Full-Length Genomic Sequence of Bovine Coronavirus (31kb)

Completion of the open reading frame 1a/1b sequences

DONGWAN YOO* and YANLONG PEI

*Department of Pathobiology, Ontario Veterinary College, University of Guelph, Ontario N1G 2W1 Canada. *Corresponding author: Email:dyoo@uoguelph.ca*

1.   INTRODUCTION

Bovine coronavirus (BCV) is an important veterinary pathogen which causes neonatal diarrhea in newborn calves and winter dysentery in adult cattle. Recent studies indicate that BCV also infects the respiratory tract of cattle producing severe respiratory problems, especially in feedlot cattle. BCV belongs to the antigenic group II of coronaviruses and shares the antigenic and genetic similarities with mouse hepatitis virus (MHV), human coronavirus strain OC43, turkey coronavirus, and hemagglutinating encephalomyelitis virus of pigs. Approximately 10 kb of the 3' most BCV genome has been sequenced, and this region is known to encode all the structural proteins plus the 32k nonstructural protein which resides immediately upstream of the hemagglutinin-esterase (HE) glycoprotein gene. The comparative studies indicate that the genome organization of BCV is similar to but distinct from MHV. In BCV, the HE gene is functional coding for a major envelop protein while it is an optional gene in MHV. Similarly, the 32k protein gene is absent in some of the MHV variants. The region between the spike (S) protein gene and the small membrane (E) protein gene also shows a significant sequence divergence among coronaviruses. Two thirds of the BCV genome remain largely undetermined.

To date, coronaviruses of which the full length genomic sequences are available include mouse hepatitis virus (MHV) (Bredenbeek et al, 1990; Lee etal, 1991; Bonilla et al, 1994), human coronavirus 229E (Herold et al,

1994), porcine transmissible gastroenteritis virus (TGEV) (Almazan et al, 2000), and avian infectious bronchitis virus (IBV) (Boursnell et al 1987). Of these viruses, human coronavirus 229E and TGEV fall within the same antigenic group (group I), and IBV is a group III virus. In the antigenic group II which BCV belongs to, MHV is the only virus that the full-length sequence has been determined. My laboratory has recently completed sequencing of the entire genome of BCV, and in this report we present the structural organization of the BCV genome.

2.   MATERIALS AND METHODS

2.1   cDNA Cloning

The Quebec strain of BCV was propagated in Mardin-Darby bovine kidney (MDBK) cells. Genomic RNA was prepared from the purified virions, and double stranded cDNA was synthesized by standard methods. After addition of Bam HI linkers, cDNAs were ligated into the Bam HI site of pTZ19R and used transform E. coli strain DH1. Specific clones were identified by colony hybridization using a ^{32}P-labeled cDNA probe prepared by randomly primed reverse transcription of viral genomic RNA or by PCR. For RT-PCR, total cellular RNA was extracted from virus-infected cells using Trizol and used as a template for cDNA synthesis.

2.2   Reverse Transcription Polymerase Chain Reaction

For first-stranded cDNA synthesis, Superscript II RNase H Negative Reverse Transcriptase (GIBCO BRL) was used. The first strand cDNA was PCR-amplified using Vent DNA polymerase (New England Biolab) in GeneAmp Thermocycler PE2400 (Perkin Elmer).

2.3   DNA Sequencing

The nucleotide sequences were determined either by manual sequencing or BY using an automated DNA sequencer. For manual sequencing, a series of overlapping clones were generated and the overlapping clones were sequenced using universal primers. For automated sequencing, new primers were generated based on the determined sequence to walk along the unsequenced region. Sequences were assembled and analyzed using the GCG Wisconsin sequence analysis package. The sequence has been deposited to the GenBank database under accession number AF220295.

3. RESULTS AND DISCUSSION

The overlapping sequences were assembled as a single contiguous sequence to represent the full-length genomic sequence of BCV. The BCV genome appeared to comprise of 31043 nucleotides, excluding the 3' terminal polyadenylation tail. The BCV sequence was 291 nucleotides shorter than the genome of MHV A59, and the two viruses shared an overall 71% sequence identity at the nucleotides level of the full-length genome. Open reading frame 1a (ORF1a) was able to code for a polypeptide of 4383 amino acids. ORF1a and 1b overlapped by 28 nucleotides, and this overlap was significantly shorter than that of MHV A59. A slippery sequence was identified within the overlapping region suggesting that a 1a/1b fusion protein of 7061 amino acids would be produced (Figure 1).

```
                                        slippery sequence        BCV ORF1a stop
                                             |                        |
+3   G  M  E  A  V  H  V  L  A  L  T  L  L  F  S  Q  K  I  L  I  F  #  T  G  S  G  Y  E  C▼R  C  P  S  R
+2   R  D  G  S  C  S  C  V  S  T  D  T  T  V  Q  S  K  D  T  N  F  L  N  G  F  G  V  R  V  #  M  P  V  S
+1   A  G  W  K  L  F  M  C  #  H  #  H  Y  C  S  V  K  R  Y  #  F  F▼K  R  V  R  G  T  S  V  D  A  R  L
5'  CGGGATGGAAGCTGTTCATGTGTTAGCACTGACACTACTGTTCAGTCAAAAGATACTAATTTtttaaacGGGTTCGGGGTACGAGTGTAGATGCCCGTCTCG
    || || || |||| || ||||| | || | |     ||||||||||| || |||||||||||||||||||| ||||| ||||||| | |||||||||| |
5'  CGAGACGGTAGCTGCTCCTGTGTAGGTACAGGCTCCCAGTTCAGTCAAAAGACACGAATTTtttaaacGGGTTCGGGGTACAAGTGTAAATGCCCGTCTTG
+1   R  D  G  S  C  S  C  V  G  T  G  S  Q  F  Q  S  K  D  T  N  F  L▲N  G  F  G  V  Q  V  #  M  P  V  L
+2   E  T  V  A  A  P  V  #  V  Q  A  P  S  F  S  Q  K  T  R  I  F  ▲T  G  S  G  Y  K  C▲K  C  P  S  C
+3      R  R  #  L  L  L  C  R  Y  R  L  P  V  S  V  K  R  H  E  F  F  K  R  V  R  G  T  S  V  N  A  R  L  V
              ORF1b start                                         |                        |
                                                         slippery sequence          JHM ORF1a stop
```

Figure 1. Sequence comparison of the ORF1a/ORF1b ribosomal frame shifting region of BCV and MHV JHM. Slippery sequences are illustrated in lower case and indicated by arrows. Stop codons for ORF1a (TAG for BCV and TAA for JHM) are depicted in boldface and indicated by arrow. Frames are indicated in numbers, and amino acid sequences encoded in the open reading frames 1a and 1b are underlined.

The BCV full-length genomic sequence was compared with that of other coronaviruses (Table 1). The BCV genome exhibited the highest sequence homology with the genome of MHV A59 (70.8%), but the homologies were only 48.3%, 52.5%, and 53%, when compared with IBV, HCV 229E, and TGEV, respectively. The 1a/1b polyprotein of BCV retained a similar level of amino acid identity to that of A59 (72.9%). The 1a sequence diversed more with only 68% homology while the 1b homology was 86%. The sequence conservation of the 1a/1b polyprotein was low in IBV, 229E, and TGEV, with an overall homology of only 36% (Table 1). Nevertheless, all the putative functional domains that were identified in all other coronaviruses were found to have been conserved in BCV, including the motifs for two separate papain-like cysteine proteinases, a 3C-like chymotrypsin proteinase, a cysteine rich murine growth factor-like, a RNA dependent RNA polymerase, a zinc finger-like, and a helicase.

Putative proteolytic cleavage sites specific for the papain-like proteinase were predicted at RG246/V and A851/G. In addition to these sites, 11 additional cleavage sites (9 sites in 1a and 4 sites in 1b) were identified as potential catalytic sites for 3C-like proteinase: Q3247/S, Q3549S, Q3836/S, Q3928/S. Q4122/N, Q4232/A, Q4369/S, Q5299/S, Q5902/C, Q6423/S, and Q6792/A.

Table 1. Sequence identities of the BCV genome with other coronaviruses

Virus	Full genomic sequence (nts)	1a/1b fusion protein sequence (aa)
MHV A59	70.8%	72.9%
TGEV	53.0%	36.4%
HCV 229E	52.5%	35.7%
IBV	48.3%	36.5%

ACKNOWLEDGMENTS

REFERENCES

Almazan, F., Gonzalez, J. M., Penzes, Z., Izeta, A., Calvo, E., Plana_Duran, J., and Enjuanes, L. 2000. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 5516_5521.
Bonilla, P. J., Gorbalenya, A. E. and Weiss, S. R. 1994. *Virology* **198**, 736-740.
Boursnell, M. E., Brown, T. D., Foulds, I. J., Green, P. F., Tomley, F. M. and Binns, M. M. 1987. *J. Gen. Virol.* **68**, 57-77.
Bredenbeek, P. J., Pachuk, C. J., Noten, A. F., Charite, J., Luytjes, W., Weiss, S. R. and Spaan, W. J. 1990. *Nucl. Acids Res.* **18**, 1825-1832.
Herold, J., Raabe, T., Schelle_Prinz, B. and Siddell, S. G. 1993. *Virology* **195**, 680-691.
Lee, H. J., Shieh, C. K., Gorbalenya, A. E., Koonin, E. V., La Monica, N., Tuler, J., Bagdzhardzhyan, A. and Lai, M. M. C. 1991. *Virology* **180**, 567-582.
Storz, J., Stine, L., Liem, A., Anderson, G. A. 1996. *J. Am. Vet. Med. Assoc.* **208**:1452-1455.