



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.

The complete sequence of the bovine torovirus genome

Ryan Draker^{a,1}, Rachel L. Roper^b, Martin Petric^c, Raymond Tellier^{a,d,*}

^a Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ont., Canada

^b Department of Microbiology and Immunology, Brody School of Medicine, East Carolina University, NC, USA

^c British Columbia Center for Disease Control, Vancouver, BC, Canada

^d Division of Microbiology, The Hospital for Sick Children, 555 University Avenue, Toronto, Ont., Canada M5G 1X8

Received 16 March 2005; received in revised form 5 July 2005; accepted 12 July 2005

Available online 30 August 2005

Abstract

Viruses in the family *Coronaviridae* have elicited new interest, with the outbreaks caused by SARS-HCoV in 2003 and the recent discovery of a new human coronavirus, HCoV-NL63. The genus *Torovirus*, within the family *Coronaviridae*, is less well characterized, in part because toroviruses cannot yet be grown in cell culture (except for the Berne virus). In this study, we determined the sequence of the complete genome of Breda-1 (BoTV-1), a bovine torovirus. This is the first complete torovirus genome sequence to be reported. BoTV-1 RNA was amplified using long RT-PCR and the amplicons sequenced. The genome has a length of 28.475 kb and consisted mainly of the replicase gene (~20.2 kb) which contains two large overlapping ORFs, ORF1a and ORF1b, encoding polyproteins pp1a and pp1b, respectively. Sequence analysis identified conserved domains within the predicted sequences of pp1a and pp1b. Sequence alignments and protein secondary structure prediction data suggest the presence of a 3C-like serine protease domain with similarity to the arterivirus 3C-like serine protease and a single papain-like cysteine protease domain with similarity to the picornavirus leader protease. The ADRP (APPR-1'') domain – unique to the *Coronaviridae* – was also located in BoTV pp1a. In addition, several hydrophobic domains were identified that are typical of a nidovirus replicase. Within the pp1b sequence the polymerase and helicase domains were identified, as well as sequences predicted to be involved in ribosomal frameshifting, including the conserved slippery sequence UUUAAC and two potential pseudoknot structures.

© 2005 Elsevier B.V. All rights reserved.

Keywords: Bovine torovirus; Complete sequence; Proteases; Nidovirales; Long PCR

1. Introduction

Toroviruses are enveloped, single-stranded, positive-sense RNA viruses with a pleomorphic virion morphology. The first report on toroviruses described the identification of an unclassified virus from diarrheic calves (Woode et al., 1982), now designated as the Breda virus (BRV), or bovine torovirus (BoTV), two serotypes of which have been identified (BoTV-1 and BoTV-2). The virus is endemic in cattle herds, with asymptomatic cows possibly acting as reservoirs.

Newborn calves generally develop symptoms of diarrhea, typically lasting 2–13 days (Koopmans et al., 1990, 1991) and BoTV has been commonly recovered from calves with diarrhea (Duckmanton et al., 1998a,b; Hoet et al., 2003a,b). Infection with BoTV has also been conclusively associated with diarrhea in veal calves (Hoet et al., 2003a,b). Interestingly, BoTV has been reported as well in nasal secretions, and may therefore be able to infect both the respiratory and gastrointestinal tract (Hoet et al., 2002), as is the case with some coronaviruses (Lai and Holmes, 2001), including the SARS coronavirus (Leung et al., 2003). The equine torovirus (EqTV), or Berne virus (BEV), has been designated as the genus prototype; it can be grown in cell culture (unlike BoTV) and consequently is better characterized. Despite the presence of anti-EqTV antibodies in horses, no disease has been firmly associated with EqTV infection.

* Corresponding author. Tel.: +1 416 813 6592; fax: +1 416 813 6257.

E-mail address: raymond.tellier@sickkids.ca (R. Tellier).

¹ Present address: Ontario Cancer Institute, Princess Margaret Hospital, and Department of Medical Biophysics, University of Toronto, Toronto, Ont., Canada.

Toroviruses have also been reported in humans with gastroenteritis (Beards et al., 1984; Duckmanton et al., 1997; Jamieson et al., 1998) and pigs (Kroneman et al., 1998; Smits et al., 2003). Antibodies reacting against BoTV and EqTV antigens have been detected in various other mammals, suggesting toroviruses may be widespread (Brown et al., 1987; Weiss et al., 1984).

In vitro studies on EqTV grown in cell culture demonstrated the generation of sub-genomic RNAs (sgRNAs) for translation of the ORFs coding for the structural proteins. Approximately 16 kb of sequence from the 3' end of the genome, and approximately 1.5 kb from the 5' end of genome have been determined. These sequence data have revealed the basic torovirus genome organization and the presence of conserved replicase domains (Snijder and Horzinek, 1993). These findings were instrumental in the revision of the *Coronaviridae* taxon to include the genus *Torovirus*, as well as the creation of the order *Nidovirales* which currently includes the families *Coronaviridae* (comprising two genera, *Coronavirus* and *Torovirus*), *Arteriviridae* and *Roniviridae* (Cavanagh, 1997; Cavanagh et al., 1994; Cavanagh and Horzinek, 1993; González et al., 2003). The complete genome sequence of EqTV has yet to be published, although Smits et al. (2003) reported the completion of the EqTV genome sequence in a recent study of torovirus field variants. The published sequence of BoTV is limited to that of the structural genes, which account for approximately 7.5 kb of the 3' end of the genome (Cornelissen et al., 1997; Duckmanton et al., 1998a,b). Consequently, the genus *Torovirus* remains the only genus within the order *Nidovirales* without a representative complete genome sequence.

In this study, we sequenced the entire genome of the bovine torovirus, BoTV-1, using long RT-PCR followed by the sequencing of amplicons. Assembly of the sequences resulted in a genome of 28,475 nucleotides. Sequence analysis revealed the presence of a nidovirus-like replicase gene with conserved domains.

2. Materials and methods

2.1. Source of BoTV-1

BoTV-1 was obtained from Dr. Gerald Woode, as described (Duckmanton et al., 1998a,b). The virus, consisting of stool specimens from gnotobiotic calves infected with the original Breda-1 virus, was aliquoted and stored at -80°C .

2.2. Extraction of viral RNA

BoTV-1 RNA was extracted from fecal specimens as described (Duckmanton et al., 1998a,b) using TRIzol Reagent (Invitrogen, Burlington, Ont., Canada). The RNA was resuspended in 10 μl of ddH₂O containing 10% 100 mM dithiothreitol (Invitrogen) and 5% 20–40 U/ μl RNasin (Promega, Mississauga, Ont.), and stored at -80°C .

2.3. Primers

Primers used in RT-PCR (synthesized by Invitrogen) were designed using Gene Runner 3.05 (Hastings Software, Inc.), based on the available BoTV-1 sequence, and EqTV sequence (GenBank accession no., X52374, and X56016). Primer names, sequences, and positions within the BoTV-1 genome are given in Table 1. One primer, APPR-RS, was designed within a conserved region of the coronavirus open reading frame (ORF) 1a of the replicase gene. This region was first identified by ORF1a amino acid alignment, followed by the corresponding nucleotide alignment of the infectious bronchitis virus (IBV), GenBank accession no. NC001451, nucleotides (nts) 3646–3699; human coronavirus 229E (HCoV 229E), NC002645, nts 4193–4246; murine hepatitis virus (MHV) A59, NC001846, nts 4239–4292; and transmissible gastroenteritis virus (TGEV), AJ271965, nts 4365–4418. Alignments were done using ClustalX 1.8 (Thompson et al., 1997). The sequence of the APPR-RS primer was based on the consensus sequence of the nucleotide alignment described above. The EEAT7 primer used in the amplification of the 3' non-coding region (3' NC) is a coxsackie virus B6 specific primer (Martino et al., 1999). The complementary sequence of the first 21 nucleotides (3'–5') of this primer were incorporated into the BRE-3'NC-CVB6 primer used in the reverse transcription reaction. Lastly, the AAP-CVB6 primer used in the 5' RACE PCR reaction was designed to bind within the poly(A) of the cDNA transcripts, and contained sequence complementary to the 3' end of the EEAT7 primer, allowing for the option of using EEAT7 in a reamplification of 1st round PCR products.

2.4. Long RT-PCR

The long RT-PCR was done essentially as described (Tellier et al., 1996a,b, 2003). Briefly, purified viral RNA was thawed on ice, incubated at 65°C for 2 min then placed back on ice. To the RNA was added 10 μl of a master mix composed of 4 μl of 5 \times 1st Strand Synthesis Buffer (Invitrogen), 0.5 μl of RNasin (20–40 U/ μl) (Promega), 1 μl of 100 mM dithiothreitol (Invitrogen), 1 μl of a 10 mM solution of deoxynucleotide triphosphates (dNTPs) (Pharmacia, Piscataway, NJ), 2.5 μl of a 10 μM solution of antisense primer, and 1 μl of Superscript II reverse transcriptase (Invitrogen). The reaction mixture was incubated at 42°C for 1 h, after which 1 μl each of RNase H (1–4 U/ μl) and RNase T1 (900–3000 U/ μl) (Invitrogen) were added and the reaction mixture was incubated at 37°C for 20 min. Ten-fold dilutions of the newly synthesized cDNA were then prepared using ddH₂O and the neat and diluted cDNA samples were kept on ice for use in PCR or stored at -80°C .

The PCR master mix was composed of 5 μl of 10 \times Advantage 2 PCR Reaction Buffer (Clontech, Palo Alto, CA), 1.25 μl of a 10 mM solution of dNTP mixture (Amersham

Table 1
RT-PCR primers used for the amplification of the BoTV-1 genome

Primer set	Primer name	Primer sequence (5' – 3')	Position in BoTV-1 genome (5' – 3')
1 (Amplicon A)	BRE-3'NC-CVB6	GTAAAAACAGCCTGTGGGTTGTTTTTTTTT	Poly-A tail
	BRE-XMA-S	GATGCCCGGGATGTTTTATGTTCTGTACCGCGCCAATTTTGTGCCTCTG	20,975–21,015
	EEAT7	CTACGGCCGAGCTAATACGACTCACTATAAGTTAAAACAGCCTGTGGGTTG	See text
2 (Amplicon B)	BRE-P-RS	ACTCTAACAGTGTGTACCGACTTTCACCCACC	21,420–21,386
	BRE-1b-RS	ACCACTAATGATTCATTTGACACAACAACGGCACCTCCTG	21,063–21,024
	*BRE-1b(2)-S	AAACCTGGAGGGACATCATCTGGTGATGCCACAACCTGC	15,978–16,015
3 (Amplicon C)	BRE-1b-RS	ACCACTAATGATTCATTTGACACAACAACGGCACCTCCTG	21,063–21,024
	POL1B-RS	GCCATAATCACCCGTAGTATAGGCGTTATGAGCAGC	16,148–16,113
	*BRE-1B-S	TAACATTGGATAATGTGGATTGAATGGTCAGCTTTATGATTTTGG	13,619–13,660
4 (Amplicon D)	1A-APPR-RS	TTGCAACACCACCACCATGAGCCAGATGCCCATTAGCAG	See text
	BRE-5'	CCAGAGAAACCGAGCAGCATTTGTCTCAAGTACC	884–918
5 (Amplicon E)	BRE-1A-RS	AAATATTCCTCATCTGTGAGTGCC	13,514–13,491
	BRE-1A-NRS	ATAAACCACTCTGGATCTTTTCACTACTTGGTATCTGCAC	13,394–13,354
	BRE-1A-S	AACTCTACAGTTATTTGGTGAGGGGGACAGTTGG	3231–3265
6 (Amplicon F)	BRE-1A-NNRS	CACTAGTACCCAACCTGCATGCCATCCAAGACAAC	12,508–12,475
	BRE-1A-NS	AATGCCGAACCTGATGAACAGCCCTTGACTGG	3922–3953
7 (Amplicon G)	BRE-1A-N3RS	GGCTGCCAAAATGAGGGTACATGAAAGTAATGAAG	11,099–11,065
	BRE-1A-NNS	ATAGGGTATATGTTCCACATGCTTTAGGCTCTCACG	5243–5279
8 (Amplicons H ₁ and H ₂)	1ASEQ-RS1	CATCTGAACTAAAACAACCTGAAGCCCTACCTACCTTAAACTC	5812–5770
	1ASEQ-S2	TACCTACATGCTAGACCACCTATGCCCATGTTT	10,281–10,314
9 (Amplicon I)	GSP1	ACAGGGTGGATGAGGTGTTT	1105–1086
	GSP2	AAACTTTGTAAGAGCTGGAGTGGGTTGATGTG	1038–1070
	GSP4	GGGTTCCAAAGTCGGTGACTCCAAGATAGATG	1032–1001
	*5'NC-BEV	CGTGTGTAGACCGTGACCATTGTGTCTAGCTATCTGTAAC	246–285
10 (Amplicon J)	*BRV-END	TCTGTGTAGGAATCTGTACAGATATATG	130–157
	GSP5	GCAAACAACCAAAACAACCTGCCAACCGTAAG	593–563
11 (5' RACE)	GSP6	GAGGTGGGTCAGATGCTGTTTGGTTGGAG	369–341
	Abridged anchor primer		See text
12 (5' RACE)	GSP6	SEE ABOVE	See above
	AAP-CVB6	AAGTACTAGTCAGCCTGTGGGTTGTTTTTTTTTTTTTTTTTTTTTTT	See text

Position in the BoTV-1 genome refers to the binding site of the primer. Where three or more primers are listed in a primer set, the first primer listed was the primer used in the RT reaction—BRE-1A-RS was used as the RT primer in production of template for PCR reactions with primer sets 6, 7 and 8, and GSP1 was used as the RT primer in production of template for PCR reactions with primer sets 10–12. Primer 1ASEQ-RS1 was used with BRE-1A-S and 1ASEQ-S2 was used with BRE-1A-NRS to generate amplicons H₁ and H₂, respectively. The names of the amplicons given in the “Primer Set” column refer to the amplicons generated with the given primer set—amplicons are displayed schematically, relative to the BoTV-1 genome, in Fig. 1. “*” indicates the primer was designed using EqTV sequence. Primer EEAT7 is a coxsackie virus B6 specific primer designed by Martino et al. (1999). The 3' end of this primer was incorporated into the BRE-3'NC-CVB6 RT primer, allowing for the use of EEAT7 in PCR. Primer 1A-APPR-RS was designed using coronavirus consensus sequence in a conserved domain within the 5' end of POL 1a (see). The abridged anchor primer was the primer supplied by the Invitrogen 5' RACE kit and binds within the poly(C) tail, whereas the AAP-CVB6 primer was designed to bind to the poly(A) tail of cDNA transcripts produced in the 5' RACE reactions (see Section 2 for details).

Biosciences, Piscataway), 1 µl each of a sense and antisense primer (10 µM solution), 1 µl of Advantage 2 DNA polymerase (Clontech), and 38.75 µl of ddH₂O for each 50 µl reaction. The mixture was aliquoted in a thin-walled PCR tube (Stratagene, La Jolla, CA) and overlaid with 40 µl of molecular grade mineral oil (Sigma). The master mix was transferred to a room dedicated for cDNA synthesis where 2 µl of the RT (or dilution) was added under the oil and the reaction tube was transferred to a Robocycler thermal cycler (Stratagene) in another room. Unless indicated otherwise, all PCR reactions were carried out as follows: 35 cycles of

denaturation at 99 °C for 35 s, annealing at 67 °C for 30 s, and elongation at 68 °C for a time specific for the target amplicon (Table 2). Fig. 1 schematically illustrates the approach used to amplify the BoTV-1 genome by long RT-PCR. Several reactions required modifications from the protocol given above. For 2nd round amplifications where 5 µl of unpurified 1st round PCR product was used as template (e.g., amplicon J), the PCR master mix was adjusted accordingly: 4.5 µl of buffer was used to account for the buffer already present in the template and the volume of ddH₂O was adjusted to bring the reaction volume to a total of 50 µl. For 2nd round

Table 2
PCR specifications for BoTV-1 amplicons used for genome sequencing

PCR	1st round primers	Template	Elongation time(s)	2nd round primers	Template	Elongation time(s)
A	BRE-XMA-S EEAT7	10 ⁻¹ RT	8 min × 25 12 min × 10	BRE-XMA-S EEAT7	2 μl Z ¹	8 min × 25 12 min × 5
B	BRE-1b(2)-S BRE-1b-RS	10 ⁻² RT	4 min 30 s × 15 5 min × 10 6 min × 10	N/A	N/A	N/A
C	BRE-1b-S POL1B-RS	10 ⁻² RT	5 min × 35	N/A	N/A	N/A
D	BRE-5' APPR-RS	10 ⁻² RT	4 min × 25 5 min × 10	N/A	N/A	N/A
E	BRE-1A-S BRE-1A-NRS	10 ⁻¹ RT	9 min 45 s × 15 11 min × 10 13 min × 10	N/A	N/A	N/A
F	BRE-1A-NS BRE-1A-NNRS	10 ⁻¹ RT	8 min × 25 12 min × 10	N/A	N/A	N/A
G	BRE-1A-NNS BRE-1A-N3RS	10 ⁻² RT	4 min 30 s × 15 5 min × 10 6 min × 10	N/A	N/A	N/A
H ₁	BRE-1A-S 1ASEQ-RS1	10 ⁻² RT	4 min × 35	N/A	N/A	N/A
H ₂	1ASEQ-S2 BRE-1A-NRS	10 ⁻² RT	4 min × 35	N/A	N/A	N/A
I	5'NC-BEV GSP2	Neat RT 10 ⁻¹ RT	3 min × 35 ²	5'NC-BEV GSP4	1 μl Y	3 min × 35
J	BRV-END GSP5	Neat RT	3 min × 35 ³	BRV-END GSP5	1 μl Z ⁴	3 min × 25
K*	AAP GSP6	Neat TdT	2 min × 35	N/A	N/A	N/A
L*	AAP-CVB6 GSP6	Neat TdT	2 min × 35	N/A	N/A	N/A

For PCR with multiple elongation times, the times and cycles are given in chronological order. The letters given in the “PCR” column refer to the amplicons shown in Fig. 1. The “Template” column provides the dilution of the RT or TdT (terminal deoxynucleotidyl transferase reaction of the 5' RACE) template used in PCR. See Table 1 for primer details. Z: 1st round PCR product was purified from an agarose gel and used as template in 2nd round PCR. Y: 1 μl of 1st round PCR was used as template in a hemi-nested PCR reaction. 1–10 μl of the first round PCR product was purified from an agarose gel; a 10⁻¹ dilution of the purified DNA was used in 2nd round PCR. 2: The annealing temperature used for both 1st and 2nd round PCR was 65 °C × 30 s. 3: The annealing temperature used in both 1st and 2nd round PCR was 50 °C × 1 min. 4–40 μl of 1st round PCR product was purified from an agarose gel; undiluted and 10⁻¹ to 10⁻³ dilutions of the purified DNA were used as template in 2nd round PCR. *: Amplicons K and L were obtained via 5' RACE reactions—see Section 2 for details.

PCR where the template used was an amplicon purified from an agarose gel, only the ddH₂O was adjusted, to bring the final reaction volume to 50 μl. The RT protocol was modified for amplicon J, using the RT primer listed in Table 1, primer set 10. After thawing on ice, the 2.5 μl aliquot of primer was added to the RNA and the mixture was heated at 99 °C for 5 min (Kostouros et al., 2003); the remainder of the reaction was carried out according to the protocol above.

2.5. 5' RACE

The 5' RACE kit (Invitrogen) was used for amplification of the extreme 5' end of the BoTV-1 genome, according to the manufacturer's protocol with modifications. The RT was performed using the protocol above, with Superscript

II, GSP1 primer, and with the adjusted RNA denaturation incubation time and temperature of 99 °C for 5 min as above. The cDNA was purified using QIAquick PCR clean-up kit (Qiagen, Mississauga, Ont., Canada) according to the manufacturer's protocol and eluted in 30 μl of supplied elution buffer. The cDNA tailing reaction was performed with the terminal deoxynucleotidyl transferase enzyme supplied with the kit, according to the manufacturer's protocol with the exception of an increased incubation time of 5 min at the protocol specified 94 °C. Tailing was performed in the presence of dCTP (provided), or dATP (Pharmacia). PCR was performed with Advantage 2 DNA polymerase using the protocol above, with modifications: 5 μl of the tailing reaction was used as template, and 2 μl of each primer were used (instead of 1 μl each); the volume of the ddH₂O was adjusted accordingly to bring the final reaction volume to 50 μl.

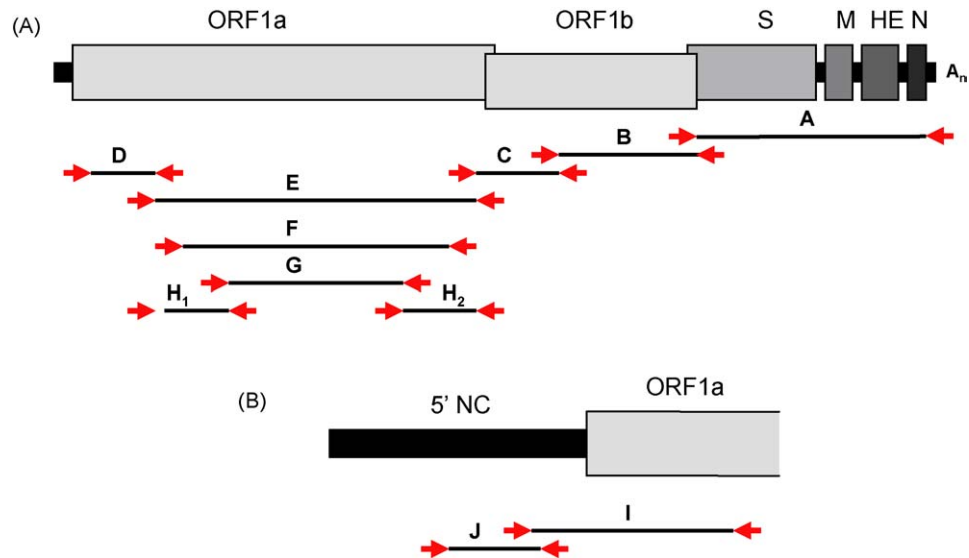


Fig. 1. Schematic Illustration of BoTV-1 Genome and of the strategy used for amplification by long RT-PCR. Shown in A and B is the strategy used to amplify the BoTV-1 genome for sequencing purposes. Amplicons A–J are named with respect to the chronological order the reactions were performed. Primers (arrows) are the primers used in PCR (RT primers are not shown here). Fig. 1B is an enlarged area of the 5' non-coding region—5' RACE products are not depicted here. The sizes of amplicons (in base pairs) are as follows: A, 7546; B, 5086; C, 2530; D, 2580; E, 10,164; F, 8587; G, 5857; H₁, 2582; H₂, 3114; I, 825; J, 464. ORF1a: replicase ORF1a, ORF1b: replicase ORF1b, S: spike (peplomer), M: membrane, HE: hemagglutinin esterase, N: nucleocapsid; A_n: poly(A) tail; the 5' and 3' non-coding regions are not labeled in A.

2.6. Sequencing

Purified amplicons were submitted for automated sequencing of both strands, which was carried out by the DNA Sequencing Facility, Centre for Applied Genomics, Hospital for Sick Children, Toronto, Ont., Canada. Amplicons were cut out of agarose gels under direct visualization on a Dark Reader transilluminator (Clare Chemical Research Inc., Dolores, CO), with a disposable scalpel and the DNA recovered in 40–50 μ l of the supplied elution buffer using Clontech Nucleospin Gel Extraction Kit according to the manufacturer's instructions. The initial sequencing reactions were done using the PCR primers. This sequence then allowed for the design of sequencing specific primers. The sequencing continued in a similar step-wise fashion until the sequences of the top and bottom strand crossed over, at which point the remaining sequencing primers could be designed in a single step.

2.7. Sequence assembly and analysis

The individual sequence fragments were aligned and assembled using Gene Runner Version 3.05. The fully assembled BoTV-1 genome sequence was analyzed using Gene Runner; multiple alignments with ClustalX 1.8 (Thompson et al., 1997), BLAST, and PSI BLAST search programs (Altschul et al., 1997); secondary structure and protein fold prediction using the 3D-PSSM Protein Fold Recognition (Threading) Server (Fischer et al., 1999; Kelley et al., 1999, 2000) (<http://www.sbg.bio.ic.ac.uk/~3dpssm/>); and

RNA pseudoknot secondary structures were identified using MFOLD Version 3.1 (Zuker, 2003) (<http://www.bioinfo.rpi.edu/applications/mfold/old/rna/>).

The Breda-1 torovirus sequence was entered into the Virus Orthologous Clusters (VOCs) database for analysis and comparison using the complete coronavirus database and suite of tools (www.sarsresearch.ca).

Composition vector trees (CVTrees) were generated using several complete genome sequence of representative viruses in the order *Nidovirales*, including members of all genera. The CVTrees were generated using the CVTree program (Qi et al., 2004a,b) (<http://cvtree.cbi.pku.edu.cn>), using a value of $K = 5$.

3. Results

3.1. Amplification and sequence analysis of the BoTV-1 genome

The BoTV-1 genome was reverse transcribed and amplified in 8 RT-PCR reactions yielding 8 overlapping products (Fig. 1) (amplicons A–E and I–K—amplicon K is the 5' RACE product, not shown in Fig. 1). Five additional amplicons, F–H₂, which overlapped the large 10,164 bp amplicon (E), were obtained to facilitate sequencing. Lastly, the 5' end of the genome was amplified for sequencing using two 5' RACE reactions.

The complete BoTV-1 sequence was assembled from the sequences of overlapping amplicons. The total genome was composed of 28,475 nts and had a G + C content of 38%. The

genome sequence has been deposited in GenBank with the accession no. AY427798.

3.2. Analysis of newly identified sequence of BoTV-1

Sequencing of the genome upstream of the structural genes yielded 20,920 nts of novel BoTV-1 sequence. Although all amplicons were sequenced on both strands, four positions were identified where two different nucleotides were equally credible at each of these four positions based on the analysis of the chromatograms on both strands. These ambiguities were considered genuine; only two of these resulted in a change of amino acid (nt positions 1895 and 9470).

Analysis of this novel sequence identified two large ORFs that overlap by 13 nts. This arrangement is common to the nidovirus replicase gene, and the two ORFs were identified as ORF1a and ORF1b (Snijder and Horzinek, 1993). The 5' end of the genome begins with a non-coding region of 858 nts. The torovirus 5' non-coding region is substantially larger than the coronavirus 5' non-coding region—the two largest being TGEV at 314 nts and IBV at 528 nts. Given the difficulty in amplifying this region of the genome (the full-length RACE amplicon of 464 bp (amplicon J, Fig. 1) was only obtained when an increased time and temperature of RNA denaturation were used), we suspect that strong secondary structures exist.

3.3. Comparison of the 3' end of BoTV-1 sequence with that previously reported

To confirm the previously reported sequence of BoTV-1 (approximately 7.5 kb comprising the 3' end of the genome) (Duckmanton et al., 1998a,b), this region was re-sequenced. Comparison of sequences showed an identity of 97% at the nucleotide and amino acid levels for the M, HE, and N genes and for the 3' non-coding region. The sequence of the S gene in this study was 93.1% in agreement at the nucleotide level, and 91.6% at the amino acid level compared to that reported by Duckmanton et al. (1998a,b). Differences were scattered except for a cluster of discrepancies between the two sequences in the small region 21,773–21,985 of the sequence reported here, which displays only 57.7% identity with the corresponding region of the previously reported sequence. Comparison of the sequence reported here with

the S gene sequence of the B145 BoTV isolate published by Smits et al. (2003) revealed an identity of 95.5% at the nucleotide level and 97.2% identity at the amino acid level.

The length of the HE gene was 9 nts longer from that reported previously for BoTV-1 and BoTV-2 (Cornelissen et al., 1997; Duckmanton et al., 1998a,b), extending the protein sequence by 3 amino acids. A stretch of seven thymidine residues (27,715–27,721), as opposed to eight in the previously reported sequence, is responsible for this difference. This resulted in a shift of the ORF leading to extension of the gene and protein sequence. The sequences of several BoTV isolates characterized by Smits et al. (2003) are in agreement with the length of the HE gene/protein sequence reported here. Of the six isolates sequenced within the structural genes, four of these had HE genes of 1260 nts, identical to what is reported here. The other two isolates contained HE genes shorter by only 3 nts, although the structural genes were identified as having undergone recombination with a porcine torovirus resulting in an HE gene of mixed origin (Smits et al., 2003).

The sequence of the 3' non-coding region matches the previously reported sequence with only two mismatches. With the primer used in the reverse transcription reaction, a portion of the poly(A) tail was amplified in PCR. Consequently, sequencing of this amplicon included poly(A) sequence ensuring that the extreme 3' end of the genome had been amplified and the last nucleotide at the 3' end of the genome had been determined.

3.4. Analysis of BoTV-1 ORFs

Fig. 2 illustrates the potential ORFs of the BoTV-1 genome. Table 3 lists the gene and protein characteristic of the BoTV-1 genome, based on currently known genes of the family *Coronaviridae*.

The putative replicase gene of BoTV-1 is comparable to the replicase gene of the coronaviruses in many aspects. Most obvious is its size, which at 20,201 nts in length is comparable to the coronavirus replicase genes with a range of 20–22 kb (Lai and Cavanagh, 1997). The putative ORF1a of 13,338 nts and ORF1b of 6876 nts encode a ORF1a/1b fusion polyprotein (pp1a/1b) of 6733 aa. A nucleotide sequence comparison of ORF1b between BoTV-1 and EqTV, the latter having been reported previously (Snijder

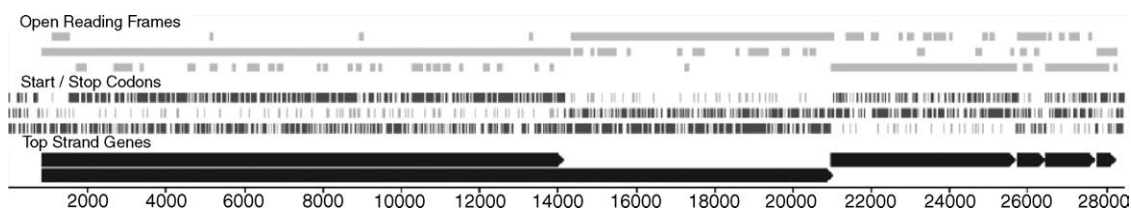


Fig. 2. Map of BoTV genome. Top light colored line bars show all possible open reading frames. Start and stop codons in all three frames are represented by vertical bars in the middle section of the figure. The annotated top strand, or rightward transcribed, genes are shown in the bottom solid arrow bars above position numbers in the genome.

Table 3

Gene and protein characteristics of the BoTV-1 genome, established with the VOCs software (Upton et al., 2003)

Gene name	ORF start	ORF stop	Molecular weight	Adenine + thymine %	pI	No of amino acids
BoTV-1 pp1a	859	14,196	50,5921	62.59	5.84	4445
BoTV-1 pp1ab	859	21,059	76,5907	62.64	6.14	6733
BoTV-1 S	20,975	25,729	177,876	61.18	5.87	1584
BoTV-1 M	25,758	26,459	26,410	60.40	9.05	233
BoTV-1 HE	26,477	27,736	46,706	61.99	5.48	419
BoTV-1 N	27,775	28,278	18,868	54.16	11.82	167

et al., 1990), revealed a high degree of identity. At the nucleotide and amino acid level the two viruses display 83.8 and 91.0% identity, respectively. Therefore, it is not surprising that similar potential secondary structures were identified in the ORF1a/ORF1b overlap, and the same conserved domains of ORF1b identified in EqTV and other nidoviruses were also found in BoTV ORF1b (Snijder et al., 2003, 1990).

The ORF1b protein product is believed to be translated via a ribosomal frameshifting event generating the pp1a/1b fusion polyprotein. The two elements required for efficient ribosomal frameshifting, a slippery sequence and a RNA pseudoknot (Brierley, 1995), were demonstrated in EqTV (Snijder et al., 1990). The heptanucleotide slippery sequence found in BoTV, UUUAAAC, (nts 14,148–14,190) is identical to the slippery sequence identified in EqTV and the coronaviruses (Lai and Cavanagh, 1997; Snijder et al., 1990). Secondary structure prediction of the ORF1a/1b overlap

region identified two potential RNA pseudoknot structures downstream of the slippery sequence (Fig. 3). Not surprisingly, these two structures are very similar to those predicted for EqTV.

Due to the relatively high degree of identity between BoTV and EqTV in pp1b, the conserved domains that have been identified in BEV, including the RNA dependant RNA polymerase (RdRp) and helicase, as well as domains with homology to cellular proteins involved in RNA processing (Snijder et al., 2003, 1990) have also been identified in BoTV (Fig. 4B).

Analysis of the BoTV-1 ORF1a revealed several interesting features. This ORF encodes a protein of 4445 aa (pp1a), larger than the pp1a of the majority of the coronaviruses, with the sole exception of the 4470 aa MHV pp1a. The pp1a of the nidoviruses also contain conserved domains with functions important to viral replication. All nidoviruses studied to date contain at least one protease domain within pp1a,

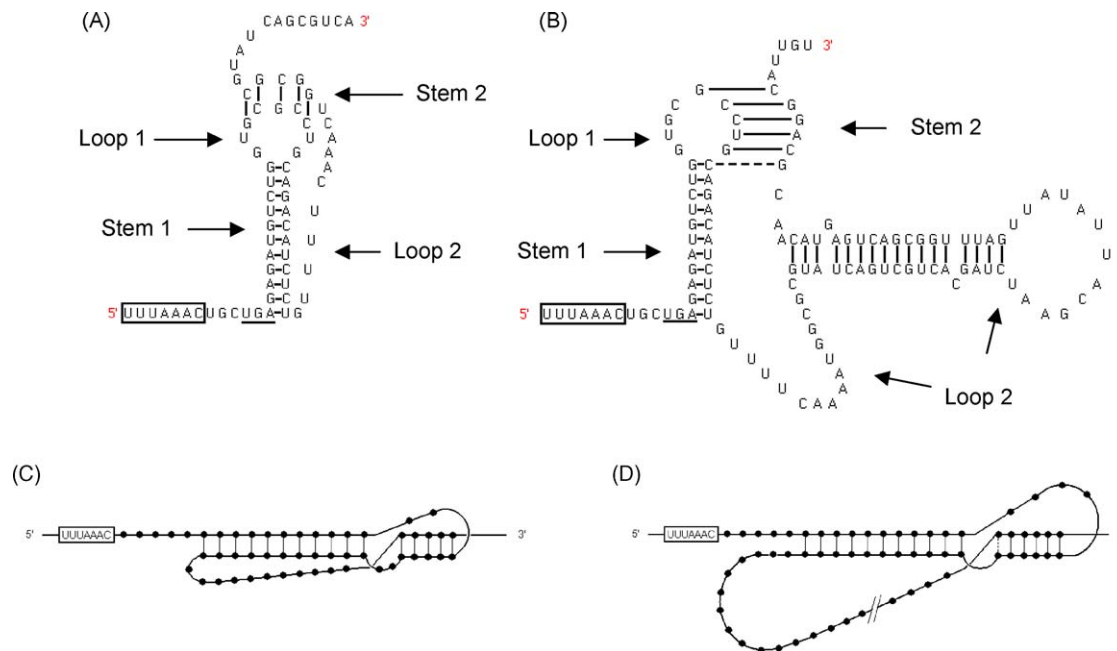


Fig. 3. Predicted pseudoknot structures in the BoTV-1 ORF1a/1b overlap region. A and B represent the two structures predicted using MFOLD software in the analysis of the ORF1a/1b overlap region. Loop 1,2 and stem 1,2 in (A) and (B) refer to structural features of the depicted pseudoknots. Note that the two structures have identical hairpin loops, but differ in the region downstream of the hairpin which folds back to bind to the loop of the hairpin structure, i.e., B has a larger loop 2 and a different stem 2 of the pseudoknot. Also note that loop 2 in B has the potential to form an internal secondary structure. C and D are schematic representations of A and B, respectively. The 5' and 3' orientations are given in each figure. The ORF1a stop codon is underlined in A and B, and the slippery sequence, UUUAAAC, is boxed in all four figures. The dashed line in B indicates that the "C" residue can potentially bind to either of the G residues, or both. A and C show BoTV-1 nucleotides 14,184–14,255; B and D show nucleotides 14,184–14,305.

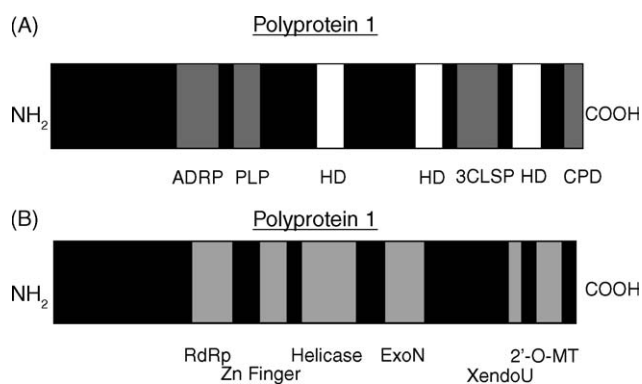


Fig. 4. Domains of the 1ab polyprotein. (A) Predicted conserved domains of the 1a polyprotein. Gray and white boxes represent the relative positions of the conserved domains of pp1a. ADRP, adenosine diphosphate-ribose 1''-phosphatase (formerly the "X" domain in the coronaviruses) (residues 1657–1781); HD, hydrophobic domain (residues ~2186–2237, 2930–3042, 3506–3616); 3CLSP, 3C-like serine protease (3129–3393); PLP, papain-like protease (1 824–1 989); CPD, cyclic phosphodiesterase (4272–4443). The ADRP domain is named after its cellular homologue. (B) Domains of the 1b Polyprotein. Shown in gray are the relative positions of the conserved domains identified in the pp1b (which, technically is only produced as a fusion protein (pp1a/1b), although only the pp1b portion is shown here for simplicity). RdRp, RNA-dependent RNA polymerase (residues 515–744); Zn Finger, Zn Finger domain containing conserved Cys and His residues—nucleic acid binding domain (residues 845–928); Helicase, viral helicase domain (residues 1 099–1 374); ExoN, 3'-5' exonuclease (residues 1 414–1 609); XendoU, homologue of poly-U-specific endoribonuclease (residues 1918–1974); 2'-O-MT, S-adenosylmethionine-dependent ribose 2'-O-methyltransferase. The ExoN, XendoU, and 2'-O-MT domains are named after their cellular homologues (Snijder et al., 2003).

known as the 'main' protease or 3C-like protease (3CLP), after the picornaviruses 3C protease, and located within the C terminus of the pp1a protein (Ziebuhr et al., 2000). Although all of the nidoviruses possess this domain, the protease is slightly different in each family. The coronavirus 3CLP contains a cysteine as its catalytic nucleophile, but does not appear to contain an acidic residue common to the catalytic triad, His-Asp(Glu)-Cys, found in this family of proteases (Ziebuhr et al., 2000). The 3CLP of the gill associated virus (GAV), genus *Oakavirus*, family *Roniviridae*, is also a cysteine protease and like the coronavirus 3CLP, possesses a Cys-His catalytic dyad (Ziebuhr et al., 2003). Finally, the arterivirus main protease is referred to as a 3C-like *serine* protease (3CLSP) due to the presence of a serine nucleophile in the triad His-Asp-Ser, more common in the cellular homologues, such as the pancreatic enzyme chymotrypsin (Snijder et al., 1996). It was, therefore, surprising to find that the main protease of BoTV appears to be a serine protease, similar to the arterivirus 3CLSP. Repeated attempts at aligning the BoTV sequence with the coronavirus 3C protease failed to provide a convincing alignment. The results presented here were obtained with supporting data from PSI BLAST analysis, the protein sequence alignment shown in Fig. 5A, and secondary structure prediction analysis using the 3D-PSSM protein fold recognition (threading) server

that identified the BoTV main protease sequence reported here as having trypsin-like or chymotrypsin-like protease folds.

In addition to the main protease, the nidoviruses, with the exception of GAV (Cowley et al., 2000), contain one or more 'accessory' protease(s) in the N-terminus of pp1a (Lai and Cavanagh, 1997; Snijder and Meulenberg, 1998). In both the coronaviruses and arteriviruses this protease is a cysteine protease with homology to papain, referred to as the papain-like protease (PLP) (Ziebuhr et al., 2000). Like the SARS human coronavirus (SARS-CoV), and IBV, the torovirus PLP is present as a single domain and although it has homology with the coronavirus PL2^{pro}, the torovirus PLP is more similar in size and has a greater degree of identity with the leader protease (L^{pro}) of the picornaviruses foot and mouth disease virus (FMDV) and equine rhinitis A virus (ERAV). This is supported by PSI BLAST analysis, protein fold prediction suggesting folds that are similar to the FMDV leader protease, and the alignment shown in Fig. 5B.

Hydrophobic domains are additional conserved domains that are common to the pp1a of all nidoviruses. These domains are conserved in the relative positions within the pp1a protein of all of the nidoviruses. The hydrophobic domains identified in BoTV maintain this configuration (Fig. 4A).

Also, a domain was identified in pp1a of BoTV-1 which has homology to the adenosine diphosphate-ribose 1''-phosphatase (ADRP) processing enzyme family mediating the removal of a phosphate from the adenosine diphosphate ribose 1''-phosphate substrate in the cellular RNA processing pathway (Martzen et al., 1999). Although this domain is present in other ssRNA viruses such as the alphaviruses, and in bacterial and eukaryotic proteins, the *Coronaviridae* are the only other nidoviruses to date to contain this domain (Snijder et al., 2003).

Sequence analysis of pp1a revealed that the 3' end of this protein displayed similarity to the non-structural protein 2a, found in coronaviruses such as MHV and HCoV-OC43. The same region of EqTV's pp1a also contains similarity to this protein (Snijder et al., 1991). Recently, this domain has been identified as a putative cyclic phosphodiesterase (Snijder et al., 2003).

3.5. Phylogeny of the Nidovirales

Composition vector trees (CVTrees) are a new method of inferring phylogeny using complete genome sequences. A distance matrix is generated from comparison of the predicted proteomes of the viruses, calculating the frequency of short amino acid sequences of a fixed length (K -tuples of amino acids, K being a parameter that can be adjusted in each analysis). Phylogenetic trees can then be built from the distance matrix using standard methods such as the Neighbor Joining (NJ) or Minimum Evolution (ME) methods (Chu et al., 2004; Qi et al., 2004a,b). This method circumvents the difficult,

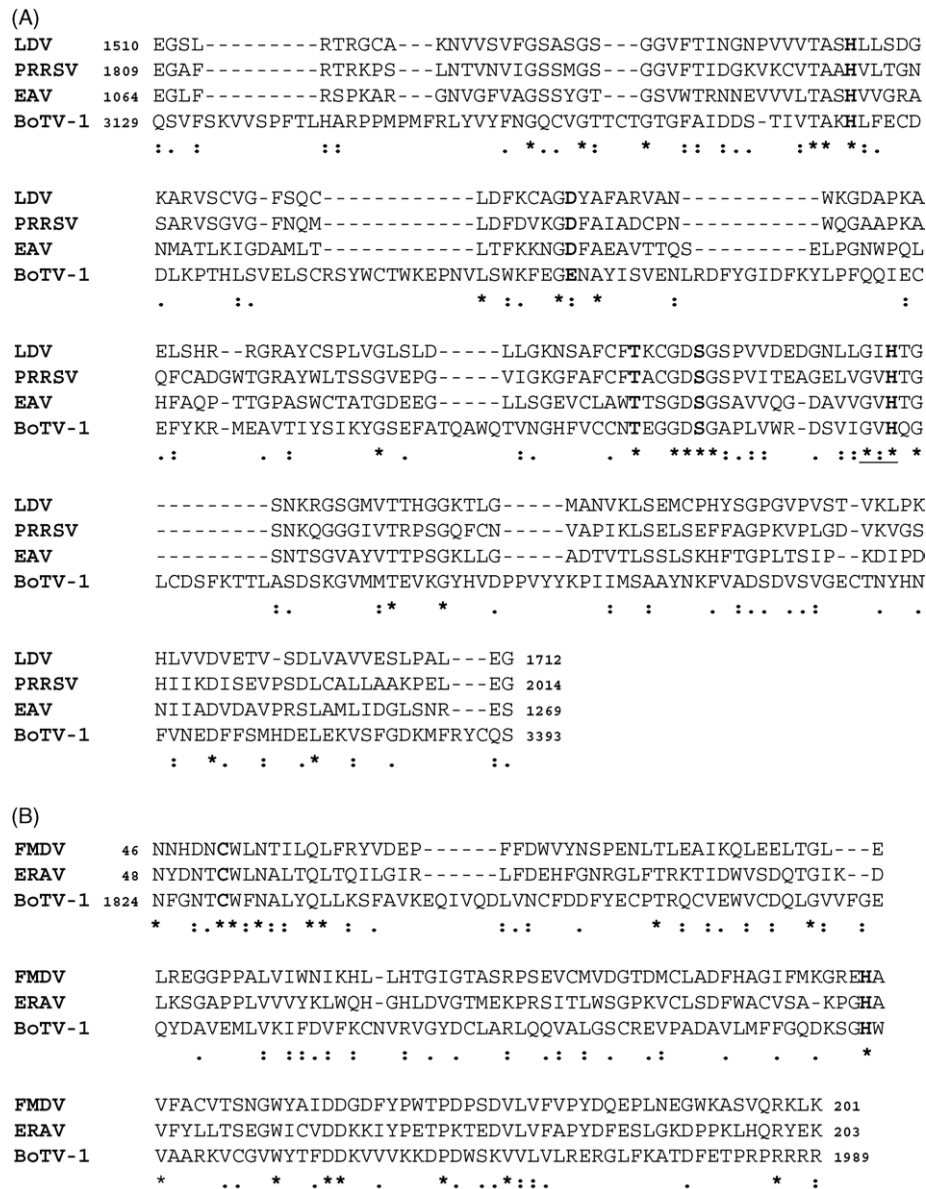


Fig. 5. (A) Alignment of the putative BoTV-1 3C-like serine protease with the arterivirus 3C-like serine protease. In bold, the residues of the catalytic triad, H–E–S, and the threonine and histidine residues believed to be involved in substrate recognition; underlined, the glycine-X-histidine substrate-binding pocket ‘core’ motif, where X is any amino acid. “*” indicates fully conserved residues; “:” indicates that one of the following ‘strong’ groups is fully conserved: STA, NEQK, NHQK, NDEQ, QHRK, MILV, MILF, HY, FYW. “.” indicates that one of the following ‘weaker’ groups is fully conserved: CSA, ATV, SAG, STNK, STPA, SGND, SNDEQK, NDEQHK, NEQHRK, HFY. LDV, lactate dehydrogenase-elevating virus, strain neuro-virulent type C, GenBank accession no. NC002534; PRRSV, porcine reproductive and respiratory syndrome virus, GenBank accession no. NC00196; EAV, equine arteritis virus, strain Bucyrus, GenBank accession no. X53459. For each viral sequence the positions of the N- and C-terminal residues, within ORF1a, are given. (B) Alignment of the putative BoTV-1 papain-like protease with the picornavirus leader protease. In bold, the residues of the catalytic dyad, C–H. “*” indicates fully conserved residues; “:” indicates that one of the following ‘strong’ groups is fully conserved: STA, NEQK, NHQK, NDEQ, QHRK, MILV, MILF, HY, FYW. “.” indicates that one of the following ‘weaker’ groups is fully conserved: CSA, ATV, SAG, STNK, STPA, SGND, SNDEQK, NDEQHK, NEQHRK, HFY. FMDV, foot and mouth disease virus C, GenBank accession no. NC002554; ERAV, equine rhinitis A virus, GenBank accession no. NC003982. For each viral sequence, the positions of the N- and C-terminal residues, within ORF1a, are given.

and often impossible task of making alignments of distantly related sequences, and analyzes the entire genomes rather than sections which may have evolved separately. Since we report here the first complete torovirus sequence it is the first time that such an analysis can be done with representatives of all genera in the order *Nidovirales* (Fig. 6). The trees in Fig. 6, calculated using a value of $K=5$, illustrate a topology only,

since calibration of branch lengths in CVTrees present several complications (Qi et al., 2004a,b). Nonetheless, interesting inferences on evolutionary relatedness can be drawn from the topology alone. It is noteworthy to observe that whereas the arteriviruses and the coronaviruses each form a cluster in both trees, BoTV-1 is quite distinct from both clusters in these trees.

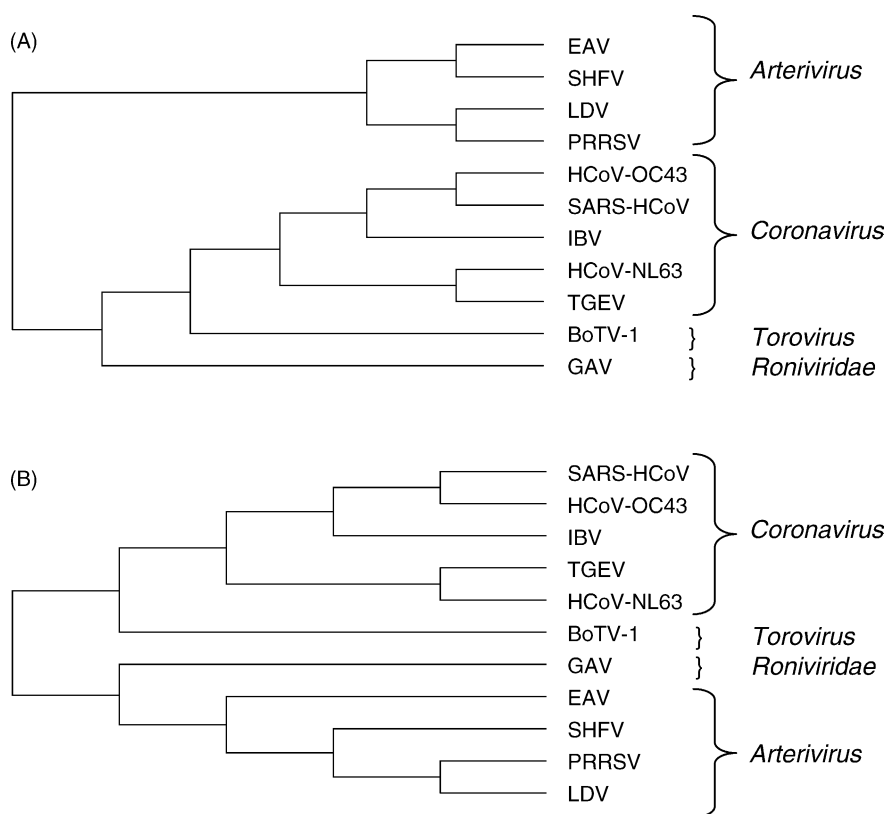


Fig. 6. CVTRees with representative viruses within *Nidovirales*. (A) Tree built with the Neighbor Joining method. (B) Tree built with the Minimum Evolution method ($K=5$ for both trees). *Torovirus*: Breda virus 1 (BoTV-1): accession no. AY427798. *Coronavirus*: SARS associated coronavirus (SARS-HCoV): NC_004718; human coronavirus OC43 (HCoV-OC43): NC_005147; human coronavirus NL63 (HCoV-NL63): AY567487; transmissible gastroenteritis virus (TGEV): NC_002306; infectious bronchitis virus (IBV): NC_001451. *Arterivirus*: equine arteritis virus (EAV): NC_002532; porcine reproductive and respiratory syndrome virus (PRRSV): NC_001961; simian hemorrhagic fever virus (SHFV): NC_003092; lactate-dehydrogenase elevating virus (LDV): NC_002534. *Roniviridae*: gill associated virus (GAV), genus *Oakavirus*: AY039647 and AF227196.

4. Discussion

Despite the evidence suggesting that toroviruses infect many different species, most of our understanding of these viruses at the biochemical and antigenic levels comes from studies of EqTV and BoTV (Koopmans and Horzinek, 1994). Our knowledge of their molecular biology has come mainly from studies using EqTV grown in cell culture. However, a complete torovirus genome sequence or ORF1a sequence has yet to be reported. This has hampered detailed characterization of this genus in terms of molecular biology, evolution, and taxonomy.

Long RT-PCR followed by direct sequencing of the amplicons had been used previously to amplify and sequence the 7.5 kb of the 3' end of the BoTV-1 genome encompassing the structural genes and 3' non-coding region (Duckmanton et al., 1998a,b). The limited BoTV sequence available for primer design posed a challenge, which was overcome by the use of available sequence from related viruses such as EqTV and the coronaviruses. Using this approach almost 21 kb of novel BoTV-1 sequence was obtained without the benefit of cultured virus. With the availability of the complete genome sequence of BoTV-1, and with more than half of the

EqTV genome sequence reported, future studies involving the molecular characterization of other toroviruses should be greatly facilitated. For example, the identification and characterization of the human coronavirus responsible for the spread of severe acute respiratory syndrome (SARS) was successful, in part, due to the use of RT-PCR assays targeting conserved domains within ORF1b of the viral replicase gene (Drosten et al., 2003; Ksiazek et al., 2003; Lee et al., 2003; Poutanen et al., 2003).

The 28,475 nts genome is reported here with confidence that the sequence is complete. The methods used to amplify the extremities of the viral genome ensured that the extreme 5' and 3' nucleotides were internal to the primers used in RT-PCR amplifying the respective areas of the genome. The sequence of first nucleotide at the 5' end of the viral genome was scrutinized further by performing two separate 5' RACE reactions using two different nucleotides in the tailing reactions. The sequences of the two 5' RACE amplicons could then be compared and the first nucleotide of the genome could be distinguished from the poly-nucleotide tail.

The sequence of the structural genes and 3' non-coding region reported here is in agreement with that previously reported for BoTV-1 (Duckmanton et al., 1998a,b). There are

a small number of discrepancies which may be due to the different sequencing methodologies used (automated sequencing was used in this study and ought to be more accurate). The differences may also point to a quasispecies in BoTV-1, which is not unlikely since it is a RNA virus, and differences in the sequencing data may arise through use of different sequencing primers. In that regard, it is remarkable that most differences occur at isolated nucleotides except for a region of 217 nts in the putative bulbous domain (S1) of the spike glycoprotein. This is highly reminiscent of a well-characterized hypervariable region in the S1 domain of the spike glycoprotein of the murine hepatitis virus (Lai and Cavanagh, 1997; Rowe et al., 1997).

Comparison of the sequences of the structural genes and 3' non-coding regions of BoTV and EqTV previously showed that the two viruses share a relatively high degree of sequence identity (Duckmanton et al., 1998a,b). Comparison of the ORF1b sequences of EqTV and BoTV further supports these findings with a nucleotide identity of 83.3% and amino acid identity of 91% and an identically sized ORF of 6876 nts. Thus, it was not surprising that the same conserved domains identified in BEV were found in BoTV pp1b including the newly identified domains with homology to cellular proteins involved in RNA processing, with potential roles in viral replication identified by Snijder et al. (2003).

The identification of two potential pseudoknot structures in the ORF1a/1b overlap region of BoTV that are virtually identical to the structures identified in EqTV (Snijder et al., 1990) reaffirms that one or both of these structures are important in the translation of ORF1b, although experimental data will be required to confirm these predictions in BoTV.

Pp1a of BoTV contains 3C-like Ser protease (3CLSP) and papain-like Cys protease (PLP) motifs that are most likely involved in the processing of the viral replicase polyprotein. The finding of a 3C-like protease with a Ser catalytic nucleophile, and substrate binding residues that more closely resembled the arterivirus 3CLSP (Snijder et al., 1996) was surprising since toroviruses are classified as members of the family *Coronaviridae*. Although a PLP motif with similarity to the coronavirus PL2^{PRO} domain was identified, sequence alignment and protein structure prediction analysis support the hypothesis that the BoTV PLP is more similar to the leader protease (L^{PRO}) of the picornaviruses FMDV and ERAV. The coronavirus PL2^{PRO} is the largest protease in this family of enzymes due to a unique zinc finger structure that connects the N-terminal alpha-helical domain to the C-terminal beta-sheet domain of the protease (Herold et al., 1999). In contrast, the N- and C-terminal domains of the picornavirus protease are joined by a short beta-strand structure (Guarne et al., 1998), which may also be the case for the torovirus protease given the size similarity.

Consistent with the nidovirus organization of domains within pp1a, several hydrophobic domains were located in the BoTV pp1a sequence. The conservation of these domains within the nidoviruses further emphasizes their significance. The hydrophobic domains in the arteriviruses have been

shown to dramatically alter the host-cell membrane architecture in addition to mediating the association of the replication complex with cellular membranes (Pedersen et al., 1999; van der Meer et al., 1998). A similar role can be conjectured in the toroviruses.

Within *Nidovirales*, the ADRP domain identified in pp1a has so far only been identified within viruses of the family *Coronaviridae*. This domain is within the same relative position in both the torovirus and coronavirus pp1a. This may suggest that an ancestor common to both toroviruses and coronaviruses acquired this domain after the other nidoviruses had split from this lineage, and before the torovirus–coronavirus split.

The presence of a region within the 3' end of the torovirus pp1a with sequence similarity to the coronavirus non-structural protein 2a has been interpreted as possible evidence for a nonhomologous recombination event, perhaps between an ancestral torovirus and a coronavirus during the co-infection of a cell (Snijder et al., 1991). Replication of these sequence results in BoTV further supports this hypothesis.

Previously published phylogenetic analysis based on the RdRp and helicase domains of several members of the *Nidovirales* including one torovirus (EqTV) suggested a rather large evolutionary distance between EqTV and the coronaviruses (González et al., 2003). Taking advantage of the BoTV sequence in this study, a new alignment of the RdRp protein sequences was done (data not shown) and the dendrogram obtained was highly similar to that obtained by González et al. (2003) but showing the two toroviruses, EqTV and BoTV, clustering together. However, the exact boundaries of the RdRp domains were difficult to establish for many sequences used in the alignment. The availability of the first complete torovirus sequence enabled the establishment of CVTRees with representatives from of all genera within *Nidovirales* (Fig. 6). The topology of the trees shows that whereas the coronaviruses and the arteriviruses form distinct clusters, in both trees the BoTV-1 sequence stands apart from these clusters. Presently, *Coronavirus* and *Torovirus* are two genera within the family *Coronaviridae*. Viral taxonomy currently uses a polythetic approach which considers numerous properties simultaneously (Condit, 2001). Based on their phylogenetic analysis, González et al. (2003) proposed that the differences between coronaviruses and toroviruses are large enough that the two genera be elevated to subfamilies or even to families within *Nidovirales*. The data presented here support further taxonomic separation between the genera *Coronavirus* and *Torovirus*.

The completion of the genome sequence of BoTV-1 will contribute to our understanding of the toroviruses at the molecular level. This sequence, combined with the findings obtained from studies on EqTV, is expected to aid in the development of sensitive molecular diagnostic assays, such as RT-PCR targeting conserved regions of the torovirus genome. With the development of such tests, the less

well-characterized toroviruses such as the human torovirus, as well any newly identified toroviruses will perhaps be more readily identified and better characterized.

Acknowledgements

This study was supported by a grant from the Canadian Institutes of Health Research (formerly the Medical Research Council of Canada). The authors wish to thank Vasily Tcherepanov and Chris Upton for help with the VOCs database and with preparing the CVTrees.

References

- Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25 (17), 3389–3402.
- Beards, G.M., Hall, C., Green, J., Flewett, T.H., Lamouliatte, F., Du Pasquier, P., 1984. An enveloped virus in stools of children and adults with gastroenteritis that resembles the Breda virus of calves. *Lancet* 1 (8385), 1050–1052.
- Brierley, I., 1995. Ribosomal frameshifting viral RNAs. *J. Gen. Virol.* 76 (Pt 8), 1885–1892.
- Brown, D.W., Beards, G.M., Flewett, T.H., 1987. Detection of Breda virus antigen and antibody in humans and animals by enzyme immunoassay. *J. Clin. Microbiol.* 25 (4), 637–640.
- Cavanagh, D., 1997. Nidovirales: a new order comprising Coronaviridae and Arteriviridae. *Arch. Virol.* 142 (3), 629–633.
- Cavanagh, D., Brian, D.A., Brinton, M., Enjuanes, L., Holmes, K.V., Horzinek, M.C., Lai, M.M.C., Laude, H., Plegemann, P.G.W., Siddell, S., Spaan, W.J.M., Taguchi, F., Talbot, P.J. (1994) Revision of the taxonomy of the Coronavirus, Torovirus and Arterivirus genera. *Archives of Virology* 135, 227–237.
- Cavanagh, D., Horzinek, M.C., 1993. Genus *Torovirus* assigned to the Coronaviridae. *Arch. Virol.* 128 (3–4), 395–396.
- Chu, K.H., Qi, J., Yu, Z.G., Anh, V., 2004. Origin and phylogeny of chloroplasts revealed by a simple correlation analysis of complete genomes. *Mol. Biol. Evol.* 21 (1), 200–206.
- Condit, R., 2001. Principles of virology. In: Knipe, D., Howley, P., Griffin, D., Lamb, R., Martin, M., Roizman, B., Straus, S. (Eds.), *Fields' Virology*, fourth ed. Lippincott Williams Wilkins, Philadelphia, pp. 19–52.
- Cornelissen, L.A., Wierda, C.M., van der Meer, F.J., Herrewegh, A.A., Horzinek, M.C., Egberink, H.F., de Groot, R.J., 1997. Hemagglutinin-esterase, a novel structural protein of torovirus. *J. Virol.* 71 (7), 5277–5286.
- Cowley, J.A., Dimmock, C.M., Spann, K.M., Walker, P.J., 2000. Gill-associated virus of *Penaeus monodon* prawns: an invertebrate virus with ORF1a and ORF1b genes related to arteri- and coronaviruses. *J. Gen. Virol.* 81 (Pt 6), 1473–1484.
- Drosten, C., Gunther, S., Preiser, W., van der Werf, S., Brodt, H.R., Becker, S., Rabenau, H., Panning, M., Kolesnikova, L., Fouchier, R.A., Berger, A., Burguiere, A.M., Cinatl, J., Eickmann, M., Escriou, N., Grywna, K., Kramme, S., Manuguerra, J.C., Muller, S., Rickerts, V., Stürmer, M., Vieth, S., Klenk, H.D., Osterhaus, A.D., Schmitz, H., Doerr, H.W., 2003. Identification of a novel coronavirus in patients with severe acute respiratory syndrome. *N. Engl. J. Med.* 348 (20), 1967–1976.
- Duckmanton, L., Carman, S., Nagy, E., Petric, M., 1998a. Detection of bovine torovirus in fecal specimens of calves with diarrhea from Ontario farms. *J. Clin. Microbiol.* 36 (5), 1266–1270.
- Duckmanton, L.M., Tellier, R., Liu, P., Petric, M., 1998b. Bovine torovirus: sequencing of the structural genes and expression of the nucleocapsid protein of Breda virus. *Virus Res.* 58 (1–2), 83–96.
- Duckmanton, L., Luan, B., Devenish, J., Tellier, R., Petric, M., 1997. Characterization of torovirus from human fecal specimens. *Virology* 239 (1), 158–168.
- Fischer, D., Barret, C., Bryson, K., Elofsson, A., Godzik, A., Jones, D., Karplus, K.J., Kelley, L.A., MacCallum, R.M., Pawowski, K., Rost, B., Rychlewski, L., Sternberg, M., 1999. CAFASP-1: critical assessment of fully automated structure prediction methods. *Proteins (Suppl. 3)*, 209–217.
- González, J.M., Gomez-Puertas, P., Cavanagh, D., Gorbalenya, A.E., Enjuanes, L., 2003. A comparative sequence analysis to revise the current taxonomy of the family Coronaviridae. *Arch. Virol.* 148 (11), 2207–2235.
- Guarne, A., Tormo, J., Kirchweger, R., Pfistermueller, D., Fita, I., Skern, T., 1998. Structure of the foot-and-mouth disease virus leader protease: a papain-like fold adapted for self-processing and eIF4G recognition. *Embo. J.* 17 (24), 7469–7479.
- Herold, J., Siddell, S.G., Gorbalenya, A.E., 1999. A human RNA viral cysteine proteinase that depends upon a unique Zn²⁺-binding finger connecting the two domains of a papain-like fold. *J. Biol. Chem.* 274 (21), 14918–14925.
- Hoet, A.E., Cho, K.O., Chang, K.O., Loerch, S.C., Wittum, T.E., Saif, L.J., 2002. Enteric and nasal shedding of bovine torovirus (Breda virus) in feedlot cattle. *Am. J. Vet. Res.* 63 (3), 342–348.
- Hoet, A.E., Nielsen, P.R., Hasoksuz, M., Thomas, C., Wittum, T.E., Saif, L.J., 2003a. Detection of bovine torovirus and other enteric pathogens in feces from diarrhea cases in cattle. *J. Vet. Diagn. Invest.* 15 (3), 205–212.
- Hoet, A.E., Smiley, J., Thomas, C., Nielsen, P.R., Wittum, T.E., Saif, L.J., 2003b. Association of enteric shedding of bovine torovirus (Breda virus) and other enteropathogens with diarrhea in veal calves. *Am. J. Vet. Res.* 64 (4), 485–490.
- Jamieson, F.B., Wang, E.E., Bain, C., Good, J., Duckmanton, L., Petric, M., 1998. Human torovirus: a new nosocomial gastrointestinal pathogen. *J. Infect. Dis.* 178 (5), 1263–1269.
- Kelley, L.A., MacCallum, R.M., Sternberg, M., 1999. Recognition of remote protein homologies using three dimensional information to generate a position specific scoring matrix in the program 3D-PSSM. In: Istrail, S., Pevzner, P., Waterman, M. (Eds.), *RECOMB 99, Proceedings of the Third Annual Conference on Computational Molecular Biology*. The Association for Computing Machinery, New York, pp. 218–225.
- Kelley, L.A., MacCallum, R.M., Sternberg, M.J., 2000. Enhanced genome annotation using structural profiles in the program 3D-PSSM. *J. Mol. Biol.* 299 (2), 499–520.
- Koopmans, M., Cremers, H., Woode, G., Horzinek, M.C., 1990. Breda virus (Toroviridae) infection and systemic antibody response in sentinel calves. *Am. J. Vet. Res.* 51 (9), 1443–1448.
- Koopmans, M., Horzinek, M.C., 1994. Toroviruses of animals and humans: a review. *Adv. Virus Res.* 43, 233–273.
- Koopmans, M., van Wuijkhuise-Sjouke, L., Schukken, Y.H., Cremers, H., Horzinek, M.C., 1991. Association of diarrhea in cattle with torovirus infections on farms. *Am. J. Vet. Res.* 52 (11), 1769–1773.
- Koustouros, E., Siu, K., Ford-Jones, E.L., Petric, M., Tellier, R., 2003. Molecular characterization of rotavirus strains from children in Toronto, Canada. *J. Clin. Virol.* 28 (1), 77–84.
- Kroneman, A., Cornelissen, L.A., Horzinek, M.C., de Groot, R.J., Egberink, H.F., 1998. Identification and characterization of a porcine torovirus. *J. Virol.* 72 (5), 3507–3511.
- Ksiazek, T.G., Erdman, D., Goldsmith, C.S., Zaki, S.R., Peret, T., Emery, S., Tong, S., Urbani, C., Comer, J.A., Lim, W., Rollin, P.E., Dowell, S.F., Ling, A.E., Humphrey, C.D., Shieh, W.J., Guarner, J., Paddock, C.D., Rota, P., Fields, B., DeRisi, J., Yang, J.Y., Cox, N., Hughes, J.M., LeDuc, J.W., Bellini, W.J., Anderson, L.J., 2003. A

- novel coronavirus associated with severe acute respiratory syndrome. *N. Engl. J. Med.* 348 (20), 1953–1966.
- Lai, M.M., Cavanagh, D., 1997. The molecular biology of coronaviruses. *Adv. Virus. Res.* 48, 1–100.
- Lai, M.M., Holmes, K.V., 2001. Coronaviridae: the viruses and their replication. In: Knipe, D., Howley, P.M., Griffin, D.E., Martin, M.A., Lamb, R.A., Roizman, B. (Eds.), *Fields Virology*, vol. 1 (2 vols.), fourth ed. Lippincott Williams & Wilkins, Philadelphia, pp. 1163–1186.
- Lee, N., Hui, D., Wu, A., Chan, P., Cameron, P., Joynt, G.M., Ahuja, A., Yung, M.Y., Leung, C.B., To, K.F., Lui, S.F., Szeto, C.C., Chung, S., Sung, J.J., 2003. A major outbreak of severe acute respiratory syndrome in Hong Kong. *N. Engl. J. Med.* 348 (20), 1986–1994.
- Leung, W.K., To, K.F., Chan, P.K., Chan, H.L., Wu, A.K., Lee, N., Yuen, K.Y., Sung, J.J., 2003. Enteric involvement of severe acute respiratory syndrome-associated coronavirus infection. *Gastroenterology* 125 (4), 1011–1017.
- Martino, T.A., Tellier, R., Petric, M., Irwin, D.M., Afshar, A., Liu, P.P., 1999. The complete consensus sequence of coxsackievirus B6 and generation of infectious clones by long RT-PCR. *Virus Res.* 64 (1), 77–86.
- Martzen, M.R., McCraith, S.M., Spinelli, S.L., Torres, F.M., Fields, S., Grayhack, E.J., Phizicky, E.M., 1999. A biochemical genomics approach for identifying genes by the activity of their products. *Science* 286 (5442), 1153–1155.
- Pedersen, K.W., van der Meer, Y., Roos, N., Snijder, E.J., 1999. Open reading frame 1a-encoded subunits of the arterivirus replicase induce endoplasmic reticulum-derived double-membrane vesicles which carry the viral replication complex. *J. Virol.* 73 (3), 2016–2026.
- Poutanen, S.M., Low, D.E., Henry, B., Finkelstein, S., Rose, D., Green, K., Tellier, R., Draker, R., Adachi, D., Ayers, M., Chan, A.K., Skowronski, D.M., Salit, I., Simor, A.E., Slutsky, A.S., Doyle, P.W., Krajden, M., Petric, M., Brunham, R.C., McGeer, A.J., 2003. Identification of severe acute respiratory syndrome in Canada. *N. Engl. J. Med.* 348 (20), 1995–2005.
- Qi, J., Luo, H., Hao, B., 2004a. CVTree: a phylogenetic tree reconstruction tool based on whole genomes. *Nucleic Acids Res.* 32 (Web Server issue), W45–W47.
- Qi, J., Wang, B., Hao, B.I., 2004b. Whole proteome prokaryote phylogeny without sequence alignment: a *K*-string composition approach. *J. Mol. Evol.* 58 (1), 1–11.
- Rowe, C.L., Baker, S.C., Nathan, M.J., Fleming, J.O., 1997. Evolution of mouse hepatitis virus: detection and characterization of spike deletion variants during persistent infection. *J. Virol.* 71 (4), 2959–2969.
- Smits, S.L., Lavazza, A., Matiz, K., Horzinek, M.C., Koopmans, M.P., de Groot, R.J., 2003. Phylogenetic and evolutionary relationships among torovirus field variants: evidence for multiple intertypic recombination events. *J. Virol.* 77 (17), 9567–9577.
- Snijder, E.J., Bredenbeek, P.J., Dobbe, J.C., Thiel, V., Ziebuhr, J., Poon, L.L., Guan, Y., Rozanov, 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.
- Snijder, E.J., den Boon, J.A., Bredenbeek, P.J., Horzinek, M.C., Rijbrand, R., Spaan, W.J., 1990. The carboxyl-terminal part of the putative Berne virus polymerase is expressed by ribosomal frameshifting and contains sequence motifs which indicate that toro- and coronaviruses are evolutionarily related. *Nucleic Acids Res.* 18 (15), 4535–4542.
- Snijder, E.J., den Boon, J.A., Horzinek, M.C., Spaan, W.J., 1991. Comparison of the genome organization of toro- and coronaviruses: evidence for two nonhomologous RNA recombination events during Berne virus evolution. *Virology* 180 (1), 448–452.
- Snijder, E.J., Horzinek, M.C., 1993. Toroviruses: replication, evolution and comparison with other members of the coronavirus-like superfamily. *J. Gen. Virol.* 74 (Pt 11), 2305–2316.
- Snijder, E.J., Meulenberg, J.J., 1998. The molecular biology of arteriviruses. *J. Gen. Virol.* 79 (Pt 5), 961–979.
- Snijder, E.J., Wassenaar, A.L., van Dinten, L.C., Spaan, W.J., Gorbalenya, A.E., 1996. The arterivirus nsp4 protease is the prototype of a novel group of chymotrypsin-like enzymes, the 3C-like serine proteases. *J. Biol. Chem.* 271 (9), 4864–4871.
- Tellier, R., Bukh, J., Emerson, S.U., Miller, R.H., Purcell, R.H., 1996a. Long PCR and its application to hepatitis viruses: amplification of hepatitis A, hepatitis B, and hepatitis C virus genomes. *J. Clin. Microbiol.* 34 (12), 3085–3091.
- Tellier, R., Bukh, J., Emerson, S.U., Purcell, R.H., 1996b. Amplification of the full-length hepatitis A virus genome by long reverse transcription-PCR and transcription of infectious RNA directly from the amplicon. *Proc. Natl. Acad. Sci. U.S.A.* 93 (9), 4370–4373.
- Tellier, R., Bukh, J., Emerson, S.U., Purcell, R.H., 2003. Long PCR amplification of large fragments of viral genomes: a technical overview. *Methods Mol. Biol.* 226, 167–172.
- 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 (24), 4876–4882.
- Upton, C., Slack, S., Hunter, A.L., Ehlers, A., Roper, R.L., 2003. Poxvirus orthologous clusters: toward defining the minimum essential poxvirus genome. *J. Virol.* 77 (13), 7590–7600.
- van der Meer, Y., van Tol, H., Locker, J.K., Snijder, E.J., 1998. ORF1a-encoded replicase subunits are involved in the membrane association of the arterivirus replication complex. *J. Virol.* 72 (8), 6689–6698.
- Weiss, M., Steck, F., Kaderli, R., Horzinek, M.C., 1984. Antibodies to Berne virus in horses and other animals. *Vet. Microbiol.* 9 (6), 523–531.
- Woode, G.N., Reed, D.E., Runnels, P.L., Herrig, M.A., Hill, H.T., 1982. Studies with an unclassified virus isolated from diarrhetic calves. *Vet. Microbiol.* 7 (3), 221–240.
- Ziebuhr, J., Bayer, S., Cowley, J.A., Gorbalenya, A.E., 2003. The 3C-like proteinase of an invertebrate nidovirus links coronavirus and potyvirus homologs. *J. Virol.* 77 (2), 1415–1426.
- Ziebuhr, J., Snijder, E.J., Gorbalenya, A.E., 2000. Virus-encoded proteinases and proteolytic processing in the Nidovirales. *J. Gen. Virol.* 81 (Pt 4), 853–879.
- Zuker, M., 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 31 (13), 3406–3415.