



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



Complete genomic sequence of turkey coronavirus

M.H. Gomaa^a, J.R. Barta^a, D. Ojkc^b, D. Yoo^{c,*}

^a Department of Pathobiology, University of Guelph, Guelph, Ontario N1G 2W1, Canada

^b Animal Health Laboratory, University of Guelph, Guelph, Ontario N1G 2W1, Canada

^c Department of Pathobiology University of Illinois at Urbana-Champaign, 2001 South Lincoln Avenue, Urbana, IL 61802, USA

ARTICLE INFO

Article history:

Received 12 December 2007

Received in revised form 26 March 2008

Accepted 28 March 2008

Available online 12 May 2008

Keywords:

Turkey coronavirus

Complete genome

Organization

Coronavirus

Nidovirus

Enteritis

ABSTRACT

Turkey coronavirus (TCoV), one of the least characterized of all known coronaviruses, was isolated from an outbreak of acute enteritis in young turkeys in Ontario, Canada, and the full-length genomic sequence was determined. The full-length genome was 27,632 nucleotides plus the 3' poly(A) tail. Two open reading frames, ORFs 1a and 1b, resided in the first two thirds of the genome, and nine additional downstream ORFs were identified. A gene for hemagglutinin-esterase was absent in TCoV. The region between the membrane (M) and nucleocapsid (N) protein genes contained three potential small ORFs: ORF-X, a previously uncharacterized ORF with an associated putative TRS within the M gene (apparently shared among all group III coronaviruses), and previously described ORFs 5a and 5b. The TCoV genome is organized as follows: 5' UTR – replicase (ORFs 1a, 1b) – spike (S) protein – ORF3 (ORFs 3a, 3b) – small envelope (E or 3c) protein – membrane (M) protein – ORF5 (ORFs X, 5a, 5b) – nucleocapsid (N) protein – 3' UTR – poly(A). TCoV genome structure and sequence was most similar, but distinct from, avian infectious bronchitis virus (IBV). This is the first complete genome sequence for a TCoV and confirms that TCoV belongs to group III coronaviruses.

© 2008 Elsevier B.V. All rights reserved.

1. Introduction

Turkey coronavirus (TCoV) is associated with highly contagious gastroenteritis in young poults. First identified in 1951 (Peterson and Hymass, 1951), TCoV causes high morbidity, some mortality, and poor long-term growth of the affected birds, resulting in significant economic losses in the turkey industry. Outbreaks have been reported in different areas in the USA including Minnesota, North Carolina, and Indiana as well as in Quebec, Canada (Dea et al., 1986). Turkey coronavirus was subsequently determined to be the causative agent of blue comb disease during an investigation of an outbreak of acute, highly contagious enteritis in a flock of young turkeys (Nagaraja and Pomeroy, 1997). In addition to turkeys, TCoV can infect a variety of avian hosts including chickens, pheasants, sea gulls, and quail (Deshmukh and Pomeroy, 1974). Recently, TCoV was incriminated as one of the most important causative agents of poult enteritis and mortality syndrome (Barnes and Guy, 1997; Teixeira et al., 2007).

Coronaviruses are divided into three groups (I, II, and III) based on the genome structure and organization (Holmes and Lai, 1996).

The coronavirus genome is a single-strand positive-sense RNA with a 5' cap and poly (A) tail at the 3' end. Two large open reading frames (ORFs) occupy the 5' proximal two thirds of the genome and are involved in polyprotein processing, genome replication, and subgenomic RNA synthesis, and the 3' one third of the genome codes for structural proteins. ORF1 consists of two overlapping ORFs (ORF1a and ORF1b) that are translated into 1a and 1a/1b polyproteins by a ribosomal frame-shifting mechanism. ORF1a encodes 2 proteases; papain-like cysteine protease (PLP) and picornavirus 3C-like chymotrypsin protease (3CLP). Both proteases cleave the polyproteins into at least 16 cleavage products (Sawicki et al., 2007). An intergenic consensus sequence of about seven bases is found immediately upstream of each gene, which plays an important role for subgenomic RNA synthesis (Sawicki et al., 2007). Most coronaviruses code for four major structural proteins; spike (S) glycoprotein, membrane (M) protein, small envelope (E) protein, and nucleocapsid (N) proteins, in addition to hemagglutinin esterase (HE) glycoprotein in some group II coronaviruses.

The genome structure and organization are not known for any TCoV, and this lack of genetic information has made it difficult to develop effective detection and control measures for TCoV. The goal of the present study was to complete the full genomic sequence of TCoV and establish a better understanding of the virus at the molecular level.

* Corresponding author. Tel.: +1 217 244 9120.

E-mail address: dyoo@uiuc.edu (D. Yoo).

2. Materials and methods

2.1. Virus

Clinical specimens consisting of intestinal tracts were collected from turkey poulters during an outbreak of diarrhea in a turkey farm in Ontario, Canada. The tissue samples were processed by washing several times with phosphate buffer saline (PBS) and cut into small pieces prior to grinding using a sterile mortar and pestle. Coarse particles and tissue debris were removed by centrifugation at $3000 \times g$ for 30 min at 4°C . The supernatants were filtered through $0.22\text{-}\mu\text{m}$ filters (Millipore, Bedford, MA). Virus was isolated from the filtrates by inoculating embryonated turkey eggs. Allantoic fluids were collected 3 days post-inoculation then clarified by centrifugation $3000 \times g$ for 30 min at 4°C and the supernatant were filtered through $0.22\text{-}\mu\text{m}$ filters and stored at -80°C until use.

2.2. RNA extraction, RT-PCR and PCR

Complementary DNA (cDNA) was synthesized using $10.5\ \mu\text{l}$ of the first strand mixture (Invitrogen, Carlsbad, CA) containing $0.2\ \mu\text{g}$ of random hexamers and $2\ \mu\text{g}$ of total RNA isolated from the intestinal specimens. The mixture was incubated at 70°C for 10 min and then quick-chilled on ice for 5 min. RT master mix was composed of $4\ \mu\text{l}$ $5\times$ RT buffer (Invitrogen), $2\ \mu\text{l}$ 10 mM DTT, $2\ \mu\text{l}$ 10 mM dNTPs (Amersham, Piscataway, NJ), $1\ \mu\text{l}$ Superscript II reverse transcriptase (Invitrogen), and $0.5\ \mu\text{l}$ RNase inhibitors (Invitrogen). This RT master mix was added to $10.5\ \mu\text{l}$ of the first strand mixture and then incubated at 42°C for 2 h. The reaction was terminated by heating at 95°C for 10 min then chilling on ice for 5 min.

Fifty microliters PCR reactions included $2\ \mu\text{l}$ of cDNA and $48\ \mu\text{l}$ of the master mix composed of $5\ \mu\text{l}$ $10\times$ buffer (100 mM Tris-HCl, 500 mM KCl), $1.5\ \mu\text{l}$ of 15 mM MgCl_2 , $1.5\ \mu\text{l}$ of 10 mM dNTPs, 2 pmol of upstream and downstream primers, two units of Taq High Fidelity DNA polymerase (Invitrogen), two units of Amplitaq gold (Roche Molecular Systems, Inc., USA) plus $38\ \mu\text{l}$ nuclease free water. PCR was performed for 35 cycles as follow; 95°C for 1 min, 65°C for 1 min, 55°C for 30 s, 72°C for 3 min, followed by final extension at 72°C for 10 min. The PCR products were analyzed by 1% agarose gel electrophoreses and visualized by staining with ethidium bromide and UV illumination.

2.3. DNA cloning and sequence analysis

Thirteen overlapped PCR fragments spanning the entire viral genome were amplified using specific primer sets (Table 1). The PCR products were purified using the QIA quick PCR purification kit (Qiagen, Valencia, CA) and cloned in pGEM-T Easy (Promega, Madison, WI). Transformants were screened by restriction enzyme digestion and sequencing using primers specific for T7 and SP6 promoter. The sequences were analyzed using the Sequencher 4.5 sequence analyses program, and a single contiguous sequence comprising the entire TCoV genome was constructed. Prediction for ORFs was conducted using Vector NTI Advanced 10 (Invitrogen), and the sequences were analyzed using Lasergene DNA STAR (version 7, Lasergene Corp, Madison, WI). The pairwise nucleotide identity was determined using Vector NTI Advanced 10 and multiple sequence alignments were generated using Clustal-W (Thompson et al., 1997). Comparative analyses of TCoV with other coronaviruses were conducted using the Coronavirus Database (CoVDB, Huang et al., 2007).

2.4. The 5' end of the genome

cDNA clone representing the 5' end of the TCoV-MG10 genome had been synthesized according to the 5' RACE System for rapid

Table 1

Oligonucleotide primers used for TCoV genome amplification and their positions in the genome

Primer ^a	Sequence (5'–3')	Genomic positions
3' UTR-F1	AGTTTAAGTTAGTTTAGAGT	27462–27481
3' UTR-R3	GCTCTAACTCTATACTAGCC	27613–27632
NF-8	ATGGCAAGCGTAAGGCCAA	26054–26072
NR-1	CGGCACTGGCATCTTTACA	27485–27503
MR-10	TTACACATTAGCAAGCCACTGA	25095–25116
MF-7	AATACACCTCAACCTAAGTT	26186–26205
G3F	CAGTTTCGATTTACAGCACA	24214–24233
G3Fa	TGGCTGACTAGTTTGGGAAG	25675–25694
SR11	TTACTAATAAGACAACCACC	21941–21960
SF-1	GTGCAAGTATTAGTTAGAG	20339–20359
SF-18	TTATTTGCAGCAATGGGTAC	22022–22041
1b-F1	ATTGAAAGCTATGCCATTC	18220–18239
1bR-15	CTCACATTACATACAAGTGACAAG	20932–20415
1bR-7	CTCACATTACATACAAGTGACAAG	20392–20415
1bR-10	GAATGGCATAGCTTTCAAT	18220–18239
1bF-12	AACCACTCCTAGTAATTATGA	12560–12580
1bR-13	CTAAAACCGAGAAATATCTGCTAC	14661–14683
1aF-26	ACCTAATGCATTACACACTG	9360–9379
1aR-4	TCAAAGGCTCGCTTTACAACAT	12412–12433
1aF-29	GCAGGTTTTGTTATTATTTG	6100–6119
1aR-30	CAGTGTGTAATGCATTAGGT	9360–9379
1aF-8	GCTTGGTGTATGCGAGTTG	4465–4484
1aR-9	CAATAATAACAAAACCTGCG	6100–6119
5' UTR-R	GTTGCTACTGTCTATTGTATG	508–528

^a Primers are designated as forward (F) or reverse (R).

amplification of cDNA ends (Invitrogen). The antisense primer had been designed based on the available TCoV-MG10 sequence (5'-CGCCAGGTGTTATTTGTCA) then cDNA was synthesized as mentioned before. The cDNA was purified using Qiagen column purification kits. Tailing of the cDNA was done using dCTP and dTd. PCR had been done to amplify the dc-tailed cDNA with the abridged anchor primer together with the designed primer (5'-GTTGCTACTGTCTATTGTATG) according to the instructions of the kits.

2.5. The 3' end of the genome

The 3' end of TCoV genome had been done using the 3' race system for rapid amplification of cDNA ends (Invitrogen) according to the instruction of the kite. cDNA was synthesized using the adaptor primer then the cDNA was amplified with PCR using the TCoV-MG10 specific primer (5' CTATCGCCAGGGAAATGTCT 3') and the universal amplification primer according to the instruction of the kits. The obtained PCR products had been cloned and sequenced in both directions as mentioned earlier. The obtained sequence aligned with the rest of the genomic sequence and run through the poly(A) tail.

2.6. TCoV-MG10 phylogenetic relationships

Phylogenetic relationships among coronaviruses were investigated using the following complete genomes representing groups I, II, and III: BCoV-ENT (NC.003045), HCoV-NL63 (DQ445912.1), HCoV-OC43 (NC.005147.1), IBV-M41 (AY851295), IBV-p65 (DQ001339.1), IBV-Cal99 (AY514485), IBV-NC (NC.001451), SARSCoV-BJ202 (AY864806.1), HCoV-229E (NC.002645.1), MHV-JHM (NC.006852.1) and FIPV WSU-79/1146 (NC.007025.1). Whole genome sequences were aligned using ClustalW (Thompson et al., 1997) and subsequently optimized by eye using the Geneious software (Version 3.0.6, Biomatters Ltd, Auckland, New Zealand) (Kumar et al., 2004). Aligned nucleotide sequences were analyzed using PAUP (version 4.0), and maximum likelihood (HKY85 model with transition/transversion ratio of 2) and maximum parsimony

(transversion/transition ratio of 2) search criteria were used with branch and bound search strategies (Swofford, 1991). Bootstrap supports for the resulting trees were determined using 100 replicate heuristic tree searches in both parsimony and likelihood analyses using the same search criteria. For more detailed ingroup analysis of relationships among the group III coronaviruses, amino acid sequences for the spike glycoprotein (S), envelope protein (E, also known as 3c) and the nucleocapsid (N) protein were aligned using ClustalW and subsequently analyzed using maximum parsimony (branch and bound search algorithm) followed by bootstrap analysis using a heuristic search method. The following additional group III coronavirus sequences used for the comparative sequence analyses: TCoV-Gh S gene (AY342356), TCoV-GI S gene (AY342357), Quail coronavirus Italy/Elvia/2005 S1 gene (EF446155.1), TCoV-NC95 N gene (AF111997), TCoV-Minnesota N gene (AF111996), TCoV-Indiana N gene (AF111995). All ingroup translation products (S, E and N proteins) from the whole genomes of the same group I and II coronaviruses used above were included in these analyses. The group I and II coronaviruses were considered functional outgroups for determining relationships among the group III coronaviruses.

2.7. GenBank accession

The TCoV full genomic sequence described in this report was deposited in the GenBank database with accession number EU095850.

3. Results

3.1. Complete TCoV genomic sequence and organization

A TCoV was isolated from an outbreak in an Ontario turkey farm and designated TCoV-MG10. Subsequently, the full-length genomic

sequence of TCoV-MG10 was determined by sequencing of overlapping PCR fragments in both directions. The sequences were assembled into one contiguous sequence to represent the entire viral genome. The sequence of 27,632 nucleotides was obtained, plus the polyadenylation tail at the 3' end. The entire genome has a GC content of 38.3%. The TCoV genome contained two large slightly overlapping ORFs in the 5' two thirds of the genome and multiple additional ORFs in the 3' one third of the genome (Fig. 1). Both termini were flanked with untranslated regions (UTRs). The TCoV genome was similar overall in its coding capacity and genomic organization to those of other coronaviruses. Eleven ORFs were identified in the genome (Table 2). Gene 1 was 19,806 nucleotides comprising ORF1a and ORF1b, located between nucleotides 529 and 20,333. This gene contained motifs common in all coronaviruses including ribosomal frameshifting and slippery sequences, as ORF1b is translated in the -1 frame. The typical coronavirus structural genes encoding the spike (S), small envelope (E, also known as 3c), membrane (M), and nucleocapsid (N) proteins were identified following Gene 1 (Table 2, Fig. 1). The TCoV genome had polycistronic genes 3 and 5 interspersed between the S and E genes, and between M and N genes, respectively. In IBV, Gene 3 is believed to be tricistronic consisting of 3a, 3b and 3c, where 3c has been shown to encode the small envelope (E) protein. Gene 5 contained a coding potential for two products, 5a and 5b, of unknown function. A third ORF of 282 nucleotides was located between the M gene and Gene 5a. This gene, designated ORF-X, contained a coding potential for a protein of 94 amino acids that had no structural or sequence homology with any known protein. BLAST searches failed to identify any protein homologs, but identified highly similar nucleotide sequences from other TCoV isolates as well as from numerous IBV isolates. ORF-X had a relatively distant but highly conserved putative transcription regulatory sequence (TRS). In summary, the genome organization for TCoV was determined as follows: 5' UTR-Gene 1 (ORF1a, 1b)-

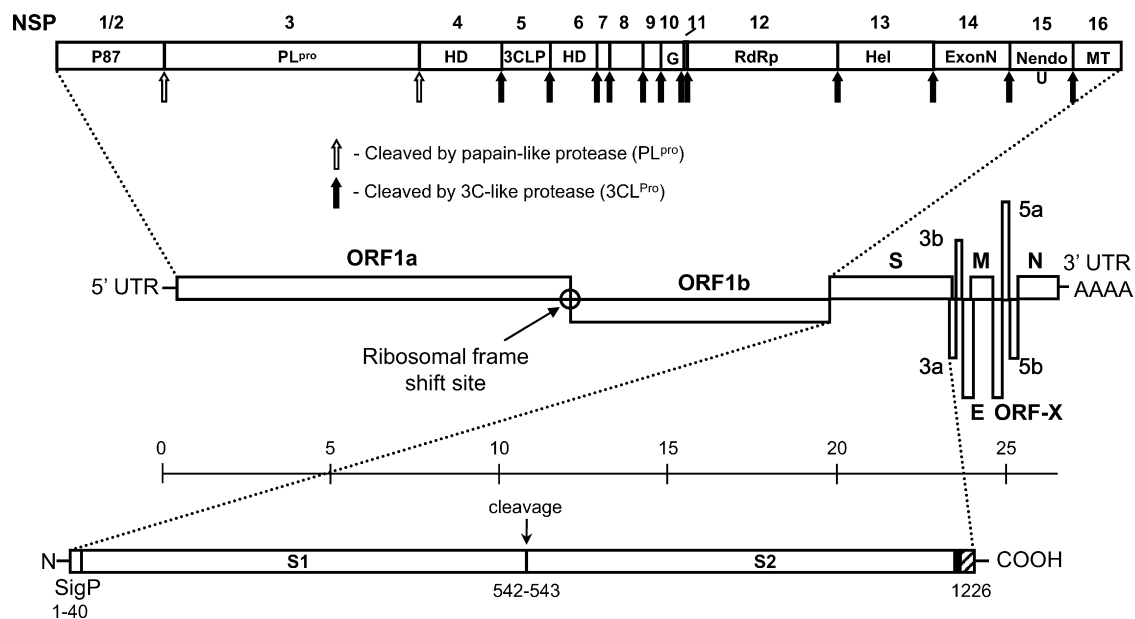


Fig. 1. Turkey coronavirus (TCov) MG10 genome organization. The full-length TCoV genome is 27,632 nucleotides excluding the polyadenylation tail at the 3' end. Center: diagrammatic representation of the genome organization shows the predicted genes and their relative sizes and positions along the TCoV genome. S, spike glycoprotein gene; 3a, 3b, and 3c (E), tricistronic gene 3; M, membrane protein gene; ORF-X, unique ORF conserved among group III coronaviruses; 5a and 5b, bicistronic gene 5; N, nucleocapsid protein gene; UTR, untranslated region. Scales indicate relative positions of the various genes within the genome in kilobases. Top: expanded representation of the two ORFs (ORF1a and ORF1b) comprising the polycistronic gene 1 and the likely cleavage products and cleavage sites after proteolytic processing of the 1a/1b polyprotein. Bottom: expanded representation of the S gene indicating the signal peptide (SigP), putative cleavage site (S1/S2), endodomain (clear bars), endodomain (hatched bar) and a short transmembrane region (solid bar).

Table 2
The genome organization and predicted viral proteins encoded by TCoV

Gene	Frame	Start	Stop	Size (nt)	Size (aa)
5' UTR		1	528	528	
ORF1a	+1	529	12381	11853	3951
ORF1ab		529	20333	19806	6601
S	+2	20360	24037	3678	1226
G3a	+3	23988	24158	171	57
G3b	+2	24161	24352	192	64
G3c (E)	+3	24336	24632	297	99
M	+2	24637	25305	669	223
ORF-X	+1	25309	25590	282	94
G5a	+1	25669	25863	195	65
G5b	+3	25863	26108	246	82
N	+2	26054	27280	1227	409
3' UTR		27281	27632	352	

Proteins	Genomic ^a positions (nt)	1a/1ab positions (aa)	C-end cleavage	Size (aa)	Identity ^b to IBV (%)	Possible ^c motif
NSP1	529–?	M1–?	?	?		
NSP2	?–2547	?–673G	AG/GK	~673	91.4	P87
NSP3	2548–7323	G674–2265G	AG/GI	1592	89.1	PLP
NSP4	7324–8865	G2266–2779Q	LQ/AG	514	87.4	
NSP5	8866–9786	A2780–3086Q	LQ/SS	307	92.8	3CLP
NSP6	9787–10665	S3087–3379Q	VQ/SK	293	92.5	HD
NSP7	10666–10914	S3380–3462Q	LQ/SV	83	96.4	
NSP8	10915–11544	S3463–3672Q	LQ/NN	210	96.2	
NSP9	11545–11877	N3673–3783Q	LQ/SK	111	99.1	
NSP10	11878–12312	S3784–3928Q	VQ/SV	145	95.2	GFL
NSP11	12313–12381	S3929–3951Q		23	91.3	
NSP12	12313–15131	S3929–4870Q	LQ/SC	941	97.2	RdRp
NSP13	15132–16931	S4871–5470Q	LQ/GT	600	97.3	HEL
NSP14	16932–18485	G5471–5988Q	LQ/SI	518	96.2	ExonN
NSP15	18486–19499	S5989–6326Q	LQ/SA	338	95.3	NendoU
NSP16	19500–20333	S6327–6604S		278	95.0	2'-O-MT

^a Not including stop codons.

^b IBV Beaudette (GenBank accession NC.001451).

^c PLP, papain-like protease; HD: hydrophobic domain; 3CLP, 3C-like proteinase; GFL: growth factor-like domain; RdRp: RNA-dependant RNA polymerase; HEL, helicase domain; ExonN, exoribonuclease; NendoU, nidoviral uridylylate-specific endoribonuclease; 2'-O-MT, 2'-O-ribose methyltransferase.

S-Gene 3-(ORFs 3a, 3b, E)-M-Gene 5 (ORFs X, 5b, 5c)-N-UTR-3' (Fig. 1).

3.2. 5' UTR

The 5' terminus of the TCoV genome was characterized by the presence of a 528 nucleotides long UTR with relatively higher GC content of 50.4% compared with the genome as a whole. The TCoV 5' UTR showed a high degree of sequence similarity with that of most IBV isolates such as IBV-NC (98% identity).

3.3. Gene 1—viral replicase

Following 5' UTR, Gene 1 of 19,806 nucleotides was located. Gene 1 included two slightly overlapping ORFs, 1a and 1b. ORF1a was 11,853 nucleotides in size enabling it to code for a protein of 3951 amino acids, and ORF1b was 7992 nucleotides in length for a coding potential of 2664 amino acids. Those two ORFs overlapped by 40 nucleotides. ORF1b did not contain a typical AUG translational initiation codon but instead started with GAA at position 12,342 (Fig. 2A). The heptanucleotide slippery sequence (UUUAAAC) was present which is conserved among all coronaviruses (Fig. 2B), and therefore the ribosomal frame-shifting mechanism seemed to be applicable to TCoV and ORF1b was believed to be translated by the ribosomal frame-shifting mechanism as a fusion with ORF1a to make the polyprotein of 1a/1b. For most coronaviruses, the polyprotein is processed into 16 cleavage products, while the IBV polyprotein is likely cleaved into 15 products (Ziebuhr et al., 2000). As with other coronaviruses, the TCoV polyprotein was also

believed to undergo proteolytic processing by viral proteinases. The TCoV replicase protein was similar to that of IBV in its processing patterns. Since the potential cleavage sites were conserved for both viruses (Table 2), the TCoV polyprotein was assumed to be processed in the similar manner. Two main proteases used by coronaviruses have been identified; PLP (papain-like proteinase) which produces 2 or 3 N-terminal products of the polyprotein and 3CLP (3-C like protease) which produces the central and the C-terminal region of the polyprotein by cleaving 11 sites (Ziebuhr et al., 2000). Similar to SARS-CoV, IBV encodes only one PLP, whereas other coronaviruses code for two PLPs. In the case of IBV, PLP equivalent to PLP2 of other coronaviruses cleaves the polyprotein at the QS dipeptide to produce the N-terminal 100 kD proteins, and TCoV PLP seemed to function in a similar fashion. TCoV 3CLP was mapped in NSP5. Based on the sequence comparisons between TCoV and IBV, the replicase cleavage products and their putative functions were predicted (Fig. 1, Table 2).

It has been suggested that NSP1 does not exist in IBV (Ziebuhr et al., 2007). Due to the high degree of similarity between TCoV and IBV in the N-terminal region of the polyprotein 1a/1b, TCoV also seemed not to contain NSP1. Thus, for TCoV, NSP2 through NSP11 would be produced from ORF1a, while NSP12 through NSP16 would be produced from ORF1b. The N-terminal most cleave product NSP2 was predicted to locate between nucleotide positions 529 and 2549. The sequence analysis for NSP1/2 reveals 44%, 55%, and 90% identity to BCoV, HCoV-229E, and IBV, respectively. PLP1 was suggested to have been lost during the IBV virus evolution (Ziebuhr et al., 2001), and our sequence data also supports that TCoV contained only PLP2. Therefore, PLP2 would be responsible for cleavage of the

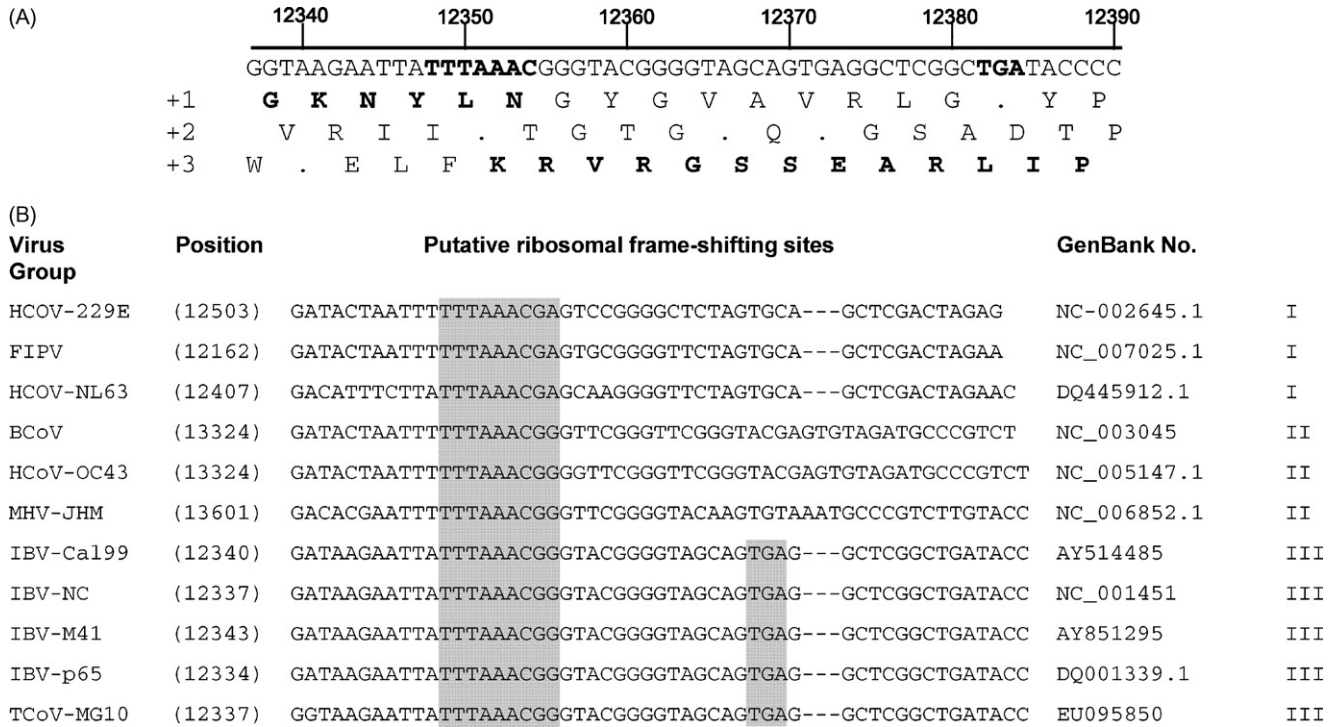


Fig. 2. TCoV putative ribosomal frameshifting regions. (A) Region of the TCoV containing the heptanucleotide slippery sequence conserved among coronaviruses and the ORF1a stop codon. The amino acids in the joining region to produce the 1a/1b fusion protein are indicated in bold. Numbers indicate the positions of these features in the TCoV genome. (B) Putative ribosomal frame-shifting sites in coronaviruses representing groups I, II, and III and ORF1a stop codons for the group III viruses are found in the shaded regions of this nucleotide alignment. The illustrated motif is highly conserved among Group III coronaviruses. HCoV, human coronavirus; FIPV, feline infectious peritonitis virus; BCoV, bovine coronavirus; MHV, murine hepatitis virus; IBV, infectious bronchitis virus; TCoV, turkey coronavirus.

N-terminal part of the polyprotein at two sites between NSP2/NSP3 (AG/GK) and NSP3/NSP4 (AG/GI). PLP2 was identified in NSP3. NSP3 was the largest subunit of the replicase cleavage products and was highly conserved among TCoV-MG10 and IBV-Baudette's strain, as both viruses share 89% identity on the nucleotide level. In contrast, 3CLP would cleave 1a/1b polyprotein at 11 sites and generate NSP5 through NSP16 (Table 2). The motif for 3CLP was found in NSP5. The 1a/1b polyprotein contained the motifs found in other nidoviruses (Fig. 1, Table 2). An earlier study identified a 87 kD (P87) protein encoded in this region (Lim and Liu, 1998), and the P87 homolog was also found in TCoV-MG10 as a 673 amino acids protein in the same region.

3CLP was located in NSP5 of 307 amino acids. NSP5 was similar by 52%, 44%, and 92% at the nucleotide sequence level and by 39%, 44%, and 93% at the amino acids level to HCoV-229E, BCoV, and IBV, respectively. This region was believed to play a critical role for ORF1b processing since the deletion of NSP5 in IBV resulted in the unprocessing of 1b protein (Liu et al., 1997). NSP3, NSP4 and NSP6 were predicted to carry a hydrophobic transmembrane domain which may play an important role in the transcription/replication process as recently discussed for other coronaviruses (Sawicki et al., 2007). NSP10 was rich in cysteine and histidine and was predicted to contain a metal binding domain as well as NTP binding helicase domain as with other coronaviruses (Gorbalenya et al., 1989). NSP11 was predicted to be a small peptide of 23 amino acids in length, which is likely the C-terminal most cleavage product of 1a. NSP12 contained the RNA-dependant RNA polymerase (RdRp) activity, which would likely be involved in the genome replication and transcription (Liu et al., 1994). The RdRp motif was highly conserved among all coronaviruses, and TCoV-MG10 also showed a high degree of sequence

identity for RdRp by 64%, 62%, 94% at the amino acid level to HCoV-229E, BCoV, IBV, respectively. NSP13 has previously been suggested to play a role for genome replication by unwinding double-strand RNA (Gorbalenya et al., 1989). NSP14 was assumed to possess the ExonN domain. This domain may be associated with RNA metabolism such as proofreading ability and recombination. Both coronaviruses and toroviruses contain one ExonN motif while roniviruses contain two copies of the ExonN motif (Snijder et al., 2003). NSP15 contained the motif for NendoU activity. In SARS-CoV and HCoV-229E, NendoU cleaved the double-strand RNA at the uridylylate-containing sequence, and this activity was essential for RNA synthesis and progeny virus production (Ivanov et al., 2004; Posthuma et al., 2006). NSP16 contained a motif for 2' O-methyltransferase (MT) in other coronaviruses (Gorbalenya et al., 2006), and TCoV also contained NSP16 as the most C-terminal cleavage product of 1ab. TCoV ORF1a showed a 46% sequence identity to both BCoV and HCoV-OC43, which are group II coronaviruses, and a 90% identity to IBV-NC, IBV-M41, and IBV-p65 which are group III coronaviruses (Table 3). TCoV ORF1b was more conserved than ORF1a, as TCoV ORF1b showed a 59% identity to groups I and II coronaviruses, and 93% identity to IBV which is a group III coronavirus.

3.4. Structural genes

As with other coronaviruses, TCoV was characterized by the presence of 4 major structural genes, located in the 3' one-third of the genome: spike (S) glycoprotein, small envelope (E) protein, membrane (M) protein, and nucleocapsid (N) protein genes (Fig. 1). The S protein gene was located immediately downstream from ORF1b for a predicted protein of 1,226 amino acids. The S gene was

Table 3
Percent (%) nucleotide identity of TCoV MG10 to other coronaviruses

Virus ^a	Entire genome	ORF1a	ORF1b	S	E	M	N
FeCoV	46	43	59	40	37	43	38
HCoV-229E	60	45	59	40	40	43	40
HCoV-NL63	50	45	60	41	42	44	39
BCoV-NC	46	46	59	45	41	50	44
HCoV-OC43	48	46	59	44	42	51	44
MHV-JHM	46	43	58	43	43	51	42
SARS-CoV	45	41	59	44	31	45	41
IBV-Beaudette	94	90	93	57	90	94	92
IBV-p65	87	90	93	57	88	92	92
IBV-M41	87	90	93	57	88	92	93
TCoV-MG10	100	100	100	100	100	100	100
TCoV-G1	n/a ^b	n/a	n/a	96	n/a	n/a	n/a
TCoV-Gh	n/a	n/a	n/a	96	n/a	n/a	n/a
TCoV-NC95	n/a	n/a	n/a	n/a	n/a	n/a	94
TCoV-Min	n/a	n/a	n/a	n/a	n/a	n/a	94
TCoV-Ind	n/a	n/a	n/a	n/a	n/a	n/a	93

^a FeCoV, feline coronavirus; HCoV, human coronavirus; BCoV, bovine coronavirus; MHV, mouse hepatitis virus; SARS-CoV, severe acute respiratory syndrome coronavirus; IBV, infectious bronchitis coronavirus; TCoV, turkey coronavirus.

^b n/a: data not available.

the most variable gene in the TCoV genome as compared to that of other coronaviruses. The S gene showed only 40% and 44% identities to that of group I and group II coronaviruses, respectively, while the similarity to different strains of IBV was 57% (Table 3). The TCoV S gene showed the highest sequence identity to that from other TCoV isolates, TCoV-Gh and TCoV-GI, by 96% for both. The sequence variability was mainly due to the hypervariable and the receptor binding regions in the S protein. The high degree of sequence identity for S among TCoV isolates suggests that TCoV is less heterogenic as with IBV than other coronaviruses (Cavanagh, 2005). TCoV S was slightly larger than IBV S as TCoV S was 3678 nucleotides, while IBV S was 3453 nucleotides in size. The coronavirus S protein is responsible for receptor binding and virus–host cell membrane fusion. For groups II and III coronaviruses, the S protein is cleaved into two subunits; the N-terminal S1 product and the C-terminal S2 product. TCoV S was also presumed to be cleaved into S1 (542 amino acids) and S2 (684 amino acids) with the putative cleavage site characterized by presence of basic amino acids (Arg-Arg-Ser) between 542 and 543 (Fig. 1) as for IBV S (Cavanagh, 2007). The receptor-binding domain in S1 is not well-identified for IBV and thus was not possible to predict for TCoV. TCoV S was likely highly glycosylated as it contained 24 potential N-linked glycosylation sites. TCoV S contained three hydrophobic transmembrane domains; two ectodomains and one endo-domain.

TCoV Gene 3 was thought to be tricistronic (ORF 3a, 3b, and 3c). The small membrane (E) protein of 99 amino acids was potentially encoded by ORF3c. Both ORF 3b and E genes overlapped by

16 nucleotides. The TCoV E gene showed a high degree of sequence identity to IBV E by 90%. The E protein was reported to be a viroporin which played a role in the membrane permeability in SARS-CoV and porcine reproductive and respiratory syndrome virus, another member of nidoviruses (Wilson et al., 2004; Lee and Yoo, 2006).

The membrane (M) protein gene was 699 nucleotides in size and was able to make a protein of 223 amino acids. The M gene seemed to be highly conserved within group III coronaviruses since TCoV M showed a 94% nucleotide identity to IBV M. The M protein contained a single putative N-linked glycosylation site at amino acid position 4 as well as 3 potential sites for O-linked glycosylation. It remains to be determined whether these sites are functional for TCoV.

The nucleocapsid (N) protein gene was 1,227 nucleotides with a coding capacity of 409 amino acids. The TCoV N gene showed a 93% sequence identity to that of various IBV strains. The N protein was shown to be a serine phosphoprotein in other coronaviruses and arteriviruses (Alexander et al., 2005; Wootton et al., 2002). The TCoV N protein contained 20 serine residues but no tyrosine residue.

3.5. Gene 3 and Gene 5

Gene 3 (ORF3) is possibly tricistronic as with other coronaviruses. ORF3 was able to code for two non-structural proteins and the small envelope (E) protein. ORF3a and 3b were 171 and 192 nucleotides in length capable of coding for 57 and 64 amino acid proteins, respectively. ORF5 is potentially bicistronic to code for 5a and 5b proteins. ORF5a was 195 nucleotides in length for 65 amino acids while ORF5b had a potential for 82 amino acids. ORFs 5a and 5b overlapped by three nucleotides, while ORF5b and the downstream N gene overlapped by 57 nucleotides. The presence of Gene 3 and Gene 5 were highly suggestive that TCoV was related to group III coronaviruses. A recent study for the role of Gene 3 and Gene 5 for IBV replication showed that deletion mutant viruses succeeded in replication in a similar manner to the wild-type virus, suggesting that those genes were non-essential for IBV replication (Casais et al., 2005). In contrast, ORF5a and ORF5b were not found in mammalian coronaviruses, and thus the presence of Gene 5 may be considered a characteristic feature of avian coronaviruses including IBV and TCoV.

3.6. TCoV ORF-X

TCoV was characterized by the presence of an additional ORF, designated ORF-X. ORF-X was 282 nucleotides in length with 33.3% GC contents. This ORF was located upstream of Gene 5 and started immediately following the M gene. This gene was able to encode a hypothetical protein of 94 amino acids. A blast search for this ORF using the amino acid sequence found no homology to described

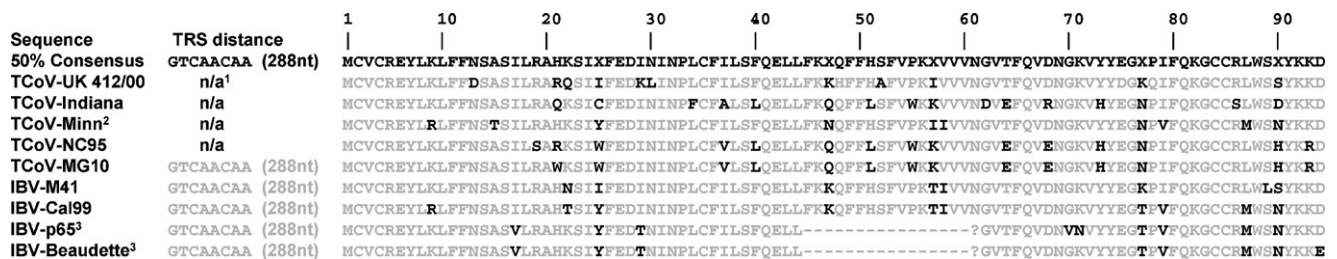


Fig. 3. Illustration of conservation of the ORF-X amino acid sequence and the putative TRS found upstream of the ORF-X gene among a number of group III coronaviruses. TCoV and IBV share a highly conserved 94 amino acids hypothetical protein, designated as ORF-X. A highly conserved TRS motif (GTCAACAA) for ORF-X is found 288 nt upstream, within the M gene, in all group III coronaviruses. The Beaudette strain of IBV and strains derived from it such as IBV-p65 have a 49 nt deletion in ORF-X but the remaining sequence aligns unequivocally with all other group III coronaviruses when this deletion is taken into account. ¹n/a, no sequence available for this region; ²single nucleotide (A) from position 446 (AF072911.1) was deleted to maintain reading frame; ³17 amino acid deletion (51 nucleotides for 17 codons) in these sequences.

Table 4
Putative transcription regulatory sequences for TCoV MG10

Gene	TRS ^a	Genomic positions	Distance (nt) ^b
ORF1a/1ab	CTTAACAA	57–64	465
S	CTGAACAA	20299–20306	53
G3	TATAACAA	23957–23964	23
M	CTTAACAA	24552–24559	77
ORF-X	GTCAACAA	25013–25020	288
G5	CGTAACAA	25652–25659	9
N	CTTAACAA	25953–25960	93

^a TRS: Transcription regulating sequence.

^b Distance between TRS and the start of the corresponding gene.

proteins; however, at the nucleotide level, this ORF was strongly conserved among group III coronaviruses including all IBV strains and all available sequences for turkey coronaviruses in the GenBank database. For several IBV strains, this region contained a well-identified ORFs including initiation (AUG) and stop codons (Fig. 3). Interestingly, a sequence, that might be the putative TRS (GUCAACAA) for this particular ORF was found 288 nucleotides upstream of the initiation codon for ORF-X within the M gene (Table 4); this putative TRS was conserved at the same relative location in virtually all IBV sequences in the GenBank database spanning the putative TRS region (Fig. 3). Further studies are warranted determine the significance of this apparently expressed protein found only in group III coronaviruses as far as is known (Fig. 3).

3.7. 3' UTR

A 3' UTR of 352 nucleotides was present immediately downstream the N gene of the genome. It has been previously shown that the 3' UTR of both TCoV Indiana and Minnesota strains were 502 nucleotides long, while TCoV-NC95 strain has a 3' UTR of 349 nucleotides which lack the first 153 nucleotides at the 5' end (Breslin et al., 1999). Also, TCoV-MG10, like IBV and some other coronaviruses, contained a conserved stem-loop structure in the 3' UTR as illustrated in Fig. 5.

3.8. Transcription regulatory sequence (TRS)

The TCoV genome contained putative TRSs located upstream of the start codon of each gene (Table 4). Although the distance between the putative TRS and the downstream initiation codon varied among the various genes in the TCoV genome, the particular distance for each gene was similar in both IBV and TCoV. In all cases, the putative TRS contained a conserved AACAA motif. The leader TRS (CUUAACAA) was found in the 5' UTR at genomic positions 57–64 which is 465 nucleotides upstream of the ORF1a initiation codon. The TRS for S gene, CUGAACAA, differed from the leader TRS by replacement of U at the 3rd position with G, and was located 53 nucleotides upstream from the S gene start. The M gene TRS was identical to the leader TRS and was located 77 nucleotides upstream of the M gene start. In the same manner, the N gene TRS

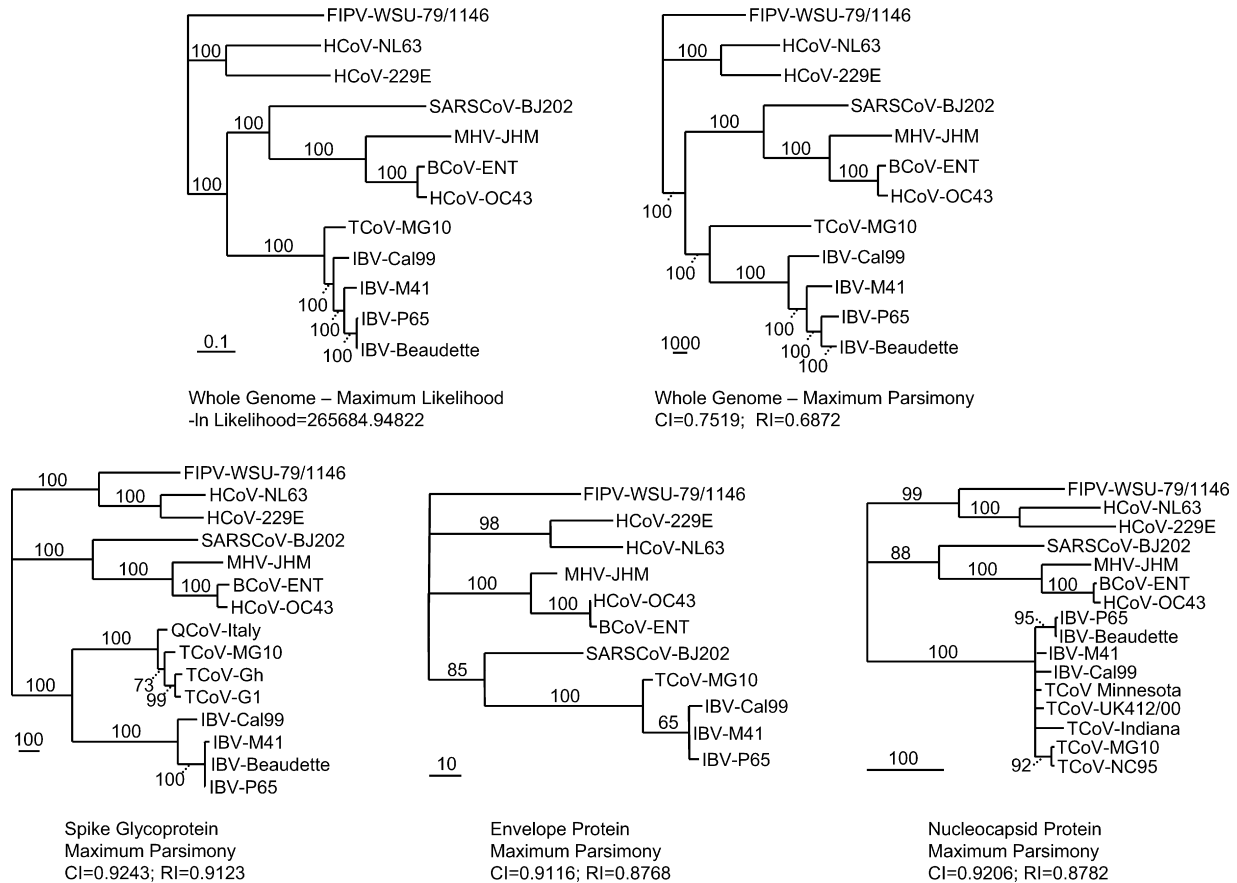


Fig. 4. Phylogenetic trees for TCoV and other coronaviruses. Aligned nucleotide sequences of complete coronavirus genomes representing coronavirus groups I, II, and III were utilized to construct maximum likelihood (ML) and maximum parsimony (MP) trees using the PAUP software package followed by bootstrap analysis using a heuristic search method. For each tree, the bootstrap support for each branch is indicated for each branch and the horizontal lengths of branches are proportional to the amount of hypothesized evolutionary change. All trees are rooted using the group I coronaviruses as a functional outgroup. The ln likelihood for the ML tree and the consistency index (CI) and retention index (RI) are provided for the MP trees. For ingroup analyses of relationships among the group III coronaviruses, aligned amino acid sequences for the spike glycoprotein (S), envelope (E) and nucleocapsid (N) proteins were analyzed using maximum parsimony (branch and bound search algorithm). HCoV, human coronavirus; FIPV, feline infectious peritonitis virus; BCoV, bovine coronavirus; MHV, murine hepatitis virus; IBV, infectious bronchitis virus; TCoV, turkey coronavirus.

Virus	Acc. No.	Position	3'-UTR stem loop structure
IBV-CAL99	AY514485	27,564	AGTGCCGAGGCCACGCGGAGTACGATCGAGGGTACAGCACT
IBV-M41	AY851295	27,346	AGTGCCGGGGCCACGCGGAGTACGACCGAGGGTACAGCACT
IBV-p65	DQ001339.1	27,483	AGTGCCGGGGCCACGCGGAGTACGATCGAGGGTACAGCACT
TCoV-MG10	EU095850	27,497	AGTGCCGGGGCCACGCGGAGTACGATCGAGGGTACAGCACA
TCoV-UK/412/00	AJ310642	104	AGTGCCGGGGCCACGCGGAGTACGATCGAGGGTACAGCACA
TCoV-NC95	AF111997	1,445	AGTGCCGGGGCCACGCGGAGTACGATCGAGGGTACAGCACA
TCoV-Indiana	AF111995	1,598	AGTGCCGGGGCCACGCGGCGTACGATCGTGGGTACAGCACT
TCoV-Minn	AF111996	1,597	AGTGCCGAGGCCACGCGGAGTACGATCGAGGGTACAGCACT
IBV-Partridge	AY646283	27,335	AGTGCCGGGGCCACGCGGAGTACGATCGAGGGTACAGCACT
Pheasant CoV	AJ619593	85	AGTGCCGAGGCCACGCGGCGTACGATCGAGGGTACAGCACT
Duck CoV	AJ871024	240	AGTGCCGGGGCCACGCGGAGTACGATCGAGGGTACAGCACT
Goose CoV	AJ871020	1,720	TGTGCCGGGGCCACGCGGAGTACGATCGAGGGTACAGCACA
Pigeon CoV	AJ871023	1,568	AATGCCGAGGCCACGCGGAGTACGATCGAGGGTACAGCATT
SARS CoV	DQ898174	29,586	TTCATCGAGGCCACGCGGAGTACGATCGAGGGTACAGTGAA
Asian Leopard CoV	EF584908	12,605	TATGCCGAGGCCACGCGGAGTACGATCGAGGGTACAGCATA

Fig. 5. Alignment of highly conserved stem-loop (s2m) sequences present in the 3' UTR of a variety of coronaviruses. This highly conserved region, flanked by highly divergent sequence (not shown), is found in TCoV, IBV and other avian coronaviruses. The s2m sequence is also found in the atypical group III coronavirus SARS-CoV and the recently sequenced coronavirus isolated from an Asian Leopard cat.

was identical to the leader TRS and located 93 nucleotides from the N gene start.

3.9. Phylogenetic analyses and classification of TCoV

Phylogenetic reconstruction of the whole genomes of 12 coronaviruses using maximum likelihood and maximum parsimony produced well-supported trees (Fig. 4) that placed TCoV-MG10 as a sister taxon to the four IBV isolates included in the analysis. Bootstrap support for both analyses was 100% at each node indicating significant support for the branching order. Group I coronaviruses (FIPV, HCoV-NL63, HCoV-229E) formed a monophyletic clade as did the group II coronaviruses (SARS-CoV, MHV, BCoV, HCoV-OC43). SARS-CoV was basal to the other group II coronaviruses in the whole genome analyses. The solid support for the monophyly of TCoV-MG10 with four IBV isolates supports the conclusion that this TCoV should be classified as a member of the group III coronaviruses.

Analyses of aligned amino acid sequences for the E, N, and S proteins of various coronaviruses produced phylogenetic trees that largely reflected the nucleic acid-based whole genome trees (Fig. 4). The N protein produced a well-supported tree that had a largely unresolved polychotomy consisting of five TCoV isolates (three USA isolates, one UK isolate, and isolate MG10 from Canada) and the four IBV isolates used in the whole genome analyses. TCoV-MG10 was most closely related to TCoV-NC95 and these two isolates formed a well-supported monophyletic group based on the N protein; however, this protein did not contain sufficient information to infer relationships reliably among group III coronavirus. Similar results were obtained with a large number of IBV sequences and five TCoV sequences with the M protein (data not shown). The E protein analysis placed TCoV-MG10 as the sister taxon to the three IBV isolates used in the analysis; however, there was only a single TCoV available for that analysis so no conclusions were drawn regarding the utility of this protein for inferring evolutionary relationships among these group III coronaviruses. Unlike the analyses based on aligned S or N protein amino acid sequences, group II coronaviruses did not form a monophyletic group in the analysis based on aligned E protein sequences; SARS-CoV did not group with the other three group II coronaviruses included in this analysis.

Using available sequences for the S protein, three TCoV isolates (Gh, G1 and MG10) formed a monophyletic group closely related to a quail coronavirus (QCoV) isolate from Italy. Together, these

isolates formed the sister group to a number of IBV isolates. The phylogenetic reconstruction based on the S protein, unlike other analyses based on the amino acid sequences of the N, E or M proteins (data not shown), produced a well-supported clade containing only TCoV strains and suggests that the S protein may be a more useful molecule for inferring relationships among the group III coronaviruses.

4. Discussion

The family *Coronaviridae* is included in the order *Nidovirales* along with the *Arteriviridae* and *Roniviridae* families. Coronaviruses are divided into three groups (I, II, and III) based on the genome structure and organization (Holmes and Lai, 1996; Lai and Cavanagh, 1997). Group I coronaviruses include porcine epidemic diarrhoea virus (PEDV), TGEV, canine coronavirus (CCoV), feline infectious peritonitis virus (FIPV), HCoV-229E, and a newly identified HCoV-NL63, whereas group II includes murine hepatitis virus (MHV), BCoV, HCoV-OC43, rat sialodacryoadenitis virus (SADV), canine respiratory coronavirus (CRCoV), and equine coronavirus (ECoV). Group III coronaviruses are IBV as well as the newly discovered pheasant coronavirus. Group III coronaviruses are characterized by modification of Gene 3 to a tricistronic structure that codes for genes 3a and 3b as well as the E gene, and the insertion of an additional unique sequence designated Gene 5. The E gene is common in all coronaviruses but is incorporated into the tricistronic Gene 3 located between S and M genes only in the group III coronaviruses as far as is known. In contrast, Gene 5 located between the M and N genes had only been reported previously from IBV. In TCoV-MG10, Gene 5 has two small ORFs (ORF5a and ORF5b) that code for products of 65 and 82 amino acids, respectively; this is identical in length to the Gene 5 products of IBV. The presence of Gene 5 has been suggested as a genetic marker for group III coronaviruses (Cavanagh et al., 2001). In vitro work with IBV has demonstrated that Gene 5 is not essential for virus replication in cell culture (Casais et al., 2005). Whether or not Gene 5 is unnecessary for in vitro replication of TCoV is unknown.

4.1. TCoV is a group III coronavirus

Since first identified in 1951 (Peterson and Hymass, 1951) and despite the economic importance to the turkey industry, TCoV has

remained one of the least characterized among all known coronaviruses. Early studies suggested that this virus could be a group II coronavirus along with BCoV and HCoV-OC43 based on serology and partial sequences for N and M genes. (Verbeek and Tijssen, 1991) reported that the TCoV N gene was 100% identical to BCoV N gene. This suggestion was supported by a study using the Minnesota strain of TCoV that demonstrated hemagglutination activity of TCoV for rabbit erythrocytes (Dea et al., 1990). A HE gene is found in many group II coronaviruses.

Despite these early confounding observations, (Guy et al., 1997), using antibodies to discriminate among coronaviruses, suggested that TCoV was more closely related to IBV than BCoV (Dea and Tijssen, 1989; Dea et al., 1990), which is in agreement with our genomic data. Our completion of the full genome sequence of a field isolate of a TCoV shows clearly that the TCoV genome structure and sequence are much closer to IBV than any other coronavirus. Like IBV, the genome structure of TCoV-MG10 is: 5' UTR – replicase (ORF1a and ORF1b) – spike (S) protein – ORF3 (ORFs 3a and 3b) – small envelop (E or 3c) protein – membrane (M) protein – ORF-X – ORF5 (ORFs 5a, and 5b) – nucleocapsid (N) protein – 3' UTR-poly(A), in order (Fig. 1). The non-structural protein gene immediately downstream of ORF1b and the further downstream HE gene that are both commonly found in group II coronaviruses were entirely absent in TCoV. Our data did not support the presence of any gene similar to HE and we concluded that TCoV does not contain the HE gene. As is obvious from Table 3, TCoV-MG10 and other TCoV isolates demonstrate much higher sequence similarity to strains of IBV that they do for any other coronavirus. For example, the N gene of TCoV was only 44% identical to BCoV and HCoV-OC43 but had 92% identity to the N gene of IBV. This apparent discrepancy with previous sequencing results (e.g. Verbeek and Tijssen, 1991) may be explained by several reasons. One possibility is genome recombination between TCoV and BCoV. It is also possible that a cell line commonly used for coronavirus cultivation, HRT-18, may harbor a latent infection with one of the human coronaviruses which was then activated upon infection with another coronavirus. Laboratory contamination of BCoV is also a possibility.

Our data shows a high degree of sequence identity between IBV and TCoV in the replicase, E, M, and N genes with greater than 90% sequence identity for each (Table 3). The S gene is most variable among IBV strains and between IBV and TCoV, perhaps reflecting the role that the S protein has in determining receptor binding in coronaviruses (Cavanagh, 2005). Despite the relatively large genetic variation in the S gene of various IBV strains, TCoV also showed a relatively higher sequence identity to IBV (up to 57%) while the sequence identity to the S gene of BCoV was only 45%. In contrast to IBV, our limited study indicates that TCoV S is relatively conserved among isolates, suggesting that the TCoV S genes might be less varied than the IBV S genes. IBV and TCoV have distinct clinical presentations in infected hosts. IBV causes respiratory disease in chickens whereas TCoV causes enteric disease in young turkeys. Perhaps the tropisms exhibited by IBV and TCoV reflect the sequence variation of the S glycoprotein and the resulting differences in receptor affinities. The TCoV S gene is larger than that of most IBV isolates. A study is required to understand the functional difference in S for IBV and TCoV, which may have a great impact on the antigenic properties and tissue tropism of both IBV and TCoV as well as the development of control measures against them.

The presence of Gene 3 and Gene 5 is a unique characteristic for group III coronaviruses as those genes do not exist in mammalian coronaviruses. As those genes are highly conserved in avian coronaviruses, they might serve as a *cis* acting elements essential for virus replication.

Based on the overall structure of the genome and sequence similarities to IBV, we conclude that TCoV should be classified as a group III coronavirus.

4.2. A newly recognized ORF characteristic of group III coronaviruses

Our sequence data revealed the presence of a novel ORF in TCoV located upstream of Gene 5 and downstream of the M gene. This ORF is unique to TCoV and IBV, the only Group III coronaviruses for which sequence data in this region of the genome are available. ORF-X was strongly conserved among group III coronaviruses including most IBV strains and all available sequences for turkey coronaviruses in GenBank as well as the TCoV-MG10 isolate. Eight (e.g. Beaudette and p65 strains) of 40 or more IBV sequences for ORF-X have a 49 base deletion in comparison with other IBV strains; all of the strains with this deletion were laboratory-adapted, cell cultured viruses originating with the Beaudette strain of IBV, suggesting that this deletion may be an artifact of cultivation outside of the natural host. The maintenance of this long sequence within a coronavirus genome seems highly unlikely if this was not a functional gene. Finding a highly conserved putative TRS, 288 bp upstream of the initiation codon for ORF-X within the M gene, reinforces the notion that this gene is functional in both IBV and TCoV despite the unusually long distance between the putative TRS and the initiation codon. This region may be a good marker for group III coronaviruses. Further studies are being carried out to further characterize this particular ORF and determine its role, if any, in virus replication. Among coronaviruses, only SARS-CoV contains an ORF immediately upstream of the N gene, which is referred to as Gene 8. Gene 8 plays an important role in the SARS-CoV replication and induction of apoptosis of its host cells (Chen et al., 2007). The function of the small ORF found in TCoV remains unknown and the lack of any homology with any known protein makes inferring its function difficult.

4.3. Regulatory features in the genome

The 3' UTR is believed to be involved in genome replication of coronaviruses (Williams et al., 1993), despite its apparent ability to possess quite variable sequence and sequence lengths. This variation within available 3' UTR sequences of TCoVs is the same as for IBV strains. The 3' UTR of IBV strains (Beaudette, KB8523 and CU-T2) are 503–505 nucleotides in length in contrast to 320 nucleotides for IBV strain M41 (Bournsnel et al., 1985). The 3' UTR of TCoV-MG10 was highly conserved with a 94–98% nucleotide identity to most published IBV and TCoV 3' UTR sequences. In contrast, the identity to BCoV was only 45%. Some viruses such as IBV, human astrovirus, and turkey astrovirus were reported to contain a stem loop-like motif (s2m) in the 3' UTR (Jonassen et al., 1998). This motif appeared to be conserved among those different viruses, suggesting that it might have resulted by RNA recombination between different viruses (Monceyron et al., 1997). TCoV was found to contain the same motif in the 3' UTR (Fig. 5). The presence of the s2m motif in the 3' UTR may also suggest that IBV and TCoV share a common ancestor. Interestingly, SARS-CoV, but not other group II coronaviruses, shares the presence of this conserved s2M motif (see Fig. 5). Sequence analyses of the 3' UTR from 19 different IBV strains revealed the presence of two distinct regions: Region I was highly variable and located immediately downstream of the N gene while region II was highly conserved and located upstream of the poly(A) tail (Dalton et al., 2001). Gobel et al. (2007) reported the presence of the octamer motif (5'-GGAAGAGC-3') within the 3' UTR hyper-variable region that was highly conserved among all coronaviruses. TCoV-MG10 had two copies of the octamer motif in its 3' UTR; the

first copy between nucleotides 25,690 and 25,697, and the second copy between nucleotides 27,553 and 27,560. The role of this motif for coronavirus replication is unknown, but (Gobel et al., 2007) suggested that it might play a role in the coronavirus replication cycle.

In summary, this is the first completion of the full-length TCoV genomic sequence. This study should lead to a better understanding of the molecular biology of TCoV and perhaps contribute to our understanding of poult enteritis mortality syndrome (PEMS) affecting young turkey flocks thought to result from co-infection of TCoV with turkey astrovirus. By completing the first genome of a TCoV, we have established the genome organization and coding strategy for the virus that unequivocally establishes that TCoV is a group III coronavirus, closely related to IBV and other avian coronaviruses. In addition, we have identified a putatively functional gene (ORF-X) shared among all sequenced IBV and TCoV strains that may be a shared feature of all group III coronaviruses.

Acknowledgements

This study was supported by NSERC Canada. MG is a scholarship recipient of the Arab Republic of Egypt. We would like to thank Janet Swinton for technical assistance.

References

- Alexander, N., Zakhartchouk, Viswanathan, S., Mahony, J.B., Gauldie, J., Babiuk, L.A., 2005. Severe acute respiratory syndrome coronavirus nucleocapsid protein expressed by an adenovirus vector is phosphorylated and immunogenic in mice. *J. Gen. Virol.* 86, 211–215.
- Barnes, H.J., Guy, J.S., 1997. Poult enteritis-mortality syndrome (spiking mortality) of turkeys. In: Saif, Y.M., Barnes, H.J., Fadly, A.M., Glison, J.R., McDougald, L.R., Swayne, D.E. (Eds.), *Diseases of Poultry*, 11th ed. Iowa State University Press, Ames, IA, pp. 1025–1031.
- Boursnell, M.E., Binns, M.M., Foulds, I.J., Brown, T.D., 1985. Sequence of the nucleocapsid genes from two strains of avian infectious bronchitis virus. *J. Gen. Virol.* 66, 573–580.
- Breslin, J., Smith, G.L., Fuller, F.J., Guy, J.S., 1999. Sequence analysis of turkey coronavirus nucleocapsid protein and 3' untranslated region identifies the virus as a close relative of infectious bronchitis virus. *Virus Res.* 65, 187–193.
- Casais, R., Davies, M., Cavanagh, D., Britton, P., 2005. Gene 5 of the avian coronavirus infectious bronchitis virus is not essential for replication. *J. Virol.* 79, 8065–8078.
- Cavanagh, D., 2005. Coronaviruses in poultry and other birds. *Avian Pathol.* 34, 439–448.
- Cavanagh, D., Mawditt, K., Sharma, M., Drury, S.E., Ainsworth, H.L., Britton, P., Gough, R.E., 2001. Detection of a coronavirus from turkey poults in Europe genetically related to infectious bronchitis virus of chickens. *Avian Pathol.* 30, 355–368.
- Cavanagh, D., 2007. Coronavirus infectious bronchitis virus. *Vet. Res.* 38, 281–297.
- Chen, C.Y., Ping, Y.H., Lee, H.C., Chen, K.H., Lee, Y.M., Chan, Y.J., 2007. Open reading frame 8a of human severe acute respiratory syndrome coronavirus not only promotes viral replication but also induce apoptosis. *J. Infect. Dis.* 196, 405–415.
- Dalton, K., Casais, R., Shaw, K., Stirrups, K., Evans, S., Britton, P., Brown, D., Cavanagh, D., 2001. Cis Acting sequence required for coronavirus infectious bronchitis virus defective RNA replication and packaging. *J. Virol.* 75, 125–133.
- Dea, S., Marsolais, G., Beaubiens, J., Ruppanner, R., 1986. Coronaviruses associated with outbreaks of transmissible enteritis of turkeys in Quebec haemagglutination properties and cell cultivation. *Avian Dis.* 30, 319–326.
- Dea, S., Tijssen, P., 1989. Isolation and trypsin enhanced propagation of turkey enteric (blue comb) coronaviruses in a continuous human rectal tumor (HRT-18) cell line. *Am. J. Vet. Res.* 5, 1310–1318.
- Dea, S., Verbeek, A.J., Tijssen, P., 1990. Antigenic and genomic relationships among turkey and bovine enteric coronaviruses. *J. Virol.* 64, 3112–3118.
- Deshmukh, D.R., Pomeroy, B.S., 1974. Physicochemical characterization of blue comb coronavirus of turkeys. *Am. J. Vet. Res.* 35, 1549–1552.
- Gobel, S.J., Miller, T.B., Bennett, C., Bernard, K.A., Masters, P.S., 2007. A hypervariable region within the 3' cis-acting element of the murine coronavirus genome is nonessential for RNA synthesis but affect pathogenesis. *J. Virol.* 81, 1274–1287.
- Gorbalenya, A.E., Enjuanes, L., Snijder, E.J., 2006. *Nidovirales*: evolving the largest RNA virus genome. *Virus Res.* 117, 17–37.
- Gorbalenya, A.E., Koonin, E.V., Donchenko, A.P., Blinov, V.M., 1989. Coronavirus genome: prediction of putative functional domains in the non-structural polyprotein by comparative amino acid sequence analysis. *Nucleic Acids Res.* 17, 4847–4861.
- Guy, J.S., Barnes, H.J., Smith, L.G., Breslin, J., 1997. Antigenic characterization of a turkey coronavirus identified in poult enteritis and mortality syndrome affected turkeys. *Avian Dis.* 41, 583–590.
- Holmes, K., Lai, M.C., 1996. Coronaviridae: the virus and their replication. In: Fields, B.N., Knipe, D.M., Howley, P.M. (Eds.), *Virology*. Lippincott-Raven Publisher, Philadelphia, pp. 1075–1093.
- Huang, H., Lau, S.K.P., Woo, P.C.Y., Yuen, K., 2007. CoVDB: a comprehensive database for comparative analysis of coronavirus genes and genomes. *Nucleic Acids Res.* 1–8.
- Ivanov, K.A., Hertzog, T., Rozanov, M., Bayer, S., Theil, V., Gobalenya, A.E., Ziebuhr, J., 2004. Major genetic marker of nidoviruses encodes a replicative endoribonuclease. *Proc. Natl. Acad. Sci. U.S.A.* 101, 12694–12699.
- Jonassen, C.M., Jonassen, T., Grinde, B., 1998. A common RNA motif in the 3' end of the genome of astroviruses, avian beonchitis virus and equine rhinovirus. *J. Gen. Virol.* 79, 715–718.
- Kumar, S., Tamura, K., Nei, M., 2004. Integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief Bioinform.* 2, 150–163.
- Lai, M.C., Cavanagh, D., 1997. The molecular biology of coronaviruses. *Adv. Virus Res.* 48, 1–100.
- Lee, C., Yoo, D., 2006. The small envelope protein of porcine reproductive and respiratory syndrome virus possesses ion channel protein-like properties. *Virology* 355, 30–43.
- Lim, K.P., Liu, D.X., 1998. Characterization of the two overlapping papain-like proteinase domains encoded in gene 1 of the coronavirus infectious bronchitis virus and determination of the C-terminal cleavage site of an 87-kDa protein. *Virology* 245, 303–312.
- Liu, D.X., Brierley, I., Tibbles, K.W., Brown, T.D.K., 1994. A 100-kilodalton polypeptide encoded by open reading frame (ORF) 1b of the coronavirus infectious bronchitis virus is processed by ORF1a products. *J. Virol.* 68, 5772–5780.
- Liu, D.X., Xu, H.Y., Brown, T.D.K., 1997. Proteolytic processing of the coronavirus infectious bronchitis virus 1a polyprotein: Identification of a 10 kilodalton polypeptide and determination of its cleavage sites. *J. Virol.* 71, 1814–1820.
- Monceyron, C., Grinde, B., Jonassen, T.Q., 1997. Molecular characterization of the 3' end of the astrovirus genome. *Arch. Virol.* 142, 99–706.
- Nagaraja, K.V., Pomeroy, B.S., 1997. Coronaviral enteritis of turkeys (bluecomb disease). In: Calnek, B.W., Barnes, H.J., Beard, C.W., McDougald, L.R., Saif, Y.M. (Eds.), *Disease of Poultry*, 11th ed. Iowa State University Press, Ames, IA, pp. 686–692.
- Peterson, E.H., Hymass, T.A., 1951. Antibiotic in treatment of unfamiliar turkey disease. *Poult. Sci.* 30, 466–468.
- Posthuma, C.C., Nedialkova, D.D., Zevenhoven-Dobbe, J.C., Blokhuis, J.H., Gorbalenya, A.E., Snijder, E.J., 2006. Site-directed mutagenesis of the nidovirus replicative endoribonuclease NendoU exerts pleiotropic effects on the arterivirus life cycle. *J. Virol.* 80, 1653–1661.
- Sawicki, S.G., Sawicki, D.L., Siddell, S.G., 2007. A contemporary view of coronavirus transcription. *J. Virol.* 81, 20–29.
- Snijder, E.J., Bredenbeek, P.J., Dobbe, J.C., Thiel, V., Ziebuhr, J., Poon, L.L., Guan, Y., Rozanov, M., Spaan, W.J.M., Gorbalenya, A.E., 2003. Unique and conserved features of genome and proteome of SARS coronavirus, an early split-off from the coronavirus group 2 lineage. *J. Mol. Biol.* 331, 991–1004.
- Swofford, D.L., 1991. PAUP: Phylogenetic Analysis Using Parsimony, Version 3.1 Computer program distributed by the Illinois Natural History Survey, Champaign, Illinois.
- Teixeira, M.C., Luvizotto, M.C., Ferrari, H.F., Mendes, A.R., da Silva, S.E., Cardoso, T.C., 2007. Detection of turkey coronavirus in commercial turkey poults in Brazil. *Avian Pathol.* 36, 29–33.
- Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The CLUSTAL X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 25, 4876–4882.
- Verbeek, A., Tijssen, P., 1991. Sequence analysis of the turkey enteric coronavirus nucleocapsid and membrane protein genes: a close genomic relationship with bovine coronavirus. *J. Gen. Virol.* 72, 1659–1666.
- Williams, A.K., Wang, L., Sneed, L.W., Collisson, E.W., 1993. Analysis of a hypervariable region in the 3' non-coding end of the infectious bronchitis virus genome. *Virus Res.* 28, 19–27.
- Wilson, L., Mckinlay, C., Gage, P., Ewart, G., 2004. SARS coronavirus E protein forms cation-selective ion channels. *Virology* 330, 322–331.
- Wootton, S.K., Rowland, R.R., Yoo, D., 2002. Phosphorylation of the porcine reproductive and respiratory syndrome virus nucleocapsid protein. *J. Virol.* 76, 10569–10576.
- Ziebuhr, J., Schelle, B., Karl, N., Minskaia, E., Bayer, S., Siddell, S.G., Gorbalenya, A.E., Theil, V., 2007. Human coronavirus 229E papin like protease have overlapping specificities but distinct functions in viral replication. *J. Virol.* 81, 3922–3932.
- Ziebuhr, J., Snijder, E., Gorbalenya, A.E., 2000. Virus-enclosed proteinases and proteolytic processing in the Nidovirales. *J. Gen. Virol.* 81, 853–879.
- Ziebuhr, J., Theil, V., Gorbalenya, A.E., 2001. The autolytic release of a putative RNA virus transcription factor from its polyprotein precursor involves two paralogous papain-like proteinase that cleave the same peptide bond. *J. Biol. Chem.* 276, 33220–33232.