Complete sequences of 3′ end coding region for structural protein genes of turkey coronavirus

Tsang Long Lin*, Chien Chang Loa, Ching Ching Wu

Department of Veterinary Pathobiology, Purdue University, ADDL, 406 South University Street, West Lafayette, Indiana 47907-2065, USA

Received 6 October 2003; received in revised form 27 May 2004; accepted 12 June 2004

Available online 3 August 2004

Abstract

Overlapping fragments of genomic RNA spanning 6963 nucleotides from 5′ end of spike (S) protein gene to 3′ end of nucleocapsid (N) protein gene of turkey coronavirus (TCoV) were amplified by reverse-transcription-polymerase chain reaction (RT-PCR). The primers were derived from the corresponding sequences of infectious bronchitis virus (IBV). The PCR products were cloned and sequenced and their nucleic acid structure and similarity to published sequences of other coronaviruses were analyzed. Sequencing and subsequent analysis revealed 9 open reading frames (ORFs) representing the entire S protein gene, tricistronic gene 3, membrane (M) protein gene, bicistronic gene 5, and N protein gene in the order of 5′–3′. The overall nucleic acid structures of these encoding regions of TCoV were very similar to the homologous regions of IBV. The consensus transcription-regulating sequence (TRS) of IBV, CT(T/G)AACAA, was highly conserved in TCoV genome at the levels of nucleotide sequence and location in regard to the initiation codon of individual genes. Pair-wise comparison of gene 3, M gene, gene 5, or N gene sequences with their counterparts of IBV revealed high levels (82.1–92.0%) of similarity. Phylogenetic analysis based on the deduced amino acid sequences of S, M, or N protein demonstrated that TCoV was clustered within the same genomic lineage as the IBV strains while all the other mammalian coronaviruses were grouped into separate clusters corresponding to antigenic groups I or II. There were substantial differences of S protein sequence between TCoV and IBV with only 33.8–33.9% of similarity.

© 2004 Elsevier B.V. All rights reserved.

Keywords: Turkey coronavirus; Infectious bronchitis virus; Coronavirus; Genomic relationship

1. Introduction

Turkey coronavirus (TCoV) was identified in the early 1970s as the major causative agent of the most costly disease of turkey encountered in Minnesota between 1951 and 1971 (Nagaraja and Pomeroy, 1997). Outbreaks of turkey poult enteritis associated with TCoV have caused severe economical losses in the turkey industry in Indiana, North Carolina, and other states for the last several years. Although the economical importance of this disease has been recognized for decades, the organization of genomic structure of TCoV is poorly understood and reports regarding the relationships of TCoV with other coronaviruses remained controversial (Van Regenmortel et al., 2000; Gonzalez et al., 2003).

Coronaviruses are pleomorphic, enveloped spherical particles surrounded by a fringe of 20nm long club-shaped spikes. The diameter of coronaviral particles are around 140–150nm. The coronavirus genome is a positive single-stranded capped RNA with a polyadenylated 3′ end. Complete genomic RNA sequences of coronaviruses has been determined for infectious bronchitis virus (IBV; 27,569 nucleotides; Boursnell et al., 1987), murine hepatitis virus (MHV; 31,092 nucleotides; Lee et al., 1991), human coronavirus (HCoV) strain 229E (27,277 nucleotides; Herold et al., 1993), and transmissible gastroenteritis virus (TGEV; 28,579 nucleotides; Eleouet et al., 1995; Penzes et al., 2001). The 5′ two-thirds of the coronavirus genome, approximately 20kb, consists of two overlapping open reading frames (ORFs) that encode non-structural proteins including the viral RNA-dependent RNA polymerase and proteases. Another one-third nucleotide sequences from 3′ end contain ORFs encode the major structural proteins:

* Corresponding author. Fax: +1 765 494 9181.
E-mail address: tllin@purdue.edu (T.L. Lin).
spike (S), membrane (M), and nucleocapsid (N) proteins in the order of 5′-3′ along the genome, respectively.

Turkey coronavirus was initially determined to be antigenically distinct from all other coronaviruses based on antigenic differences revealed by immunoelectron microscopy (Ritchie et al., 1973) and hemagglutination-inhibition (Dea et al., 1986). This unique antigenicity was questioned when the close relationship between TCoV and bovine coronavirus (BCoV) was demonstrated in a series of antigenic studies (Dea et al., 1990) and by sequence analysis of TCoV M and N genes (Verbeek and Tijssen, 1991). In contrast, recent antigenic (Guy et al., 1997; Loa et al., 2000) and genomic (Breslin et al., 1999a,b; Akin et al., 2001; Cavanagh et al., 2001, 2002; Lin et al., 2002) analysis of TCoV, however, demonstrated that TCoV and IBV, two avian coronaviruses, are closely related. The causes for these discrepant results regarding the relationships of TCoV with BCoV or IBV remained unclear. Further analysis of genomic structure of TCoV is important to clarify this enigma. Thus, the purpose of the present study was to determine the sequences of the 3′ end coding region for structural protein genes of TCoV.

2. Materials and methods

2.1. Turkey coronavirus

The TCoV isolate (isolate 540) used in the present study were recovered from fecal contents and intestines of turkey poults with acute coronaviral enteritis in Indiana, US in 1994. The viruses were passaged 5 times in 22-day-old embryonating turkey eggs. The presence of TCoV in the intestines of embryos were confirmed by TCoV-specific immunofluorescence antibody assays and electron microscopy at the Indiana State Animal Disease Diagnostic Laboratory in West Lafayette, Indiana, US.

2.2. RNA isolation and reverse transcription

Total RNA was extracted from the intestines and intestinal content of turkey embryo infected with TCoV by a modified method using guanidinium thiocyanate and acid-phenol (Chomczynski and Sacchi, 1987; Akin et al., 1999). Conversion of total RNA to cDNA was essentially performed according to a protocol supplied by the manufacturer of the reverse transcriptase (Superscript II system, Life Technologies, Gaithersburg, MD).

2.3. PCR amplification

Three microliters of cDNA were used in PCR amplifications with the primers designed from IBV genomic sequences. The locations and sequences of primers for the amplification of 4 fragments I–IV for 3′ end coding region of TCoV structural protein genes are outlined in Fig. 1. PCR was performed with a mixture (64:1, v:v) of Taq (Promega Corp., Madison, WI) and Pfu polymerases (Stratagene, La Jolla, CA) in a 96-well thermal cycler (GeneAmp, Perkin-Elmer Cetus Corp., Norwalk, CT) (Barnes, 1994; Akin et al., 1999). The cyclic parameters of the PCR was as follows: 94 °C for 1 min for denaturation, 37 °C for 2 min for annealing, and 72 °C for 5 min for extension for 40 cycles followed by 72 °C for 10 min for final extension.

2.4. Molecular cloning and sequencing

One microliter of the amplification product was used to ligate with pcRII plasmid vector according to the manufacturer’s instructions (Invitrogen, San Diego, CA). Determination of the nucleotide sequences of the selected clone with amplified sequences was performed by dyeoxy-cycle sequencing method with the corresponding sequencing primers for both strands (DAVIS Sequencing, Davis, CA).

2.5. Sequence analysis

The nucleotide and deduced amino acid sequences between the TCoV and other coronaviruses were analyzed by DNAstar program (Lasergene Corp, Madison, WI), respectively. Percent similarities were calculated to find nucleic
Fig. 2. Nucleotide sequence of the amplified fragments containing entire spike (S) protein gene, gene 3, membrane (M) protein gene, gene 5, and nucleocapsid protein gene region of turkey coronavirus (TCoV) and their similarity to those of infectious bronchitis (IBV) strain Beaudette (GenBank accession number AJ311317). The positions where nucleotide bases are missing are indicated as (-) and identical nucleotides as (.). Heavy underlines below the sequence of TCoV indicate the putative start codons. Light lines above the sequence of TCoV indicate the stop codons. The conserved transcription-regulating nucleotide sequence (A/C)T(T/G)AACAA, which is located upstream from the start codons of individual genes, is boxed. The start codon of IBV M protein gene is also underlined because it is at different position from that of TCoV M protein gene.
Fig. 2. (Continued)
Fig. 2. (Continued)
acid and amino acid pair distances. Based on the obtained sequences of TCoV and previously published sequences of different coronaviruses, phylogenetic trees were constructed according to the coding sequences for S, M, and N genes.

3. Results

3.1. Complete nucleotide sequences of 3′ end coding region for structural protein genes of turkey coronavirus

Cloning and sequencing of the 4 overlapping fragments revealed a total of 6963 nucleotides in a region containing the entire S protein gene, tricistronic gene 3, M protein gene, bicistronic gene 5, and N protein gene of TCoV in the present study. The primary structures of the coding sequences for these genes of TCoV in the present study were very similar to those found in the corresponding genomic regions of IBV strain Beaudette as shown in Fig. 2 and Table 1. The canonical consensus transcription-regulating sequence (TRS) of IBV, CT(T/G)AACAA, was also found in TCoV in the present study. Both the nucleotide sequence of the TRS and the distance between the 3′ end of the TRS and the initiation codon of the downstream adjacent ORF were highly conserved between TCoV in the present study and IBV (Table 1).

<table>
<thead>
<tr>
<th>Virus</th>
<th>Gene</th>
<th>ORF size (nucleotides)</th>
<th>TRS sequence</th>
<th>RS distance (nucleotides)</th>
</tr>
</thead>
<tbody>
<tr>
<td>TCoV</td>
<td>Spike</td>
<td>3612</td>
<td>ctagaacaa</td>
<td>52</td>
</tr>
<tr>
<td></td>
<td>3a</td>
<td>174</td>
<td>ctagaacaa</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>3b</td>
<td>195</td>
<td>ctagaacaa</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>3c</td>
<td>318</td>
<td>ctagaacaa</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>Membrane</td>
<td>669</td>
<td>ctagaacaa</td>
<td>74</td>
</tr>
<tr>
<td></td>
<td>5a</td>
<td>198</td>
<td>ctagaacaa</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>5b</td>
<td>243</td>
<td>ctagaacaa</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>Nucleocapsid</td>
<td>1230</td>
<td>ctagaacaa</td>
<td>93</td>
</tr>
<tr>
<td>IBV</td>
<td>Spike</td>
<td>3489</td>
<td>ctagaacaa</td>
<td>52</td>
</tr>
<tr>
<td></td>
<td>3a</td>
<td>174</td>
<td>ctagaacaa</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>3b</td>
<td>195</td>
<td>ctagaacaa</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>3c</td>
<td>330</td>
<td>ctagaacaa</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>Membrane</td>
<td>678</td>
<td>ctagaacaa</td>
<td>77</td>
</tr>
<tr>
<td></td>
<td>5a</td>
<td>198</td>
<td>ctagaacaa</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>5b</td>
<td>249</td>
<td>ctagaacaa</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>Nucleocapsid</td>
<td>1230</td>
<td>ctagaacaa</td>
<td>93</td>
</tr>
</tbody>
</table>

* ORF: open reading frame.
* TRS: transcription-regulating sequence.
* The distance is calculated as nucleotides between 3′ end of TRS and the ATG start codon of the corresponding first downstream ORF.

Fig. 2. (Continued).
Table 2
Sequence pair distances for nucleic acid and deduced amino acid sequence of the entire spike protein gene region of turkey coronavirus (TCoV) with other coronaviruses

<table>
<thead>
<tr>
<th>Nucleotide identity (%)</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
</tr>
</thead>
<tbody>
<tr>
<td>TCeCoV</td>
<td>100</td>
<td>95.1</td>
<td>95.4</td>
<td>95.7</td>
<td>95.0</td>
<td>94.9</td>
<td>94.8</td>
<td>94.8</td>
<td>94.3</td>
<td>94.2</td>
<td>94.2</td>
<td>94.2</td>
<td>94.2</td>
<td>94.1</td>
<td>94.1</td>
<td>93.9</td>
</tr>
<tr>
<td>TCeCoV-Gb</td>
<td>98.2</td>
<td>100</td>
<td>96.8</td>
<td>96.1</td>
<td>96.2</td>
<td>96.4</td>
<td>96.4</td>
<td>96.4</td>
<td>96.4</td>
<td>96.4</td>
<td>96.4</td>
<td>96.4</td>
<td>96.4</td>
<td>96.4</td>
<td>96.4</td>
<td>96.4</td>
</tr>
<tr>
<td>IBV-KB</td>
<td>35.8</td>
<td>35.1</td>
<td>35.2</td>
<td>35.7</td>
<td>35.7</td>
<td>35.7</td>
<td>35.7</td>
<td>35.7</td>
<td>35.7</td>
<td>35.7</td>
<td>35.7</td>
<td>35.7</td>
<td>35.7</td>
<td>35.7</td>
<td>35.7</td>
<td>35.7</td>
</tr>
<tr>
<td>IBV-Beau</td>
<td>33.9</td>
<td>33.1</td>
<td>33.2</td>
<td>33.6</td>
<td>33.7</td>
<td>33.7</td>
<td>33.7</td>
<td>33.7</td>
<td>33.7</td>
<td>33.7</td>
<td>33.7</td>
<td>33.7</td>
<td>33.7</td>
<td>33.7</td>
<td>33.7</td>
<td>33.7</td>
</tr>
<tr>
<td>IBV-Dund</td>
<td>25.4</td>
<td>24.9</td>
<td>24.5</td>
<td>23.8</td>
<td>23.2</td>
<td>23.2</td>
<td>23.2</td>
<td>23.2</td>
<td>23.2</td>
<td>23.2</td>
<td>23.2</td>
<td>23.2</td>
<td>23.2</td>
<td>23.2</td>
<td>23.2</td>
<td>23.2</td>
</tr>
<tr>
<td>IBV-KB</td>
<td>25.2</td>
<td>24.8</td>
<td>24.4</td>
<td>24.0</td>
<td>23.5</td>
<td>23.5</td>
<td>23.5</td>
<td>23.5</td>
<td>23.5</td>
<td>23.5</td>
<td>23.5</td>
<td>23.5</td>
<td>23.5</td>
<td>23.5</td>
<td>23.5</td>
<td>23.5</td>
</tr>
<tr>
<td>IBV-Beau</td>
<td>25.9</td>
<td>25.2</td>
<td>25.6</td>
<td>24.7</td>
<td>24.8</td>
<td>24.8</td>
<td>24.8</td>
<td>24.8</td>
<td>24.8</td>
<td>24.8</td>
<td>24.8</td>
<td>24.8</td>
<td>24.8</td>
<td>24.8</td>
<td>24.8</td>
<td>24.8</td>
</tr>
<tr>
<td>IBV-Dund</td>
<td>22.3</td>
<td>21.8</td>
<td>21.7</td>
<td>20.3</td>
<td>20.2</td>
<td>20.1</td>
<td>20.1</td>
<td>20.1</td>
<td>20.1</td>
<td>20.1</td>
<td>20.1</td>
<td>20.1</td>
<td>20.1</td>
<td>20.1</td>
<td>20.1</td>
<td>20.1</td>
</tr>
<tr>
<td>IBV-Beau</td>
<td>26.4</td>
<td>25.4</td>
<td>25.9</td>
<td>25.1</td>
<td>25.4</td>
<td>25.3</td>
<td>25.3</td>
<td>25.3</td>
<td>25.3</td>
<td>25.3</td>
<td>25.3</td>
<td>25.3</td>
<td>25.3</td>
<td>25.3</td>
<td>25.3</td>
<td>25.3</td>
</tr>
<tr>
<td>IBV-Dund</td>
<td>19.5</td>
<td>18.6</td>
<td>18.8</td>
<td>18.6</td>
<td>18.6</td>
<td>18.6</td>
<td>18.6</td>
<td>18.6</td>
<td>18.6</td>
<td>18.6</td>
<td>18.6</td>
<td>18.6</td>
<td>18.6</td>
<td>18.6</td>
<td>18.6</td>
<td>18.6</td>
</tr>
</tbody>
</table>

Amino acid identity (%)

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>95.1</td>
<td>95.4</td>
<td>95.7</td>
<td>95.0</td>
<td>94.9</td>
<td>94.8</td>
<td>94.8</td>
<td>94.3</td>
<td>94.2</td>
<td>94.2</td>
<td>94.2</td>
<td>94.2</td>
<td>94.2</td>
<td>94.2</td>
<td>94.2</td>
</tr>
<tr>
<td>98.2</td>
<td>100</td>
<td>96.8</td>
<td>96.1</td>
<td>96.2</td>
<td>96.4</td>
<td>96.4</td>
<td>96.4</td>
<td>96.4</td>
<td>96.4</td>
<td>96.4</td>
<td>96.4</td>
<td>96.4</td>
<td>96.4</td>
<td>96.4</td>
<td>96.4</td>
</tr>
<tr>
<td>35.8</td>
<td>35.1</td>
<td>35.2</td>
<td>35.7</td>
<td>35.7</td>
<td>35.7</td>
<td>35.7</td>
<td>35.7</td>
<td>35.7</td>
<td>35.7</td>
<td>35.7</td>
<td>35.7</td>
<td>35.7</td>
<td>35.7</td>
<td>35.7</td>
<td>35.7</td>
</tr>
<tr>
<td>33.9</td>
<td>33.1</td>
<td>33.2</td>
<td>33.6</td>
<td>33.7</td>
<td>33.7</td>
<td>33.7</td>
<td>33.7</td>
<td>33.7</td>
<td>33.7</td>
<td>33.7</td>
<td>33.7</td>
<td>33.7</td>
<td>33.7</td>
<td>33.7</td>
<td>33.7</td>
</tr>
<tr>
<td>25.4</td>
<td>24.9</td>
<td>24.5</td>
<td>23.8</td>
<td>23.2</td>
<td>23.2</td>
<td>23.2</td>
<td>23.2</td>
<td>23.2</td>
<td>23.2</td>
<td>23.2</td>
<td>23.2</td>
<td>23.2</td>
<td>23.2</td>
<td>23.2</td>
<td>23.2</td>
</tr>
<tr>
<td>25.2</td>
<td>24.8</td>
<td>24.4</td>
<td>24.0</td>
<td>23.5</td>
<td>23.5</td>
<td>23.5</td>
<td>23.5</td>
<td>23.5</td>
<td>23.5</td>
<td>23.5</td>
<td>23.5</td>
<td>23.5</td>
<td>23.5</td>
<td>23.5</td>
<td>23.5</td>
</tr>
<tr>
<td>25.9</td>
<td>25.2</td>
<td>25.6</td>
<td>24.7</td>
<td>24.8</td>
<td>24.8</td>
<td>24.8</td>
<td>24.8</td>
<td>24.8</td>
<td>24.8</td>
<td>24.8</td>
<td>24.8</td>
<td>24.8</td>
<td>24.8</td>
<td>24.8</td>
<td>24.8</td>
</tr>
<tr>
<td>22.3</td>
<td>21.8</td>
<td>21.7</td>
<td>20.3</td>
<td>20.2</td>
<td>20.1</td>
<td>20.1</td>
<td>20.1</td>
<td>20.1</td>
<td>20.1</td>
<td>20.1</td>
<td>20.1</td>
<td>20.1</td>
<td>20.1</td>
<td>20.1</td>
<td>20.1</td>
</tr>
<tr>
<td>26.4</td>
<td>25.4</td>
<td>25.9</td>
<td>25.1</td>
<td>25.4</td>
<td>25.3</td>
<td>25.3</td>
<td>25.3</td>
<td>25.3</td>
<td>25.3</td>
<td>25.3</td>
<td>25.3</td>
<td>25.3</td>
<td>25.3</td>
<td>25.3</td>
<td>25.3</td>
</tr>
<tr>
<td>19.5</td>
<td>18.6</td>
<td>18.8</td>
<td>18.6</td>
<td>18.6</td>
<td>18.6</td>
<td>18.6</td>
<td>18.6</td>
<td>18.6</td>
<td>18.6</td>
<td>18.6</td>
<td>18.6</td>
<td>18.6</td>
<td>18.6</td>
<td>18.6</td>
<td>18.6</td>
</tr>
</tbody>
</table>

* TCeV: a US, Indiana, isolate of TCoV
* TCeCoV: an isolate of TCeV
* TCeCoV-Gb: an isolate of TCeV
* IBV: infectious bronchitis virus
* IBV-KB: a Japanese strain
* IBV-Beau: a US strain
* BCoV: bovine coronavirus
* CCoV: canine coronavirus
* FCoV: feline coronavirus
* HCoV: human coronavirus
* HCoV-Oc43: human coronavirus
* HCoV-229E: human coronavirus
* HCoV-NL63: human coronavirus
* HCoV-HKU1: human coronavirus
* MHV: murine hepatitis virus
* PEDV: porcine epidemic diarrhea virus
* TGEV: porcine transmissible gastroenteritis virus
* SARS: severe acute respiratory syndrome coronavirus
3.2. Sequence comparison and phylogenetic analysis

Pair-wise comparison of nucleotide and deduced amino acid sequence distances between TCoV S protein gene in the present study and the homologous gene sequences of other known coronaviruses is summarized in Table 2. The similarity score between TCoV and other non-TCoV coronaviruses within the S protein gene region ranged from 38.3% to 52.5% at the nucleotide level or from 19.5% to 33.9% at the amino acid level. The similarity score between TCoV in the present study and IBV strains within the M or N protein gene region were high (>80%). In contrast, the similarity score between TCoV in the present study and other mammalian coronaviruses within the M or N protein gene region ranged from 24.8% to 30.8% for nucleotide sequence and from 16.9% to 29.1% for deduced amino acid sequence. The tricistronic gene 3 with 3 overlapping ORFs, 3a–c, in between S and M genes as well as the dicistronic gene 5 with 2 overlapping ORFs, 5a and 5b, in between M and N genes of TCoV in the present study all shared high similarity with the corresponding genomic sequences of IBV strains.

Phylogenetic analysis according to the deduced amino acid sequence of S, M, or N proteins indicated that TCoV in the present study was clustered within the same genomic lineage with IBV strains while all the other mammalian coronaviruses were grouped into separate clusters corresponding to antigenic groups I and II (data not shown).

4. Discussion

Total 6963 nucleotides of TCoV genome were cloned and sequenced in the present study. This region is likely to include all of the viral genes excluding the polymerase gene and, thus, provides substantial genetic information of the virus for comparison with other coronaviruses. The genomic structures of ORFs for S protein, 3a–c, M protein, 5a–b, and N protein were very similar to those of IBV. The phylogenetic analysis based on the deduced amino acid sequences of S, M, or N protein all showed that the TCoV in the present study was classified within the same genomic lineage with IBV strains while all the other mammalian coronaviruses including BCoV were grouped to separate clusters. The nucleotide sequences of ORFs for 3a–c, M protein, 5a–b, and N protein of TCoV shared high similarity (82.1–92.0%) with the corresponding sequences of IBV. These results clearly demonstrated the close relationship of TCoV in the present study to avian IBV.

The presence of tricistronic gene 3 in between 3’ end of S gene and 5’ end of M gene as well as the presence of dicistronic gene 5 in between 3’ end of M gene and 5’ end of N gene are particular features of avian coronaviruses, TCoV and IBV. These particular genomic structures are not found in any other mammalian coronaviruses as determined to date. These distinct features of genome structure implied that TCoV shares a relatively close evolutionary relationship with IBV.

The predicted proteins of ORF 3a–c, 5a and 5b were small (about or less than 10kd). The functions of these gene products are not known. Several ORFs encoding non-structural proteins have been recognized in coronavirus genomes (Boursnell et al., 1987; Lee et al., 1991; Herold et al., 1993; Eleouet et al., 1995). The number, nucleotide sequence, and gene order of these ORFs varied remarkably among different coronaviruses. It is speculated that these genes were inserted into different sites in the coronavirus genomes due to the RNA recombination-prone discontinuous transcription mechanism and were not essential for virus replication and pathogenesis. However, sequence analysis in the present study indicated that both nucleotide sequences and locations of these ORFs and their consensus TRS of TCoV are highly conserved with those of IBV. Given such a highly conserved sequences and structures within avian coronaviruses, genes 3 and 5 may play important roles in coronavirus pathogenesis to avian species.

One of the characteristic features for coronavirus replication is the synthesis of a 3’ coterminal nested set of polycistronic subgenomic mRNAs by a discontinuous transcription mechanism. Several conserved TRS have been identified for different coronavirus proximal to the initiation codon of the first ORF for each particular subgenomic mRNA. The consensus sequences of the TRS sites are CTG(AG/C)CAA for IBV, ATC(T/C/AC for BCoV, AACTAACAC for TGEV, AACT(CT/CA/(T/A)AC for MHV, and AAATTAAC for FIPV (Spaan et al., 1988; Stirrups et al., 2000). The distance between the TRS and the first ORF is different for each subgenomic mRNA of different coronaviruses. Both the nucleotide sequence of TRS and the distance between the TRS 3’ end and the initiation codon of first ORF are suggested to play important role in the transcription of mRNAs. As shown in the present study, the TRS sequences of TCoV were found highly conserved with those of the corresponding genes of IBV except one nucleotide substitution in that of gene 3. The TRS of gene 3 is ATGAACAA for TCoV and CTGAACAA for IBV. The distances between TRS and initiation codon of S gene, gene 3, gene 5, and N gene of TCoV were all the same as those of IBV while the distances for TCoV or IBV M gene are 74 or 77, respectively. The highly conserved sequences and structures of TRS between TCoV and IBV provide further evidence that these two avian coronaviruses share close evolutionary relationship. These highly conserved TRS sequences of IBV has been shown to be recombination “hot spot” and may serve as the template switching sites for the viral encoded RNA dependent RNA polymerase (Lee and Jackwood, 2000). These recombination events play important role to the emergence of new IBV variants responsible for continuous outbreaks in the chicken flocks vaccinated with live attenuated viruses due to failure of cross protection. It is possible that the similar recombination events of IBV in chicken may contribute to the origin and evolution of TCoV in turkey and merit further investigation.

Even though the close genetic relationship between TCoV and IBV was clearly demonstrated as discussed above,
these two avian coronaviruses are dramatically different at the S protein gene level. The similarity of S protein sequences between TCoV in the present study and IBV strains (33.8–33.9%) is much lower than that among IBV strains (83.2–94.0%). The difference of nucleotide between TCoV in the present study and IBV seems to be randomly distributed throughout the entire S gene except a stretch of 225 nucleotides from the 3′ end that shared high similarity (88.9%) with the corresponding sequences of IBV. These observations suggested that cross-over homologous recombination, very likely by a template switching mechanism, occurred around the consensus TRS site of S gene and within the 3′ end 225 nucleotides region (involving the TRS site of gene 3) and resulted in a whole new codon reading frame for S protein of TCoV with conserved TRS and other genomic structure features of IBV. Spike protein of coronaviruses has been well known as the major structural protein responsible for attachment, fusion, and penetration of virions to the target cells. The substantial difference of S protein gene between TCoV and IBV well explains the different host tropism and different tissue pathogenicity of these two avian coronaviruses.

Turkey coronavirus is associated with enteric disease of turkey while IBV is usually associated with respiratory disease in chicken.

Two group-specific monoclonal antibodies, which reacted with a broad spectrum of homologous and heterologous IBV serotypes, were tested for reactivity with TCoV in a previous study (Loa et al., 2000). The antibody specific to M protein (Mab 919) of IBV had strong cross-reactivity with TCoV but the antibody specific to S protein (Mab 94) of IBV did not react with TCoV. In line with these previous observations of antigenicity, the sequence analysis in the present study revealed a high homology of M protein gene between TCoV in the present study and IBV. On the other hand, the difference of S protein gene between TCoV in the present study and IBV is substantial. Therefore, molecular diagnostic assay or antigenic analysis using antibody specific to S protein or gene will be useful tools to differentiate TCoV from IBV.

The results of sequence analysis in the present study stress the close relationship of TCoV to IBV. Coronavirus genomes are dynamic with high frequency of recombination, insertion, and deletion, subsequently, may result in significant genetic differences. Further cloning and sequencing analysis of full-length genomic sequences of more TCoV isolates are under way for revealing a faithful picture of the TCoV genome.

Acknowledgements

The authors thank the financial support provided by US Department of Agriculture.

References


Penzes, Z., Gonzalez, J.M., Calvo, E., Izeta, A., Smerdou, C., Mendez, A.,
genome sequence of transmissible gastroenteritis coronavirus PUR46-
MAD clone and evolution of the purdue virus cluster. Virus Genes
23, 105–118.
tron microscopy of coronavirus-like particles characteristic of turkey
Stimpfs, K., Show, K., Evans, S., Dalton, K., Casais, R., Cavanagh, D.,
Virol. 81, 1667–1689.
Verbeek, A., Tijssen, P., 1991. Sequence analysis of the turkey en-
teric coronavirus nucleocapsid and membrane protein genes: a close
genomic relationship with bovine coronavirus. J. Gen. Virol. 72,
1659–1666.
Van Regenmortel, M.H.V., Fauquet, C.M., Bishop, D.H.L., Carstens, E.B.,
Estes, M.K., Lemon, S.M., Maniloff, J., Mayo, M.A., McGeoch, D.J.,
and nomenclature of viruses. The Seventh Report of the International
Committee on Taxonomy of Viruses. Academic Press, San Diego,
California.