

Molecular characterization of a Chinese vaccine strain of transmissible gastroenteritis virus: mutations that may contribute to attenuation

Chengbao Wang · Jianfei Chen · Hongyan Shi · Huaji Qiu ·
Fei Xue · Changming Liu · Yuanmao Zhu · Shengwang Liu ·
Fernando Almazán · Luis Enjuanes · Li Feng

Received: 23 November 2009 / Accepted: 24 February 2010 / Published online: 13 March 2010
© Springer Science+Business Media, LLC 2010

Abstract Transmissible gastroenteritis virus (TGEV), the etiological agent of transmissible gastroenteritis (TGE), is the major cause of viral enteritis and fetal diarrhea in swine neonates, resulting in significant economic losses to the swine industry. The Chinese vaccine strain H165 of TGEV was derived from a virulent field strain H16 by serial passage in vitro. Strain H165 has been proven to be safe in piglets and pregnant sows and displays efficacy against TGEV infection. In this study, we report the complete genome sequences of strains H165 and H16, obtained by sequencing several overlapping fragments amplified from viral RNA and our findings from sequence and phylogenetic analyses. The genomes were 28,569 nucleotides in length, including the poly (A) tail. No deletions or insertions were detected in the H16 genome sequence after continuous passage in vitro; however, we found 27 nucleotide mutations in strain H165 compared with strain H16, resulting in 16 amino acid changes distributed among the genes 1, S, 3, and sM. An A to G nucleotide mutation was found in the intergenic region between the 3a and 3b genes. Furthermore, six unique nucleotides identified in the genome sequence of H165 could be used as makers to differentiate the H165 vaccine strain from wild-type TGEV

strains. Our findings from phylogenetic analysis may enhance our understanding of the evolution of TGEV, as well as the other coronaviruses.

Keywords Transmissible gastroenteritis virus · Complete genomic sequence · Phylogenetic analysis · Mutations · Attenuation

Introduction

Members of the *Coronaviridae*, within the order *Nidovirales*, contain positive-stranded RNA genomes that range from 27 kb to 31 kb in size [1]. Transmissible gastroenteritis virus (TGEV), together with porcine epidemic diarrhea virus (PEDV), feline infectious peritonitis virus (FIPV), and the human coronaviruses HCoV-229E and HKU1, belongs to the genus *Alphacoronaviruses* (<http://www.ictvonline.org/>). TGEV is an enveloped virus that contains a positive-sense, single-stranded RNA genome of about 28.5 kb in length. Its genome includes nine open reading frames (ORFs) that encode four structural proteins (spike [S], envelope [E], membrane [M], and nucleoprotein [N]), and five nonstructural proteins (the replicases 1a and 1ab, proteins 3a and 3b, and protein 7) in the order 5'-replicase (1a/1b)-S-3a-3b-E-M-N-7-3' [2]. TGEV was initially identified as the etiological agent of transmissible gastroenteritis (TGE) in swine in 1946 in the United States [3, 4]. In neonates, TGEV infects the epithelial cells of the small intestines, leading to potentially fatal gastroenteritis. The virus can also lead to infection in the upper respiratory tract and less often, in the lungs [5]. In adults, TGEV causes mild disease. In swine, it is the major cause of viral enteritis and fetal diarrhea in neonates, resulting in significant economic losses [6].

C. Wang · J. Chen · H. Shi · H. Qiu · F. Xue · C. Liu ·
Y. Zhu · S. Liu · L. Feng (✉)

Division of Swine Infectious Diseases, National Key Laboratory of Veterinary Biotechnology, Harbin Veterinary Research Institute of the Chinese Academy of Agricultural Sciences, No. 427 Maduan Street, Nangang District, Harbin 150001, People's Republic of China
e-mail: fl@hvri.ac.cn

F. Almazán · L. Enjuanes
Department of Molecular and Cell Biology, Centro Nacional de Biotecnología, Campus Universidad Autónoma, Cantoblanco (Madrid), Spain

TGEV was reported in many swine-producing countries between the late 1980s and the 1990s [7–10]. TGEV strains of varying virulence have been isolated and characterized worldwide [11–14]. Some strains have been used to develop modified live vaccines with limited success. In China, a TGE outbreak was first reported in the 1970s. Since then it has been prevalent in many provinces and has become one of the most important viral diarrhea diseases in China. The Chinese TGEV vaccine strain H165 was derived from a virulent field strain H16 by 165 passages in PK15 cells. H165 virus was proven to be safe in piglets and pregnant sows and efficacious against TGEV infection [15]. Vaccines based on the H165 strain are currently commercially available to prevent and control TGEV infections in China. To genetically characterize the Chinese vaccine strain H165 and its parental strain H16 at the molecular level, we determined the complete nucleotide (nt) sequences of the H165 and H16 viruses. To determine the molecular basis for attenuation of the Chinese TGEV vaccine strain, the deduced amino acid (aa) sequences of the structural and non-structural proteins of the H165 and H16 viruses were compared with those of TGEV reference strains, as well as with that of the porcine respiratory coronavirus (PRCV) strain PRCV-ISU-1.

Materials and methods

Cells and viruses

The pig kidney cell line PK-15 (ATCC, CCL-33) was regularly maintained in Eagle's Minimum Essential Medium (MEM) supplemented with 10% fetal bovine serum, penicillin (100 U/ml), streptomycin (100 µg/ml), and 1 mM Na-pyruvate. The Chinese vaccine strain H165 and its parental strain H16 were obtained by our laboratory [15].

Extraction of genomic TGEV RNA

Total RNA was extracted from PK-15 cells infected with H165 virus and from purified H16 virus, using the QIAamp Viral RNA Mini kit (QIAGEN). The resulting RNA pellet was resuspended in 10 µl DNase-free, RNase-free double-distilled water and was used as a template for RT-PCR.

RT-PCR and DNA sequencing

The complete genomes of the H165 virus and its parental strain H16 were amplified by RT-PCR and sequenced. The reverse transcriptase step was performed with Superscript® III reverse transcriptase (Invitrogen) using a combined random priming or oligo (dT) priming strategy and the

resulting cDNA was amplified by PCR using Primer-STAR® HS DNA Polymerase (TaKaRa). Primers were designed based on conserved regions of the TGEV strains (their sequences are available on request). For amplification of the TGEV subclones, reactions were carried out in a total volume of 50 µl containing 10 µl of 5 × PCR Buffer, 4 µl of 2.5 mmol/l dNTPs, 0.5 µl of Taq polymerase (TaKaRa), 1 µl of each specific primer (20 µM), about 2.5 ng of template, and sterile deionized water. The PCR protocols was carried out for 30 cycles at 98°C for 10 s, 55°C for 15 s, and 68°C for several minutes depending on the size of the amplicons. The 5' and 3' ends of the viral genome were confirmed by rapid amplification of cDNA ends using the SMART™ RACE cDNA Amplification Kit (Clontech). The PCR amplicons were purified using the AxyPrep™ Gel Extraction Kit (Axygen) and cloned into pGEM-T Easy Vector (Promega). Three to five independent clones of each TGEV amplicon were isolated and sequenced using the M13 universal primers.

Sequence analysis

Sequences were assembled and edited to obtain the complete genome sequences of the viral strains H165 and H16. The complete nt sequences of the H165 and H16 strains were submitted to the GenBank sequence database and assigned accession nos. EU074218 and FJ755618. ORFs were predicted using the Gene Runner program, version 3.00 (<http://www.generunner.com>). The nt sequences of both genomes and the deduced aa sequences of the ORFs were compared to several TGEV strains and the PRCV strain PRCV-ISU-1. Multiple alignments and phylogenetic trees were constructed using the MegAlign program from the DNASTar software package (Version 7.1.0, DNASTAR Inc., USA). The sequences used for analysis were the available strains on 4 February 2010 in the GenBank. Prediction of signal peptides and their cleavage sites was performed using SignalP software [16]. Prediction of transmembrane domains was performed using TMHMM software [17]. Potential *N*-glycosylation sites were predicted using ScanProsite software [18]. The secondary structures of the 5' and 3'-untranslated regions (UTRs) of the H165 and H16 viruses were predicted using RNAdraw 1.1 software [19].

Results

Genomic sequences of the strains H165 and H16

Full-length genome sequences were generated by several overlapping cDNA fragments to encompass the entire RNA genomes of strains H165 and H16. The findings indicated

Table 1 Sequence comparisons of the Chinese vaccine strain H165 and its parental strain H16 with other TGEV strains and the PRCV-ISU-1 strain

Strains	H165	H16	Purdue P115	Virulent Purdue	PUR46-MAD	Miller M60	Miller M6	TS	SC-Y	PRCV-ISU-1
H165	***	99.9	98.5	98.7	98.6	99.8	99.8	99.7	98.3	97.9
H16	0.1	***	98.6	98.7	98.6	99.8	99.8	99.7	98.4	97.9
Purdue P115	1.5	1.4	***	99.8	99.9	98.7	98.7	98.6	99.6	97.5
Virulent Purdue	1.3	1.3	0.2	***	99.8	98.8	98.9	98.8	99.5	97.6
PUR46-MAD	1.4	1.4	0.1	0.2	***	99.7	98.8	98.7	99.7	97.5
Miller M60	0.2	0.2	1.3	1.2	0.3	***	99.8	99.7	98.4	96.2
Miller M6	0.2	0.2	1.3	1.1	1.2	0.2	***	99.8	98.5	98
TS	0.3	0.3	1.4	1.2	1.3	0.3	0.2	***	98.4	97.9
SC-Y	1.7	1.6	0.4	0.5	0.3	1.6	1.5	1.6	***	97.2
PRCV-ISU-1	2.1	2.1	2.5	2.4	2.5	3.8	2	2.1	2.8	***

Sequence distance was determined with DNASTAR software using the ClustalW program. Percent similarities are indicated in the upper right portion of the table (above stars) and percent divergency in the lower left portion (below stars)

that the genome sequence was 28,569 nt in length, including a poly A tail of 30 nt. The genome encodes nine ORFs, characteristic of the genus Alphacoronavirus, and 5' and 3' untranslated sequences of 314 and 274 nt, respectively (which do not include the poly A tail). The predicted ORFs 1a (nt 315–12,368) and 1ab (nt 12,326–20,368) of the Chinese vaccine strain H165 contained 12,051 and 8,040 nt, encoding the non-structural proteins 1a and 1ab, respectively. ORF 1ab contained a 43 nt region that overlapped with ORF 1a and included a typical coronavirus “slippery site,” 5'-UUUAAAC-3' (nt 12,332–12,338). Based on evidence from other TGEV coronaviruses [20], this sequence causes a –1 frameshift during the translation of ORF1a, which results in a portion of the proteins (protein 1ab) avoiding translation termination and containing an additional 2,680 aa. The structural proteins S, E, M and N were found to be encoded by ORFs S (nt 20,365–24,711), E (nt 25,815–26,063), M (nt 26,074–26,862) and N (nt 26,875–28,023), respectively. Furthermore, the three non-structural proteins 3a, 3b and 7, were identified to be encoded by ORFs 3a (nt 24,814–25,023), 3b (nt 25,094–25,828) and 7 (nt 28,029–28,265), respectively. Upstream of eight of the genes of the Chinese vaccine strain H165 and its parental strain H16, there was a repeated intergenic sequence, 5'-CUAAAC-3', called a transcription regulating sequence (TRS) [21].

Sequence analysis showed that strains H165 and H16 were more closely related to strains Miller M6 and TS than the other reference strains (Table 1). Strains H165 and H16 shared between 98.3 and 99.9% nt identity to the other TGEV reference strains, being slightly less closely related to strains Purdue and SC-Y with around 98.5–98.7% and 98.3% nt identity, respectively (Table 1). They were also less closely related to strain PRCV-ISU-1 with around 97.9% of nt identity. Nearly all the ORFs of strains H165

and H16 were highly similar to those of strains Miller M6 and TS (98.4–100% nt and 95.2–100% aa identity), except for 88.9% and 52.9% nt identity with strain Miller M6 genes 3a and 3b, respectively, which are attributable to the fact that Miller M60 has a 531 nt in frame deletion in gene 3 [22]. In order to investigate the relationships among TGEV strains and PRCV-ISU-1, a phylogenetic analysis was performed on the genome sequences of strains H165 and H16, strain PRCV-ISU-1, and the available TGEV strains in the GenBank. As shown in Fig. 1, strains H165 and H16 were confirmed to be more closely related to the TGEV Miller strains than the Purdue strains or strain PRCV-ISU-1. To further explore the evolutionary relationship among strains, the S protein sequences were used to investigate genetic relatedness among TGEV strains and strain PRCV-ISU-1. Phylogenetic tree constructed using the available S protein sequences showed that the reference strains were grouped into five distinct clusters (Fig. 2). Strains H165 and H16 formed a cluster with strains Miller M60, HN2002, Miller M6, TS, TSX, 96-1933, TFI, FS772/70, and TO14. Taken together these results, it appears that strains H165 and H16 are more closely related to strains Miller M60, HN2002, Miller M6, and TS than to the other Chinese TGEV strains, even though they are of Chinese origin.

Characterization of 5' and 3'-UTR within the strains H165 and H16

Compared with its parental strain H16, there was no mutation, deletion, or insertion in the 5' and 3'-UTR. Both the 5' and 3' ends of the genome of the two strains contain short UTRs. The 5'-UTR comprises 314 nt and includes a potential short AUG-initiated ORF (nt 117–123), begins in a suboptimal Kozak context (GCCAUGG) for translation

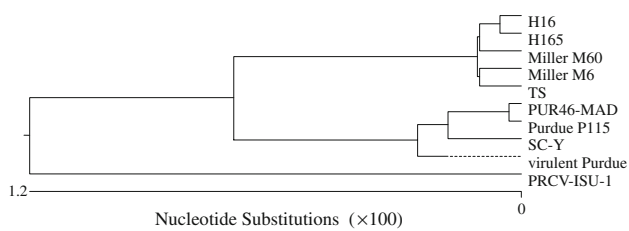


Fig. 1 Phylogenetic analysis of the entire genome sequences of the Chinese vaccine strain H165, its parental strain H16, several TGEV reference strains, and PRCV-ISU-1 strain. The phylogenetic tree was constructed using the ClustalW method with DNASTAR software

[23] and potentially encodes peptides of three aa (MKS). Analysis of the 5'-UTR showed a high level of secondary structure, with three simple and two complex stem-loop structures (data not shown), as predicted using the software RNAdraw 1.1. The 3'-UTR comprises 274 nt (28,266–28,539) and possess an octameric sequence of GGAA-GAGC beginning at nt 198–205 upstream from the poly (A) tail, and possess a poly (A) tail of 30 nt. A high level of secondary structure, with six simple and one complex stem-loop structures (data not shown) predicted using RNAdraw 1.1., which has been shown to be important in enterovirus replication [24].

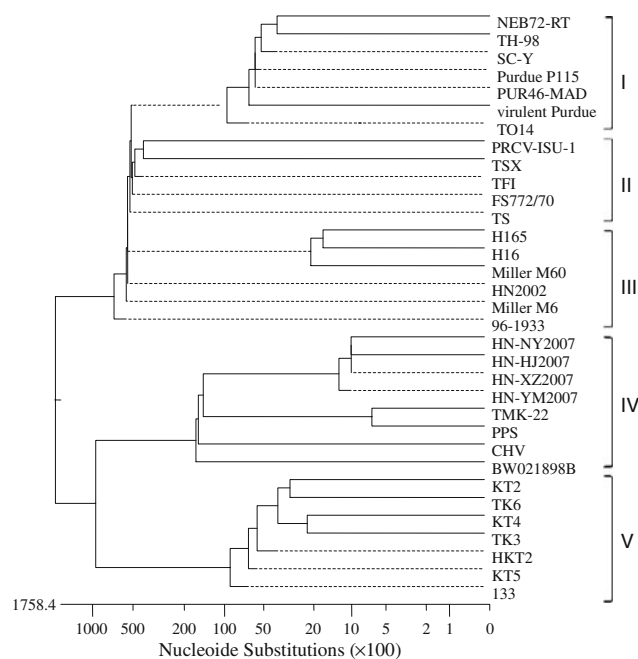


Fig. 2 Phylogenetic analysis of the S protein of the Chinese vaccine strain H165, its parental strain H16, other TGEV reference strains, and PRCV-ISU-1 strain. The phylogenetic tree was constructed using the Jotun Hein method with DNASTAR software

Nucleotide and amino acid mutations of the Chinese vaccine strain H165

As shown in Table 2, there were a total of 27 nt mutations identified in strain H165, resulting in a total of 16 aa mutations mainly located within proteins 1a, 1ab, S, 3a, 3b, and E. Moreover, a point mutation ($A_{25074}G_{ORF3}$) was found in the intergenic region between the genes 3a and 3b. In brief, there were 13, 8, 2, and 2 nt mutations in the genes 1, S, 3, and sM, resulting in 7, 5, 2, and 2 aa differences from its parental strain H16, respectively (Table 2). Furthermore, six nt mutations ($G_{6014}T_{ORF1a}$, $T_{12388}C_{ORF1b}$, $T_{21937}C_S$, $T_{21969}A_S$, $A_{26025}C_E$, and $C_{27507}T_N$) could be the markers used to differentiate the Chinese vaccine strain from other strains of TGEV in the GenBank. Comparisons with the other two pairs of attenuated and virulent TGEV strains (virulent Purdue and P115, Miller M6, and Miller M60) revealed no common mutations among them.

Bioinformatics analysis of the structural proteins of the strains H165 and H16

ScanProsite analysis of the proteins S, M, and N of strains H165 and H16 showed that there were thirty-three (seven NXS and twenty-six NXT), three (two NXS and one NXT), and four (three NXS and one NXT) potential *N*-linked glycosylation sites, respectively. No *N*-linked glycosylation site was identified in the E proteins of strains H165 and H16. SignalP analysis of the proteins S of strains H165 and H16 showed that there was a potential cleavage site between aa 16 and 17, revealing a signal peptide probability of 0.995. It was predicted to contain a signal anchor (probability 0.998) at position 30 with a cleavage site between aa 37 and 38 in E proteins of strains H165 and H16. An *N*-terminal signal peptide was also identified in the M proteins of strains H165 and H16 with a potential cleavage site between aa 16 and 17, revealing a signal peptide probability of 0.931. The S proteins of strains H165 and H16 were the typical type I membrane protein with the *N*-terminal 1,411 aa residues exposed on the outside of the cell surface or virus particle and a transmembrane domain near the C-terminus (aa 1,412–1,434) followed by a cytoplasmic tail rich in cysteine residues (aa 1,435–1,472). There was also a stretch of highly hydrophobic residues at position 1,394–1,412, with a maximum value of 3.467 at position 1,398 and a minimum value of -2.967 at position 953–954. One transmembrane domain was predicted at position 15–37 of proteins E of strains H165 and H16 by TMHMM analysis, which predicted the *N*-terminus of the E protein to be external to the cell surface or viral envelope. Three transmembrane domains were predicted to be present at positions 46–68, 78–100, and 112–134 of proteins M of strains H165 and H16 by TMHMM analysis,

Table 2 Nucleotide and amino acid differences unique to strain H165 or to both attenuated and virulent viruses (strains H165 and H16) compared with all other TGEV strains or strain PRCV-ISU-1 listed in the GenBank database

Gene	Nucleotide positions	Amino acid positions	Nucleotide differences			Amino acid differences			Number in GenBank
			H16	H165	GenBank viruses	H16	H165	GenBank viruses	
5'-UTR	153		<u>G</u>	<u>G</u>	T				8
ORF1a	759	149	<u>T</u>	<u>T</u>	C	<u>Y</u>	<u>Y</u>	H	8
	1199	295	<u>C</u>	T	T	N	N	N	
	1200	296	<u>A</u>	T	T	<u>S</u>	C	C	
	2272	653	<u>G</u>	A	A	<u>R</u>	K	K	
	2273	653	<u>A</u>	G	G	<u>R</u>	K	K	
	2548	745	<u>C</u>	T	T	<u>A</u>	V	V	
	4948	1545	<u>G</u>	<u>G</u>	A	<u>G</u>	<u>G</u>	D	
	5757	1815	<u>G</u>	<u>G</u>	A	<u>D</u>	<u>D</u>	N	
	6014	1900	<u>G</u>	<u>T</u>	G	L	<u>F</u>	L	
	8153	2613	<u>T</u>	<u>T</u>	C	C	C	C	
	9125	2937	<u>C</u>	<u>C</u>	T	S	S	S	
	9608	3098	<u>G</u>	<u>G</u>	A	T	T	T	
	11914	3867	C	<u>G</u>	C	P	<u>R</u>	P	
	11915	3867	C	<u>A</u>	C	P	<u>R</u>	P	
ORF1b	12388	21	T	<u>C</u>	T	D	D	D	8
	15062	913	<u>A</u>	G	G	<u>T</u>	A	A/V	
	16069	1248	C	<u>A</u>	C	N	<u>K</u>	N	
	16126	1267	C	T	C/T	Y	Y	Y	
	16753	1476	<u>G</u>	A	A	T	T	T	
	17105	1594	<u>T</u>	<u>T</u>	C	L	L	L	
	18502	2057	<u>G</u>	<u>G</u>	A	T	T	T	
	19012	2229	<u>T</u>	<u>T</u>	G	<u>H</u>	<u>H</u>	Q	
	19825	2500	<u>G</u>	<u>G</u>	A	E	E	E	
S	20506	48	C	T	A/C/T	P	S	I/P/S	30
	21616	418	<u>G</u>	<u>G</u>	A	<u>V</u>	<u>V</u>	I/S	
	21905	514	T	<u>C</u>	T	L	<u>P</u>	F/L	
	21937	525	T	<u>C</u>	T	F	<u>P</u>	F	
	21938	525	T	<u>C</u>	T	F	<u>P</u>	F	
	21969	535	T	<u>A</u>	T	I	I	I	
	22048	562	<u>G</u>	<u>G</u>	A/C	<u>D</u>	<u>D</u>	H/K/N	
	22208	615	<u>C</u>	<u>C</u>	A	<u>A</u>	<u>A</u>	D	
	22376	671	A	<u>G</u>	A	D	<u>G</u>	D	
	22583	740	<u>T</u>	G	G	<u>I</u>	S	S	
	23686	1108	<u>G</u>	<u>G</u>	A	<u>A</u>	<u>A</u>	I/T	
	24022	1220	<u>T</u>	<u>T</u>	G	<u>L</u>	<u>L</u>	V	
	24501	1379	C	T	C/T	L	L	L	
	ORF3a	24976	55	G	T	G/T	A	S	A/S
25074			A	G	A/G				
ORF3b	25523	144	G	A	A/G	A	T	A/T/deleted	18
sM	25881	23	T	<u>G</u>	T	F	<u>V</u>	F	9
	25986	58	<u>T</u>	<u>T</u>	G	<u>L</u>	<u>L</u>	A/V	
	26025	71	A	<u>C</u>	A	I	<u>L</u>	I	
M	26209	46	<u>T</u>	<u>T</u>	G	<u>Y</u>	<u>Y</u>	D	12

Table 2 continued

Gene	Nucleotide positions	Amino acid positions	Nucleotide differences			Amino acid differences			Number in GenBank
			H16	H165	GenBank viruses	H16	H165	GenBank viruses	
N	27507	211	C	<u>T</u>	C	R	R	R	23

Underlined nucleotides and amino acids are unique to strain H165 or strain H16 or to both strains H165 and H16 (as indicated by comparison with all strains listed in the GenBank database)

with a stretch of eight aa (SEESFNPE) directly adjacent to the third hydrophobic domain.

Discussion

Since TGEV was identified in 1946, researchers have tried to attenuate the virulence of field TGEV in order to develop an appropriated attenuated vaccine. Some TGEV isolates gradually lose their virulence and developed to be commercially attenuated TGEV vaccines after continuous passages in cell culture. Strain H165, derived from a virulent strain H16 by 165 passages in PK-15 cells, has been proven to be safe in piglets and pregnant sows and displays efficacy against TGEV infection [15]. In this study, sequence analysis showed that passaging of strain H16 in PK-15 cells resulted in 27 nt mutations that caused 16 aa mutations in strain H165 relative to strain H16. The mutation rate observed was lower than that previously attained in studies by Zhang [14] where twenty and thirty-two aa mutations were found in strains Miller M60 and Purdue P115, respectively, when compared to their virulent counterparts. Up to now, there were only seven TGEV strains and one PRCV strain PRCV-ISU-1 had been fully sequenced, though partial sequences of TGEV strains were available in the GenBank. Moreover, only two virulent and attenuated TGEV pairs were reported [14]. Whole genome sequences of strains H165 and H16 will help us to understand the genetic basis of TGEV coronavirus attenuation and enhance the geographic differentiation information among TGEV strains.

The 5' and 3'-UTR regions of TGEV genome are crucial for viral replication and transcription [25]. Furthermore, the 5' and 3'-UTR regions had been demonstrated to be targeted for the attenuation of some viruses [26]. In this study, no mutation, deletion, or insertion was detected in the 5' and 3'-UTR regions of strains H165 and H16. Moreover, the single nt difference from the two pairs of virulent/attenuated TGEV strains found in the 5'-UTR of strains H165 and H16 may not abolish the role in viral replication and transcription. The second conserved region involves the gene 1a, a frameshift region that forms a typical coronavirus “slippery” sequence (5'-UUUAAAC-3') and a pseudo-knot structure, which is proposed to be critical for

the transcription of gene 1ab that involves ribosomal frame shifting [20]. Though a nt mutation was found in the TRS of the gene 3b for virulent Purdue and attenuated P115 strains [14], no mutation was found in the conserved region TRS, 5'-CUAAAC-3', located upstream of each gene serve as signals for the transcription of the sgRNAs.

In this study, of the fifteen nt mutations identified in ORFs, 1a, 1ab, 3a, and 3b, eleven led to aa mutations in nonstructural proteins of strain H165. Seven of eleven mutations were present in proteins 1a and 1ab, with two aa mutations present in proteins 3a and 3b, respectively. The proteins 1a and 1ab are expressed by ribosomal frame-shifting and polyprotein cleavage. A comparative analysis of replicative polyproteins of coronaviruses and arteriviruses identified the most variable regions in the N-terminal half of protein 1a [27–29], and four aa mutations (S₂₉₆C_{ORF1a}, R₆₅₃K_{ORF1a}, A₇₄₅V_{ORF1a}, and L₁₉₀₀F_{ORF1a}) in protein 1a were found to occur in the N-terminal half of protein 1a, while 1 aa mutation (P₃₈₆₇R_{ORF1a}) occurred in the C-terminal of protein 1a. A point mutation within the replicase gene 1a had been demonstrated to affect coronavirus genome versus minigenome replication differentially [30], although, attempts to delete the replicase non-conserved domains to determine whether they are essential have not yet been made. Deletion of genes 3a and 3b reduced TGEV virulence very little [31], further studies will be needed to ascertain the roles of the two aa mutations (A₅₅S_{ORF3a} and A₁₄₄T_{ORF3b}) in proteins 3a and 3b, as well as one nt mutation (A₂₅₀₇₄G_{ORF3}) identified in the intergenic region between the genes 3a and 3b.

No aa mutations were present in proteins M and N, but two aa mutations occurred in protein E and five aa mutations occurred in protein S, which had been demonstrated to be an antigen for a neutralizing antibody for coronaviruses and play vital roles in viral entry, cell-to-cell spread, and determining tissue tropism [32–35]. All of the five aa mutations (P₄₈S_S, L₅₁₄P_S, F₅₂₅P_S, D₆₇₁G_S, and I₇₄₀S_S) in protein S had also been found to occur in the N-terminal 1,411 aa residues, which had been predicted to expose on the outside of the cell surface or virus particle. Antigenic site A/B has been mapped from aa 506–706 of protein S [36], previously studies showed that there were two aa mutations for M60 within this region as compared to M6 and four aa mutations for Purdue P115 as compared to

virulent Purdue [14]. In this study, there were three aa mutations for H165 as compared to H16 in the corresponding region. It has been implied that protein S may be related to the virulence of TGEV [35, 37]. However, whether these five aa mutations in protein S had influenced the virulence of H16 will require additional work by using a reverse genetics system and the characterization of these mutations in animal experiments. There are also one aa mutation (F₂₃V_E) to be found in position 15–37 of protein E, which showed that the N-terminus of protein E to be external to the cell surface or viral envelope.

In summary, most of the aa mutations identified were located in the functional regions of the TGEV genome, such as in the most variable regions of the protein 1a of the TGEV (S₂₉₆C_{ORF1a}, R₆₅₃K_{ORF1a}, A₇₄₅V_{ORF1a}, and L₁₉₀₀F_{ORF1a}), at the N-terminal (P₄₈S_S, L₅₁₄P_S, F₅₂₅P_S, D₆₇₁G_S, and I₇₄₀S_S) of protein S and its antigenic site (L₅₁₄P_S, F₅₂₅P_S, and D₆₇₁G_S), as well as in the putative transmembrane domain (F₂₃V_E), all of which are hypothesized to have the potential to affect virus replication, signal transport, and antibody neutralization processes associated with TGEV. Future studies could involve generating specific mutants via reverse genetics and characterizing these mutants in animal experiments. In addition, the findings from our study also revealed that six unique nt mutations (G₆₀₁₄T_{ORF1a}, T₁₂₃₈₈C_{ORF1b}, T₂₁₉₃₇C_S, T₂₁₉₆₉A_S, A₂₆₀₂₅C_E, and C₂₇₅₀₇T_N) in the genome sequence of Chinese vaccine strain H165 that could be the markers used to differentiate strain 165 from the other strains of TGEV. Furthermore, a rapid differentiation method to differentiate H165 from wild-type viruses of TGEV had been established by restriction fragment length polymorphism of the N gene based on the nt mutation (C₂₇₅₀₇T_N) (data not shown). Whole genome sequences of strains H165 and H16 may enhance our understanding of the evolution of TGEV coronavirus, as well as the other coronaviruses.

Acknowledgments This study was supported by grants from the National Key Laboratory of Veterinary Biotechnology, Harbin Veterinary Research Institute, Chinese Academy of Agricultural Sciences, Harbin, China.

References

1. D. Cavanagh, Arch. Virol. **142**, 629–633 (1997)
2. L. Enjuanes, A.E. Gorbalenya, R.J. de Groot, J.A. Cowley, J. Ziebuhr, E.J. Snijder, in *Encyclopedia of Virology*, 3rd edn, ed. by B.W.J. Mahy, M. Van Regenmortel, P. Walker, D. Majumder-Russell (Elsevier Ltd., Oxford, UK, 2008), pp. 419–430
3. L.P. Doyle, L.M. Hutchings, J. Am. Vet. Med. Assoc. **108**, 257–259 (1946)
4. L.P. Doyle, North Am. Vet. **32**, 477–478 (1951)
5. L.J. Saif, Rev. Sci. Technol. **23**, 643–660 (2004)
6. L. Enjuanes, C. Smerdou, J. Castilla, I.M. Anton, J.M. Torres, I. Sola, J. Golvano, J.M. Sanchez, B. Pintado, Adv. Exp. Med. Biol. **380**, 197–211 (1995)
7. L.J. Kemeny, R.D. Woods, Am. J. Vet. Res. **38**, 307–310 (1977)
8. E.N. Wood, G.C. Pritchard, E.A. Gibson, Vet. Rec. **108**, 41 (1981)
9. R.D. Woods, R.D. Wesley, Am. J. Vet. Res. **47**, 1239–1242 (1986)
10. G.C. Pritchard, Vet. Rec. **120**, 226–230 (1987)
11. C.M. Chen, D. Cavanagh, P. Britton, Virus Res. **38**, 83–89 (1995)
12. S.J. Kim, J.H. Han, H.M. Kwon, Vet. Microbiol. **94**, 195–206 (2003)
13. J.H. Park, J.H. Han, H.M. Kwon, Virus Genes **36**, 71–78 (2008)
14. X.S. Zhang, M. Hasoksuz, D. Spiro, R. Halpin, S.L. Wang, S. Stollar, D. Janies, N. Hadya, Y.X. Tang, E. Ghedin, L.J. Saif, Virology **358**, 424–435 (2007)
15. S.Q. Ma, Y.C. Wang, F.X. Wei, Q. Dong, W.T. Yu, C.M. Liu, W.K. Wang, Chin. J. Prev. Vet. Med. **23**, 4–10 (1985)
16. H. Nielsen, J. Engelbrecht, S. Brunak, G. von Heijne, Protein Eng. **10**, 1–6 (1997)
17. E.L. Sonnhammer, G. von Heijne, A. Krogh, Proc. Int. Conf. Intell. Syst. Mol. Biol. **6**, 175–182 (1998)
18. A. Gattiker, E. Gasteiger, A. Bairoch, Appl. Bioinform. **1**, 107–108 (2002)
19. O. Matzura, A. Wennborg, Comput. Appl. Biosci. **12**, 247–249 (1996)
20. I. Brierley, J. Gen. Virol. **76**, 1885–1892 (1995)
21. S. Alonso, A. Izeta, I. Sola, L. Enjuanes, J. Virol. **76**, 1293–1308 (2002)
22. H.M. Kwon, L.J. Saif, D.J. Jackwood, J. Vet. Med. Sci. **60**, 589–597 (1998)
23. M. Kozak, J. Biol. Chem. **266**, 19867–19870 (1991)
24. M.H. Mirzomoni, P.J. Hughes, G. Stanway, J. Virol. **71**, 2363–2370 (1997)
25. D.A. Brian, R.S. Baric, Curr. Top. Microbiol. Immunol. **287**, 1–30 (2005)
26. A.A.F. de Vries, E.D. Chrinside, P.J. Bredenbeek, L.A. Gravestein, M.C. Horzinek, W.J.M. Spaan, Nucleic Acids Res. **18**, 3241 (1990)
27. P.J. Bonilla, A.E. Gorbalenya, S.R. Weiss, Virology **198**, 736–740 (1994)
28. H.J. Lee, C.K. Shieh, A.E. Gorbalenya, E.V. Koonin, N. La Monica, J. Tuler, A. Bagdzyahdzhyan, M.M.C. Lai, Virology **180**, 567–582 (1991)
29. C.J. Nelsen, M.P. Murtaugh, K.S. Faaberg, J. Virol. **73**, 270–280 (1999)
30. A. Galán, L. Enjuanes, F. Almazán, J. Virol. **79**, 15016–15026 (2005)
31. I. Sola, S. Alonso, S. Zuniga, M. Balasch, J. Plana-Duran, L. Enjuanes, J. Virol. **77**, 4357–4369 (2003)
32. C.L. Yeager, R.A. Ashmun, R.K. Williams, C.B. Cardellicchio, L.H. Shapiro, A.T. Look, K.V. Holmes, Nature **357**, 420–422 (1992)
33. B.D. Zelus, J.H. Schickli, D.M. Blau, S.R. Weiss, K.V. Holmes, J. Virol. **77**, 830–840 (2003)
34. S. Matsuyama, F. Taguchi, J. Virol. **76**, 11819–11826 (2002)
35. C.M. Sanchez, A. Izeta, J.M. Sanchez-Morgado, S. Alonso, I. Sola, M. Balasch, J. Plana-Duran, L. Enjuanes, J. Virol. **73**, 7607–7618 (1999)
36. M. Godet, J. Grosclaude, B. Delmas, H. Laude, J. Virol. **68**, 8008–8016 (1994)
37. S. Bernard, H. Laude, J. Gen. Virol. **76**, 2235–2241 (1995)