

GENOME ANNOUNCEMENTS

Complete Genome Sequence of a Chinese Virulent Porcine Epidemic Diarrhea Virus Strain

Jianfei Chen,¹ Chengbao Wang,¹ Hongyan Shi,¹ Hua-Ji Qiu,¹
Shengwang Liu,² Da Shi,¹ Xin Zhang,¹ and Li Feng^{1*}

Division of Swine Infectious Diseases¹ and Division of Avian Infectious Diseases,² State Key Laboratory of Veterinary Biotechnology, Harbin Veterinary Research Institute, Chinese Academy of Agricultural Sciences, Harbin 150001, People's Republic of China

Received 16 August 2011/Accepted 22 August 2011

CH/S is a virulent porcine epidemic diarrhea virus (PEDV) strain and is used as the virulent strain to evaluate the protection rates of vaccines against PEDV infection in China. Here, we report the complete genome sequence of strain CH/S, which may aid in understanding the molecular characteristics of this strain.

Porcine epidemic diarrhea virus (PEDV), a member of genus *Alphacoronavirus* (family *Coronaviridae*, order *Nidovirales*), is a single-stranded positive-sense RNA virus that causes an acute and highly contagious enteric disease characterized by severe enteritis, vomiting, and watery diarrhea in swine. Strain CH/S occurred in a swine breeding farm in 1986 in Shanghai and was confirmed as PEDV by electron microscopy, direct immunofluorescence testing, and serum neutralization testing (4). To date, the complete genome sequence of only Belgian strain CV777 has been reported (9). It is thus necessary to analyze the complete genome sequence of CH/S and understand its molecular characteristics.

The 5' and 3' ends of the genome of CH/S were confirmed by using a Smarter rapid amplification of cDNA ends (RACE) kit (Clontech, Japan), and the other parts were generated by 12 overlapping cDNA fragments to encompass the entire genome and determined by genome walking sequencing. The coronavirus genome is a single-stranded positive-sense RNA genome of 27 to 32 kb that is infectious (2, 10). The complete genome sequence of CH/S is 28,026 nucleotides (nt) in length [excluding the poly(A) tail]. The 5' untranslated region (UTR) existing on the extreme 5' end of the genome has 292 nt and contains a positioned short AUG-initiated open reading frame (ORF), which consists of 39 nt (nt 95 to 133) and encodes a peptide of 12 amino acids. Such small ORFs are present in all coronaviruses sequenced so far (1, 3, 5, 8, 11, 12, 13, 15, 16, 17). The 3' UTR existing on the extreme 3' end of the genome is 334 nt in length and possesses an octameric sequence, GGA AGAGC, beginning at base 73 upstream from the poly(A) tail. The genome of CH/S contains six genes, the replicase (Rep), the spike (S), ORF3, envelope (E), membrane (M), and nucleoprotein (N) genes, arranged in the order 5'-Rep-S-ORF3-E-M-N-3'. Two long ORFs (ORF1a and ORF1b) are of

12,351 nt (nt 293 to 12643) and 8,037 nt (nt 12598 to 20634) in length and overlap by 46 nt. At the overlapping region, there is a specific seven-nucleotide "slippery" sequence (UUUAAAC) and a pseudoknot structure (ribosomal frameshifting signal), which are required for the translation of ORF1b (7).

Sequences at the 5' end of each gene represent signals for the transcription of subgenomic mRNAs of coronavirus (10, 14). These sequences, known as transcription-regulating sequences (TRSs), include a stretch of a highly conserved sequence designated the core sequence (CS), located at sites immediately upstream of most of the genes (7). The CSs in the genome of CH/S are the hexameric motifs 5'-XUA(A/G)AC-3'. The sequence CUAAAC found at the starting sites of gene 1 and the N gene, the GUAAAC sequence found at the starting site of the S gene, the CUAGAC sequence found at the starting sites of gene 3 and the E gene, and the AUAAAC sequence found at the starting site of the M gene are identical to those found in the same sites in the CV777 strain (6).

It is hoped that these data will not only facilitate future investigations of the molecular pathogenesis of PEDV but also contribute to the elucidation of the structure and function of the coronavirus gene-derived products.

Nucleotide sequence accession number. The complete genome sequence of PEDV strain CH/S was deposited in GenBank under accession no. JN547228.

This work was supported by grants from the National Key Laboratory of Veterinary Biotechnology (SKLVBP201028), National Natural Science Foundation of China (30901081), Basic Scientific Research Fund of Central Research Academies and Institutes (ZGKJ201012), and Technology Research and Development Program of Harbin (2009AA6AN029).

REFERENCES

1. Bournsnel, M. E. G., et al. 1987. Completion of the sequence of the genome of the coronavirus avian infectious bronchitis virus. *J. Gen. Virol.* **68**:57–77.
2. Brian, D. A., and R. S. Baric. 2005. Coronavirus genome structure and replication. *Curr. Top. Microbiol. Immunol.* **287**:1–30.
3. Chouljenko, V. N., X. Q. Lin, J. Storz, K. G. Kousoulas, and A. E. Gorbalenya. 2001. Comparison of genomic and predicted amino acid sequences of respiratory and enteric bovine coronaviruses isolated from the same animal with fatal shipping pneumonia. *J. Gen. Virol.* **82**:2927–2933.
4. Cui, X., et al. 1990. Studies on the detection of porcine epidemic diarrhea

* Corresponding author. Mailing address: Division of Swine Infectious Diseases, State Key Laboratory of Veterinary Biotechnology, Harbin Veterinary Research Institute, Chinese Academy of Agricultural Sciences, Harbin 150001, People's Republic of China. Phone: 86 18946066048. Fax: 86 45182733132. E-mail: fl@hvri.ac.cn.

- virus by immunofluorescent techniques. *Chin. J. Prev. Vet. Med.* **5**:20–24. (In Chinese).
5. **Das Sarma, J., L. Fu, S. T. Hingley, and E. Lavi.** 2001. Mouse hepatitis virus type-2 infection in mice: an experimental model system of acute meningitis and hepatitis. *Exp. Mol. Pathol.* **71**:1–12.
 6. **Duarte, M., J. Gelfi, P. Lambert, D. Rasschaert, and H. Laude.** 1993. Genome organization of porcine epidemic diarrhoea virus. *Adv. Exp. Med. Biol.* **342**:55–60.
 7. **Enjuanes, L., et al.** 2001. Coronavirus derived expression systems. *J. Biotechnol.* **88**:183–204.
 8. **Herold, J., T. Raabe, B. Schelle-Prinz, and S. G. Siddell.** 1993. Nucleotide sequence of the human coronavirus 229E RNA polymerase locus. *Virology* **195**:680–691.
 9. **Kocherhans, R., A. Bridgen, M. Ackermann, and K. Tobler.** 2001. Completion of the porcine epidemic diarrhoea coronavirus (PEDV) genome sequence. *Virus Genes* **23**:137–144.
 10. **Lai, M. M. C., and D. Cavanagh.** 1997. The molecular biology of coronaviruses. *Adv. Virus. Res.* **48**:1–100.
 11. **Leparc-Goffart, L., et al.** 1997. Altered pathogenesis of a mutant of the murine coronavirus MHV-A59 is associated with a Q159L amino acid substitution in the spike protein. *Virology* **239**:1–10.
 12. **Penzes, Z., et al.** 2001. Complete genome sequence of transmissible gastroenteritis coronavirus PUR46-MAD clone and evolution of the Purdue virus cluster. *Virus Genes* **23**:105–118.
 13. **Rota, P. A., et al.** 2003. Characterization of a novel coronavirus associated with severe acute respiratory syndrome. *Science* **300**:1394–1399.
 14. **Sawicki, S. G., and D. L. Sawicki.** 1998. A new model for coronavirus transcription. *Adv. Exp. Med. Biol.* **440**:215–218.
 15. **Thiel, V., J. Herold, B. Schelle, and S. G. Siddell.** 2001. Infectious RNA transcribed in vitro from a cDNA copy of the human coronavirus genome cloned in vaccinia virus. *J. Gen. Virol.* **82**:1273–1281.
 16. **Woo, P. C., et al.** 2005. Characterization and complete genome sequence of a novel coronavirus, coronavirus HKU1, from patients with pneumonia. *J. Virol.* **79**:884–895.
 17. **Yoo, D., and Y. Pei.** 2001. Full-length genomic sequence of bovine coronavirus (31 kb): completion of the open reading frame 1a/1b sequences. *Adv. Exp. Med. Biol.* **494**:73–76.