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 nucleotide sequence of polyprotein gene 1 and genome organization of turkey coronavirus

Jianzhong Cao, Ching-Ching Wu, Tsang Long Lin

Department of Comparative Pathobiology, 406 South University Street, Purdue University, West Lafayette, IN 47907, United States

A R T I C L E   I N F O

Article history:
Received 27 December 2007
Accepted 18 April 2008
Available online 2 June 2008

Keywords:
Turkey coronavirus
Polyprotein
Polymerase
coronavirus
Genome
Subgenomic

A B S T R A C T

The complete nucleotide sequence of polyprotein gene 1 and the assembled full-length genome sequence are presented for turkey coronavirus (TCoV) isolates 540 and ATCC. The TCoV polyprotein gene encoded two open reading frames (ORFs), which are translated into two products, pp1a and pp1ab, the latter being produced via −1 frameshift translation. TCoV polyprotein pp1a and pp1ab were predicted to be processed to 15 non-structure proteins (nsp2–nsp16), with nsp1 missing. ClustalW analysis revealed 88.99% identity and 96.99% similarity for pp1ab between TCoV and avian infectious bronchitis virus (IBV) at the amino acid level. The whole genome consists of 27,749 nucleotides for 540 and 27,816 nucleotides for ATCC, excluding the poly(A) tail. A total of 13 ORFs were predicted for TCoV. Five subgenomic RNAs were detected from ATCC-infected turkey small intestines by Northern blotting. The whole genome sequence had 86.9% identity between TCoV and IBV, supporting that TCoV is a group 3 coronavirus.

© 2008 Elsevier B.V. All rights reserved.

1. Introduction

Turkey coronavirus (TCoV) is a causative agent for bluecomb disease in turkey poults. The outbreak of the disease was first reported more than 40 years ago, and the viral agent responsible for the disease was identified as turkey coronavirus in 1973 (Ritchie et al., 1973). TCoV infects the small intestine of turkey poults and causes disruption of the infected tissue resulting in reduced surface area of intestine, reduced consumption of food and apparent decrease in body weight of infected turkeys. The mortality rate is low, however. The outbreaks of TCoV were mostly reported from turkey farms in the US and Europe (Cavanagh, 2005; Nagaraja and Pomeroy, 1997). Based on the antigenic relationship between TCoV and other coronaviruses, TCoV was classified with avian infectious bronchitis virus (IBV), which infects chicken, as a group 3 coronavirus within the Genus Coronaviruses, Family Coronaviridae, and Order Nidovirales (Gonzalez et al., 2003).

Coronavirus genome contains a single, positive-strand RNA ((+ ssRNA) molecule, which is about 27–33 kilobases (kb) and has a cap at the 5′ end and poly(A) tail at the 3′ end (Boursnell et al., 1987; Lai and Stohlman, 1981). There are four structural genes encoded by all coronavirus genomes so far sequenced; these are spike protein (S), envelope protein (E), matrix protein (M), and nucleocapsid protein (N). The genome organization of coronavirus is 5′-polymerase-S-E-M-N-3′. An untranslated region (UTR) is located at both the 5′ and 3′ ends of the genome. The production of structural proteins is through transcription of a set of co-terminal subgenomic mRNA (sgRNA). The molecular mechanisms of genome replication and transcription are not fully understood, but the discontinuous negative-strand extension model has gained wide acceptance (Sawicki and Sawicki, 1995; Sawicki et al., 2007).

The polymerase gene accounts for about two-thirds of the genome (20–22 kDa) and consists of two open reading frames (ORFs): ORF1a and ORF1b (Boursnell et al., 1987). The polymerase is necessary and sufficient for genome replication and transcription because purified viral RNA or in vitro transcribed viral RNA from cDNA construct are infectious when transfected into permissive cells (Yount et al., 2000). However, nucleocapsid protein greatly enhanced coronavirus genome replication (Almazan et al., 2004; Schelle et al., 2005), suggesting that nucleocapsid protein may have a regulatory role for coronavirus replication. When viral genomic RNA enters the host cell, ORF1a (pp1a) and polyprotein 1ab (pp1ab) are translated first, the latter being translated through a −1 frameshift translation mechanism (Bredenbeek et al., 1990; Brierley et al., 1989; Herold and Siddell, 1993; Lee et al., 1991). Coronavirus pp1ab contains a 3C-like proteinase (3CLpro)
and a papain-like protease (PLP) that automatically cleaves themselves from polyprotein and further process the polyprotein into four accessory ORFs designated as 3a, 3b, 5a, and 5b (Breslin et al., 1999a; Lin et al., 2004). Sequence analysis indicated that the sequences of M and N of TCoV shared over 80% sequence identity with that of IBV. However, the S gene shared less than 40% sequence similarity to any known coronavirus S genes (Lin et al., 2004). These results suggest that TCoV may have diverged from IBV during evolution. In this study, we continue to determine and analyze the nucleotide sequence of the polyprotein gene of TCoV and use bioinformatics to predict potential functional domains encoded by TCoV polyprotein 1ab. The polymerase gene sequence is then combined with structure gene sequence to assemble the full-length genome sequence for TCoV.

2. Materials and methods

2.1. Viruses

TCoV isolate ATCC (Minnesota strain) was obtained from American Type Culture Collection (ATCC, Mallasas, VA). The TCoV isolate 540 used in the present study was recovered from fecal contents of embryos from turkey poult with acute coronaviral enteritis in Indiana, USA in 1994. The viruses were propagated in 22-day-old embryonating turkey eggs. The presence of TCoV in the intestines of embryos was confirmed by TCoV-specific immunofluorescence antibody assays and electron microscopy at the Indiana State Animal Disease Diagnostic Laboratory in West Lafayette, IN, USA (Lin et al., 2004). Viruses were purified from small intestine following published method (Loa et al., 2002) and either used immediately or stored at −80°C for further use.

2.2. RNA isolation and cDNA synthesis

Viral genomic RNA was purified with RNApure reagent (Gen-Hunter). Briefly, 0.2 ml of virus suspension was mixed with 1 ml of RNApure reagent followed by chloroform extraction. RNA was finally precipitated by isopropanol and washed with 70% ethanol. RNA pellet was air dried and dissolved in 30 μl of DEPC–H2O and used for cDNA synthesis by SuperScript RT II system with random hexamer or oligo dT18 (for 3′ RACE) (Invitrogen). The synthesized cDNA was treated with RNase A to digest viral RNA and then served as template for PCR.

2.3. PCR amplification

To clone the whole 1ab gene, the following strategies were employed. The first was to amplify a 900-bp conserved RdRp and spike gene. Based on sequence results, bioinformatics analysis was used to design PCR primers to amplify the remaining sequence of ORF 1a gene. Expand LA PCR system (Roche) was used for all PCR amplification. The PCR reaction consisted of 1× PCR buffer, 1.7 mM MgCl2, 500 nM each of dNTPs, 200 pmol of each primer, 2 μl of cDNA, and 0.25 unit of DNA polymerase in a final volume of 50 μl. The PCR was performed on a Tetra machine (MJ Research) with the following conditions: initial denaturation at 94°C for 3 min; denaturation at 93°C for 10 s, annealing at 55°C for 30 s, extension at 68°C for 5–6 min; total of 30 cycles. The final extension at 68°C was 10 min. PCR product was purified by Qiagen PCR purification Kit (Qiagen), cloned into pCRII-TOPO vector, and transformed into TOP10F cells (Invitrogen). The plasmid was prepared by QIAquick Spin Miniprep Kit (Qiagen) and submitted for DNA sequencing at Purdue Genomic Center (Purdue University, West Lafayette, IN, USA). At least two independent colonies were sequenced for each sequence. All PCR primers are listed in Supplementary Table S1 and are available upon request.

2.4. Amplification of 5′ and 3′ ends by RACE

To amplify the 5′ end of TCoV genome, 5′ RACE system for rapid amplification of cDNA ends (Invitrogen) was employed except that Expand LA polymerase was used in the PCR. Random primers were used to synthesize cDNA from ATCC and 540 RNA. The cDNA was treated with RNase mix and purified by GlassMax spin cartridge according to manufacturer’s protocol (Invitrogen). The 3′ end of cDNA was tailed with dCTP by TdT. After tailing with dCTP, PCR was performed with primers AAP (GCC CAC GCC TCG ACT AGT ACG GCI IGG GII GG GIG IIGG IIG) and IBPR2 (TGG CAC TAC CCC CTA CAA AC). The amplified PCR product was analyzed and cloned for sequencing in the same way as described in previous section for PCR amplification.

To amplify the 3′ end of TCoV genome, oligo dT18 was used to synthesize cDNA from genomic RNA of ATCC and 540. After degradation of RNA with RNase mix, the cDNA was used as template for PCR with primers oligodT15 and AT3endF (TGAATTTGTATGATGATGGI IGG GII GIG) and IBPR2 (TGG CAC TAC CCC CTA CAA AC). The amplified PCR product was analyzed and cloned for sequencing in the same way as described in previous section for PCR amplification.

To obtain the leader–body junction sequence of each subgenomic mRNA (sgRNA), primers TCVF (ACTAAAGATAGATTAAATATATCTACTGACTACG) and TCVsgR1 (AAACAACAGTTGACATTTC) were used to amplify the 5′ end of sgRNA for 3, M, and N. For amplifying the 5′ end of sgRNA for S gene, TCVF and AT5174 (TCTG-GCCGTCCTACAATGCTGGA) were used in PCR. PCR products were cloned into pCRII-TOPO for sequencing as described in previous section.

2.5. Sequence analysis

All DNA sequences were analyzed by DNAStar software (Madison, WI, USA) and ClustalW program (Thompson et al., 1994) or online softwares as indicated in the results. Frameshift pseudoknot was predicted using M-fold (Mathews et al., 1999).

2.6. Polyprotein mapping

Polyprotein mapping of TCoV 1ab polyprotein was based on predicted 3Cpro and PLP and their substrate preferences as described for IBV (Liu et al., 1998) and other coronaviruses (Hegyi and Ziebuhr, 2002; Kiemer et al., 2004). BLASTp program (NCBI: http://www.ncbi.nlm.nih.gov/blast/blast.cgi) and pfman (www.expasy.org) were used to find sequence similarity and conserved domains in database. TMHMM was used to predict transmembrane domains (http://www.cbs.dtu.dk/services/TMHMM-2.0/). The nomenclature for pp1a and pp1ab mapping product (nsp) was according to Ziebuhr, 2005 and Ziebuhr et al., 2000.
Table 1

Open reading frames encoded in TCoV-ATCC genome

| ORF | Location | Size (nt) | Size (aa/kDa) |
|-----|----------|-----------|---------------|
| 1a  | 529–12,402 | 11,874 | 3957/441.130 |
| 1b  | 12,477–20,441 | 7,965 | 2654/300.788 |
| S (2) | 20,392–24,003 | 3,612 | 1203/132.168 |
| 3a  | 24,003–24,176 | 174 | 57/6.680 |
| 3b  | 24,176–24,370 | 195 | 64/7.395 |
| E (3c) | 24,351–24,662 | 312 | 103/11.451 |
| M (4a) | 24,652–25,323 | 672 | 223/25.153 |
| 4b  | 25,324–25,608 | 285 | 94/11.180 |
| 5a  | 25,684–25,881 | 198 | 65/7.502 |
| 5b  | 25,878–26,126 | 249 | 82/9.354 |
| N (6a) | 26,069–27,298 | 1,230 | 409/45.069 |
| 6b  | 27,307–27,531 | 225 | 74/8.744 |

2.7. Phylogenetic analysis

The alignments were performed using CLUSTALW (Thompson et al., 1994), and phylogenetic trees were drawn by DNASTar and program at http://www.genebee.msu.su/services/phtree_full.html.

Coronavirus sequences used in this article were from NCBI. Their GenBank accession numbers were:

- BCoV, NC_003045
- BtCoV, NC_008315
- FCoV, NC_007025
- HCoV-229E, NC_002645
- HCoV-NL63, NC_005147
- HCoV-OC43, NC_005147
- HCoV-HKU1, NC_006577
- IBV, NC_001451
- MHV, NC_001846
- PEDV, NC_003436
- SARS-CoV, NC_004718
- TGEV, NC_002306

2.8. Northern blotting

About 10 µg of isolated total RNA from mock and ATCC-infected turkey small intestines were separated on 1% agarose gel and transferred onto nitrocellulose membrane. 32P-CTP-(GE healthcare) labeled N gene probe was prepared using High Prime DNA Labeling Kit (Roche) with N gene primers N102F and N102R. Membrane was prehybridized for 2 h at 68 °C and then hybridized overnight at 68 °C with 32P-labeled N gene probe. After hybridization, membranes were wrapped with Saran Wrap and exposed to X-ray film for signal development.

3. Results and discussion

3.1. Nucleotide sequence accession number

The sequences reported in this work have been deposited in the GenBank database under accession number EU022526 for TCoV-ATCC and EU022525 for TCoV-540.

Fig. 1. (a) Genome organization of TCoV. Diagram shows putative ORFs, UTRs, leader (L), and TRS are not to scale. Above, the genome organization of TCoV are shown the predicted five sgRNA in relative sizes. Genome organization of IBV-Beaudette (NC_001451) is displayed below for comparison with that of TCoV. (b) Mapping of TCoV polyprotein. Predicted non-structure proteins (nsp2–nsp16) for pp1a are shown in relative sizes (bottom panel). Nsp1 is missing from TCoV and nsp11 contains only 23 aa. The sequence is for ATCC isolate (accession number EU022526).
3.2. Polyprotein gene of TCoV

The sequence of polymerase gene 1 of TCoV isolate ATCC contained 20,441 nts, excluding the 5' UTR. Two ORFs were encoded by gene 1. ORF1a contained 11,874 nts (529–12,402) encoding a protein of 3957 aa (pp1a); ORF1b contained 7955 nt (12,477–20,441) encoding a protein of 2654 aa (pp1b) (Table 1). The polyprotein gene of the TCoV 540 isolate consisted of 20,411 nts excluding pp1a of 3945 aa, and ORF 1b was 7955 nts (12,443–20,401), encoding a protein of 2652 aa (Table 2). Through −1 frameshift translation, pp1ab was predicted to contain 6637 aa for ATCC and 6623 aa for 540. The 3' end of ORF1b overlapped with the 50 nts on the 5' end of spike gene. There were 14 aa missing in 540 pp1ab when compared with ATCC pp1ab. They were distributed at 7 positions on pp1ab of the ATCC isolate, i.e. positions 922–923 (2 aa, nsp3); 930 (1 aa, nsp3); 971–973 (3 aa, nsp3); 2306–2307 (2 aa, nsp4); 3226–3229 (4 aa, nsp6); 4234 (1 aa, nsp12); 5095 (1 aa, nsp13). ClustalW comparison of the protein sequence between 540 and ATCC showed that the sequence identities of pp1a, pp1b, and pp1ab were 89.92%, 95.86%, and 92.26%, respectively. The overall similarities for pp1a, pp1b, and pp1ab were 97.4%, 98.91%, and 97.97%, respectively.

The frameshift "slippery sequence" UUUAAAC (Brierley et al., 1989) was identified for both ATCC and 540. Both sequences were located before the end of ORF1a. The sequences downstream of UUUAAAC were predicted to form a pseudoknot to support the translational frameshift (Brierley et al., 1989) (Supplementary Fig. S1). The frameshift position was predicted at C of UUUAAAC.

Comparison of pp1a and pp1ab of TCoV with those of other coronaviruses revealed that the TCoV polyprotein was predicted to be processed into 15 non-structure proteins (nsp2–nsp16; Fig. 1(b) and Table 3) by polyprotein-encoded viral proteinases. One 3C-like proteinase (3CLpro) was predicted to reside in nsp5 due to its conserved residues responsible for 3CLpro activity (Supplementary Fig. S2) (Ziebuhr et al., 2000); one papain-like proteinase (PLpro) was identified in nsp3 due to its conserved PLP residues (CHD) (Supplementary Fig. S3). Like another group 3 coronavirus, IBV, only one active PLpro was predicted for TCoV. The structure of TCoV nsp3 bears similar organization to nsp3 of IBV in that the Ac domain, X domain (ADPR), and Y domain were all present and arranged in the same order (Fig. 1(b)). Comparison of amino acid sequences of each nsp of TCoV with those of other coronaviruses predicted several putative enzymatic activities: among them, the enzymatic activity and potentials of nsp2, nsp5 (Supplementary Fig. S3), nsp13 (Supplementary Fig. S5), nsp14 (Supplementary Fig. S6), and nsp15 (Supplementary Fig. S7) were confirmed in other coronaviruses by

3.2. Polyprotein gene of TCoV

The sequence of polymerase gene 1 of TCoV isolate ATCC contained 20,441 nts, excluding the 5' UTR. Two ORFs were encoded by gene 1. ORF1a contained 11,874 nts (529–12,402) encoding a protein of 3957 aa (pp1a); ORF1b contained 7955 nt (12,477–20,441) encoding a protein of 2654 aa (pp1b) (Table 1). The polyprotein gene of the TCoV 540 isolate consisted of 20,411 nts excluding pp1a of 3945 aa, and ORF 1b was 7955 nts (12,443–20,401), encoding a protein of 2652 aa (Table 2). Through −1 frameshift translation, pp1ab was predicted to contain 6637 aa for ATCC and 6623 aa for 540. The 3' end of ORF1b overlapped with the 50 nts on the 5' end of spike gene. There were 14 aa missing in 540 pp1ab when compared with ATCC pp1ab. They were distributed at 7 positions on pp1ab of the ATCC isolate, i.e. positions 922–923 (2 aa, nsp3); 930 (1 aa, nsp3); 971–973 (3 aa, nsp3); 2306–2307 (2 aa, nsp4); 3226–3229 (4 aa, nsp6); 4234 (1 aa, nsp12); 5095 (1 aa, nsp13). ClustalW comparison of the protein sequence between 540 and ATCC showed that the sequence identities of pp1a, pp1b, and pp1ab were 89.92%, 95.86%, and 92.26%, respectively. The overall similarities for pp1a, pp1b, and pp1ab were 97.4%, 98.91%, and 97.97%, respectively.

The frameshift "slippery sequence" UUUAAAC (Brierley et al., 1989) was identified for both ATCC and 540. Both sequences were located before the end of ORF1a. The sequences downstream of UUUAAAC were predicted to form a pseudoknot to support the translational frameshift (Brierley et al., 1989) (Supplementary Fig. S1). The frameshift position was predicted at C of UUUAAAC.

Comparison of pp1a and pp1ab of TCoV with those of other coronaviruses revealed that the TCoV polyprotein was predicted to be processed into 15 non-structure proteins (nsp2–nsp16; Fig. 1(b) and Table 3) by polyprotein-encoded viral proteinases. One 3C-like proteinase (3CLpro) was predicted to reside in nsp5 due to its conserved residues responsible for 3CLpro activity (Supplementary Fig. S2) (Ziebuhr et al., 2000); one papain-like proteinase (PLpro) was identified in nsp3 due to its conserved PLP residues (CHD) (Supplementary Fig. S3). Like another group 3 coronavirus, IBV, only one active PLpro was predicted for TCoV. The structure of TCoV nsp3 bears similar organization to nsp3 of IBV in that the Ac domain, X domain (ADPR), and Y domain were all present and arranged in the same order (Fig. 1(b)). Comparison of amino acid sequences of each nsp of TCoV with those of other coronaviruses predicted several putative enzymatic activities: among them, the enzymatic activity and potentials of nsp2, nsp5 (Supplementary Fig. S3), nsp13 (Supplementary Fig. S5), nsp14 (Supplementary Fig. S6), and nsp15 (Supplementary Fig. S7) were confirmed in other coronaviruses by
experimentation (Bhardwaj et al., 2004; Eckerle et al., 2007; Fang et al., 2006; Graham et al., 2005; Kiemer et al., 2004). Nsp8, nsp9, and nsp10 were predicted to have RNA binding activity (Egloff et al., 2004; Matthes et al., 2006; Zhai et al., 2005). Nsp12 was predicted to be the major RdRp (Supplementary Fig. S4), though its activity has not been experimentally confirmed. Nsp16 was predicted to be 2′-O-methyltransferase (Supplementary Fig. S8).

3.3. Genome organization of TCoV

The first two full-length genome sequences were reported for TCoV prototype ATCC and field isolate 540. The complete genome sequences were obtained by assembly of polyprotein gene sequences that were determined by direct sequencing of cloned RTPCR products in this report and published structure gene sequences of the same isolates from our lab (Lin et al., 2004; Loa et al., 2006). Both 5′ and 3′ UTR sequences were determined by RACE and used to assemble the full-length genomic sequence. The reported genomic sequences were 27,817 nucleotides (nt) for ATCC and 27,749 nt for 540, excluding poly(A) tail. For both TCoV isolates, the percentage of nucleotide composition was 29% for A, 33% for U, 22% for G, and 16% for C. A + U was 62%, indicating that the genome of TCoV was AU rich. The genome nucleotide sequence identity between 540 and ATCC was 92.8% by Clastal W. 540 and ATCC shared nucleotide sequence identity of 86.9% and 87.5% with that of IBV, respectively. Analysis of genome organization of the TCoV-ATCC isolate revealed that there was a 64-nt (1–64) leader sequence within the 5′ UTR of 530 nt. As found in other coronaviruses (Brian and Baric, 2005), the 5′ UTR of TCoV encoded an ORF of 11 amino acids (Supplementary Table S2). Using the ORF finder at NCBI, it was revealed that there were 13 putative ORFs in the genomes of TCoV isolates ATCC and 540. These ORFs were 1a, 1b, 2 (spike), 3a, 3b, 3c (envelope), 4a (matrix), 4b, 4c, 5a, 5b, 6a (nucleocapsid), and 6b (Fig. 1(a); Tables 1 and 2). 4b came immediately after the matrix 3c (envelope), 4a (matrix), 4b, 4c, 5a, 5b, 6a (nucleocapsid), and 6b revealed that there were 13 putative ORFs in the genomes of TCoV (Supplementary Table S2). Using the ORF finder at NCBI, it was revealed that there were 13 putative ORFs in the genomes of TCoV isolates ATCC and 540. These ORFs were 1a, 1b, 2 (spike), 3a, 3b, 3c (envelope), 4a (matrix), 4b, 4c, 5a, 5b, 6a (nucleocapsid), and 6b (Fig. 1(a); Tables 1 and 2). 4b came immediately after the matrix gene. 6b was immediately following N gene. By comparison with another group 3 coronavirus, IBV-Beaudette, it was found out that 4c and 6b were not present in IBV-Beaudette (Fig. 1(a)). The prediction of 6b was not expected. After N gene, the nucleotide sequences of TCoV and IBV-Beaudette were highly conserved (Supplementary Fig. S9). However, there was no ORF in this region of IBV, so the 3′ UTR of IBV was over 500 nt. In both isolates of TCoV, a 74-aa ORF (6b) was predicted in this region irrespective of nucleotide sequence conservation between TCoV and IBV. The prediction of ORF 6b reduced the potential 3′ UTR of TCoV to less than 301-nats as compared with 506-nats in IBV. Determination of whether or not proteins of 4b, 4c, and 6b were produced requires further experimental confirmation. A consensus octanucleotide motif GGAAGAGC was found 72-nt upstream of the poly(A) tail in 540 and ATCC genomes of the TCoV. In mouse hepatitis virus, the octanucleotide motif was found to be unnecessary for virus replication in vitro, but a deletion mutant showed reduced replication in mouse brain, suggesting that the octanucleotide motif affects pathogenesis (Goebel et al., 2007). A consensus transcriptional regulated sequence (TRS) (CUUAACAAA) was found located at the 3′ end of the genome leader (1–64 nt) and in front of each structure gene and major accessory gene with either an exact match (sg3–6) or one mismatch (sg2). A total of five sgRNA were predicted for production of structure and accessory proteins in the genome of TCoV.

3.4. Phylogenetic analysis of 1ab

Clastal W program was used to analyze the relationship between TCoV pp1ab and other coronavirus pp1ab. Table 4 was a summary of the amino acid sequence identity of nsps between TCoV ATCC and other coronaviruses. It was noticed that TCoV and IBV shared highest sequence identity for all nsps when compared with other coronaviruses. Tree-top software was used to draw phylogenetic trees (http://www.genebee.msu. su/services/phtree_full.html). Fig. 2 shows the result of phylogenetic analysis of pp1ab. The TCoV was grouped with the IBV in the group 3. A close examination of TCoV and IBV polyprotein pp1ab showed that the matrix distance within the two TCoV strains was longer (0.047) than that of TCoV and IBV (0.045). ClustalW analysis of pp1ab, 3Clpro, RdRp, and helicase of TCoV and IBV showed sequence similarity of 97.97%, 94.09%, 98.5%, and 97.66%, respectively.

3.5. Subgenomic mRNA detection for TCoV

Based on the location of TRS on the genome, it was predicted that 5 subgenomic mRNA would be produced for structure and accessory gene translation (Fig. 1).
Fig. 4. Sequences flanking TRS region for TCoV sgRNA. The partial sequences display each sgRNA for ATCC isolate. For each sgRNA, partial genomic leader (gL) and body (gS, gE, gM, g5, and gN) sequences are displayed above and below sgRNA. The star (*) indicates identical nucleotide and the box indicates TRS region where template switch is assumed to occur.

To confirm predicted sgRNA production for TCoV, total RNA was isolated from mock or ATCC-infected turkey small intestines and used for Northern blotting with 32P-labeled PCR probe specific for the N gene. Fig. 3 shows 7 RNA bands detected in the ATCC infected sample, but not in the mock-infected sample, indicating the specificity of the probe. Based on predicted sizes for genomic and subgenomic RNA, one band was assigned to genomic RNA and five bands were assigned to sgRNA 2–6 for expression of S, E, M, 5, and N proteins. One extra band whose size was smaller than genomic RNA was assumed to be a defect interfering (DI) RNA. DI RNA has been detected in other coronavirus-infected cells and was assumed to be the template switch products during replication.

Because TRS in sgRNA could be derived from template switch between leader and body TRS, we aimed to determine potential switch position by analyzing sequences flanking the TRS region in each sgRNA. Fig. 4 is a summary of partial sequences flanking the TRS region for each sgRNA. It was noticed that the TRS (CUUAA-CAAA) of the S gene sgRNA was identical to the TRS of the leader, but different from the body TRS by one nucleotide (CUgAACAAA). This suggested that the template switch was downstream of CUU on the leader TRS. The TRS of the remaining sgRNA was the same as for the leader and the body TRS, implying the template switch could have occurred anywhere within CUUAAACAAA. As expected, genes 3a, 3b, and 3c (E) shared the same sgRNA for translation; genes 4a (M), 4b, and 4c shared the same sgRNA; genes 5a and 5b shared the same sgRNA; genes 6a (N) and 6b shared the same sgRNA. Determination of weather or not the predicted 3a, 3b, 4b, 4c, 5b, and 6b were expressed require experimental confirmation and hence their biological functions during replication and pathogenesis.

4. Conclusion

In conclusion, our data of completed TCoV polyprotein gene sequence and the assembly of the first full-length genome of TCoV support the classification of TCoV as a group 3 coronavirus. The completed genome sequences of two TCoV isolates will aid our understanding of coronavirus in terms of molecular evolution and molecular pathogenesis. It will also provide a strong basis for the
development of up-dated molecular diagnostics and recombinant or DNA-based vaccines for the control and prevention of TCoV infection in turkey flocks.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.viruses.2008.04.015.

References

Almazan, F., Galan, C., Enjuanes, L., 2004. The nucleoprotein is required for efficient coronavirus genome replication, J. Virol. 78 (22), 12683–12688.

Bhardwaj, K., Guarino, L., Rao, C.C., 2004. The severe acute respiratory syndrome coronavirus Nsp15 protein is an endoribonuclease that prefers manganese as a cofactor, J. Virol. 78 (22), 12218–12224.

Boursnell, M.E., Brown, T.D., Foulds, I.J., Green, P.F., Tomley, F.M., Binns, M.M., 1987. Completion of the sequence of the genome of the coronavirus avian infectious bronchitis virus, J. Gen. Virol. 68 (Pt 1), 57–77.

Bredenbeek, P.J., Bauwens, B., Noten, A.F., Charite, J., Luytjes, W., Weiss, S.R., Spaan, W.J., 1990. The primary structure and expression of the second open reading frame of the polymerase gene of the coronavirus MHV-A59: a highly conserved polymerase is expressed by an efficient ribosomal frameshifting mechanism, Nucleic Acids Res. 18 (7), 1825–1832.

Brelsin, J.J., Smith, L.G., Fuller, F.J., Guy, J.S., 1999a. Sequence analysis of the matrix/nucleocapsid gene region of turkey coronavirus, Interivirology 42 (1), 22–29.

Brelsin, J.J., Smith, L.G., Fuller, F.J., Guy, J.S., 1999b. Sequence analysis of the turkey coronavirus nucleocapsid protein gene and 3 untranslated region identifies the virus as a close relative of infectious bronchitis virus, Virus Res. 65 (2), 187–193.

Brian, D.A., Baric, R.S., 2005. Coronavirus genome structure and replication, Curr. Top. Microbiol. Immunol. 287, 1–30.

Brietley, I., Digard, P., Inglese, S.C., 1989. Characterization of an efficient coronavirus ribosomal frameshifting signal: requirement for an RNA pseudoknot, Cell 57 (4), 537–547.

Cavanagh, D., 2005. Coronaviruses in poultry and other birds, Avian Pathol. 34 (6), 661–671.

Eckerle, L.D., Lu, X., Spero, S.M., Choi, L., Denison, M.R., 2005. Coronavirus genome replication. Curr. Top. Microbiol. Immunol. 287, 1–30.

Egloff, M.P., Ferron, F., Campanacci, V., Longhi, S., Rancurel, C., Dutartre, H., Snijder, E.J., Gorbalenya, A.E., Cambillau, C., Canard, B., 2004. The severe acute respiratory syndrome-coronavirus replicase protein nsp9 is a single-stranded RNA-binding subunit unique in the RNA virus world, Proc. Natl. Acad. Sci. U.S.A. 101 (11), 3792–3796.

Fang, S., Chen, B., Tay, F.P., Ng, B.S., Liu, D.X., 2006. An arginine-to-proline mutation in the nsp2 replicase proteins of murine hepatitis virus and severe acute respiratory syndrome-coronavirus replicative protein nsp9 is a single-stranded RNA-binding subunit unique in the RNA virus world, Proc. Natl. Acad. Sci. U.S.A. 103 (13), 5108–5113.

Ivanov, K.A., Ziebuhr, J., 2004. Human coronavirus 229E nonstructural protein 13: characterization of duplex-unwinding, nucleoside triphosphatase, and RNA 5′-triphosphatase activities, J. Virol. 78 (14), 7833–7838.

Lee, H.J., Shieh, C.K., Gorbalenya, A.E., Koomin, E.V., La Monica, N., Tuler, J., Bagdhadzhyan, A., Lai, M.M., 1991. The complete sequence (22 kilobases) of murine coronavirus genome 1 encoding the putative proteases and RNA poly- merase, Virology 180 (2), 567–582.

Lin, T.L., Loa, C.C., Wu, C.C., 2004. Complete sequences of 3′ end coding region for structural protein genes of turkey coronavirus, Virus Res. 106 (1), 61–70.

Liu, D.X., Chen, B., Xiu, H.F., Wang, S.F., 1998. Proteolytic mapping of the coronavirus infectious bronchitis virus 1b polyprotein: evidence for the presence of four cleavage sites of the 3C-like protease and identification of two novel cleavage products, Virology 246 (2), 288–297.

Loa, C.C., Lin, T.L., Wu, C.C., Bryan, T., Hooper, T., Schrader, D., 2002. Specific mucosal IgA immunity in turkey pouls infected with turkey coronavirus, Vet. Immunol. Immunopathol. 88 (1/2), 57–64.

Loa, C.C., Wu, C.C., Lin, T.L., 2006. Comparison of 3′-end encoding regions of turkey coronavirus isolates from Indiana, North Carolina, and Minnesota with chicken infectious bronchitis coronavirus strains, Intervirology 49 (4), 230–238.

McGinty, P.R., Jaffrey, S.R., Baltimore, D., 2000. A structural role for the guanylyltransferase activity of the nsp15 endoribonuclease that is essential for coronavirus replication, J. Virol. 74 (22), 10360–10367.

Matthes, N., Meister, J.R., Coudert, B., Canard, B., Snijder, E.J., Moll, R., Hilgenfeld, R., 2006. The non-structural protein Nsp10 of mouse hepatitis virus binds zinc ions and nucleic acids, FEBS Lett. 580 (17), 4143–4149.

Miki, K., Hasegawa, M., Takei, Y., Sakagami, S., Watanabe, K., Tomida, K., Ikeda, S., 1997. The 3C-like proteinase of coronavirus 229E: a novel cysteine protease in the coronavirus family, J. Virol. 71 (1), 6620–6630.

Nagaraja, K.V., Pomeroy, B.S., 1997. Coronavirus enteritis of turkeys (bluecomb disease), In: Calnek, B.W., Barnes, H.J., Beard, C.W., Reid, W.M., Yoda, H.W. (Eds.), Disease of Poultry. Iowa State University Press, Ames.

Ritchie, A.E., Deshmukh, D.R., Larsen, C.T., Pomeroy, B.S., 1973. Electron microscopy of coronavirus-like particles characteristic of turkey bluecomb disease, Avian Dis. 17 (3), 546–558.

Snijder, E.J., 2005. The coronavirus replicase, Curr. Top. Microbiol. Immunol. 287, 1–30.

Snijder, E.J., Bredenbeek, P.J., Dobbe, J.C., Thiels, V., Ziebuhr, J., Poon, L.L., Guan, Y., Roan, M., Spaan, W.J., 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 (5), 991–1004.

Sawicki, S.G., Sawicki, D.L., 1995. Coronaviruses use discontinuous extension for synthesis of subgenome-length strands, Adv. Exp. Med. Biol. 380, 499–506.

Sawicki, S.G., Sawicki, D.L., Siddell, S.G., 2007. A contemporary view of coronavirus replication, J. Virol. 81 (1), 20–29.

Schelle, B., Karl, N., Luedewig, B., Siddell, S.G., Thiels, V., 2005. Selective replication of coronavirus genomes that express nucleocapsid protein, J. Virol. 79 (11), 6620–6630.

Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity and speed of multiple sequence alignment by deleting gaps, Nucleic Acids Res. 22, 4673–4680.

Weiss, S.R., Hughes, S.A., Bonilla, P., Turner, J.D., Leibowitz, J.L., Denison, M.R., 1994. Coronavirus polyprotein processing, Arch. Virol. Suppl. 9, 349–358.

Yount, B., Burns, K.M., Baric, R.S., 2000. Strategy for systematic assembly of large DNA and RNA genomes: transmissible gastroenteritis virus model, J. Virol. 74 (22), 10600–10611.

Ziebuhr, J., 2005. The coronavirus replicase, Curr. Top. Microbiol. Immunol. 287, 57–94.

Ziebuhr, J., Snijder, E.J., Gorbalenya, A.E., 2000. Virus-encoded proteases and proteolytic processing in the Nidovirales, J. Gen. Virol. 81 (Pt 4), 853–879.