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.

VIROLOGY 226, 412–417 (1996)
ARTICLE NO. 0670

SHORT COMMUNICATION

Novel Variation in the N Protein of Avian Infectious Bronchitis Virus[1]

S. I. SAPATS,*,†,[2] F. ASHTON,* P. J. WRIGHT,† and J. IGNJATOVIC*

*CSIRO, Australian Annual Health Laboratory, Geelong 3220; and
†Department of Microbiology, Monash University, Clayton 3168, Australia

Received March 27, 1996; revised September 18, 1996; accepted October 1, 1996

The nucleocapsid protein of coronaviruses has been considered highly conserved, showing greater than 94% conservation within strains of a given species. We determined the nucleotide sequence of the N gene and the 3′ untranslated region (UTR) of eight naturally occurring strains of IBV which differed in pathogenicity and tissue tropism. In pairwise comparisons, the deduced amino acid sequences of N of five strains Vic S, N1/62, N9/74, N2/75, and V5/90 (group I) shared 92.3–98.8% identity. The three strains N1/88, Q3/88, and V18/91 (group II) shared 85.8–89.2% identity with each other, but only 60.0–63.3% identity with viruses of group I. Amino acid substitutions, deletions, and insertions occurred throughout the N protein and involved regions previously identified as being conserved. Despite the considerable variation observed between the two virus groups, all N proteins contained a high proportion of basic residues, 80% of which were conserved in position. In addition, all strains contained approximately 30 serine residues of which 10 were conserved, the majority occurring between positions 168 and 194. As for all other coronaviruses, the region between positions 92 and 103 was highly conserved. Hence, a large number of amino acid changes can be tolerated within the N protein without affecting its integrity or functioning. The 3′ UTR immediately downstream from the N gene was highly heterogeneous with extensive deletions occurring in the group II strains. © 1996 Academic Press, Inc.

Infectious bronchitis virus (IBV), a member of the family *Coronaviridae*, causes an acute highly contagious disease of chickens resulting in significant economic losses to poultry industries throughout the world. The IBV genome consists of a single strand of positive sense RNA measuring 27.6 kb in length excluding the poly(A) tail (*1*). The genes encoding the three major structural proteins are situated within an 8-kb region located at the 3′ end of the genome. These proteins are the spike glycoprotein (S), the membrane glycoprotein (M), and the phosphorylated nucleocapsid protein (N). The N protein plays a role in viral replication, assembly, and immunity. It interacts with leader RNA sequences facilitating viral mRNA synthesis and also binds to the viral RNA forming a helical nucleocapsid (2). The N protein of all coronaviruses is overall very basic and in IBV contains 409 amino acids with a predicted M_r of approximately 50,000 (*3*); it also contains a high proportion of serine residues which act as sites for phosphorylation (*2*).

The N protein of 27 strains of IBV isolated over a period of 60 years from diverse locations such as the U.S.A., the UK, Holland, Saudi Arabia, and Japan has been shown to be highly conserved differing by only 2 to 6% at the amino acid level (*4, 5*). The high level of conservation in the N protein has resulted in the widely held view that the S1 glycoprotein, which may show up to 49% variation (*6*), is the only relevant structural element in assessing the genetic diversity and evolutionary direction of IBV.

Immediately downstream of the N gene is the 3′ untranslated region (UTR) which is presumably important in the initiation of negative-strand RNA synthesis. The organization of the 3′ UTR differs among the coronaviruses. In porcine, canine, and feline coronaviruses this region is conserved among strains within a species and contains at least one open reading frame (ORF) (*7, 8*). To the present time, functional ORFs have not been detected in the 3′ UTR of IBV. However, the virulent M41 virus lacks a sequence of 184–196 nucleotides that has been detected in five other IBV strains. This sequence has been termed a hypervariable region (HVR) and was proposed to be exogenous in origin (*3, 9*). The region located downstream of the HVR (315 nucleotides ending at the poly(A) tail) is highly conserved in strains of IBV, probably indicative of its role in the synthesis of negative-strand RNA.

Recently we reported the isolation of three IBV strains, N1/88, Q3/88, and V18/91, which differed from strains isolated previously in tissue tropism and pathogenicity. In pairwise comparisons, the S1 glycoprotein of these strains differed by 38–46% from the S1 glycoprotein of

[1] The nucleotide sequence data reported in this paper has been submitted to the GenBank nucleotide database and has been assigned Accession Nos. U52594–U52601.

[2] To whom correspondence and reprint requests should be addressed. Fax: 61 3 5227 5555. E-mail: sapats@aahl.dah.csiro.au.

0042-6822/96 $18.00
Copyright © 1996 by Academic Press, Inc.
All rights of reproduction in any form reserved.

TABLE 1

Comparison of Nucleotide and Deduced Amino Acid Sequences for N

Amino acid identity (%)	Nucleotide identity (%)							
	Vic S	V5/90	N1/62	N9/74	N2/75	N1/88	Q3/88	V18/91
Vic S	100	99.2	94.1	91.7	94.5	65.2	64.5	64.9
V5/90	98.8	100	94.0	91.8	94.5	66.0	65.5	65.6
N1/62	95.0	94.5	100	91.0	93.2	66.2	65.0	66.0
N9/74	92.5	92.3	92.5	100	90.7	65.1	64.2	66.1
N2/75	95.5	95.3	94.3	92.8	100	65.4	65.5	65.8
N1/88	61.5	61.6	61.9	63.1	61.6	100	89.4	84.6
Q3/88	61.2	61.2	61.7	63.2	61.1	89.2	100	86.4
V18/91	61.4	60.9	62.3	63.3	60.0	85.8	88.3	100

other Australian IBV strains. Based on S1 sequences, these strains formed a distant and novel genetic group of IBV (*10*). In addition, the N protein of these strains failed to react with five monoclonal antibodies directed against different epitopes on the N protein. These epitopes are conserved in other Australian strains of IBV, indicating unusual changes in the N genes of N1/88, Q3/88, and V18/91 (*11*). We have now sequenced the N gene and 3′ UTR of eight Australian IBV strains isolated over the period of thirty years from both vaccinated and unvaccinated flocks. The results demonstrated less conservation in the N protein and 3′ UTR of IBV than previously detected.

The IBV strains and methods used for their propagation have been described (*10, 11*). Vic S is a commercial vaccine (Arthur Webster Pty Ltd, Castle Hill, Australia). Strain N1/62 was isolated from unvaccinated chicks in 1962, whereas N9/74, N2/75, N1/88, Q3/88, V5/90, and V18/91 were isolated from vaccinated commercial chicks between 1974 and 1991. Strains were cloned either in chicken embryo kidney cells (by plaque assay) or tracheal organ cultures (by limiting dilutions) and passaged 1–2 times in embryonated chicken eggs. All strains replicated in the trachea and strains Vic S, V5/90, N1/62, N9/74, and N2/75 also replicated in the kidneys, the latter three causing 32 to 96% mortality (*10*).

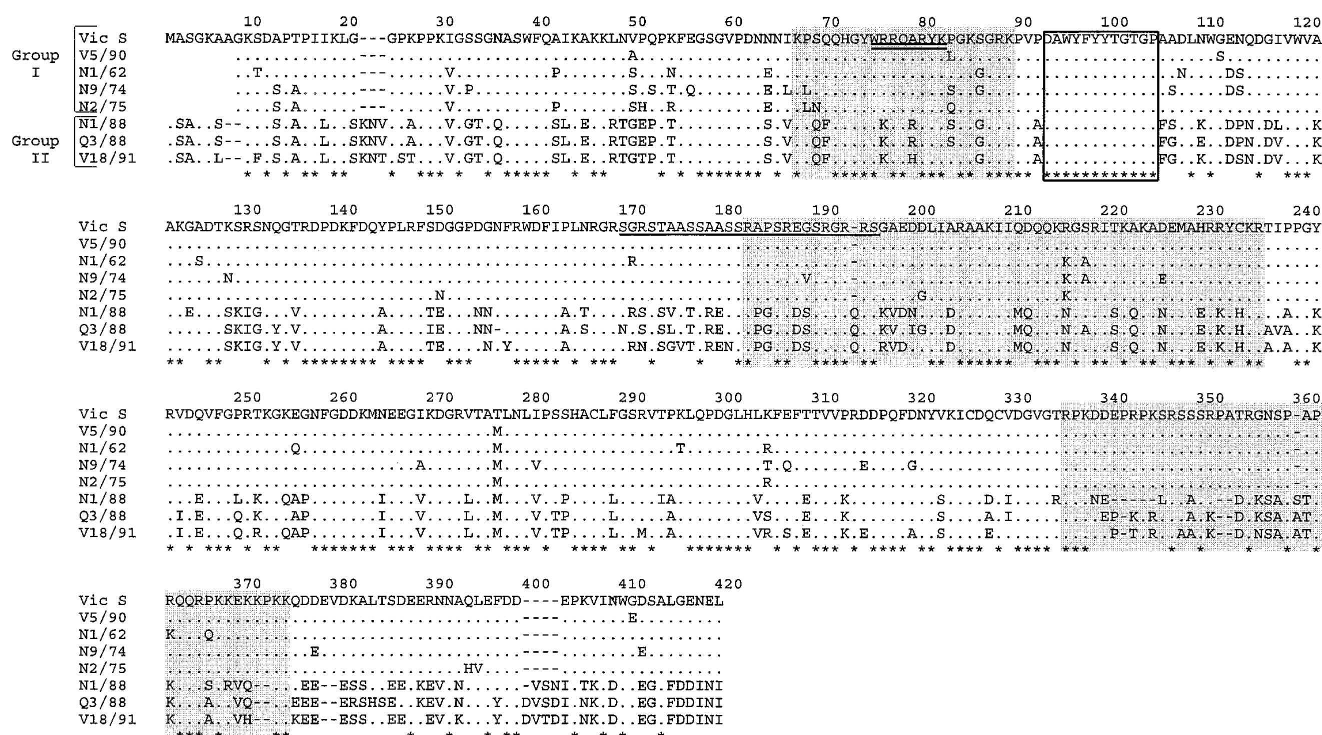


FIG. 1. Sequence alignment of the N protein of Australian strains of IBV. The complete deduced amino acid sequence of N of Vic S is shown. Gaps (dashes) were introduced to align the sequences. Dots indicate residues identical to Vic S. Asterisks indicate residues conserved in all strains. The longest region of complete conservation is boxed. The serine-rich region is underlined; the T cell epitope is double underlined. Clusters of basic residues are shaded. The Clustal V program was used for all sequence alignments (*17*).

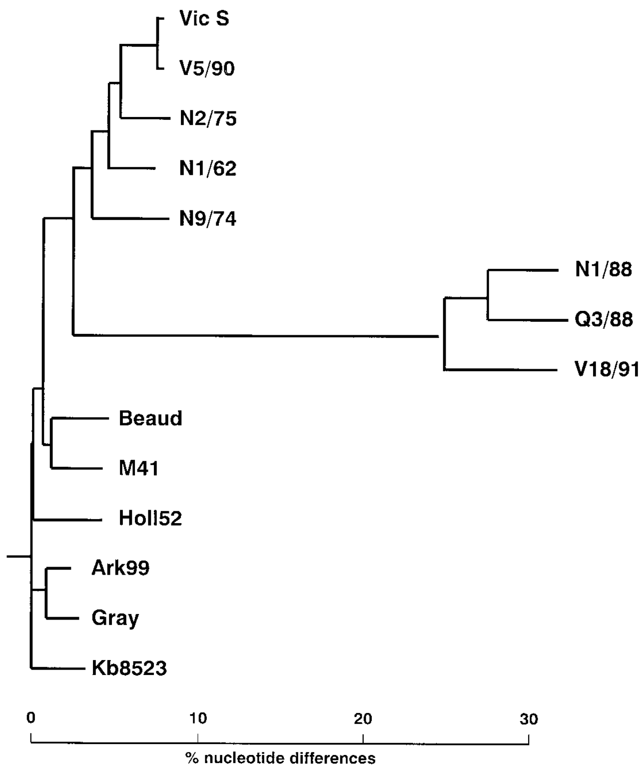


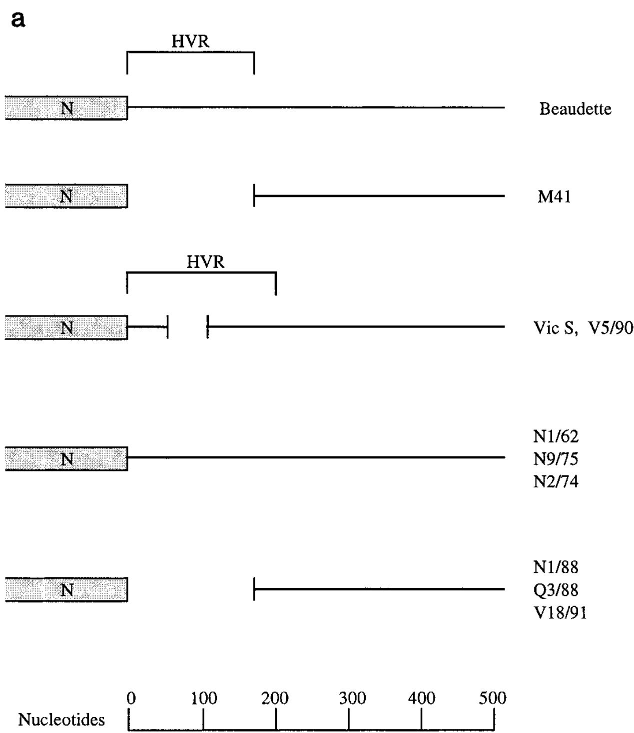
FIG. 2. Phylogenetic relationship of Australian strains to viruses with published sequences for N (*3, 4, 18*). The tree was constructed using the neighbor-joining method (*19*).

Viral RNA was purified using methods described (*10, 12*). Vic S, N1/88, and Q3/88 cDNA was synthesized using random primers, and sequences obtained from the cloned cDNA were used to design primers for amplification of the N gene of all strains by reverse transcription and polymerase chain reaction. All cDNAs were cloned into pUC series plasmids. For each virus two or more independent cDNA clones were sequenced using the Pharmacia T7 sequencing kit. Pairwise comparisons of the nucleotide and deduced amino acid sequences of the N genes of the eight strains Vic S, V5/90, N1/62, N9/74, N2/75, N1/88, Q3/88, and V18/91 (excluding the first 17 nucleotides of V5/90, N1/62, N9/74, and N2/75) are shown in Table 1. The strains formed two distinct genotypic groups based on the level of nucleotide and amino acid identities. The first five strains in Table 1, namely Vic S, V5/90, N1/62, N9/74, and N2/75 (group I) were related, showing 92.3–98.8% identity at the amino acid level. Within this group, Vic S and V5/90 were the most closely related (98.8% identity), whereas N9/74 and V5/90 were the least related (92.3% identity). The N proteins of these five strains were also similar to those of geographically distant strains isolated in the U.S.A., Europe, and Japan with which they shared 91.4–92.9% identity at the amino acid level (results not shown). This confirmed the previous observation of the tendency for conservation of the N protein over a long time (~30 years) irrespective of geographical distances and immunological pressures. Contrary to this, however, the N protein of three other strains N1/88, Q3/88, and V18/91 (group II) shared only 60.0–63.3% amino acid identity with the N proteins of strains in the first group, while sharing 85.8–89.2% amino acid identity with each other. This lack of conservation of the N protein has not been reported before for any other coronavirus.

An alignment of the deduced amino acid sequence of the N gene of Vic S with sequences of the other Australian strains is shown in Fig. 1. The N proteins of Vic S and V18/91 consisted of 409 amino acids while those of N1/88 and Q3/88 consisted of 404 and 408 amino acids, respectively. Among the group I strains most variation occurred between positions 49 and 85, 105 and 127, and 275 and 318, the remaining residues being almost entirely conserved. The group II strains varied from group I strains throughout the N protein. Most variation was located in the carboxy terminus downstream from position 337. The group II strains also contained a number of insertions and deletions relative to group I strains, the majority of which were at the amino and carboxy termini between positions 7 and 23 and 339 and 401, respectively. Overall, only 54% (221/411) of the residues were conserved in all strains, the longest region of complete conservation occurring between positions 92 and 103, corresponding to the part of the N protein previously found to have the highest degree of conservation in IBV and all other coronaviruses (*13*). Published sequences for the N proteins of IBV also contain a region of complete conservation between positions 242 and 296 (*4*). As evident from Fig. 1, the corresponding region is not conserved among the Australian strains. The precise location and role of functional domains within the N protein are not well understood (*2, 4, 14*). However, a T cell epitope has been identified in IBV corresponding to positions 74–81 in Fig. 1 (*15*). Examination of these sequences reveals that they are completely conserved in group I strains but not in group II strains.

Regardless of the amino acid variability just described, the N proteins of both groups of Australian strains shared features noted previously for the N proteins of other coronaviruses (*2, 3, 14*). (i) All strains possessed a high pro-

FIG. 3. (a) Diagrammatic representation of 3′ UTR of Australian strains compared to Beaudette and M41. The position of the HVR in Beaudette and Australian strains is marked. (b) Sequence alignment of the 3′ UTR of Australian strains of IBV compared to Beaudette and M41. The 3′ UTR of Vic S is shown. Nucleotides are numbered from the stop codon for the N gene. Gaps (dashes) were introduced to align the sequences. Dots indicate nucleotides identical to Vic S. Asterisks indicate nucleotides that are conserved in all strains. The HVR of Australian strains (boxed) starts immediately after the stop codon and continues to position 219. The additional ORFs of N1/62 and N9/74 (positions 12–238) are shaded.



b

```
          END N
          ***        10        20        30        40        50        60        70        80        90       100       110       120
Vic S     UGA---ACAUAAUGGACUUGCCGCAUUUGCUGGCACAUUUUGUUAAACACUA---------------------------------------------------------------AGUUAUUU
V5/90     ...---...............................................------------------------------------------------------------..........
N1/62     .A.GUA................U......UA.A.U.............UU...UUUUGUCACUUUCCUAUCAAUUAUUACAGGCAUUG-AUUGUGAUUAUGUGCAA-UAUUUA....C.....
N9/74     ...GUAU...............U......U..A.U............U.....UUUCGUUACUUUCCUAUCAAUUAUUACAGGCAUUG-AUUGUGAUUAUGUCCAA-UACUUA....C.UC..
N2/75     ...GUA................U......C..A.U............U..U..UUUUGUUACUUUCCUAUCAAUUAUUACAGGCAUUGGAUUGUGAUUAUGUUCAACUACUUA....C.UC..
N1/88     .A.------------------------------------------------------------------------------------------------------------------------
Q3/88     .A.------------------------------------------------------------------------------------------------------------------------
V18/91    .A.------------------------------------------------------------------------------------------------------------------------
Beaud     ...GUA............C..UU.--...C....U..................UUUCUGUGCUUUCCUAUCAAUUAUUACAGGCAUUG-AUUGUGAUUAUGUUCAA-UACUUA....C.UC..
M41       ......---------------------------------------------------------------------------------------------------------------------
          * *
                 130       140       150       160       170       180       190       200       210       220       230       240
Vic S     UUGGUUGCUCUUUGCUUGUUGUGUUGUAGCUGUGCUCUUUAUUAUUGUUGUUCUCCUUUCUUUGCUUUAUAGAAAAAGUUCAAUAGUAAGAGUUAAGGCAGAUAGGCAUGUAGCUUGAUU
V5/90     ........................................................................................................................
N1/62     .........U.................U.....................A..............................U................AA.....................
N9/74     ........CU............A....U...............G.....A....U...........................................A.....................
N2/75     .........U.................U.........................C............................................A.................A...
N1/88     ----------------------------------------------------------AU.AGUAG...UUC....CCCCUU.AG.G---UU.GUG.C........UG..C....GU---
Q3/88     ----------------------------------------------------------AU.AGUAG...UUC....CCCCUU.AG.----UU.GUG.C........UG..C....GU.A.
V18/91    -----------------------------------------------------------------------------------------..A.C....U................-U.GC
Beaud     .........U.......A....A....U........U......G....GA.....A..AG..........C.U.G..A....................A.....................
M41       ---------------------------------------------------------------------------..A....................A.....................
                                                                                                     *   *  *******  ** ****
                 250       260       270       280       290       300       310       320       330       340       350       360
Vic S     ACCUACAUGUCUAUCGCCAGGGAAAUGUCUAAUCUGUCUACUUAGUAGCCUGGAAACGAACAGUAGACCCUUAGAU-UUUAAUUUAGUUUAAUUUUUAGUUUAGUUUAAGUUAGUUUAGA
V5/90     ............................................................................-...........................................
N1/62     .....U.......................................................G..............-...........................................
N9/74     .............................A..............................................-...........................................
N2/75     ............................................................................-...........................................
N1/88     U...GU...C............................U...G..............U..AG..............-.C............--A.........U................
Q3/88     U...GU...C................................G..............U..AG..............-.C............--A.........U................
V18/91    .............................................................G...........U..C..G.U..........-.....U.....................
Beaud     ..........G..................................................C..............-...........................................
M41       ..........G......................U...........................C..............-...........................................
           ***  ***  ****************** *** *** **** ************* ***  *********** ** *  * ********   ***** ****  ***************
                 370       380       390       400       410       420       430       440       450       460       470       480
Vic S     GUAGGUUUAAAGAUGCCAGUGCCGGGGCCACGCGGAGUACGAUCGAGGGUACAGCACUUGGACGCCCAUUAUGGGAAGAGCUAAAUUUUAGUUUAAGUUAAGUUUAAUUGGCUAAGUAUA
V5/90     ........................................................................................................................
N1/62     ......A...................................................A.........C..G................................C...............
N9/74     ....U.A...................................................A............G................................................
N2/75     ......A.................A.................................A.........C..G.........................................-......
N1/88     ...AU.AG.....A............................................A.........C..G.............-.G.........U..U.AG...G.AAA.U..C...
Q3/88     ....U.AG.....A..........A......U..........................A.........C..G.............-.G.........U..U.AG...G.AAA.U......
V18/91    ....U.AG.....C............................................A............G.............-..............U.AGG..G..A.........
Beaud     ......A...................................................A............G................................................
M41       ......A...................................C...............A............G..........................................U....
          ***  *  ***** ********** ****** ********** *************** ********* ** ************* * ********* ** *  **  *   *  * ***
                 490       500       510
Vic S     GUUAAAAUUUUUAGGCUAGUAUAGAGUUAGAGC
V5/90     .................................
N1/62     ..........A......................
N9/74     ..........A......................
N2/75     ..........A......................
N1/88     ....GUU.CA.......................
Q3/88     ....GUU.CA.......................
V18/91    ....GG...AA......................
Beaud     ..........A......................
M41       ..........A......................
          ****   *   **********************
```

portion of basic residues (17.2–19.6%). The majority of these (80%) were conserved in position and generally clustered in three regions between positions 66 and 88, 181 and 234, and 334 and 373 (Fig. 1). This basic character probably facilitates protein–RNA interactions. (ii) The carboxy terminus contained a clustering of acidic residues located between positions 375 and 415, a region which otherwise showed very little conservation. However, the acidic residues were less conserved (66%) in position than the basic residues. (iii) All strains contained approximately 30 serine residues, their precise locations varying considerably. Only 10 were totally conserved in position. The majority of these conserved serine residues occurred between positions 168 and 194. Serine residues are potential sites for phosphorylation (2). Hence it appears that the N protein is able to tolerate some variability in the distribution of basic, acidic, and serine residues, although the basic residues were the most conserved consistent with the function of the protein as a ribonucleocapsid protein. These residues are expected to directly influence protein charge, state of phosphorylation, and secondary structure.

The phylogenetic relationship between Australian and other strains of IBV based upon the published nucleotide sequences for the N gene is shown in Fig. 2. The group I and group II Australian strains formed two clusters distinct from previously published strains. The five group I strains formed one cluster with Vic S and V5/90 being the most closely related and showing approximately the same degree of relatedness to N1/62 and N2/75, while N9/74 appeared to be the most divergent. The group II strains formed another cluster having diverged markedly from all other clusters. The phylogenetic relationships among the Australian IBV strains based on the N nucleotide sequences were similar to those based on S1 (*10*). This suggests that the S1 and N genes of strains in group II have evolved in parallel. Interestingly, a comparison of the Dutch strain D1466 with other European strains showed that the D1466 S1 protein had significantly diverged from the S1 of the other strains, similar to the distance between the Australian group I and II strains. However, this divergence was not matched by the N protein, which was conserved between D1466 and the other European strains (*5*).

The 3′ UTR, located immediately downstream of the N gene, was sequenced for Vic S and the other Australian strains (Figs. 3a and 3b). A HVR was identified, ranging from 10 to 216 nucleotides in length and ending 24 nucleotides further toward the 3′ end of the genome (positions 196–219 in Fig. 3b) than previously reported (*9*). In the group I strains N1/62, N9/74, and N2/75 the HVR was 214, 214, and 216 nucleotides, respectively, while for Vic S and V5/90 it was 153 nucleotides. In the group II strains N1/88, Q3/88, and V18/91 the HVR was 38, 37, and 10 nucleotides in length, respectively. Pairwise comparisons of nucleotides within the HVR (ignoring deletions) revealed values of identity ranging from 22.2 to 100%. Within group I the identity was 92.1–100%; within group II it was 22.2–100% identity (results not shown). The HVR of group I viruses (i) contained a high U content which was evenly distributed, and (ii) was similar to that of Beaudette, Ark99, and Gray viruses (*9*). Little similarity to Kb8523 and Holl52 viruses was observed (results not shown).

This HVR downstream of the N gene was originally identified in the vaccine strain Beaudette by comparison with the virulent M41 strain and considered to be an insert of exogenous origin acquired through recombination during adaptation in eggs (*3, 9*). However, the presence of the long HVR (153 to 216) in group I strains, which received a low number of passages in eggs, suggests that the HVR region may have been present in an ancestral strain and subsequently deleted in some strains. In addition, in contrast with the Beaudette and M41 comparison, the shorter forms (10 to 38) of the HVR in the Australian strains N1/88, Q3/88, and V18/91 (group II) are associated with a decrease in virulence. Thus there is no clear association between the length of the HVR and virulence. The remaining 294 nucleotides downstream of the HVR (from position 220 in Fig. 3b) showed considerable conservation with values of identity ranging from 84.6–100.0% in pairwise comparisons. Within group I strains, the identity was greater than 97.3%.

An ORF was detected within the 3′ UTR of N1/62 and N9/74 with the potential to encode hydrophobic proteins of 8793 M_r and 8711 M_r, respectively. A similar ORF has been identified in four other IBV strains (*9*). Although the initiation codons of these ORFs fit the Kozak consensus sequence (*16*) with an A at position −3 and a G at position +4, there are no IBV mRNA transcription motifs (CTT/GAACAA) directly upstream. Other coronaviruses such as porcine transmissible gastroenteritis virus also contain an ORF within the 3′ UTR. The latter virus directs the synthesis of a hydrophobic and membrane-associated protein of 9101 M_r (*7*). At present it is unknown whether the corresponding protein for IBV is synthesized in infected cells and what its function in virus replication may be. More experiments are required to detect the protein and to examine possible mechanisms for the initiation of its translation.

ACKNOWLEDGMENT

S. Sapats was the recipient of a Junior Research Fellowship from the Chicken Meat and Egg Industry Research and Development Councils.

REFERENCES

1. Boursnell, M. E. G., Brown, T. D. K., Foulds, I. J., Green, P. F., Tomley, F. M., and Binns, M. M., *J. Gen. Virol.* **68,** 57–77 (1987).

2. Laude, H., and Masters, P. S., *In* "The Coronaviridae" (S. G. Sidel, Ed.), pp. 141–163. Plenum Press, New York, 1995.

3. Boursnell, M. E. G., Binns, M. M., Foulds, I. J., and Brown, T. D. K., *J. Gen. Virol.* **66,** 573–580 (1985).

4. Williams, A. K., Wang, L., Sneed, L. W., and Collisson, E. W., *Virus Res.* **25,** 213–222 (1992).
5. Zwaagstra, K. A., Van Der Zeijst, B. A. M., and Kusters, J. G., *J. Clin. Microb.* **30,** 79–84 (1992).
6. Kusters, J. G., Niesters, H. G. M., Lenstra, J. A., Horzinek, M. C., and Van Der Zeijst, B. A. M., *Virology* **169,** 217–221 (1989).
7. Tung, F. Y. T., Abracham, S., Sethna, M., Hung, S-L., Sethna, P., Hogue, B. G., and Braian, D. A., *Virology* **186,** 676–683 (1993).
8. Vennema, H., Rossen, J. W. A., Wesseling, J., Horzinek, M. C., and Rottier P. J. M., *Virology* **191,** 134–140 (1992).
9. Williams, A. K., Wang, L., Sneed, L. W., and Collisson, E. W., *Virus Res.* **28,** 19–27 (1993).
10. Sapats, S. I., Ashton, F., Wright, P. J., and Ignjatovic, J., *J. Gen. Virol.,* **77,** 413–418 (1996).
11. Ignjatovic, J., and McWaters, P. G., *J. Gen. Virol.* **72,** 2915–2922 (1991).
12. Niesters, H. G. M., Lenstra, J. A., Spaan, W. J. M., Zijderveld, A. J., Bleumink-Pluym, N. M. C., Hong, F., Van Scharrenburg, G. J. M., Horzinek, M. C., and Van Der Zeijst, B. A. M., *Virus Res.* **5,** 253–263 (1986).
13. Masters, P. S., Koetzner, C. A., Kerr, C. A., and Heo, Y., *J. Virol.* **68,** 328–337 (1994).
14. Parker, M. M., and Masters, P. S., *Virology* **179,** 463–468 (1990).
15. Boots, A. M. H., Kusters, J. G., Van Noort, J. M., Zwaagstra, K. A., Rijke, E., Van Der Zeijst, B. A. M., and Hensen, E. J., *Immunology* **74,** 8–13 (1991).
16. Kozak, M., *J. Mol. Biol.* **196,** 947–950 (1987).
17. Higgins, D. J., and Sharp, P. M., *Gene* **73,** 237–244 (1988).
18. Sutou, S., Sato, S., Okabe, t., Nakai, M., and Sasaki, N., *Virology* **165,** 589–595 (1988).
19. Saitou, N., and Nei, M., *Mol. Biol. Evol.* **4,** 406–425 (1987).