.2019.01.022

414   Chen X, Zhang XX, Li C, Wang H, Wang H, Meng XZ, Ma J, Ni HB, Zhang X, Qi Y, and Sun D. 2019b. Epidemiology
415         of porcine epidemic diarrhea virus among Chinese pig populations: A meta-analysis. *Microb Pathog* 129:43-49.
416         10.1016/j.micpath.2019.01.017
417   Choudhury B, Dastjerdi A, Doyle N, Frossard JP, and Steinbach F. 2016. From the field to the lab - An European view
418         on the global spread of PEDV. *Virus Res* 226:40-49. 10.1016/j.virusres.2016.09.003
419   Darling AC, Mau B, Blattner FR, and Perna NT. 2004. Mauve: multiple alignment of conserved genomic sequence with
420         rearrangements. *Genome Res* 14:1394-1403. 10.1101/gr.2289704
421   Gatto L, Catanzaro D, and Milinkovitch MC. 2007. Assessing the applicability of the GTR nucleotide substitution model
422         through simulations. *Evol Bioinform Online* 2:145-155.
423   Gong L, Lin Y, Qin J, Li Q, Xue C, and Cao Y. 2018. Neutralizing antibodies against porcine epidemic diarrhea virus
424         block virus attachment and internalization. *Virol J* 15:133. 10.1186/s12985-018-1042-3
425   Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, and Tiedje JM. 2007. DNA-DNA hybridization
426         values and their relationship to whole-genome sequence similarities. *Int J Syst Evol Microbiol* 57:81-91.
427         10.1099/ijs.0.64483-0
428   Guo J, Fang L, Ye X, Chen J, Xu S, Zhu X, Miao Y, Wang D, and Xiao S. 2019. Evolutionary and genotypic analyses of
429         global porcine epidemic diarrhea virus strains. *Transbound Emerg Dis* 66:111-118. 10.1111/tbed.12991
430   Han DP, Lohani M, and Cho MW. 2007. Specific asparagine-linked glycosylation sites are critical for DC-SIGN- and L-
431         SIGN-mediated severe acute respiratory syndrome coronavirus entry. *J Virol* 81:12029-12039.
432         10.1128/jvi.00315-07
433   Hou Y, Lin CM, Yokoyama M, Yount BL, Marthaler D, Douglas AL, Ghimire S, Qin Y, Baric RS, Saif LJ, and Wang Q.
434         2017. Deletion of a 197-Amino-Acid Region in the N-Terminal Domain of Spike Protein Attenuates Porcine
435         Epidemic Diarrhea Virus in Piglets. *J Virol* 91. 10.1128/jvi.00227-17
436   Hsu TH, Liu HP, Chin CY, Wang C, Zhu WZ, Wu BL, and Chang YC. 2018. Detection, sequence analysis, and antibody
437         prevalence of porcine deltacoronavirus in Taiwan. *Arch Virol* 163:3113-3117. 10.1007/s00705-018-3964-x
438   Hsueh FC, Lin CN, Chiou HY, Chia MY, Chiou MT, Haga T, Kao CF, Chang YC, Chang CY, Jeng CR, and Chang HW.
439         2020. Updated phylogenetic analysis of the spike gene and identification of a novel recombinant porcine
440         epidemic diarrhoea virus strain in Taiwan. *Transbound Emerg Dis* 67:417-430. 10.1111/tbed.13365
441   Katoh K, and Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in
442         performance and usability. *Mol Biol Evol* 30:772-780. 10.1093/molbev/mst010
443   Kumar S, Stecher G, Li M, Knyaz C, and Tamura K. 2018. MEGA X: Molecular Evolutionary Genetics Analysis across
444         Computing Platforms. *Mol Biol Evol* 35:1547-1549. 10.1093/molbev/msy096
445   Letunic I, and Bork P. 2019. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res*
446         47:W256-w259. 10.1093/nar/gkz239
447   Li D, Feng H, Liu Y, Chen Y, Wei Q, Wang J, Liu D, Huang H, Su Y, Wang D, Cui Y, and Zhang G. 2018. Molecular
448         evolution of porcine epidemic diarrhea virus and porcine deltacoronavirus strains in Central China. *Res Vet*
449         *Sci* 120:63-69. 10.1016/j.rvsc.2018.06.001
450   Li F. 2012. Evidence for a common evolutionary origin of coronavirus spike protein receptor-binding subunits. *J Virol*
451         86:2856-2858. 10.1128/jvi.06882-11

452    Li W, Li H, Liu Y, Pan Y, Deng F, Song Y, Tang X, and He Q. 2012. New variants of porcine epidemic diarrhea virus,
453        China, 2011. *Emerg Infect Dis* 18:1350-1353. 10.3201/eid1808.120002

454    Lole KS, Bollinger RC, Paranjape RS, Gadkari D, Kulkarni SS, Novak NG, Ingersoll R, Sheppard HW, and Ray SC. 1999.
455        Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India,
456        with evidence of intersubtype recombination. *Journal of virology* 73:152-160.

457    Masuda T, Murakami S, Takahashi O, Miyazaki A, Ohashi S, Yamasato H, and Suzuki T. 2015. New porcine epidemic
458        diarrhoea virus variant with a large deletion in the spike gene identified in domestic pigs. *Arch Virol* 160:2565-
459        2568. 10.1007/s00705-015-2522-z

460    Peng Z, Liang W, Liu W, Wu B, Tang B, Tan C, Zhou R, and Chen H. 2016. Genomic characterization of Pasteurella
461        multocida HB01, a serotype A bovine isolate from China. *Gene* 581:85-93. 10.1016/j.gene.2016.01.041

462    Puranaveja S, Poolperm P, Lertwatcharasarakul P, Kesdaengsakonwut S, Boonsoongnern A, Urairong K, Kitikoon P,
463        Choojai P, Kedkovid R, Teankum K, and Thanawongnuwech R. 2009. Chinese-like strain of porcine epidemic
464        diarrhea virus, Thailand. *Emerg Infect Dis* 15:1112-1115. 10.3201/eid1507.081256

465    Reed LJ, and Muench H. 1983. A simple method of estimating fifty per cent endpoints 12. *Am J Epidemiol* 27:493-497.

466    Rocha EP, Smith JM, Hurst LD, Holden MT, Cooper JE, Smith NH, and Feil EJ. 2006. Comparisons of dN/dS are time
467        dependent for closely related bacterial genomes. *J Theor Biol* 239:226-235. 10.1016/j.jtbi.2005.08.037

468    Sagesser R, Martinez E, Tsagris M, and Tabler M. 1997. Detection and isolation of RNA-binding proteins by RNA-
469        ligand screening of a cDNA expression library. *Nucleic Acids Res* 25:3816-3822. 10.1093/nar/25.19.3816

470    Su Y, Hou Y, Prarat M, Zhang Y, and Wang Q. 2018. New variants of porcine epidemic diarrhea virus with large
471        deletions in the spike protein, identified in the United States, 2016-2017. *Arch Virol* 163:2485-2489.
472        10.1007/s00705-018-3874-y

473    Su Y, Hou Y, and Wang Q. 2019. The enhanced replication of an S-intact PEDV during coinfection with an S1 NTD-del
474        PEDV in piglets. *Vet Microbiol* 228:202-212. 10.1016/j.vetmic.2018.11.025

475    Sun J, Li Q, Shao C, Ma Y, He H, Jiang S, Zhou Y, Wu Y, Ba S, Shi L, Fang W, Wang X, and Song H. 2018. Isolation and
476        characterization of Chinese porcine epidemic diarrhea virus with novel mutations and deletions in the S gene.
477        *Vet Microbiol* 221:81-89. 10.1016/j.vetmic.2018.05.021

478    Sun RQ, Cai RJ, Chen YQ, Liang PS, Chen DK, and Song CX. 2012. Outbreak of porcine epidemic diarrhea in suckling
479        piglets, China. *Emerg Infect Dis* 18:161-163. 10.3201/eid1801.111259

480    Sun Y, Chen Y, Han X, Yu Z, Wei Y, and Zhang G. 2019. Porcine epidemic diarrhea virus in Asia: An alarming threat
481        to the global pig industry. *Infect Genet Evol* 70:24-26. 10.1016/j.meegid.2019.02.013

482    Suzuki T, Terada Y, Enjuanes L, Ohashi S, and Kamitani W. 2018. S1 Subunit of Spike Protein from a Current Highly
483        Virulent Porcine Epidemic Diarrhea Virus Is an Important Determinant of Virulence in Piglets. *Viruses* 10.
484        10.3390/v10090467

485    Tamura K, and Nei M. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial
486        DNA in humans and chimpanzees. *Mol Biol Evol* 10:512-526. 10.1093/oxfordjournals.molbev.a040023

487    Temeeyasen G, Sinha A, Gimenez-Lirola LG, Zhang JQ, and Pineyro PE. 2018. Differential gene modulation of pattern-
488        recognition receptor TLR and RIG-I-like and downstream mediators on intestinal mucosa of pigs infected
489        with PEDV non S-INDEL and PEDV S-INDEL strains. *Virology* 517:188-198. 10.1016/j.virol.2017.11.024

490  Vlasova AN, Marthaler D, Wang Q, Culhane MR, Rossow KD, Rovira A, Collins J, and Saif LJ. 2014. Distinct
491      characteristics and complex evolution of PEDV strains, North America, May 2013-February 2014. *Emerg Infect*
492      *Dis* 20:1620-1628. 10.3201/eid2010.140491
493  Wang D, Fang L, and Xiao S. 2016a. Porcine epidemic diarrhea in China. *Virus Res* 226:7-13.
494      10.1016/j.virusres.2016.05.026
495  Wang E, Guo D, Li C, Wei S, Wang Z, Liu Q, Zhang B, Kong F, Feng L, and Sun D. 2016b. Molecular Characterization
496      of the ORF3 and S1 Genes of Porcine Epidemic Diarrhea Virus Non S-INDEL Strains in Seven Regions of
497      China, 2015. *PLoS One* 11:e0160561. 10.1371/journal.pone.0160561
498  Wang L, Byrum B, and Zhang Y. 2014. Detection and genetic characterization of deltacoronavirus in pigs, Ohio, USA,
499      2014. *Emerg Infect Dis* 20:1227-1230. 10.3201/eid2007.140296
500  Woo PC, Lau SK, Lam CS, Lau CC, Tsang AK, Lau JH, Bai R, Teng JL, Tsang CC, Wang M, Zheng BJ, Chan KH, and
501      Yuen KY. 2012. Discovery of seven novel Mammalian and avian coronaviruses in the genus deltacoronavirus
502      supports bat coronaviruses as the gene source of alphacoronavirus and betacoronavirus and avian
503      coronaviruses as the gene source of gammacoronavirus and deltacoronavirus. *J Virol* 86:3995-4008.
504      10.1128/jvi.06540-11
505  Xu X, Zhang H, Zhang Q, Dong J, Liang Y, Huang Y, Liu HJ, and Tong D. 2013. Porcine epidemic diarrhea virus E
506      protein causes endoplasmic reticulum stress and up-regulates interleukin-8 expression. *Virol J* 10:26.
507      10.1186/1743-422x-10-26
508  Yang X, Huo JY, Chen L, Zheng FM, Chang HT, Zhao J, Wang XW, and Wang CQ. 2013. Genetic variation analysis of
509      reemerging porcine epidemic diarrhea virus prevailing in central China from 2010 to 2011. *Virus Genes* 46:337-
510      344. 10.1007/s11262-012-0867-x
511  Zhang L, Liu X, Zhang Q, Zhou P, Fang Y, Dong Z, Zhao D, Li W, Feng J, Zhang Y, and Wang Y. 2019. Biological
512      characterization and pathogenicity of a newly isolated Chinese highly virulent genotype GIIa porcine
513      epidemic diarrhea virus strain. *Arch Virol* 164:1287-1295. 10.1007/s00705-019-04167-3
514  Zhang X, Pan Y, Wang D, Tian X, Song Y, and Cao Y. 2015. Identification and pathogenicity of a variant porcine
515      epidemic diarrhea virus field strain with reduced virulence. *Virol J* 12:88. 10.1186/s12985-015-0314-4
516  Zhou YJ, Wu YL, Zhu JP, Tong W, Yu H, Jiang YF, and Tong GZ. 2012. Complete genome sequence of a virulent porcine
517      epidemic diarrhea virus strain. *J Virol* 86:13862. 10.1128/jvi.02635-12
518

# Figure 1

Phylogenetic and genetic characteristics of PEDV strain HB2018.

(**A**) Phylogenetic analysis of HB2018 and the other PEDV strains based on the whole genome sequence; (**B**) Nucleotide similarity of the complete genome sequences between PEDV strains HB2018 and CV777; (**C**) Sequence alignment of the S-NTD regions of PEDV strains HB2018 and CV777; (**D**) Modelling the 3D structure of the S protein of CV777; (**E**) Modelling the 3D structure of the S protein of HB2018.

**A.**

**B.**

**C.**

```
                    *           20          *           40          *           60          *
CV777-S  :  MRSLIYFWLLLPVLPTLSLPQDVTRCQSTTNFRRFFSKFNVQAPAVVVLGGYLPSMNS----SSWYCGTGIET  :  69
HB2018-S :  MKSLTYFWLFLPVLSTLSLPQDVTRCSANTNFRRFFSKFNVQAPAVVVLGGYLPIGENQGVNSTWYCAGQHPT  :  73

                80          *           100         *           120         *           140
CV777-S  :  ASGVHGIFLSYIDSGQGFEIGISQEPFDPSGYQLYLHKATNGNTNAIARLRICQFPDNKTLGPTVN-DVTTGR  :  141
HB2018-S :  ASGVHGIFLSHIRGGHGFEIGISQEPFDPSGYQLYLHKATNGNTNATARLRICQFPGIKTLGPTADNDVTTGR  :  146

                    *           160         *           180         *           200         *           22
CV777-S  :  NCLFNKAIPAYMRDGKDIVVGITWDNDRVTVFADKIYHFYLKNDWSRVATRCYNRRSCAMQYVYTPTYYMLNV  :  214
HB2018-S :  NCLFNKAIPAHMSEHS--VVGITWDNDRVTVFSDKIYHFYFKNDWSRVATKCYNSGGCAMQYVYKPTYYMLNV  :  217

                0           *           240
CV777-S  :  TSAGEDGIYYEPCTANCTGYA  :
HB2018-S :  TSAGEDGISYQPCTANCIGYA  :
```

**D.**

**E.**

CV777                                   HB2018

# Figure 2

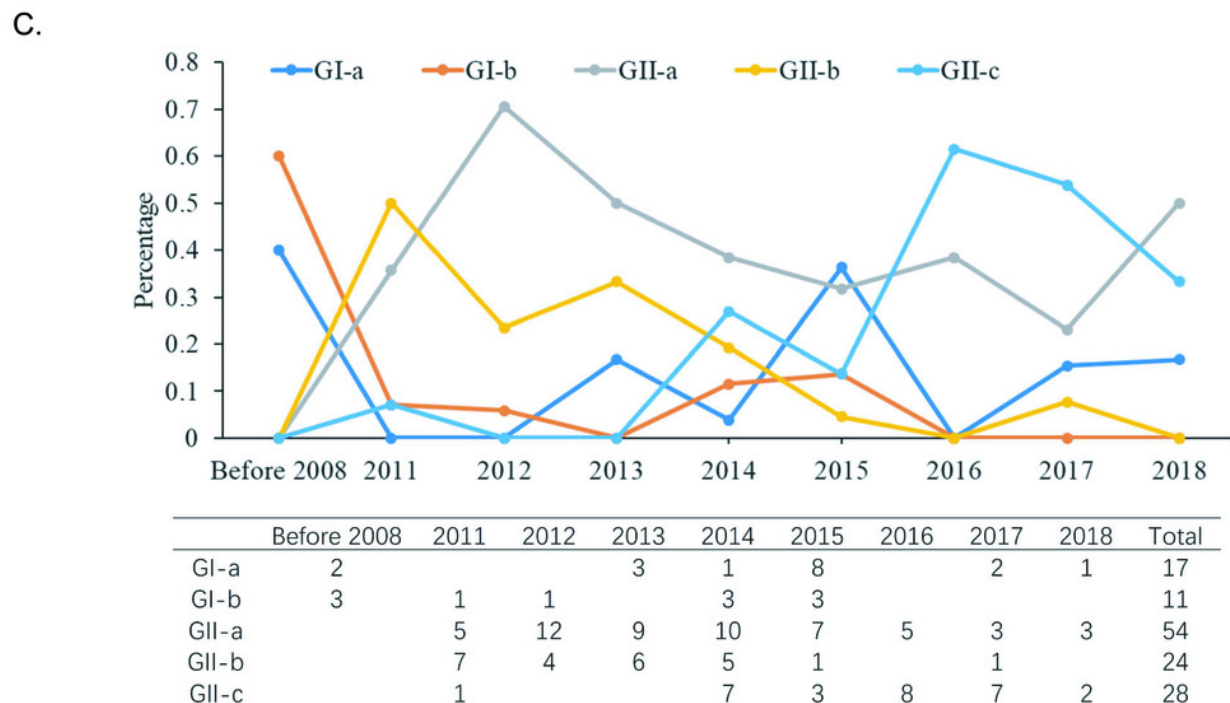
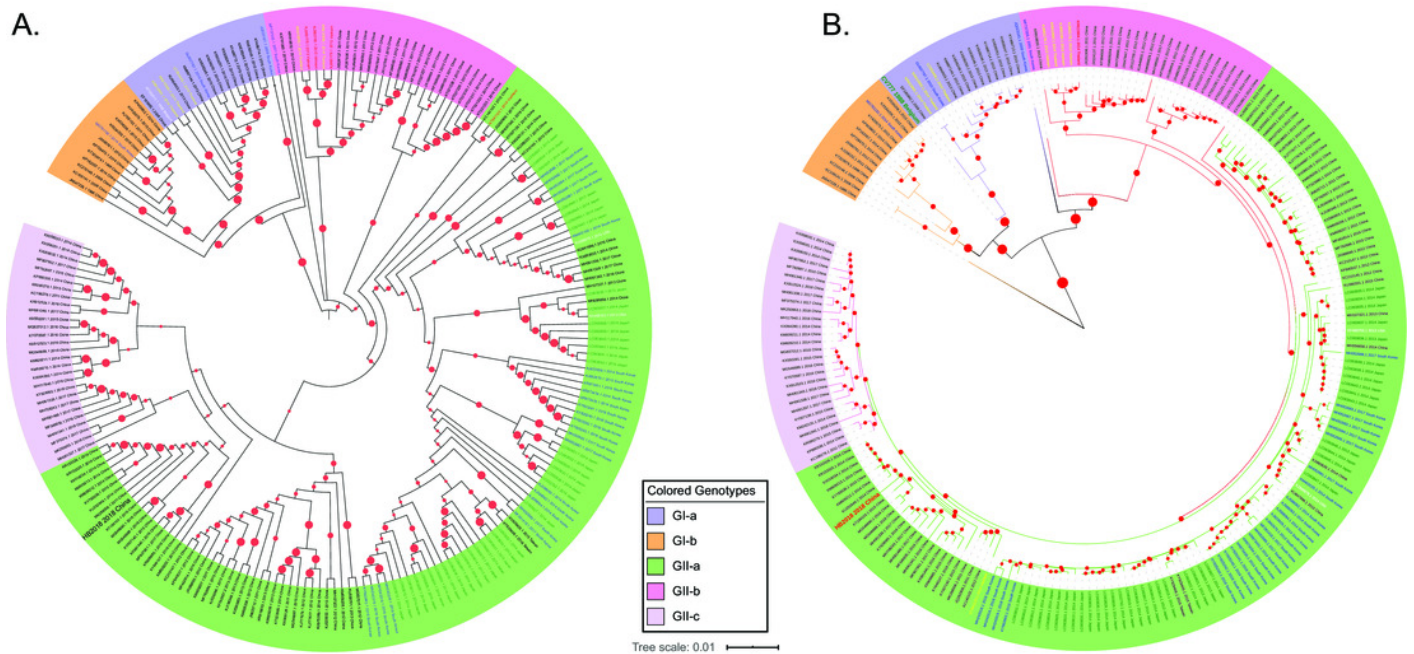
Genotyping of the 208 Asian PEDV strains and the non-Asian reference strains based on full-length genomic sequences.

**(A)** A phylogenetic tree was generated by using the MEGAX package. The evolutionary history was inferred by using the Maximum Likelihood method based on the TN93 model. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. GenBank accession numbers of strains, years, places of isolation, genogroups, and subgroups are shown. **(B)** A Maximum Likelihood tree was generated by using the BEAST 2 package. Gamma correction for site heterogeneity and the GTR model were selected for the tree generation. GenBank accession numbers of strains, years, places of isolation, genogroups, and subgroups are shown. **(C)** Line chart shows the number of PEDV sequences obtained by gene subgroup and year of sampling. Yearly percentages of samples positive for PEDV are indicated by different colored lines.

Colored Genotypes

| | |
|---|---|
| | GI-a |
| | GI-b |
| | GII-a |
| | GII-b |
| | GII-c |

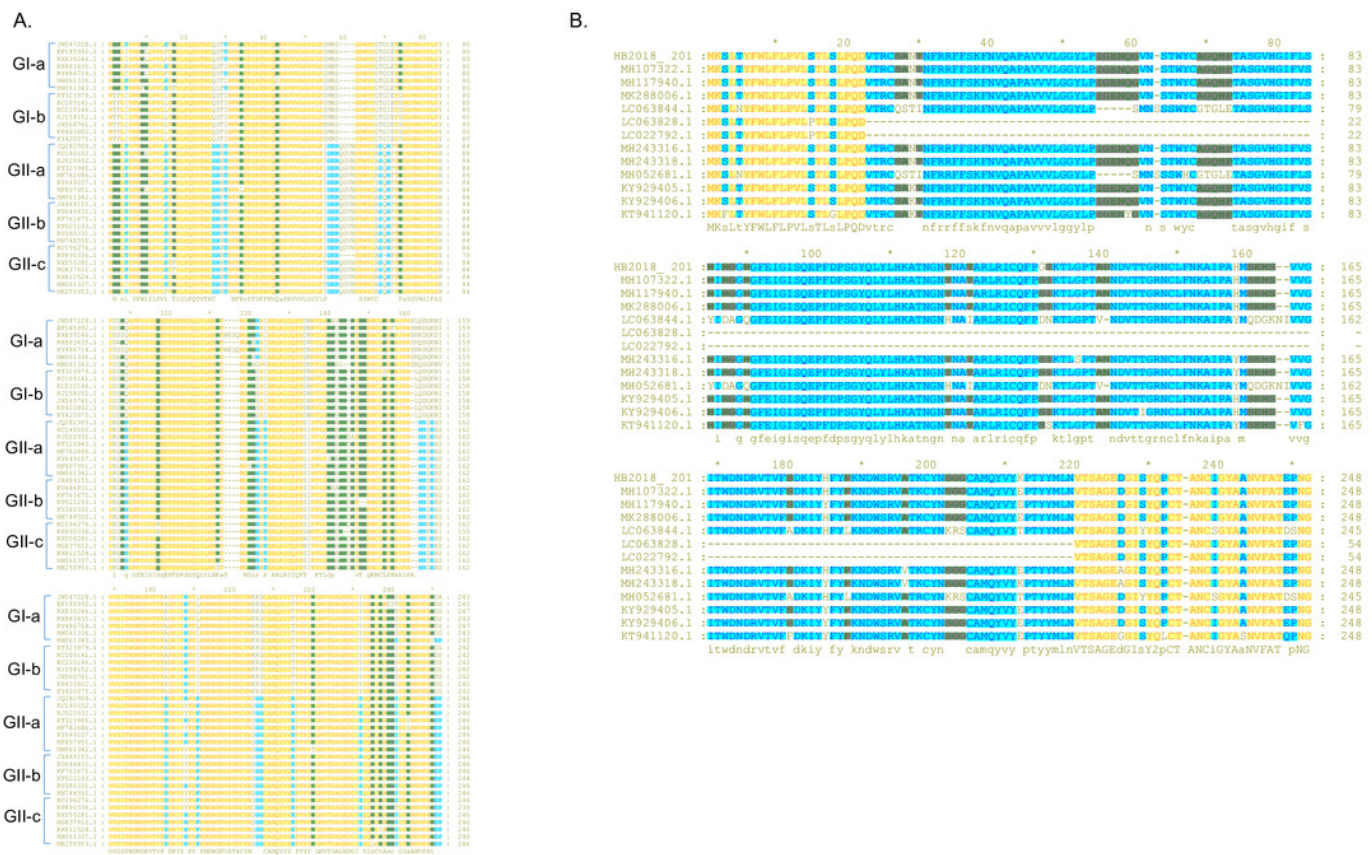
Tree scale: 0.01

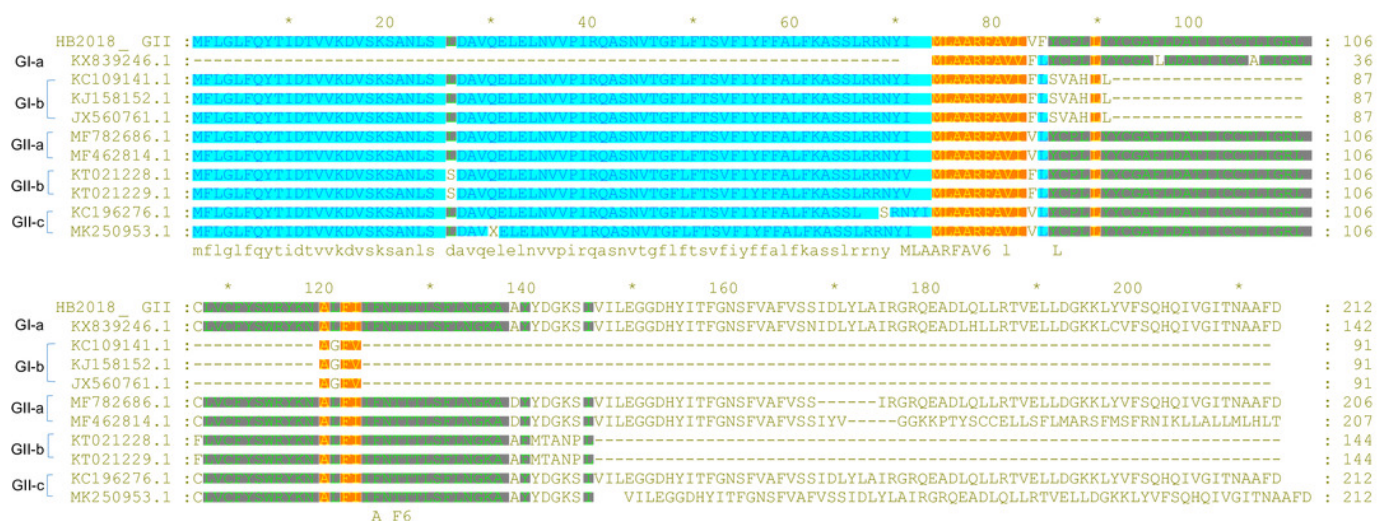
| | Before 2008 | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 | Total |
|---|---|---|---|---|---|---|---|---|---|---|
| GI-a | 2 | | | 3 | 1 | 8 | | 2 | 1 | 17 |
| GI-b | 3 | 1 | 1 | | 3 | 3 | | | | 11 |
| GII-a | | 5 | 12 | 9 | 10 | 7 | 5 | 3 | 3 | 54 |
| GII-b | | 7 | 4 | 6 | 5 | 1 | | 1 | | 24 |
| GII-c | | 1 | | | 7 | 3 | 8 | 7 | 2 | 28 |

# Figure 3

Sequence alignment of the S-NTD regions of PEDV strains.

**(A)** Sequence alignment of the S-NTD regions of the strains from China; **(B)** Sequence alignment of the S-NTD regions of the strains from Asia.

A.

B.

# Figure 4

Sequence alignment of ORF3 of PEDV strains.

# Figure 5

Sequence alignment of E, M and N protein of PEDV strains.

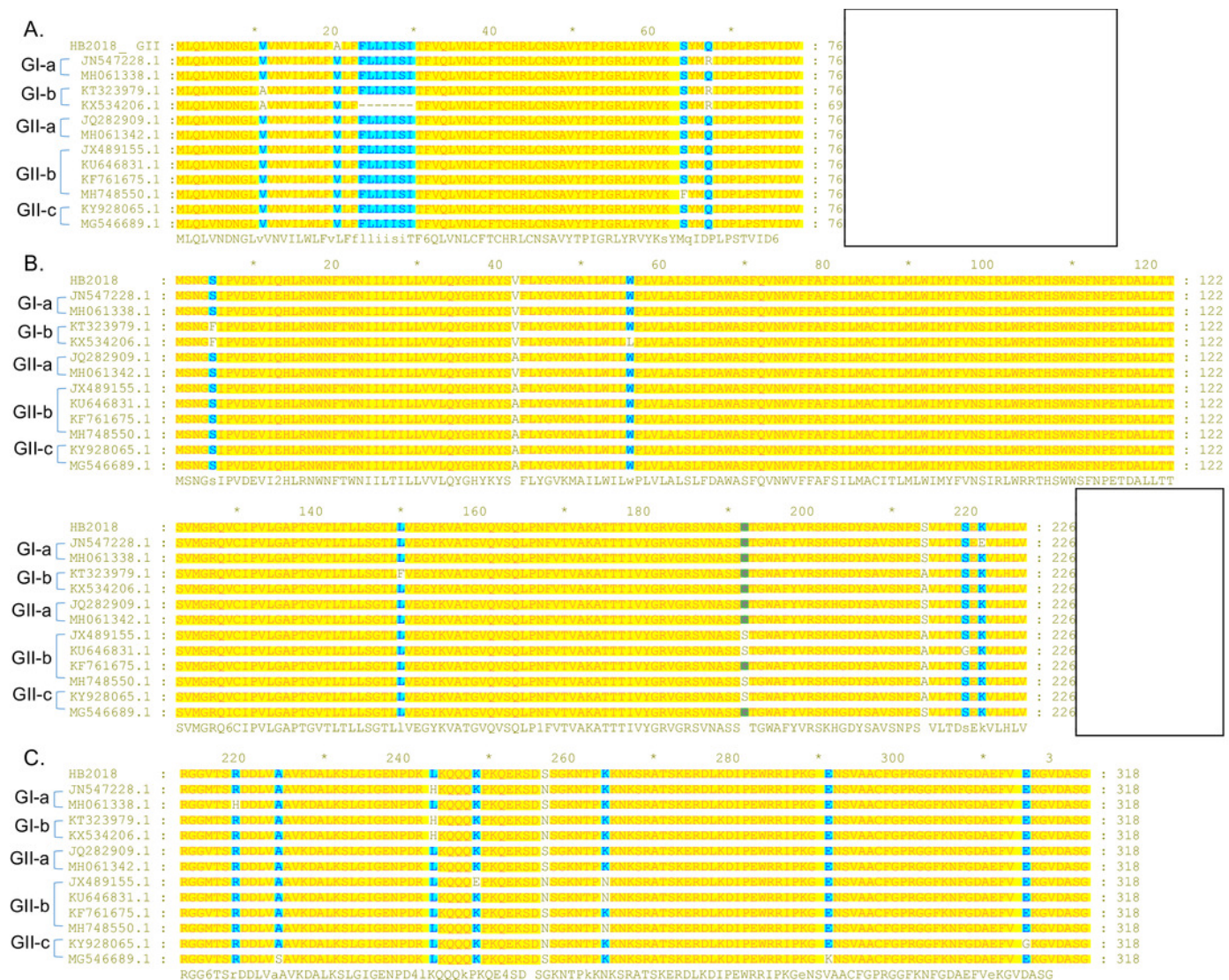
**(A)** E protein; **(B)** M protein; **(C)** N protein.

Sequence comparisons of different ORF regions between HB2018 and CV777.

**Table 1.** Sequence comparisons of different ORF regions between HB2018 and CV777.

| ORFs | HB2018 vs. CV777 | |
|---|---|---|
| | Amino acid similarity (%) | DNA identity (%) |
| ORF1 | 97.73 | 97.11 |
| S | 93.44 | 93.76 |
| ORF3 | 95.98 | 96.44 |
| E | 97.40 | 96.97 |
| M | 99.12 | 97.80 |
| N | 96.60 | 95.48 |

Single nucleotide polymorphism (SNP) analysis and dN/dS ratios of PEDV strains HB2018 and CV777.

1 **Table 2 Single nucleotide polymorphism (SNP) analysis and dN/dS ratios of PEDV strains**
2 **HB2018 and CV777.**

|  | ORFs | Sum | Non-synonymous | Synonymous | dN/dS |
|---|---|---|---|---|---|
|  | Total | 925 | 262 | 663 | 0.395 |
|  | ORF1a | 397 | 118 | 279 | 0.423 |
|  | ORF1b | 188 | 26 | 162 | 0.160 |
| HB2018 vs. | S | 233 | 88 | 145 | 0.607 |
| CV777 | ORF3 | 24 | 9 | 15 | 0.600 |
|  | E | 8 | 2 | 6 | 0.333 |
|  | M | 15 | 2 | 13 | 0.154 |
|  | N | 60 | 17 | 43 | 0.395 |

High-specificity N-glycosylation sites predicted in Asian strains

**Table 3.** High-specificity N-glycosylation sites predicted in Asian strains.

| Country/ Regions | Strain | 57 NSTW | 112 NATA | 127 NKTL | 212 NVTS | 320 NDTS | 347 NSSD | 510 NITV | 552 NVTN | 777 NISI | 1245 NKTL | 1257 NRTG | 22 | 347 | 382 | 421 | 524 | 739 | 869 | 1198 | 1274 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| China | CV777 | - | NTSA | • | • | - | • | • | • | • | - | - | - | - | - | - | - | - | - | - | - |
| | HB2018 | • | • | - | • | • | • | - | - | • | • | - | - | - | - | - | - | - | - | - | - |
| | LZC | - | - | • | • | • | • | • | • | • | • | • | - | - | - | - | - | - | - | - | NLTG |
| | DR13 | • | • | • | • | • | NSSN | - | • | • | • | • | - | - | - | - | - | - | - | - | - |
| | SD-M | NSSS | NTSA | • | • | • | • | • | • | • | - | - | - | - | - | - | - | - | - | - | - |
| | AJ1102 | • | • | • | • | • | • | - | • | • | - | - | - | - | - | - | - | - | - | - | - |
| | FJZZ1 | • | • | • | • | • | • | - | • | • | - | - | - | - | - | - | - | - | - | - | - |
| | LS | - | - | - | • | • | • | - | • | • | • | NVTR | - | - | - | - | - | - | - | - | - |
| | CHS | - | • | • | • | • | • | • | • | • | • | - | - | NFTD | - | - | - | - | - | - | - |
| | CHHNQX-314 | - | • | • | • | • | • | - | • | • | • | - | - | - | - | NLTA | NCTE | - | - | - | - |
| Thailand | CHYJ130330 | • | • | - | • | • | NSTN | • | - | NITI | • | • | - | - | - | - | - | - | - | - | - |
| | CBR1 | • | • | • | • | - | • | • | - | • | • | • | - | - | - | - | - | - | - | - | - |
| | AVCT12 | - | - | • | • | • | • | • | • | • | - | - | - | - | - | - | - | - | - | NYTA | - |
| Taiwan | PT-P5 | • | • | - | • | • | NSSN | • | - | • | NKTR | • | - | - | - | - | - | - | - | - | - |
| South Korea | KNU-1709 | • | • | - | • | • | NSSN | - | - | • | • | • | - | - | NSTV | - | - | - | NISS | - | - |
| | KNU-1702 | NSSS | - | • | • | • | • | - | • | • | • | - | - | - | - | - | - | - | - | - | - |
| | KNU-1305 | • | • | - | • | • | NSSN | - | - | • | • | - | - | NFSL | - | - | - | - | - | - | - |
| Vietnam | HUA- | • | • | - | • | • | NSS | - | - | • | • | • | - | - | - | - | - | - | NIS | - | - |

| Country | Strain | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
|  | **14PED96** |  |  |  |  |  | N |  |  |  |  |  |  |  |  |  |  |  |  | S |  |
|  | **VN/JFP1013** | • | • | • | • | - | • | - | - | • | • | • | - | - | - | - | - | - | - | - | - |
|  | **Tottori2** | - | - | • | • | • | • | • | • | • | • | - | - | - | - | - | - | - | - | NYT A | - |
| **Japan** | **OKY-1** | - | - | • | • | • | • | • | • | • | • | - | - | - | - | - | - | - | - | - | - |
|  | **IBR-7** | • | • | - | • | • | NSS N | - | - | • | • | • | - | - | - | - | - | - | - | - | - |

[1] The high-specificity N-glycosylation sites and their amino acids are summarized, the representative strains from each country/regions are listed." •" means the strain has this high-specificity N-glycosylation site; "-" means the strain has no this high-specificity N-glycosylation site; the amino acid sequences means the strain has a high-specificity N-glycosylation in this sits , but the amino acid sequences are different with the common sequences.