Virus Genes (2014) 49:58–67
DOI 10.1007/s11262-014-1055-y

# Bioinformatics insight into the spike glycoprotein gene of field porcine epidemic diarrhea strains during 2011–2013 in Guangdong, China

**Jianwei Hao · Chunyi Xue · Liangliang He · Yang Wang · Yongchang Cao**

Received: 28 November 2013 / Accepted: 6 March 2014 / Published online: 26 April 2014
© Springer Science+Business Media New York 2014

**Abstract** Three strains of porcine epidemic diarrhea virus (PEDV) were isolated from dead or diseased pigs at different swine farms in Guangdong during 2011–2013, and their S genes were sequenced. In the same period, seven PEDV strains were also isolated in Guangdong by other laboratories. The spike sequences of 10 Guangdong isolates were compared with vaccine strains and reference pathogenic isolates using six bioinformatics tools. The results revealed that 10 Guangdong strains, excluding strain GDS03, had distinct characteristics in terms of primary structure, secondary structure, high-specificity N-glycosylation sites, potential phosphorylation sites, and palmitoylation sites. Phylogenetic analysis also confirmed these findings and revealed that all PEDV strains were clustered into three distinct groups. Ten Guangdong strains, not including GDS03, belong to Group 1, whereas four vaccine strains and GDS03 belong to Group 3, which is evolutionarily distant from Group 1. Alignment analysis of the neutralizing region amino acid sequences indicated that the amino acid substitutions of Y/D766S, T549S, and G594S that are present in the Guangdong strains, not including GDS03, were a sign of predominant genetic changes among the isolated strains. GDS03 is closely related to the 83P-5 vaccine strain, which suggests that it might represent re-isolation of the vaccine strain or vaccine variants. Taken together, these results indicate that there have been predominant new strains circulating in Guangdong from 2011 to 2013, and the circulating PEDV strains have a genetic composition that is distant from reference strains, especially the vaccine strains; however, the vaccinations might also provide some level of cross-protection, as there have been no changes in the neutralizing epitopes of SS2 and 2C10. This explains why there have been constant but infrequent outbreaks recently in comparison to late 2010 in which PEDV outbreaks were more frequent and severe. In addition, the USA-Colorado-2013 strain had the same amino acid substitutions in the neutralizing regions as the Guangdong strains except GDS03, which suggests that the information and strategies in this study may play role in PEDV variant research in other countries.

**Keywords** Porcine epidemic diarrhea virus · Spike gene · Bioinformatics analysis · Phylogenetic relationship

## Introduction

Porcine epidemic diarrhea (PED), which is characterized by vomiting, watery diarrhea, and dehydration, is a highly contagious enteric disease. When PED occurs in a swine farm, nursery piglets usually have high morbidity and mortality, although all of the pigs are susceptible. The etiological agent of PED is porcine epidemic diarrhea virus (PEDV), which was first recognized in Belgium and the UK in 1978, and it quickly spread in Europe and Asia, resulting in severe economic losses [1–4]. In China, PEDV infection is a continuing problem, although a dual-combination kill vaccine against PEDV and transmissible gastroenteritis virus (TGEV) is in use. Since late 2010, PEDV variant infections have become a severe problem in the Chinese swine-production industry. Guangdong, as one of the major swine-producing provinces in China, has suffered economic loss from these new PEDV outbreaks

J. Hao · C. Xue · L. He · Y. Wang · Y. Cao (✉)
State Key Laboratory of Biocontrol, School of Life Sciences, Sun Yat-sen University, Guangzhou 510006, People's Republic of China
e-mail: caoych@mail.sysu.edu.cn

[5–8]. Seven strains were isolated in these outbreaks. In the course of our continuous PEDV surveillance activities in Guangdong, we also isolated three Vero-adapted PEDV strains (GDS01–3) from Guangdong in 2012 and 2013. Apparently, PEDV infection status has greatly changed since late 2010. Although new variants were isolated, we have less knowledge and understanding of the reason for the circulation and of the molecular profiles of these new isolates.

PEDV is a positive-sense single-strand RNA virus belonging to Group 1 of the genus *Alphacoronavirus*, within the family *Coronaviridae* [9]. Its genome organization is similar to other coronaviruses and is arranged into orderly 5′-untranslated regions (5′-UTRs), the polymerase gene, spikes (S), open reading frame 3 gene (ORF3), envelope (E), membrane (M), nucleocapsid (N), and 3′-UTR [10]. The structural protein genes, including S, E, M, and N, are located downstream, with the ORF3 gene among the genes as the only accessory gene. Among the structural proteins encoded by the genes mentioned above, the S protein, a glycoprotein peplomer on the viral surface, plays an important role in the process of inducing neutralizing antibodies and specific receptor binding [11]. Several neutralizing peptides (SS2, SS6, 2C10, and COE) have been identified in the spike protein of PEDV [12–14]. In addition, mutations within the S gene are also associated with the adaption process in vitro and the induction of attenuation in vivo [15]. Some studies of CoVs S have revealed that the differences within the S protein influence the induction of syncytia in infected cells [15, 16]. To better understand the newly isolated PEDV strains in Guangdong and effectively prevent related outbreaks, it is necessary to investigate the molecular changes of the spike gene of the new variants and compare them with early reference isolates and glycoprotein peplomer vaccine strains. The evaluation of sequence property, primary structure, secondary structure, and leucine-rich repeat (LRR) regions could reveal whether the mutations change the structure and further impact the protein function. The aim of those tests is to form the basic structure profile of PEDV variants with reference strains. Post-translational modifications (PTMs) often strongly affecting the protein functions, N-glycosylation sites, and potential phosphorylation are studied as they relate to virus biology such as survival and virulence. In addition, palmitoylation as a special class of PTMs could enhance the surface hydrophobicity and membrane affinity of protein substrates and then may influence the virulence. Thus, palmitoylation predictions of both early strains and recent strains are needed. In this study, 10 Guangdong strains isolated from 2011 to 2013 and reference strains were evaluated for those bioinformatics features, and meanwhile, phylogenetic analysis of those strains was also performed. These findings provide useful evidence as to the nature of the circulating PEDV strains, and the implications of this study in terms of the strategies for future protection of PEDV infections are also discussed.

## Materials and methods

### Isolation of the PEDV strains GDS01–3

The PEDV GDS01 strain was isolated from an intestine sample obtained from a single farm located in western Guangdong province in 2012. The PED outbreak in this farm was characterized by vomiting, watery diarrhea, and dehydration, with 90 % mortality in suckling piglets. The GDS02 strain was isolated from a farm located in central Guangdong province in 2013, and the piglets' mortality was 60 % lower than that caused by the GDS01 strain, although both cases had the same breed of sows and the same immune program with the killed vaccines. The infected pigs of other ages exhibited diarrhea and the loss of appetite to different degrees at both farms. Affected sows always recovered within 1 week.

The GDS03 strain was also isolated from a farm in central Guangdong province in 2013; however, this farm displayed lesser diarrhea severity for pigs of all ages in comparison to the GDS01 and GDS02 cases. An intestine sample was obtained from a diseased pig. The PEDV vaccinations also used killed vaccines.

Virus isolations were performed according to a method described previously with minor modifications. Briefly, the intestinal samples that were positive for PEDV but negative for TGEV, classical swine fever virus, porcine circovirus 2, porcine kobuvirus, porcine reproductive and respiratory syndrome virus, porcine bocavirus, and porcine reovirus were homogenized with phosphate-buffered saline (PBS). After centrifugation, the homogenized samples were further filtered through a 0.22-µm syringe drive filter (Millipore, USA) and were then used as inoculum. The growth medium (Dulbecco's modified Eagle's medium, DMEM supplemented with 10 % heat-inactivated fetal calf serum and antibiotics) was removed from confluent monolayer Vero cells, which were washed twice with the PBS. The washed cells were infected with a mixture of the inoculum and ''infection medium'' (DMEM, 0.3 % tryptose phosphate broth and 10 µg $ml^{-1}$ trypsin). After adsorption for 60 min at 37 °C, the cells were washed with PBS, and the infection medium was then added. The Vero cell cultures were observed for 5 days for cytopathic effects (CPEs). The PEDV isolates were also identified using specific primers in a reverse transcription polymerase chain reaction (RT-PCR).

Table 1 Primers used for sequencing reactions

| Primers | Locations[a] | Sequences |
|---|---|---|
| 1F | 18,874–18,896 | CGTAGCTTTTGAGTTGTATGCCA |
| 1R | 21,330–21,309 | GCAATTAGCTGTACAGGGTTCA |
| 2-1F | 21,080–21,101 | CCATTCCAGCTTATATGCGTGA |
| 2-1R | 23,487–23,465 | GTACATGTGAAGCTTCTCAGCGT |
| 2-2F | 21,188–21,205 | ATGATTGGTCCCGTGTTG |
| 2-2R | 23,386–23,365 | AGCGCTTATAGTCTTCATCAAC |
| 3F | 23,272–23,292 | GTGTACGATCCTGCAAGTGGC |
| 3R | 25,715–25,694 | TCACCTCATCAACGGGAATAGA |

[a] Location corresponds to position within the CV777 (AF353511) genome

## Determination of S gene of PEDV strains GDS01–3

For spike gene sequencing, the viral RNA was extracted from 200 µl of the GDS01–3 viral stocks using the TRIzol reagent (Invitrogen, NY, USA) according to the manufacturer's instructions. Three overlapped PCR fragments spanning the entire spike gene of PEDV were amplified using the specific primer sets (Table 1). Briefly, the first and the third fragment that harbor the N-terminal and C-terminal of the S gene of the three strains were amplified using two pairs of universal primers (1F and 1R for the first fragment, 3F and 3R for the third fragment). The second fragment of the S gene of GDS01 and GDS02 was amplified using 2-1F and 2-1R, whereas the second fragment for GDS03 was amplified with the primer pair of 2-2F and 2-2R. The PCR products were cloned into the pMD-18T vector (TaKaRa) after the purification procedure, and each fragment was sequenced at least five times, and the consensus sequence was determined. The sequences were analyzed using a Lasergene DNAStar (version 7, Lasergene Corporation, Madison, WI, USA).

## Phylogenetic analysis of PEDV S genes

The Guangdong PEDV sequence set included the strains GDS01–3 kept in our laboratory, along with the CHGD-01, CH-GD-2011, CH-GDHY-2011, GD-1, GD-A, GD-B, and LC strains that were isolated from Guangdong by different laboratories. All of these strains were isolated during the period 2011–2013. In this study, a total 31 PEDV reference strains for which the spike sequences were available in the GenBank database were selected. The selected PEDV reference strains and their accession numbers are shown in Table 2.

Nucleotide (nt) sequences were translated with the EditSeq program, and both the nt and peptide sequences were aligned and analyzed with the MegAlign program [17]. Phylogenetic trees, based on the nt sequence of the S gene,

Table 2 Ten Guangdong PEDV isolations and reference strains used in this study

| Virus strains | Country and year of isolation | Accession numbers | References |
|---|---|---|---|
| **GDS01**[a] | China, 2012 | AB857233 | In this study |
| **GDS02**[a] | China, 2013 | AB857234 | In this study |
| **GDS03**[a] | China, 2013 | AB857235 | In this study |
| **CH-GD-2011**[a] | China, 2011 | JQ638915 | Unpublished |
| **CHGD-01**[a] | China, 2011 | JX261936 | Pan et al. [34] |
| **CH-GDHY-2011**[a] | China, 2011 | JX145339 | Unpublished |
| **GD-1**[a] | China, 2011 | JX647847 | Wei et al. [8] |
| **GD-A**[a] | China, 2012 | JX112709 | Fan et al. [6, 37] |
| **GD-B**[a] | China, 2012 | JX088695 | Luo et al. [7] |
| **LC**[a] | China, 2012 | JX489155 | Chen et al. [5] |
| CV777 | Belgium, 1988 | AF353511 | Kocherhans et al. [39] |
| Br1/87 | UK, 1993 | Z25483 | Duarte et al. [40] |
| USA/Colorado/2013 | USA, 2013 | KF272920 | Marthaler et al. [41] |
| 83P-5 | Japan | AB548618 | Sato and Takeyama [15] |
| 83P-5 34th passage | Japan | AB548619 | Sato and Takeyama [15] |
| 83P-5 61st passage | Japan | AB548620 | Sato and Takeyama [15] |
| 83P-5 100th passage | Japan | AB548621 | Sato and Takeyama [15] |
| KH | Japan | AB548622 | Sato and Takeyama [15] |
| MK | Japan | AB548624 | Sato and Takeyama [15] |
| NK | Japan | AB548623 | Sato and Takeyama [15] |
| Attenuated DR13 | South Korea, 2001 | JQ023162 | Park et al. [42] |
| Virulent DR13 | South Korea, 1999 | JQ023161 | Park et al. [42] |
| Chinju99 | South Korea, 1999 | AY167285 | Yeo et al. [43] |
| SM98 | South Korea | GU937797 | Unpublished |
| CNU-091222-01 | South Korea, 2009 | JN185634 | Unpublished |
| CNU-091222-02 | South Korea, 2009 | JN184635 | Unpublished |
| KNU-0801 | South Korea, 2008 | GU180142 | Lee et al. [44] |
| KNU-0802 | South Korea, 2008 | GU180143 | Lee et al. [44] |
| KNU-0901 | South Korea, 2009 | GU180144 | Lee et al. [44] |
| KNU-0902 | South Korea, 2009 | GU180145 | Lee et al. [44] |

**Table 2** continued

| Virus strains | Country and year of isolation | Accession numbers | References |
|---|---|---|---|
| KNU-0903 | South Korea, 2009 | GU180146 | Lee et al. [44] |
| KNU-0904 | South Korea, 2009 | GU180147 | Lee et al. [44] |
| KNU-0905 | South Korea, 2009 | GU180148 | Lee et al. [44] |
| CH/S | China, 1986 | JN547228 | Chen et al. [45] |
| JS-2004-2 | China, 2004 | AY653204 | Unpublished |
| LJB/03 | China, 2006 | DQ985739 | Unpublished |
| LZC | China, 2006 | EF185992 | Unpublished |
| DX | China, 2007 | EU031893 | Unpublished |
| JS2008 | China, 2008 | KC210146 | Unpublished |
| BJ-2011-1 | China, 2011 | JN825712 | Yue et al. [46] |
| CH-FJND-3-2011 | China, 2011 | JQ282909 | Chen et al. [45] |

[a] The 10 Guangdong field isolates from this study are indicated in boldface type

were constructed using PHYLIP software [18], applying the neighbor-joining method.

### Bioinformatics analysis

To fully understand the bioinformatics characters for the Guangdong sequence set and the 31 reference strains, six bioinformatics tools were applied to test all the above strains. N-glycosylation sites were predicted by services available on http://www.cbs.dtu.dk/services/NetNGlyc [19]. For identifying high-specificity N-glycosylation sites, any potential crossing 0.5 and Jury agreement (9/9) or potential greater than 0.75 for asparagines that occur within the Asn-Xaa-Ser/Thr triplet was used. The potential phosphorylation sites were determined using http://www.cbs.dtu.dk/services/NetPhos [19]. LRR regions were identified by LRRfinder, which is available at http://www.lrrfinder.com/result.php [20]. The primary structures of the spike glycoprotein were predicted by http://expasy.org/tools [21]. Palmitoylation sites were predicted with the medium threshold frequency using services at http://csspalm.biocuckoo.org/prediction.php [22]. The secondary structures (alpha helices, beta strands, and random coils) of the protein were predicted using the bioinformatics tools available on the following website http://npsa-pbil.ibcp.fr. The method of GOR4 at http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=n-psa_gor4.html was used to identify the alpha helices, beta strands, and coil residues. At least a minimum of three or more turns were taken into account for one helix, strand, or coil in the spike glycoprotein structure [23].

## Results

### Sequence determination of the S gene of GDS01–3

Three PEDV strains, designated as GDS01–3, were successfully isolated, and all of them displayed typical PEDV CPEs, e.g., the cell fusion and syncytia formation in Vero cells. The spike genes of the three PEDV strains were amplified successfully by RT-PCR. The sequenced results revealed that the S gene length of GDS01 and GDS02 was 4,158 nts, which was 9 nts longer than that of GDS03, and revealed that mutations, insertions, and/or deletions presented in S genes, resulting in different lengths of nts and consequently different numbers of amino acids being encoded.

The genomic sequences of the PEDV strains GDS01–3 have been submitted to the DDBJ database and were assigned the accession numbers AB857233, AB857234, and AB857235, respectively.

### Sequence properties of the S gene of Guangdong PEDVs

The S gene sequences of 10 Guangdong isolates and four PEDV vaccine strains (CV777, SM98, attenuated DR13, and 83P-5-serial strains) were analyzed. The results revealed that (excluding GDS03) nine strains of the Guangdong sequence set displayed high levels of sequence identity of 97.8–99.9 % nt (97.9–99.8 % aa), whereas they displayed low sequence identity 93.7–94 % nt (91.6–92.3 % aa) with GDS03. The homology of the GDS03 spike sequence to vaccine strains was 96.8–99.8 % nt (95–99.6 % aa), whereas the identity of the rest of the strains in the Guangdong sequence set to those vaccine strains was 93.7–95.2 % nt (92.5–94.5 % aa).

Previous studies on the S protein of PEDV identified neutralizing epitopes of SS2 (aa positions 748–755) and SS6 (aa positions 764–771) in the S1 domain and 2C10 (aa positions 1,368–1,374) in the cytoplasmic domain. We found that nine strains (i.e., strains excluding GDS03) of the Guangdong sequence set had the same mutation Y/D766S in SS6 (Fig. 1). There were amino acid substitutions in the neutralizing region of COE (499–638 aa; Table 3). The nine strains (i.e., strains excluding GDS03) of the Guangdong sequence set had the same mutations T549S and G594S in COE.

The Chinese early strains of LJB/03, DX, and JS-2004-02 present all three mutations, and the JS2008 and CH/S strains present Y/D766S mutation. The KNU-0802, KNU-0902, KNU-0903, and KNU-0905 present the G594S mutation; the KNU-0802 and KNU-0902–5 present Y/D766S substitution. The CNU-serial strains contained all three substitutions.

```
                          SS2                        SS6                                                 2C10
                          750            760              770          780                  790     1370
CV777 *             TEPVLV|YSNIGVCK|SGSIGYVP|SQYGQVKI|APTVTGNISIPTNFSMSIR|GPRLQPY
attenuated DR13 *   ......|........|........|L.D.....|..M................|.......
83-P5 100th *       ......|........|........|L.D.....|..M................|.......
SM98 *              ......|........|........|P.......|...................|.......
CHGD-01             ......|........|.......R|..S.....|...................|.......
CH-GD-2011          ......|........|........|..S.....|...................|.......
CH-GDHY-2011        ......|........|........|..S.....|...................|.......
GD-A                ......|........|.......R|..S.....|...................|.......
GD-B                ......|........|........|..S.....|...................|.......
GD-1                ......|........|.......R|..S.....|...................|.......
LC                  ......|........|........|..S.....|...................|.......
GDS01               ......|........|.......R|..S.....|...................|.......
GDS02               ......|........|........|..S.....|...................|.......
GDS03               ......|........|........|L.D.....|..M................|.......
USA-Colorado-2013   ......|........|........|..S.....|...................|.......
```

**Fig. 1** Alignment of the deduced amino acid sequences of partial S protein of Guangdong isolates with that of vaccine strains. The *dots* represent amino acids that are identical to those in the CV777. *Boxes* indicate the neutralizing epitopes (748–755, SS2; 764–771, SS6; and 1,368–1,374, 2C10). The vaccine strains are indicated with *asterisk*

**Table 3** Analysis of amino acid mutations in COEs of Guangdong sequences and reference strain CV777

| Strains | Amino acids in COE of CV777 | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 511N | 517A | 521L | 523S | 527V | 549T | 567S | 577N | 594G | 605A | 606F | 611K | 612L | 621K | 633E | 635I |
| Attenuated DR13* | – | – | H | G | I | – | – | – | – | E | – | – | F | – | Q | V |
| 83P-5 100th* | – | – | H | G | I | – | – | – | – | E | – | – | F | – | Q | V |
| SM98* | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – | – |
| CH-GDHY-2011 | – | – | H | G | I | S | – | S | S | E | – | E | F | – | – | V |
| CH-GD-2011 | I | S | H | G | I | S | T | – | S | D | Y | – | F | – | – | V |
| CHGD-01 | – | – | H | G | I | S | – | – | S | E | – | – | F | – | – | V |
| LC | – | – | H | G | I | S | – | – | S | E | – | – | F | – | – | V |
| GDS01 | – | – | H | G | I | S | – | – | S | E | – | – | F | – | – | V |
| GDS02 | – | – | H | G | I | S | – | – | S | E | – | – | F | – | – | V |
| GDS03 | – | – | H | G | I | – | – | – | – | E | – | – | F | – | Q | V |
| GD-A | – | – | H | G | I | S | – | – | S | E | – | – | F | – | – | V |
| GD-B | – | S | R | G | I | S | – | – | S | D | – | – | F | T | – | V |
| GD-1 | – | – | H | G | I | S | – | – | S | E | – | – | F | – | – | V |
| USA/Colorado/2013 | – | S | – | G | I | S | – | – | S | E | – | – | F | – | – | V |

Number indicates the position for amino acids, and the dashes indicate the amino acids are identical to those of strain CV777. The vaccine strains used in other country are indicated with asterisk

Interestingly, the USA-Colorado-2013 strain, isolated recently from PEDV-affected swine in the USA, has the same mutations as the nine Guangdong isolates in both neutralizing regions.

Previous reports suggested that the five amino acid changes ($Thr_2$, $Tyr_{152}$ deletion, $Leu_{153}$, $Met_{774}$, and $Leu_{963}$ in the S of 83P-5 100th) may influence the virulence of the cell-adapted PEDV virus. The GDS03, JS2008, attenuated DR13, and 83P-5 100th strains fulfill those changes. The CH/S strain presents the $Leu_{153}$ and $Met_{774}$ changes. The CH-FJND-3-2011 presents the $Leu_{963}$ change. Other strains in this study do not contain those changes.

The strains of the Guangdong group possessed similar molecular weights, whereas the reference strains had differences of approximately 0.01–1.17 kDa. The isoelectric points (p*I*) 5.00–5.22 of the Guangdong sequence group were lower than that of KNU-serial, 5.24–5.45, but were similar to the 5.19 of the other reference isolates. There were 80–83 positively charged residues (Arg + Lys) in the Guangdong sequence sets, which are 1–6 residues less than the strains isolated in Japan. The computed instability index (II) and aliphatic index had various values of 30.57–33.75 and 89.21–93.56, respectively, in all analyzed strains. The grand hydropathicity (GRAVY) was found to have positive values in all strains. The primary structures of the spike glycoprotein were found to be relatively different between the groups defined by time and districts.

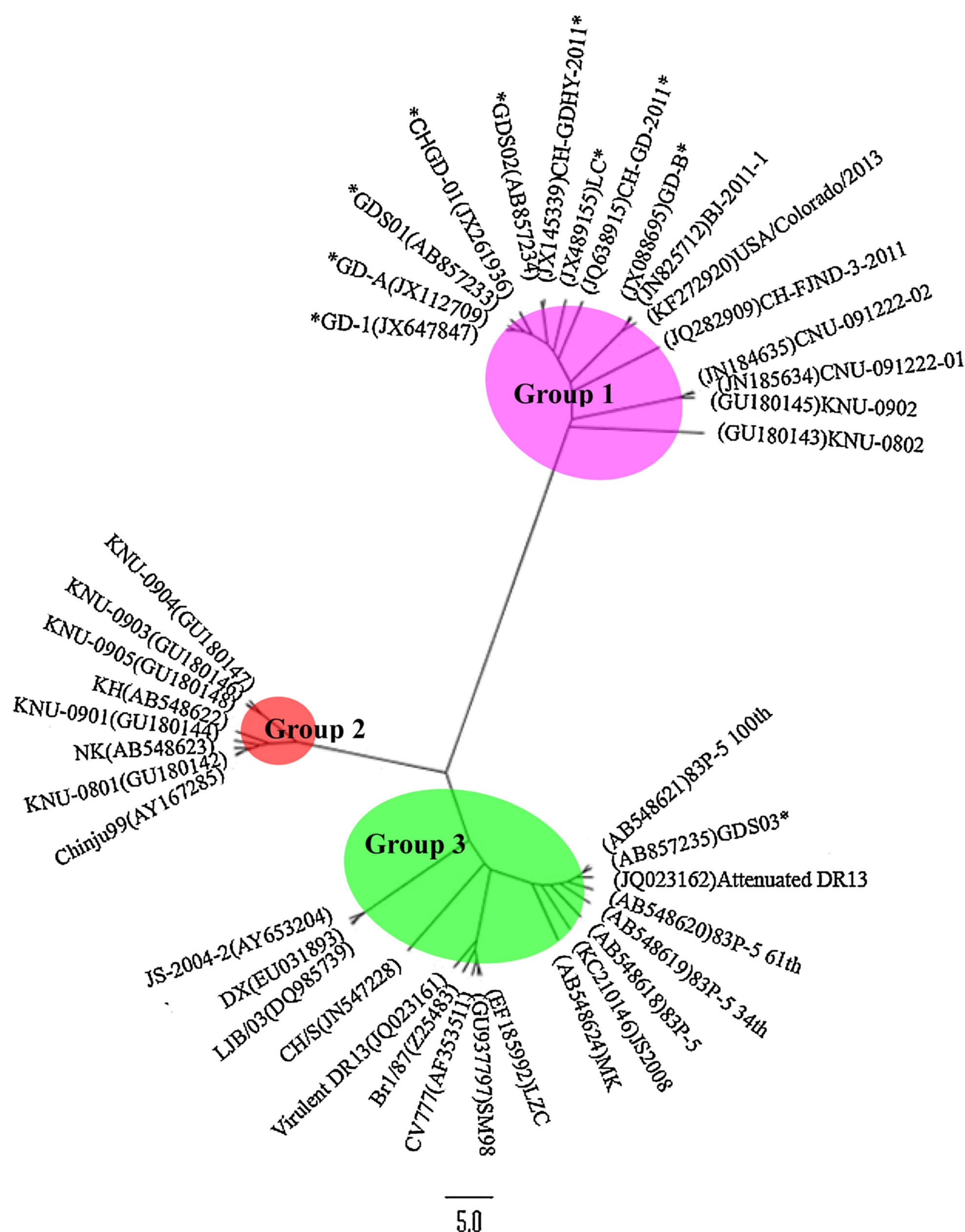


**Fig. 2** Relationships among the Guangdong sequence set and reference strains determined using a neighbor-joining method phylogeny *tree* and based on complete S nucleotides. The tree reliability/robustness of the hypothesis was evaluated by bootstrap of 1,000 replicates. The Guangdong sequences are indicated with an *asterisk*. Groups 1–3 are shown in *pink*, *red*, and *green*, respectively (Color figure online)

## Forty-one PEDV strains grouped into three genotypes by phylogenetic analysis

The phylogenetic analysis of the complete spike gene based on the PEDV indicated that the sequences of the analyzed 41 strains were divided into three distinct groups (Fig. 2). The Guangdong PEDV sequence set strains, except for GDS03, were clustered with four South Korean strains (CNU-091222-01, CNU-091222-02, KNU-0802, and KNU-0902), two Chinese 2011s PEDV isolates (BJ-2011-1 and CH-FJND-3-2011), and a field strain from the USA (Fig. 2). Group 2 contained two Japanese strains (KH and NK), five KNU-serial strains (KNU-0801, KNU-0901, KNU-0903–5), and Chinju99. Group 3 included the rest of the reference strains isolated from various areas, which were mainly characterized during early isolation. The vaccine strains (CV777, attenuated DR13, SM98, and 83P-5) used in different countries were clustered in Group 3 (Fig. 2).

## Bioinformatics analysis revealed Guangdong PEDV strains have different molecular characteristics

### *Prediction of N-glycosylation*

All analyzed strains contained 9–10 high-specificity N-glycosylation sites, three of which are conserved. The CV777-based strain is the only vaccine source in China. An analysis comparing CV777 and all of other strains was

**Table 4** The variation of highly specific N-glycosylation sites of 40 strains in comparison to CV777

| Strains | High-specificity N-glycosylation sites of CV777 | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | 127NKTL | 213NVTS | 321NDTS | 348NSSD | 511NITV | 553NVTN | 778NISI | 1246NKTL | 1258NRTG |
| Attenuated DR13* | – | – | N | – | – | – | – | – | N |
| Virulent DR13 | – | – | – | – | – | – | – | – | – |
| 83P-5 | – | – | N | – | – | – | – | – | – |
| 83P-5 34th | – | – | N | – | – | – | – | – | – |
| 83P-5 61th | – | – | N | – | – | – | – | – | N |
| 83P-5 100th* | – | – | N | – | – | – | – | – | N |
| SM98* | – | – | – | – | – | – | – | – | N |
| Br1/87 | – | – | – | – | – | – | – | – | – |
| USA/Colorado/2013 | N | – | – | – | N | N | – | – | – |
| KH | – | – | – | – | – | N | – | – | – |
| MK | – | – | – | – | – | – | – | – | – |
| NK | – | – | – | – | N | – | – | – | – |
| Chinju99 | N | – | – | – | N | – | – | – | – |
| CNU-091222-01 | N | – | – | – | N | – | – | – | N |
| CNU-091222-02 | N | – | – | – | N | – | – | – | N |
| KNU-0801 | N | – | – | – | N | N | – | – | – |
| KNU-0802 | N | – | – | – | N | N | – | – | – |
| KNU-0901 | N | – | – | – | N | N | – | – | – |
| KNU-0902 | N | – | – | – | N | N | – | – | N |
| KNU-0903 | – | – | – | – | N | N | – | – | N |
| KNU-0904 | – | – | – | – | N | N | – | – | – |
| KNU-0905 | – | – | – | – | N | N | – | – | N |
| CH/S | – | – | – | – | – | – | – | – | – |
| JS-2004-2 | – | – | – | – | N | N | – | – | N |
| LJB/03 | – | – | – | – | N | N | – | – | – |
| LZC | – | – | – | – | – | – | – | – | – |
| DX | – | – | – | – | N | N | – | – | – |
| JS2008 | – | – | N | – | – | – | – | – | N |
| BJ-2011-1 | N | – | – | – | N | N | – | – | – |
| CH-FJND-3-2011 | N | – | – | – | N | N | – | – | – |
| **CH-GD-2011** | N | – | – | – | N | N | – | – | – |
| **CHGD-01** | N | – | – | – | – | N | – | – | – |
| **CH-GDHY-2011** | N | – | – | – | – | N | – | – | – |
| **GD-1** | N | – | – | – | – | N | – | – | – |
| **GD-A** | N | – | – | – | – | N | – | – | – |
| **GD-B** | N | – | – | – | N | N | – | – | – |
| **LC** | N | – | – | – | – | N | – | – | – |
| **GDS01** | N | – | – | N | – | N | – | – | – |
| **GDS02** | N | – | – | – | – | N | – | – | – |
| **GDS03** | – | – | N | – | – | – | – | – | N |

Guangdong isolates are indicated in boldface type, with the dash indicating possession of the same sites, and the "N" indicating that the strain did not contain the site, or the sites were not high-specificity sites. The vaccine strains used in other country are indicated with asterisk

conducted. According to the results, all Guangdong strains, except GDS03, lost the high-specificity N-glycosylation sites of Asn-Xaa-Ser/Thr-127NKTL and −553NVTN, which are conserved in vaccine strains (CV777, SM98, DR13, and 83P-5; Table 3). There was one additional N-glycosylation site, −62, present in CNU-serial, KNU-0802, KNU-0902, and Guangdong isolates, except GDS03, compared with all other reference strains.

Within the Guangdong sequence set, the nine strains (i.e., strains excluding GDS03) had similar N-glycosylation

conformation. The N-glycosylation sites of GDS03 were the same as 83P-5 61st and 83P-5 100th conformations. There were two strains, CH-GD-2011 and GD-B, that lost the site −511 contained in other Guangdong strains (Table 4).

*Prediction of palmitoylation sites*

Seven of 10 Guangdong strains (CHGD-01, CH-GD-2011, CH-GDHY-2011, GD-B, LC, GD-1, and GDS02) contained an additional palmitoylation site 230 (DGISYQPCTANCITG) compared with Chinese early strains (CH/S, JS2008, JS-2004-2, LJB/03, LZC, and DX). In addition, 8 of 10 Guangdong strains (CHGD-01, CH-GDHY-2011, GD-B, LC, GD-1, GDS01, GDS02, and GD-A) lost the site 122 (AIARLRCQFPDNKT) that was previously conserved. The isolate GD-B contained additional the palmitoylation site 1,362 (CGCCGCCCACFSGCC) compared with other Guangdong strains. The result revealed that most palmitoylation site changes are in the S1 part of the spike protein. The 2008–2009 Korea PEDV isolates (KNU-serial, CNU-serial) and pre-2011 Chinese strains had more palmitoylation site changes than other isolates.

*Prediction of potential phosphorylation sites*

The prediction results revealed that there were differences in numbers and locations of potential phosphorylation peptides in different strains. In all 41 strains, phosphorylation sites at positions Ser92, -216, -433, -698, -720, -788, -842, -878, -892, 918, -1013, -1062, -1089, -1108, -1124, and -1203 were found to be conserved. In the Guangdong sequence set, nine strains did not contain phosphorylation sites at positions Ser28 and -60, whereas there was an additional site at Ser177 in comparison to GDS03. In addition, compared with all other strains used in this study, the GDS01 strain contained one additional potential site at position Ser352.

*Secondary structure prediction*

The data regarding the secondary structures revealed that all the analyzed isolates contained 17–21 alpha helices, 63–68 beta sheets, and 80–89 residual coils. In the Guangdong sequence set, there was an addition alpha helix in the nine Guangdong isolates other than GDS03.

*Detection of LRR regions*

We searched for LRR regions in the spike gene; however, the results revealed that there were no LRR regions in all tested sequences.

## Discussion

The genetic analysis of the complete spike glycoprotein of the Guangdong strains isolated during 2011–2013 and selected reference isolates was conducted to predict the molecular characteristics of the strains in Guangdong and also the differences between the Guangdong sequence set and the selected reference isolates. According to the Guangdong sequences property analysis, GDS03 had lower identity to other Guangdong isolates, whereas the strain had higher sequence identity with early reference strains. These results indicated that there were at least two types of PEDV strains within the Guangdong set.

Excluding the GDS03 strain, for the other nine Guangdong isolates, there were several amino acid substitutions compared with vaccine strains that changed the amino acid constitution of neutralizing epitope SS6 and neutralizing region COE. Interestingly, we noticed that both serine amino acid substitutions in COE (T549S and G594S) and in SS6 (Y/D766S) were not only present in the nine Guangdong variants but also present in other newly reported PEDV strains in China [24] and other countries such as USA. The retrospective study of the three amino acid substitutions revealed that the shift process not only existed in Chinese early strains but also presented in 2008–2009 Korea PEDV strains. From retrospective study and our analysis, the three substitutions could be seen as the marker, which make PEDV virus easy to live. Both China and Korea control the disease on pig farms with periodic vaccination strategy, and thus, we speculate that those mutations present in neutralizing regions may represent the virus evolution through escaping from the antibodies. Further experiments are needed to determine whether they produce any antigenicity changes.

In another aspect, we also noticed that the SS2 and 2C10 epitopes had no mutations. In our study, though GDS01 and -2 had similar genetic variation trends, and the backgrounds of both strains were also similar, they caused different mortality in nursing piglets. In addition, the USA-Colorado-2013 has the same mutations as most Guangdong variants in COE and SS6; meantime, recent research revealed that USA strains may originate from Chinese variants [25], however, the PEDV circulation status is quite different. To sum up, there is a hint that following a large-scale vaccination in last 2–3 years in China, the CV777-based vaccine appears to provide some level of cross-protection, and the influences of PEDV infection were lower than at the beginning of the epidemic for China.

Since 2010, Chinese PEDV epidemic status has changed a lot; meanwhile, we noticed that previous attenuation marker of the five amino acid changes needs to be determined whether its contribution to the viral attenuation for virulent strains also presents some changes in our analysis;

the residue- or motif-distinguished newly strains from early Chinese strains or attenuated strains are investigated in our study. Six softwares were used to form the spike protein profile and to reveal changes of virus biology. The prediction result of primary structure of the 41 PEDV strains indicated a relative lower variable trend in physical properties such as p*I* and molecular weight of the Guangdong sequence set in comparison with other groups such as the Korea isolates and Japan isolates. The II values of all strains indicated that the spike glycoprotein is predicted to be stable. The grand average of the hydropathicity index results suggested a hydrophobic nature of the spike glycoprotein in all variants.

The secondary structure prediction revealed that there are no new structural domains and that no evolution has occurred in the structure of all 41 isolates. However, secondary structural variation in the alpha helices and beta strands has been observed within the Guangdong sequence set and between strains isolated in different years. In addition, the results of LRR investigation suggested that PEDV spike protein may not interact with PAMPs through LRR.

Three types of PTMs were predicted. The variations in N-glycosylation sites may affect survival and transmission of the virus and also affect the interaction with receptors and result in lower virus recognition by antibodies, hence influencing virus replication and infectivity [26–28]. In our study, the high-specific site-553NVTN is a promising marker of virus evolvement for it locates in the COE region which means small change perhaps leads to disturbance in folding and conformation of the molecule. The 127NKTL site loss in after-2010 strains except GDS03 is a unique motif change in China according to our analysis. For the loss or acquisition of N-glycosylation sites in the protein of lactate dehydrogenase-elevating virus, a virus belonging to the *Coronaviridae* has been reported to result in changes of virulence and cellular tropism [29]. We speculate that this motif perhaps influences PEDV virulence and could be useful to distinguish PEDV variants from vaccine or Chinese pre-2011 strains.

For palmitoylation motif, there is an interesting trend that newly site 230 present in most after-2010 strains accompanies with the 122 site lost which was conserved in vaccine and Chinese pre-2011 strains. We noticed that this change did not suddenly appear since the shift process could be observed in the CNU-serial, KNU-serial, KH, and NK strains. Palmitoylation plays roles in modulating protein trafficking, stability, and numerous cellular processes, including signaling and apoptosis, and it has been reported that the palmitoylation site changes influence the influenza virus hemagglutinin cellular trafficking [30] and HIV-1 infectivity [31]. Thus, different palmitoylation site conformations between PEDV variants and pre-2011 Chinese strains may lead to changes in virus–host interactions and then influence the virus virulence. However, further experimental studies are required to identify the effects of the newly motif changes on virus biology.

Potential phosphorylation prediction results revealed that Guangdong strains (i.e., strains excluding GDS03) have lost two potential phosphorylated peptides because of the amino acid mutations. However, for coronaviruses, the N protein is the only structural protein with phosphorylation activity. In this context, the observed potential phosphorylated peptides seem to be of no consequence for phosphorylation [32].

The result of PTM predictions revealed some motifs could be considered as the points distinguished between the variants and the vaccine strains or Chinese pre-2011 strains. These motifs perhaps are related to the virus virulence, however, the virus virulence may be decided by not only the changes present in spike protein but also ORF3 protein according to previous reports [33]. Thus, we also analyzed the ORF3 protein of Guangdong strains and the result showed that the ORF3 of GDS03 presents the characteristic continuous deletion of attenuated vaccine. Previously, the CHGD-01 ORF3 revealed notably five amino acid substitutions from other PEDV isolates [34]; our study found GDS02, GD-1, and GD-A have the same substitutions with the CHGD-01 strain. Furthermore, GDS01, LC, CH-GD-2011, and particularly DX strain present four of five amino acid substitutions; the phylogenetic analysis showed those eight strains cluster in one group (data not shown). We speculated that the ORF3 of Chinese early strain DX maybe the origin of the seven Guangdong strains. These unique amino acid substitutions are further needed to test the virulence.

There have been several phylogenetic studies performed on the spike gene of both partial sequences of S1 and complete spike sequences [35, 36]. The N, M, and ORF3 gene phylogenetic analysis also has been investigated in different areas [37, 38]. In this study, phylogenetic analysis revealed an interesting trend that the after-2010 strains in China or other areas appear to have closer relationships between each other [36]. The findings of the present study suggest that the genetic diversity and molecular characteristics in the sequences of the spike glycoprotein of new isolates may be the basis of the trend.

The characteristics of GDS03 are nearly the same as the 83P-5 100th strain; however, it was isolated from a diseased piglet with diarrhea. We speculate that it may be a re-isolation of vaccine strains or vaccine variations after several passages in pigs. As the CV777-based vaccine is the only choice used for vaccination, there is lack of information about the use of other vaccine strains in China.

In conclusion, the bioinformatics predictions confirm previous data on phylogenetic analysis and indicate that the

spike glycoprotein of the early predominant isolates and Guangdong strains, except GDS03, is molecularly different. As the spike glycoprotein has an important role in cell recognition, attachment, and neutralizing antibody induction, the identification of these characteristics will be the foundation of attaining a better understanding of epidemics, molecular mechanisms of infections, evolution and the basis for the development of effective vaccines.

## References


1. M.B. Pensaert, P. de Bouck, Arch. Virol. **58**, 243–247 (1978)
2. D. Chasey, S.F. Cartwright, Res. Vet. Sci. **25**, 255–256 (1978)
3. A. Pijpers, A.P. Van Nieuwstadt, C. Terpstra, J.H. Verheijden, Vet. Rec. **132**, 129–131 (1993)
4. K. Van Reeth, M. Pensaert, Vet. Rec. **135**, 594–597 (1994)
5. F. Chen, Y. Pan, X. Zhang, X. Tian, D. Wang, Q. Zhou, Y. Song, Y. Bi, J. Virol. **86**, 12448 (2012)
6. H. Fan, J. Zhang, Y. Ye, T. Tong, K. Xie, M. Liao, J. Virol. **86**, 10248–10249 (2012)
7. Y. Luo, J. Zhang, X. Deng, Y. Ye, M. Liao, H. Fan, J. Virol. **86**, 9551 (2012)
8. Z.Y. Wei, W.H. Lu, Z.L. Li, J.Y. Mo, X.D. Zeng, Z.L. Zeng, B.L. Sun, F. Chen, Q.M. Xie, Y.Z. Bee, J.Y. Ma, J. Virol. **86**, 13824–13825 (2012)
9. M.B. Pensaert, S.G. Yeo, in *Porcine Epidemic Diarrhea*, ed. by B.E. Straw, J.J. Zimmerman, S. D'Allaire, D.J. Taylor (Blackwell Publishing, Ames, 2006), pp. 367–372
10. A.A.F. de Vries, M.C. Horzinek, P.J.M. Rottier, R.J. de Groot, Semin. Virol. **8**, 33–47 (1997)
11. B.J. Bosch, R. Van Der Zee, C.A. De Haan, P.J. Rottier, J. Virol. **77**, 8801–8811 (2003)
12. S.H. Chang, J.L. Bae, T.J. Kang, J. Kim, G.H. Chung, C.W. Lim, H. Laude, M.S. Yang, Y.S. Jiang, Mol. Cells **14**, 295–299 (2002)
13. D.J. Cruz, C.J. Kim, H.J. Shin, Virus Res. **132**, 192–196 (2008)
14. D.B. Sun, L. Feng, H.Y. Shi, J.F. Chen, X.C. Cui, H.Y. Chen, S.W. Liu, Y.E. Tong, Y.F. Wang, G.Z. Tong, Vet. Microbiol. **131**, 73–81 (2008)
15. T. Sato, N. Takeyama, Virus Genes **43**, 72–78 (2011)
16. K. Shirato, M. Maejima, S. Matsuyama, M. Ujike, A. Miyazaki, N. Takeyama, H. Ikeda, F. Taquchi, Virus Res. **161**, 188–193 (2011)
17. T.G. Burland, Methods Mol. Biol. **132**, 71–91 (2000)
18. J. Felsenstein, Mol. Biol. Evol. **23**, 691–700 (2006)
19. N. Blom, S. Gammeltoft, S. Brunak, J. Mol. Biol. **294**, 1351–1362 (1999)
20. V. Offord, D. Werling, Innate Immun. **19**, 398–402 (2013)
21. E. Gasteiger, C. Hoogland, A. Gattiker, S. Duvaud, M.R. Wilkins, R.D. Appel, A. Bairoch, in *Protein Identification and Analysis Tool on the ExPASy Server*, ed. by J.M. Walker (Humana Press, Totowa, 2005)
22. J. Ren, L.P. Wen, X.J. Gao, C.J. Jin, Y. Xue, X.B. Yao, Protein Eng. Des. Sel. **21**, 639–644 (2008)
23. J. Garnier, J.-F. Gibrat, B. Robson, Methods Enzymol. **266**, 540–553 (1996)
24. Y.F. Tian, Z.J. Yu, K.H. Cheng, Y.X. Liu, J. Huang, Y. Xin, Y.G. Li, S.T. Fan, T.C. Wang, G. Huang, N. Feng, Z.G. Yang, S.T. Yang, Y.W. Gao, X.Z. Xia, Viruses **5**, 1991–2004 (2013)
25. Y.W. Huang, A.W. Dickerman, P. Pineyro, L. Li, L. Fang, R. Kiehne, T. Opriessnig, X.J. Meng, MBio. (2013). doi:10.1128/mBio.00737-13
26. T. Slater-Handshy, D.A. Droll, X. Fan, A.M. Di Bisceglie, T.J. Chambers, Virology **319**, 36–48 (2004)
27. J.C. Meunier, A. Fournillier, A. Choukhi, A. Cahour, L. Cocquerel, J. Dubuisson, C. Wychowski, J. Gen. Virol. **80**, 887–896 (1999)
28. D.J. Vigerust, V.L. Shepherd, Trends Microbiol. **15**, 211–218 (2007)
29. K. Li, T. Schuler, Z. Chen, G.E. Glass, J.E. Childs, P.G. Plagemann, Virus Res. **67**, 153–162 (2000)
30. J. Zhang, A. Pekosz, R.A. Lamb, J. Virol. **74**, 4634–4644 (2000)
31. I. Rousso, M.B. Mixon, B.K. Chen, P.S. Kim, Proc. Natl. Acad. Sci. USA **97**, 13523–13525 (2000)
32. D.J. Wilbur, V.L. Shepherd, Biochem. Biophys. Res. Commun. **141**, 7–12 (1986)
33. S.J. Park, H.J. Moon, Y.Z. Luo, H.K. Kim, E.M. Kim, J.S. Yang, D.S. Song, B.K. Kang, C.S. Lee, B.K. Park, Virus Genes **36**, 95–104 (2008)
34. Y.F. Pan, X.Y. Tian, W. Li, Q.F. Zhou, D.D. Wang, Y.Z. Bi, F. Chen, Y.H. Song, Virol. J. **9**, 195 (2012). doi:10.1186/1743-422X-9-195
35. S.J. Park, H.J. Moon, J.S. Yang, C.S. Lee, D.S. Song, B.K. Kang, B.K. Park, Virus Genes **35**, 321–332 (2007)
36. X.M. Wang, B.B. Niu, H. Yan, D.S. Gao, X. Yang, L. Chen, H.T. Chang, J. Zhao, C.Q. Wang, Arch. Virol. (2013). doi:10.1007/s00705-013-1767-7
37. J.H. Fan, Y.Z. Zuo, J.H. Li, L.H. Pei, Virus Genes **45**, 113–117 (2012)
38. X. Yang, J.Y. Huo, L. Chen, F.M. Zheng, H.T. Chang, J. Zhao, X.W. Wang, C.Q. Wang, Virus Genes **46**, 337–344 (2012)
39. R. Kocherhans, A. Bridgen, M. Ackermann, K. Tobler, Virus Genes **23**, 137–144 (2001)
40. M. Duarte, K. Tobler, A. Bridgen, D. Rasschaert, M. Ackermann, H. Laude, Viroligy **198**, 466–476 (1994)
41. D. Marthaler, Y. Jiang, T. Otterson, S. Goyal, K. Rossow, J. Collins, Genome Announc. **1**, e00555–13 (2013)
42. S.J. Park, H.K. Kim, D.S. Song, D.J. An, B.K. Park, J. Virol. **86**, 5964 (2012)
43. S.G. Yeo, M. Hernandez, P.J. Krell, E.E. Nagy, Virus Genes **26**, 239–246 (2003)
44. D.K. Lee, C.K. Park, S.H. Kim, C. Lee, Virus Res. **149**, 175–182 (2010)
45. J. Chen, C. Wang, H. Shi, H.J. Qiu, S. Liu, D. Shi, X. Zhang, L. Feng, J. Virol. **85**, 11538–11539 (2011)
46. G. Yue, Q. Kou, X. Ge, L. Zhou, X. Guo, H. Yang, Arch. Virol. **158**, 711–175 (2012)