

Short communication

Sequence analysis of the turkey coronavirus nucleocapsid protein gene and 3' untranslated region identifies the virus as a close relative of infectious bronchitis virus

Jamie J. Breslin, Lynda G. Smith, Frederick J. Fuller, James S. Guy *

Department of Microbiology, Pathology, and Parasitology, North Carolina State University, College of Veterinary Medicine, 4700 Hillsborough Street, Raleigh, NC 27606, USA

Received 7 July 1999; received in revised form 1 September 1999; accepted 1 September 1999

Abstract

The 3' end of the turkey coronavirus (TCV) genome (1740 bases) including the nucleocapsid (N) gene and 3' untranslated region (UTR) were sequenced and compared with published sequences of other avian and mammalian coronaviruses. The deduced sequence of the TCV N protein was determined to be 409 amino acids with a molecular mass of approximately 45 kDa. The TCV N protein was identical in size and had greater than 90% amino acid identity with published N protein sequences of infectious bronchitis virus (IBV); less than 21% identity was observed with N proteins of bovine coronavirus and transmissible gastroenteritis virus. The 3' UTR showed some variation among the three TCV strains examined, with two TCV strains, Minnesota and Indiana, containing 153 base segments which are not present in the NC95 strain. Nucleotide sequence identity between the 3' UTRs of TCV and IBV was greater than 78%. Similarities in both size and sequence of TCV and IBV N proteins and 3' UTRs provide additional evidence that these avian coronaviruses are closely related. © 1999 Elsevier Science B.V. All rights reserved.

Keywords: Turkey coronavirus; Infectious bronchitis virus; Nucleocapsid gene

1. Introduction

Coronaviruses are enveloped, pleomorphic viruses, 80–220 nm in diameter with club shaped surface projections, and a positive-sense single-

stranded RNA genome of 20–30 kilobases (kb; Siddell, 1995; Holmes and Lai, 1996). The virion contains at least three major structural proteins, the surface (S) glycoprotein (90–180 kilodaltons (kDa)), an integral membrane (M) protein (20–35 kDa), and a nucleocapsid (N) protein (50–60 kDa; Siddell, 1995; Holmes and Lai, 1996). Additionally, some coronaviruses also contain a fourth major structural protein, the hemagglutinin-es-

* Corresponding author. Tel.: +1-919-5136287; fax: +1-919-5136455.

E-mail address: jim_guy@ncsu.edu (J.S. Guy)

terase (HE) protein (120–140 kDa; Siddell, 1995; Holmes and Lai, 1996).

Coronaviruses have been subdivided into three major antigenic groups based on differences identified by serological analyses, and these findings have been substantiated by nucleotide sequence analyses (Pedersen, 1978; Robb and Bond, 1979; Wege et al., 1982; Williams et al., 1992). Human coronavirus (HCV) 229E, transmissible gastroenteritis virus (TGEV), canine coronavirus and feline infectious peritonitis virus are members of group I, HCV OC43, murine hepatitis virus (MHV) and bovine coronavirus (BCV) are members of group II, and infectious bronchitis virus (IBV) belongs to group III. Previous studies by Dea et al. (1990) and Verbeek and Tijssen (1991) indicated that turkey coronavirus (TCV) is a member of group II; however, recent antigenic studies and nucleotide sequence analyses indicate that TCV belongs to group III (Guy et al., 1997; Breslin et al., 1999; Stephensen et al., 1999).

The N protein has been shown to be highly variable in size and amino acid composition between the viruses that comprise the three coronavirus antigenic groups, but highly conserved within these groups (Williams et al., 1992; Siddell, 1995). Lapps et al. (1987) examined the degree of amino acid identity between the BCV N protein and other coronaviruses; they found that BCV and MHV, group II coronaviruses, had 70% identity, but only 29% identity with TGEV (group I) and IBV (group III). The N protein tends to be highly conserved among different strains of the same coronavirus (Williams et al., 1992; Laude and Masters, 1995). N proteins of 27 different infectious bronchitis virus (IBV) strains isolated in USA, UK, Holland, Saudi Arabia, and Japan were shown to have greater than 94% identity at the amino acid level (Williams et al., 1992; Zwaagstra et al., 1992).

The 3' untranslated region (UTR) of coronaviruses is found downstream of the N gene and is believed to be important in initiation of negative-strand RNA synthesis. This region, like the N gene, has been shown to be highly conserved among different strains of the same coronavirus (Collisson et al., 1990; Williams et al., 1993; Hsue and Masters, 1997).

The avian coronaviruses, IBV and TCV, cause several host specific diseases of economic importance. Infectious bronchitis virus is the cause of an acute, highly contagious respiratory disease in chickens with potential involvement of kidney and reproductive tract (Cavanagh and Naqi, 1997). TCV is the cause of an acute, highly contagious enteric disease of turkeys referred to as bluecomb disease (Nagaraja and Pomeroy, 1997). Early studies indicated that IBV and TCV were antigenically unrelated based on immune electron microscopy, hemagglutination inhibition and virus-neutralization studies (Ritchie et al., 1973; Dea et al., 1986); however, these findings are inconsistent with recent antigenic and nucleotide sequence analyses (Guy et al., 1997; Breslin et al., 1999; Stephensen et al., 1999). Antigenic analyses by Guy et al. (1997) indicated that IBV and TCV were closely related based on cross-immunofluorescent studies using both polyclonal and monoclonal antibodies. Subsequent studies indicated that IBV and TCV were closely related based on the nucleotide sequence of a 1.1-kb segment of the TCV genome spanning portions of both the M and N genes (Breslin et al., 1999). Additionally, Stephensen et al. (1999) demonstrated a close genetic relationship between TCV and IBV based on sequence analysis of a highly conserved region of the polymerase gene. In the present study, the complete nucleocapsid gene and the 3' UTR of three epidemiologically distinct TCV strains were sequenced and compared with those of previously published IBV strains and representative members of groups I and II coronaviruses, TGEV and BCV, respectively.

2. Materials and methods

TCV strains examined in the present study included Minnesota strain (American Type Culture Collection, ATCC VR-911), Indiana strain (Tom Hooper, Purdue University), and NC95 strain (Guy et al., 1997). These viruses were isolated from turkeys in Minnesota in 1974, Indiana in 1994 and North Carolina in 1995, respectively. TCV strains were propagated by amniotic inoculation of embryonated turkey eggs (Guy et al.,

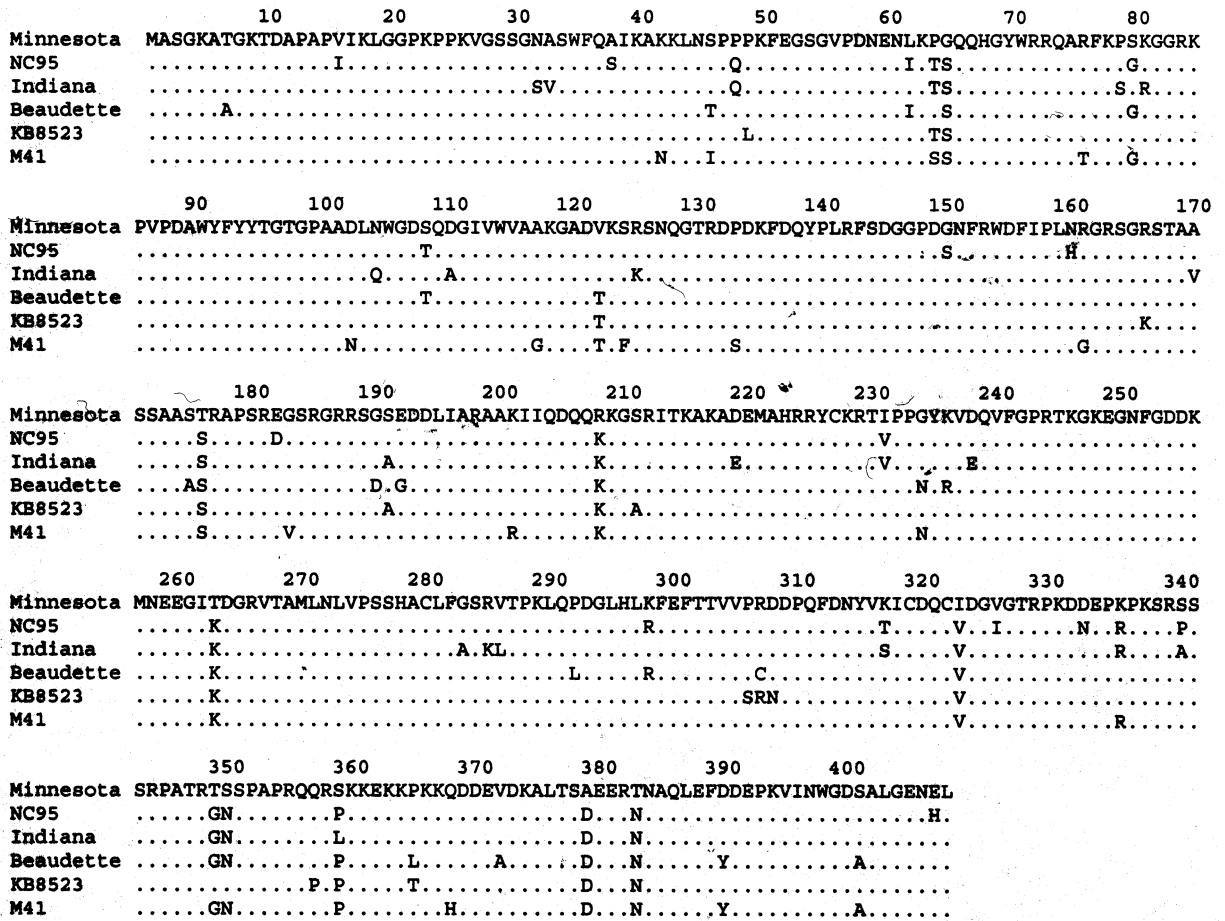


Fig. 1. Comparison of amino acid sequences of the nucleocapsid protein of TCV (Minnesota, Indiana and NC95 strains) and published sequences of IBV (Beaudette, KB8523 and M41 strains; Bournsnel et al., 1985; Sutou et al., 1988). The positions where amino acids are identical are indicated as (·).

Table 1

Percent sequence identity between N proteins of TCV (Minnesota, Indiana and NC95 strains) and published sequences of IBV (Beaudette, KB8523 and M41 strains), BCV (Mebus) and TGEV (Purdue; Bournsnel et al., 1985; Kapke and Brian, 1986; Lapps et al., 1987; Sutou et al., 1988)

	TCV (Minnesota)	TCV (NC95)	TCV (Indiana)	IBV (Beaudette)	IBV (KB8523)	IBV (M41)	BCV (Mebus)
TCV (NC95)	93.2%						
TCV (Indiana)	92.7%	92.9%					
IBV (Beaudette)	93.2%	93.2%	90.5%				
IBV (KB8523)	95.6%	93.2%	92.9%	93.4%			
IBV (M41)	93.2%	92.4%	91.2%	93.9%	93.4%		
BCV (Mebus)	18.6%	20.2%	20.0%	20.0%	18.3%	19.3%	
TGEV (Purdue)	20.2%	20.2%	19.4%	20.4%	20.4%	19.1%	20.2%

	10	20	30	40	50	60	70	80	
Minnesota	GTAACATAATGGACCTGC-ATTTCTTAG-ACATTTTGTAAACACTATTTTGTACTTTCCCTATCAATTATTACAAGCATTGATTG								
IndianaT.G.....GC.....C.G.....G.....								
NC95	-----								
M41	-----								
BeaudetteTTG...C.G.T.....C.G.....G.....								
	90	100	110	120	130	140	150	160	170
Minnesota	TGATTTTGGTTCAATATTTAAGCTTCTTTTGGTTGCTTTTGGCTTGTGTTGATTTGTTGCTGTGCTTTTATTATTGTGATTCTCATTAA								
Indiana	.C.A.....								
NC95	-----A.C..C..C.....								
M41	-----								
BeaudetteA.....C.....A.....G.....								
	180	190	200	210	220	230	240	250	
Minnesota	CTTTGCTTTATCGTAGAAGTTCAATAGTAAGAGTTAAGGCAGATAGGCATGTAGCTTAGC-ACCTACATGTCTATCGCCAGGGAAA								
Indiana	G.....A.....								
NC95	.C.T.....T.....A..T..T.....A.....								
M41	-----A.....A.....GATT.....								
Beaudette	G.....A.....A.....GATT.....G.....								
	260	270	280	290	300	310	320	330	340
Minnesota	TGTCTAATCTGTCTACCTAGTAGCCTGGAAACGAACGGTAGACCCTTAGATTTTAAATTTAGTPTAATTTTGTAGTTAGTTAAGTT								
IndianaT.....G.....C.....								
NC95T.....								
M41T.....								
BeaudetteT.....C.....								
	350	360	370	380	390	400	410	420	
Minnesota	AGTTTAGAGTAGGTATAAAGATGCCAGTGCCGAGGCCACGCCGAGTACGATCGAGGGTACAGCACTAGGACGCCCATAGGGGAAG								
IndianaG.....C.....T.....C.....								
NC95G.....A.....								
M41G.....C.....								
BeaudetteG.....								
	430	440	450	460	470	480	500	510	
Minnesota	AGCTAAATTTTAGTTAAGTTAAGTTAATTTGGCTAAGTATAGTTAAAATTTATAGGCTAGTATAGAGTTAGAGC								
IndianaC.G.....								
NC95	-----								
M41T.....								
Beaudette	-----								

Fig. 2. Comparison of nucleotide sequences of the 3' untranslated region of TCV (Minnesota, Indiana, and NC95 strains) and published sequences of IBV (Beaudette and M41 strains; Bourns et al., 1985; Sutou et al., 1988). The positions where nucleotides are missing are indicated as (-) and identical nucleotides are indicated as (·).

1997). Viral RNA was extracted from sucrose gradient-purified virus, and cDNA synthesis was accomplished by reverse transcriptase polymerase chain reaction (RT-PCR) as described (Breslin et al., 1999). The RT reaction was primed with an oligo (dT) primer 15 bases in length; this primer along with a specific oligonucleotide primer designed from the 5' end of the N gene (TGAATTCTAAATTCACCTCAACCTAAGT) (Breslin et al., 1999) was used in the PCR procedure. Primers possessed *Eco*RI restriction sites that facilitated cloning of cDNA into pUC19;

clones were used to transform competent *Escherichia coli* strain DH5 α (Gibco BRL) as described (Sambrook et al., 1989). DNA was sequenced at the University of North Carolina (Chapel Hill) Automated DNA Sequencing Facility on a Model 373A DNA Sequencer using the Taq Dye Deoxy™ Terminator Cycle Sequencing Kit (Applied Biosystems). All sequences were confirmed by sequencing of both strands. Preliminary nucleotide sequence data allowed the design of primers to amplify internal N gene and 3' UTR regions to confirm the nucleotide sequence.

Primers again were designed with *EcoRI* restriction sites, used in a RT-PCR procedure and cDNA fragments were cloned and sequenced as described above. The complete nucleotide sequence of the N gene and the 3'UTR of each TCV have been deposited in GenBank; accession numbers are TCV (Indiana) AF111995, TCV (Minnesota) AF111996, TCV (NC95) AF111997.

3. Results and discussion

Nucleotide sequences were entered into the Translate Tool of the ExPASy molecular biology www server of the Swiss Institute of Bioinformatics and used to determine possible open reading frames (ORF). One ORF was identified for each TCV strain that predicted a protein, 409 amino acids in length with a molecular mass of approximately 45 kDa. Previous sequence analyses of the N proteins of IBV strains (Beaudette, M41, and KB8523; Bournsnell et al., 1985; Sutou et al., 1988; Jia et al., 1995) resulted in similar findings: IBV N proteins were determined to be 409 amino acids in length with a molecular mass of approximately 45 kDa. In contrast, groups I and II coronaviruses have been shown to possess N proteins of 378–389 amino acids and 448–455 amino acids, respectively (Kapke and Brian, 1986; Lapps et al., 1987; Parker and Masters, 1990; Williams et al., 1992).

The deduced amino acid sequences of the N protein of TCV is shown in Fig. 1. The TCV (Minnesota) strain is used as the reference strain with amino acid differences noted for Indiana strain, NC95 strain and three strains of IBV (Beaudette, M41, KB8523; Bournsnell et al., 1985; Sutou et al., 1988). A comparison of the percent identity of TCV N protein amino acid sequences with published sequences of three IBV strains (Beaudette, M41, KB8523), BCV (Mebus) and TGEV (Purdue) is presented in Table 1 (Bournsnell et al., 1985; Kapke and Brian, 1986; Lapps et al., 1987; Sutou et al., 1988). Overall the N protein sequence of TCV and IBV strains have greater than 90% identity. TCV strains have 92.7–93.2% sequence identity when compared to each other, and 90.5–95.6% identity to IBV. In contrast, TCV and IBV N proteins have less than 21% identity

with BCV (Mebus) and TGEV (Purdue). Previous studies comparing the TCV M protein (70 amino acids at carboxy-terminus) and N protein (55 amino acids at amino-terminus) and published sequences from other coronaviruses indicated greater than 90% identity of both protein segments with IBV, and less than 30% identity with groups I and II coronaviruses (Breslin et al., 1999).

Differences among TCV strains were observed in the 3' UTRs (Fig. 2). TCV (Minnesota) and TCV (Indiana) contained a 153-nucleotide segment that was not present in the NC95 strain. 3' UTRs of TCV (Indiana) and TCV (Minnesota) were 502 bp in length, compared with a 349-bp 3' UTR of TCV (NC95). The missing nucleotide segment in the NC95 3' UTR occurs immediately downstream of the N gene (Fig. 2). Similar differences in structure of 3' UTRs have been observed among IBV strains (Williams et al., 1993; Sapats et al., 1996). IBV strains (Beaudette, KB8523, CU-T2) have been shown to possess 3' UTRs of 503–505 nucleotides (Bournsnell et al., 1985; Sutou et al., 1988). IBV (M41) was found to differ from other strains in that it possessed a 3' UTR of 320 bases; compared with other IBV strains, M41 lacked a 183–196 nucleotide sequence occurring four bases downstream of the N gene (Bournsnell et al., 1985; Sapats et al., 1996). The significance of these differences in the 3' UTRs has not been determined (Collisson et al., 1990; Sapats et al., 1996).

Comparison of 3' UTRs of TCV strains (Minnesota, Indiana, NC95) and published sequences of three IBV strains (Beaudette, KB8523, M41) demonstrated a sequence identity greater than 78% (data not shown). TCV strains had a 90.8–96.0% sequence similarity when compared to each other, and 78.5–94.4% similarity to IBV. TCV 3' UTRs had less than 30% similarity with those of BCV (Mebus) and TGEV (Purdue). In these comparisons the large deletions in the 5' portion of the 3' UTR of TCV (NC95) and IBV (M41) were counted as single differences, rather than considering each missing base as a separate difference.

Williams et al. (1993) previously compared 3' UTR sequences of several different IBV strains, comparing them in two sections: the 5' region not found in IBV (M41) (184–196 bases) and the remaining bases downstream of this sequence.

Results of the study indicated that the 5' end of the IBV 3' UTR is quite variable between strains, ranging from 53.2 to 92.8% identity. In contrast, the sequence downstream of this hypervariable region was highly conserved with 94.3–97.8% identity (Williams et al., 1993). In the present study we performed similar comparisons with the TCV 3' UTR sequence data (data not shown). The 5' region of the 3' UTR (153 bases) of TCV (Minnesota) and TCV (Indiana) had 94.7% identity with each other and 57.8–90.4% identity with IBV strains (Beaudette, KB8523). The remaining bases of the 3' UTR of TCV strains (Minnesota, Indiana, NC95) had 96.5–97.8% sequence identity when compared to each other, and 91.7–95.2% identity to IBV strains (Beaudette, KB8523, M41).

In summary, the amino acid sequence of TCV N protein and the nucleotide sequence of the 3' UTR were compared with published sequences of other avian and mammalian coronaviruses. The size and sequence characteristics of the TCV N protein and 3' UTR closely resembled those of IBV strains, thus supporting previous antigenic analyses and nucleotide sequence studies that indicated a close relationship between TCV and IBV (Breslin et al., 1999; Guy et al., 1997; Stephensen et al., 1999). Together, these findings refute previous studies that indicated a close relationship between TCV and group II coronaviruses (Dea et al., 1990; Verbeek and Tijssen, 1991). These findings instead indicate that the avian coronaviruses, IBV and TCV, share a close phylogenetic relationship and together comprise group III of the coronavirus major antigenic groups.

References

- Bournell, M.E., Binns, M.M., Foulds, I.J., Brown, T.D., 1985. Sequences of the nucleocapsid genes from two strains of avian infectious bronchitis virus. *J. Gen. Virol.* 66, 573–580.
- Breslin, J.J., Smith, L.G., Fuller, F.J., Guy, J.S., 1999. Sequence analysis of the matrix/nucleocapsid gene region of turkey coronavirus. *Intervirology* 42, 22–29.
- Cavanagh, D., Naqi, S.A., 1997. Infectious bronchitis. In: Calnek, B.W., Barnes, H.J., Beard, C.W., McDougald, L.R., Saif, Y.M. (Eds.), *Diseases of Poultry*, 10th ed. Iowa State University Press, Ames, IA, pp. 511–526.
- Collisson, E.W., Williams, A.K., Vonder Harr, R., Li, W., Sneed, L.W., 1990. Sequence comparisons of the 3' end of the genomes of five strains of avian infectious bronchitis virus. In: Cavanagh, D., Brown, T.D.K. (Eds.), *Coronaviruses and Their Diseases*. Plenum, New York, pp. 373–377.
- Dea, S., Marsolais, G., Beaubien, J., Ruppner, R., 1986. Coronaviruses associated with outbreaks of transmissible enteritis of turkeys in Quebec: hemagglutination properties and cell cultivation. *Avian Dis.* 30, 319–326.
- Dea, S., Verbeek, A.J., Tijssen, P., 1990. Antigenic and genomic relationships among turkey and bovine enteric coronaviruses. *J. Virol.* 64, 3112–3118.
- Guy, J.S., Barnes, H.J., Smith, L.G., Breslin, J., 1997. Antigenic characterization of a turkey coronavirus identified in poult enteritis- and mortality syndrome-affected turkeys. *Avian Dis.* 41, 583–590.
- Holmes, K.V., Lai, M.M.C., 1996. Coronaviridae: the viruses and their replication. In: Fields, B.N., Knipe, D.M., Howly, P.M. (Eds.), *Fundamental Virology*, 3rd ed., Vol. 1. 2 vols. Lippincott-Raven, Philadelphia, PA, pp. 1075–1093.
- Hsue, B., Masters, P.S., 1997. A bulged stem-loop in the 3' untranslated region of the genome of the coronavirus mouse hepatitis virus is essential for replication. *J. Virol.* 71, 7567–7578.
- Jia, W., Karaca, K., Parrish, C.R., Naqi, S.A., 1995. A novel variant of avian infectious bronchitis virus resulting from recombination among three different strains. *Arch. Virol.* 140, 259–271.
- Kapke, P.A., Brian, D.A., 1986. Sequence analysis of the porcine transmissible gastroenteritis coronavirus nucleocapsid protein gene. *Virology* 151, 41–49.
- Lapps, W., Hogue, B.G., Brian, D.A., 1987. Sequence analysis of the bovine coronavirus nucleocapsid and matrix protein genes. *Virology* 157, 47–57.
- Laude, H., Masters, P.S., 1995. The coronavirus nucleocapsid protein. In: Siddell, S.G. (Ed.), *The Coronaviridae*. Plenum, New York, pp. 141–163.
- Nagaraja, K.V., Pomeroy, B.S., 1997. Coronaviral enteritis of turkeys (bluecomb disease). In: Calnek, B.W., Barnes, H.J., Beard, C.W., McDougald, L.R., Saif, Y.M. (Eds.), *Diseases of Poultry*, 10th ed. Iowa State University Press, Ames, IA, pp. 686–692.
- Parker, M.M., Masters, P.S., 1990. Sequence comparison of the N genes of five strains of the coronavirus mouse hepatitis virus suggests a three-domain structure for the nucleocapsid protein. *Virology* 179, 463–468.
- Pedersen, N.C., 1978. Antigenic relationship of feline infectious peritonitis virus to coronaviruses of other species. *Arch. Virol.* 58, 45–53.
- Ritchie, A.E., Desmukh, D.R., Larsen, C.T., Pomeroy, B.S., 1973. Electron microscopy of coronavirus-like particles characteristic of turkey bluecomb disease. *Avian Dis.* 17, 546–558.
- Robb, J.A., Bond, C.W., 1979. Coronaviridae. In: Fraenkel-Conrat, H., Wagner, R.R. (Eds.), *Comprehensive Virology*, vol. 14. Plenum, New York, pp. 193–247.

- Sambrook, J., Fritsch, E.F., Maniatis, T. (Eds.), 1989. *Molecular Cloning, A Laboratory Manual*, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
- Sapats, S.I., Ashton, F., Wright, P.J., Ignjatovic, J., 1996. Novel variation in the N protein of avian infectious bronchitis virus. *Virology* 226, 412–417.
- Siddell, S.G., 1995. The coronaviridae an introduction. In: Siddell, S.G. (Ed.), *Coronaviridae*. Plenum, New York, pp. 1–9.
- Stephensen, C.B., Casebolt, D.B., Gangopadhyay, N.N., 1999. Phylogenetic analysis of a highly conserved region of the polymerase gene from 11 coronaviruses and development of a consensus polymerase chain reaction assay. *Virus Res.* 60, 181–189.
- Sutou, S., Sato, S., Okabe, T., Nakai, M., Sasaki, N., 1988. Cloning and sequencing of genes encoding structural proteins of avian infectious bronchitis virus. *Virology* 165, 589–595.
- Verbeek, A., Tijssen, P., 1991. Sequence analysis of the turkey enteric coronavirus nucleocapsid and membrane protein genes: a close genomic relationship with bovine coronavirus. *J. Gen. Virol.* 72, 1659–1666.
- Wege, H., Siddell, S., ter Meulen, V., 1982. The biology and pathogenesis of coronaviruses. *Curr. Top. Microbiol. Immunol.* 99, 165–200.
- Williams, A.K., Wang, L., Sneed, L.W., Collisson, E.W., 1992. Comparative analyses of the nucleocapsid genes of several strains of infectious bronchitis virus and other coronaviruses. *Virus Res.* 25, 213–222.
- Williams, A.K., Wang, L., Sneed, L.W., Collisson, E.W., 1993. Analysis of a hypervariable region in the 3' non-coding end of the infectious bronchitis virus genome. *Virus Res.* 28, 19–27.
- Zwaagstra, K.A., van der Zeijst, B.A.M., Kusters, J.G., 1992. Rapid detection and identification of avian infectious bronchitis virus. *J. Clin. Microbiol.* 30, 79–84.