Molecular identification and characterization of novel coronaviruses infecting graylag geese (*Anser anser*), feral pigeons (*Columbia livia*) and mallards (*Anas platyrhynchos*)

Christine Monceyron Jonassen,[1] Tone Kofstad,[1] Inger-Lise Larsen,[1] Atle Løvland,[2] Kjell Handeland,[3] Arne Follestad[4] and Atle Lillehaug[3]

[1–3]Section for Virology and Serology[1], Section for Pathology[2] and Section for Wildlife Diseases[3], National Veterinary Institute, PO Box 8156 Dep., N-0033 Oslo, Norway

[4]Norwegian Institute for Nature Research, Tungasletta 2, N-7485 Trondheim, Norway

Correspondence
Christine Monceyron Jonassen
christine.monceyron-jonassen@vetinst.no

In light of the finding of a previously unknown coronavirus as the aetiology of the severe acute respiratory syndrome (SARS), it is probable that other coronaviruses, than those recognized to date, are circulating in animal populations. Here, the results of a screening for coronavirus are presented, using a universal coronavirus RT-PCR, of the bird species graylag goose (*Anser anser*), feral pigeon (*Columbia livia*) and mallard (*Anas platyrhynchos*). Coronaviruses were found in cloacal swab samples from all the three bird species. In the graylag goose, 40 of 163 sampled birds were coronavirus positive, whereas two of 100 sampled pigeons and one of five sampled mallards tested positive. The infected graylag geese showed lower body weights compared with virus-negative birds, suggesting clinical significance of the infection. Phylogenetic analyses performed on the replicase gene and nucleocapsid protein sequences, indicated that the novel coronaviruses described in the present study all branch off from group III coronaviruses. All the novel avian coronaviruses harboured the conserved s2m RNA structure in their 3′ untranslated region, like other previously described group III coronaviruses, and like the SARS coronavirus. Sequencing of the complete nucleocapsid gene and downstream regions of goose and pigeon coronaviruses, evidenced the presence of two additional open reading frames for the goose coronavirus with no sequence similarity to known proteins, but with predicted transmembrane domains for one of the encoded proteins, and one additional open reading frame for the pigeon coronavirus, with a predicted transmembrane domain, downstream of the nucleocapsid gene.

Received 9 February 2005
Accepted 16 March 2005

INTRODUCTION

Coronaviruses are enveloped viruses with a positive-strand RNA genome (27–32 kb). They have been associated with disease in several warm-blooded animals, including humans, and in 2003, a previously unknown coronavirus was evidenced as the aetiological agent of the severe acute respiratory syndrome (SARS) that killed nearly 800 people from November 2002 to July 2003 (Drosten *et al.*, 2003; Ksiazek *et al.*, 2003; Kuiken *et al.*, 2003).

The coronaviruses comprise three serologically and genetically distinct groups (Enjuanes *et al.*, 2000; Gonzalez *et al.*, 2003). Coronaviruses from group I and II have been found to infect several mammalian species, including humans, pigs, cows, dogs, horses, cats and rodents. Group III coronaviruses have been found to infect poultry, and include the chicken infectious bronchitis virus (IBV), the turkey coronavirus (TCoV) and the pheasant coronavirus. IBV causes an acute respiratory disease in chickens, while TCoV causes an acute enteric disease in turkeys. The poultry coronaviruses are antigenically similar and phylogenetically related (Breslin *et al.*, 1999; Cavanagh *et al.*, 2001, 2002). IBV has also been isolated from a peafowl and a teal, in a recent coronavirus screening of domestic birds (Liu *et al.*, 2005), as well as from a flock of racing pigeons (Barr *et al.*, 1988). Wild-bird species have been thought to be susceptible to coronavirus infection, as a coronavirus-like virus was found in ticks feeding on seabirds in Norway. However, this virus was not antigenically characterized further (Traavik *et al.*, 1977).

Published online ahead of print on 22 March 2005 as DOI 10.1099/vir.0.80927-0

The GenBank/EMBL/DDBJ accession numbers of the sequences reported in this paper are AJ854114–AJ854159 and AJ871017–AJ871024.

The SARS coronavirus has been postulated to be an early split-off from the group II coronaviruses (Snijder *et al.*, 2003), while some authors have argued that there are evidences for recombination events in the evolution of this virus, involving both mammalian and avian coronaviruses (Rest & Mindell, 2003; Stanhope *et al.*, 2004; Stavrinides & Guttman, 2004). Also, antigenic reactivity of the SARS coronavirus nucleocapsid with polyclonal antisera against group I coronaviruses, but not with polyclonal antisera against group II or group III coronaviruses, has recently been reported (Sun & Meng, 2004).

An interesting feature of the SARS coronavirus genome is that it harbours a short motif (s2m) at its 3′ end, also found in group III coronaviruses, but not in other coronaviruses (Ksiazek *et al.*, 2003; Marra *et al.*, 2003; Wang *et al.*, 2003). s2m is also present in all but one of the characterized astroviruses, and in one picornavirus (equine rhinitis virus B), and seems to be an element that is readily transferred between RNA virus genomes (Jonassen *et al.*, 1998). The origin of the SARS coronavirus has still not been elucidated, but it has been proposed that the virus had an animal origin (Bell *et al.*, 2004; Guan *et al.*, 2003; Klempner & Shapiro, 2004; Poon *et al.*, 2004). The emergence of the SARS coronavirus has made it important to characterize further the coronavirus family to understand better which viruses are circulating in animal populations as potential zoonotic agents.

In 2003, a Norwegian survey, originally aimed at investigating the presence of enteropathogenic bacteria as well as influenza A virus and Newcastle disease virus in wild birds, was carried out. The sampling included fresh droppings, cloacal and tracheal swabs as well as a limited number of organs, from three bird species; the graylag geese (*Anser anser*), feral pigeons (*Columbia livia*) and mallards (*Anas platyrhynchos*). The same samples, as well as cloacal swabs sampled from graylag geese in August 2004, were investigated for the presence of coronaviruses, using RT-PCR. We report here the detection and characterization of novel coronaviruses infecting geese, pigeons and mallards.

METHODS

Droppings, swabs and organ sampling

Graylag geese. A total of 119 fresh faecal samples from graylag geese were gathered in early April 2003 at eight different locations adjacent to lakes in the Klepp municipality (Rogaland county, south-western Norway). This is an important resting area for the graylag goose during its migration from wintering quarters in southern Europe to the nesting areas in Norway. The samples were kept at −70 °C until analysis, and pooled six by six, into 20 pools, prior to analysis.

Cloacal swab samples were collected from a total of 163 graylag geese shot during the hunting season on the islands of Smøla, Møre and Romsdal county, mid-Norway, (100 birds in early August 2003 and 19 in early August 2004), and Vega, Nordland county, North-Norway (44 birds in early August 2004). Samples from the liver and kidneys were collected from a limited number of geese, and kept frozen until analysis. The sampling was usually carried out within 1 day after killing. The swabs were placed in tubes containing virus transport medium, and all samples were kept on ice until they reached the laboratory (2003) or they were kept frozen until analysis (2004). For all geese sampled, the age (juvenile or adult) was recorded, and for most of them, sex and body weight were recorded, together with some other physical parameters.

Feral pigeons. One hundred pigeons (55 juveniles and 45 adults) were caught by net or cage traps at 11 different locations in Oslo, between late June and early September 2003. The birds were brought to the laboratory alive, euthanized and autopsied. Cloacal and tracheal swabs were taken from each bird and processed fresh, or kept at −70 °C until analysis. Samples from several organs, including liver, lungs and spleen were taken from some of the birds and kept at −70 °C.

Mallards. Cloacal swabs were taken from five mallards (two juveniles and three adults) caught by hand, in Oslo, on three occasions from early June to mid-October 2003. The swabs were placed in tubes containing virus transport medium, and the samples were kept on ice until they reached the laboratory, and processed fresh or kept at −70 °C until analysis.

RNA isolation and RT-PCR for coronavirus screening. RNA from swabs was isolated with QIAamp Viral RNA mini kit (Qiagen), and RNA from tissues was isolated with RNeasy mini kit (Qiagen) according to the manufacturer's instructions.

RT-PCR was performed using consensus coronavirus primers, 4Bm [5′-TCACA(C/T)TT(A/T)GGATA(A/G)TCCCA-3′] and 2Bp [5′-ACTCA(A/G)(A/T)T(A/G)AAT(C/T)TNAAATA(C/T)GC-3′], designed to amplify 250 bp of the replicase gene of all groups of coronaviruses (Stephensen *et al.*, 1999). The samples were analysed with a two-step RT-PCR. cDNA synthesis was performed using SuperScript III RNase H^- Reverse Transcriptase (Invitrogen) according to the manufacturer's protocol, with 5 pmole of the 4Bm primer. The RT reaction was performed in a thermocycler (MJ Research) at 55 °C for 45 min, followed by an inactivation step at 70 °C for 15 min.

cDNA (5 μl) was added to a 50 μl PCR reaction and amplified by using HotStar Taq DNA polymerase (Qiagen). The primers used were 2Bp (3·75 μM) and 4Bm (0·75 μM). The concentration of Mg^{2+} in the reaction was 1·5 mM.

The amplification programme consisted of an initial 15 min step at 95 °C, followed by 44 cycles with the following conditions: 2 cycles with 94 °C for 60 s, 58 °C for 25 s and 72 °C for 50 s, 2 cycles with 94 °C for 60 s, 56 °C for 25 s and 72 °C for 50 s, and 40 cycles with 94 °C for 50 s, 55 °C for 25 s and 72 °C for 50 s. A final elongation step at 72 °C for 5 min was performed, followed by chilling to 4 °C.

RNA isolated from a bovine faeces sample positive for coronavirus was used as a positive control in all RT-PCR set-ups, while negative controls consisted of RNase/DNase-free water.

As sequence information became available for the degenerated positions in the 2Bp and 4Bm primers, new primers were designed and used further in the screening, using the same conditions for the RT and PCR. The new primer sequences were 4Bm-mod: 5′-TCACATTT(A/T)GGATAGTCCCA-3′ and 2Bp-mod: 5′-ACTCAG-(A/T)TGAATTTGAAATA(C/T)GC-3′. On some occasions, unspecific PCR products were observed, but only PCR products that could be verified by sequencing were considered positive. To increase further the specificity, another primer, 4Bm-7gaas [5′-T(A/T)GGATAGT-CCCAGCCCATA-3′], partly overlapping 4Bm-mod, was used instead

of 4Bm-mod in the cycle sequencing reaction, as well as in the PCR step of the RT-PCR performed on different tissues in the virus propagation experiments.

Virus propagation. Faecal dropping and swab samples were routinely passaged twice in 9-day-old embryonated chicken eggs, inoculated in the allantoic cavity. Mortality was recorded, and a haemagglutination test was performed on all the samples after both passages. Non-inoculated embryonated eggs were used as controls.

After detection of coronavirus RNA in a large part of the geese samples, new inoculations were performed in a total of 26 embryonated chicken eggs, in several experiments. Inoculations were performed in the allantoic cavity of 16 9–10-day-old embryonated eggs, and in the amniotic cavity of three 15-day-old embryonated eggs (Adams & Hifstad, 1971; Jordan & Nassar, 1973). Several organs were sampled from the infected birds for coronavirus RT-PCR, including intestine, lungs, trachea, liver, kidneys, brain and spleen from the amniotic cavity inoculated eggs, and including the corioallantoic membrane (CAM) as well, from most eggs inoculated in the allantoic cavity. Organ sampling was performed 5 days post-infection for the eggs inoculated in the amniotic cavity and both 5 and 10 days post-infection for the eggs inoculated in the allantoic cavity. A second passage was performed later on, in the allantoic cavity of three 10-day-old eggs using kidney homogenates from the first passage as the source, and a new inoculation with geese cloacal swabs on four 10-day-old eggs was performed as well. RT-PCR for coronavirus detection and observation of histological changes were performed on kidneys, liver and brain of these birds, 10 days post-infection.

Similarly, organ samples positive for coronavirus from a pigeon were used to inoculate the allantoic cavity of a total of ten 10-day-old embryonated eggs. Several organs were collected from the embryos after 5 or 10 days of incubation, and tested for the presence of coronavirus by RT-PCR. Kidneys, spleen and liver were included in all cases, as well as intestine, trachea, CAM and lungs, from the chicken.

Amplification of the 3′ end and 5′ RACE for sequencing. All other coronaviruses known to infect birds possess s2m in their 3′-untranslated regions (UTR). RT-PCR was performed using a primer designed in s2m [5′-CCGAGTA(C/G)GATCGAGGG-3′] as a sense primer and oligo$(dT)_{18}$ as the reverse primer, producing a PCR product for sequencing. For all novel coronaviruses described here, electrophoresis of the PCR products generated a smear, consistent with the presence of s2m, and the products were sequenced using the s2m primer, giving sequence information between s2m and the poly(A) tail. The sequence information allowed the design of two primers, coronarace1 (5′-TTTTTTTTTTTTGCTCTAACTCTAT-3′) and coronarace2 (5′-TTGCTCTAACTCTATACTAGCCTA-3′), which were used as a primer in RT and an antisense primer in PCR, respectively, followed by sequencing, using a 5′ RACE strategy. Several specific sequencing primers were then designed and used for primer walking, as described previously (Jonassen *et al.*, 1998). When sequence information spanning the nucleocapsid gene had been obtained for one goose and one pigeon coronavirus isolate, PCR products for sequencing of other isolates were obtained using the nucleocapsid sense primers GC7 (5′-TGCAGTCACGTCTG-GAGAG-3′) for goose coronavirus, and DC5 (5′-CAGCACCAGT-CCTAAAACTT-3′) for pigeon coronavirus, towards the coronarace2 antisense primer.

Sequencing and sequence analyses. Sequencing was performed on purified PCR products, using the ABI PRISM BigDye Terminator Cycle Sequencing Ready Reaction kit v3.1 (Applied Biosystems) according to manufacturer's instructions, and analysed on an ABI PRISM 3100-*Avant* Genetic Analyser (Applied Biosystems).

Software used for sequence analysis were Sequencher version 4.1.4 (Gene Codes Corporation; http://www.genecodes.com), FASTA similarity search and CLUSTALW Multiple Sequence Alignment Program (http://www.ebi.ac.uk/), BioEdit Sequence Alignment Editor version 5.0.9 (Tom Hall, Department of Microbiology, North Carolina State University, North Carolina, USA; http://www.mbio.ncsu.edu/BioEdit/bioedit.html), PUZZLE 4.02 (ftp://ftp.ebi.ac.uk/pub/software/dos/puzzle/), PHYLIP package version 3.6 (Joe Felsenstein, Department of Genome Sciences, University of Washington, Seattle, Washington, USA; http://evolution.gs.washington.edu/phylip.html) and TREEVIEW (Win32) version 1.6.6 (Roderick D. M. Page, Division of Environmental and Evolutionary Institute of Biomedical and Life Sciences, University of Glasgow, Glasgow, UK; http://taxonomy.zoology.gla.ac.uk/rod/treeview.html). ORF analyses and protein domain predictions were performed using JustBio (http://www.justbio.com), PSORTII prediction (http://psort.nibb.ac.jp/form2.html), TMHMM server version 2 (http://www.cbs.dtu.dk/services/TMHMM/) and HMMTOP (http://www.enzim.hu/hmmtop/index.html).

Data processing and statistical analyses. The statistical analyses of the influence of the different physical parameters on body weight were performed in SAS-PC 8.3 for Windows with enterprise Guide (SAS Institute).

The values of the weight of the birds (gram) were normally distributed, and therefore the association to the various independent variables was assessed in a multiple general least square model based regression analysis. The independent variables included in the analysis were the categorical variables; population (referring to the different islands and years, i.e. Smøla 2003, Smøla 2004 and Vega 2004), sex, age (juvenile/adult) and coronavirus detection (negative/positive). Sex was not recorded for 31 geese, and only the 132 birds with registrations for all variables were included in the analysis. The elimination criterion for variables in the model was the type-III F-test, a *P*-value of 0·05 being used as the level for exclusion. All variables were below this level, and they were all included.

RESULTS

Coronavirus screening with RT-PCR

Only samples in which the presence of coronavirus could be confirmed by sequencing were considered positive. The findings are summarized in Table 1.

Of the 20 pooled goose dropping samples collected in April 2003, three were found positive for coronavirus, and of the 163 cloacal swabs collected from graylag geese in August 2003 and August 2004, 40 (25 %) were found positive for coronavirus. From four of the virus-positive geese (three adults and one juvenile), organ samples were available, and three of them (one juvenile, two adults) tested positive for coronavirus in kidney samples, while all four liver samples were negative. The characteristic of being virus positive contributed significantly to the variation in body weight of the geese ($P<0·001$), with an estimated weight difference of 177 grams (Table 2).

Cloacal swabs from two adult pigeons were found to be positive for coronavirus. One of them also tested positive from the tracheal swab sample, while all other tracheal swab samples examined were negative. Organ samples were available only from one of the two positive pigeons, the one that was positive in both tracheal and cloacal

Table 1. Results of testing for coronavirus by RT-PCR on samples from graylag geese, feral pigeons and mallards in Norway, 2003–2004

Juv., Juveniles; Ad., adults.

Bird species	Sample material	Place and year of sampling	Age	No. positive/analysed samples	Positive rate (%)
Graylag geese	Droppings 6-sample pools	Klepp 2003	–	3/20	–
	Cloacal swabs ($n=163$)	Smøla 2003	Juv.	4/31	12·9
			Ad.	12/69	17·4
		Smøla 2004	Juv.	4/6	66·6
			Ad.	2/13	15·4
		Vega 2004	Juv.	11/26	42·3
			Ad.	7/18	38·9
Feral pigeons ($n=100$)	Cloacal swabs	Oslo 2003	Juv.	0/55	0·0
			Ad.	2/45	4·4
	Tracheal swabs		Juv.	0/55	0·0
			Ad.	1/45	2·2
Mallards	Cloacal swabs	Oslo 2003	Juv.	0/2	0·0
			Ad.	1/3	33·3

swabs. RT-PCR evidenced coronavirus in the liver and spleen, but not in the lungs of this bird. Both of the virus-positive pigeons were in a normal body condition, and no pathological lesions were found at necropsy.

One of the mallard cloacal swab samples was positive for coronavirus by RT-PCR. This sample was also positive for low-pathogen influenza A virus (H3N8).

Virus propagation

All samples screened, except the one containing influenza A virus, were negative in the haemagglutination test. None of the attempts to grow goose or pigeon coronaviruses in embryonated chicken eggs was successful, as assessed by RT-PCR on sampled organs. Besides, there were no specific histological findings on organs sampled, and mortality was reported in only one of the eggs, inoculated with goose coronavirus, 10 days post-infection. No morphological changes, like dwarfing, were observed for any of the embryos.

Prior to coronavirus screening, very little material was left of the original cloacal sample from the positive mallard, due to the influenza virus investigations. Therefore, no further attempts to grow duck coronavirus in embryonated eggs were initiated. Still, coronavirus RT-PCR was negative, when performed on allantoic fluid samples collected from the dead eggs inoculated in the allantoic cavity when investigated for influenza virus.

Sequence analyses

Fig. 1(a) shows the phylogenetic analysis of the coronavirus family, including the novel bird coronaviruses, on

Table 2. The least square means estimates from the final multiple least square regression model of the individual weights of 132 graylag geese at the time of sampling in 2003 and 2004, in relation to the coronavirus findings

The final model explains 56·7 % of the observed variation in the bird weights.

Independent variable	Category (n)	Least square means	Type 3 P-value for variable (F-value)
Population	Smøla 2003 (69)	2728	$<0·01$
	Smøla 2004 (19)	2706	
	Vega 2004 (44)	2890	
Age	Adult (85)	2944	$<0·001$
	Juvenile (47)	2605	
Sex	Female (52)	2606	$<0·001$
	Male (80)	2943	
Virus detection by RT-PCR	Negative (97)	2863	$<0·001$
	Positive (35)	2686	

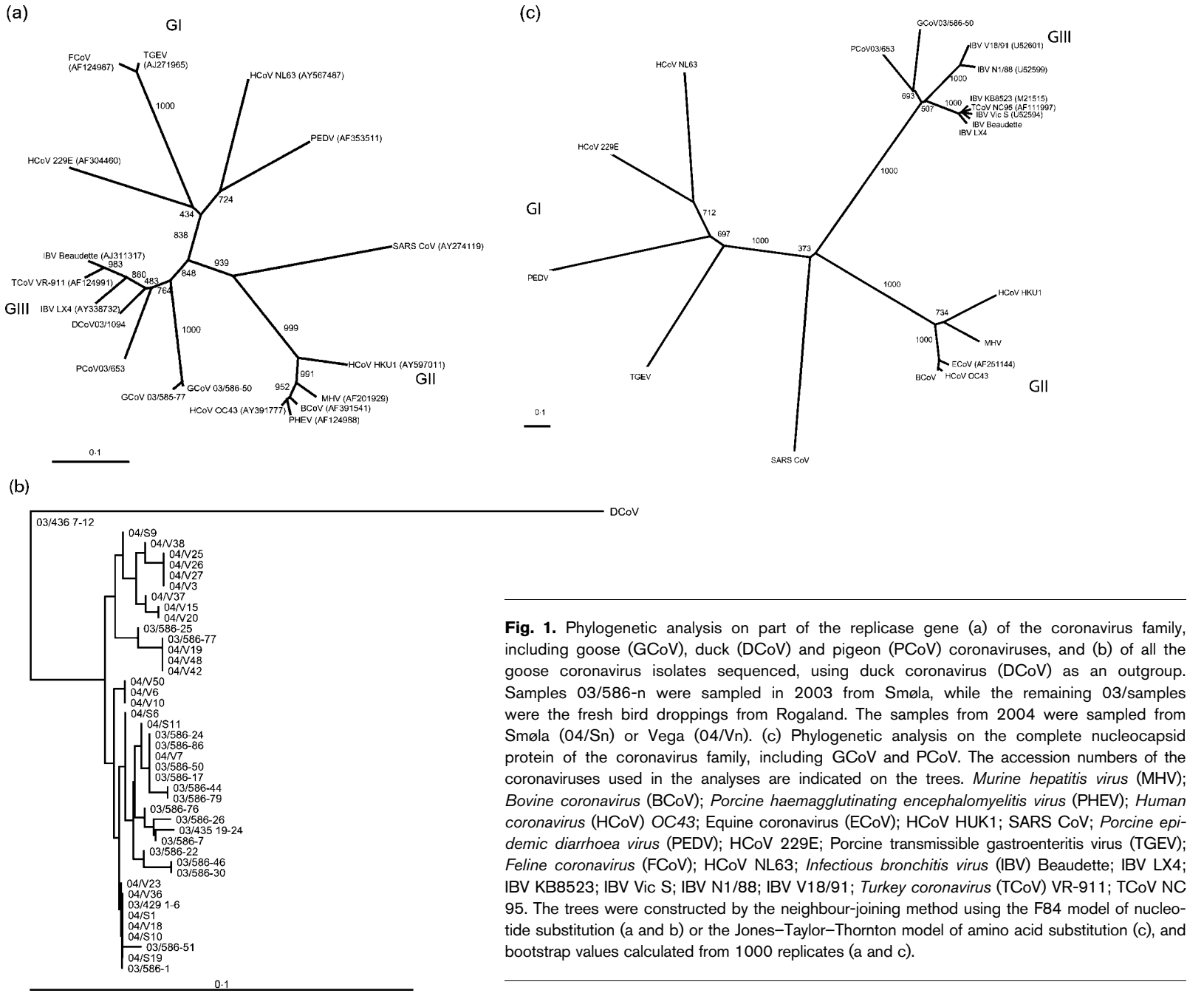
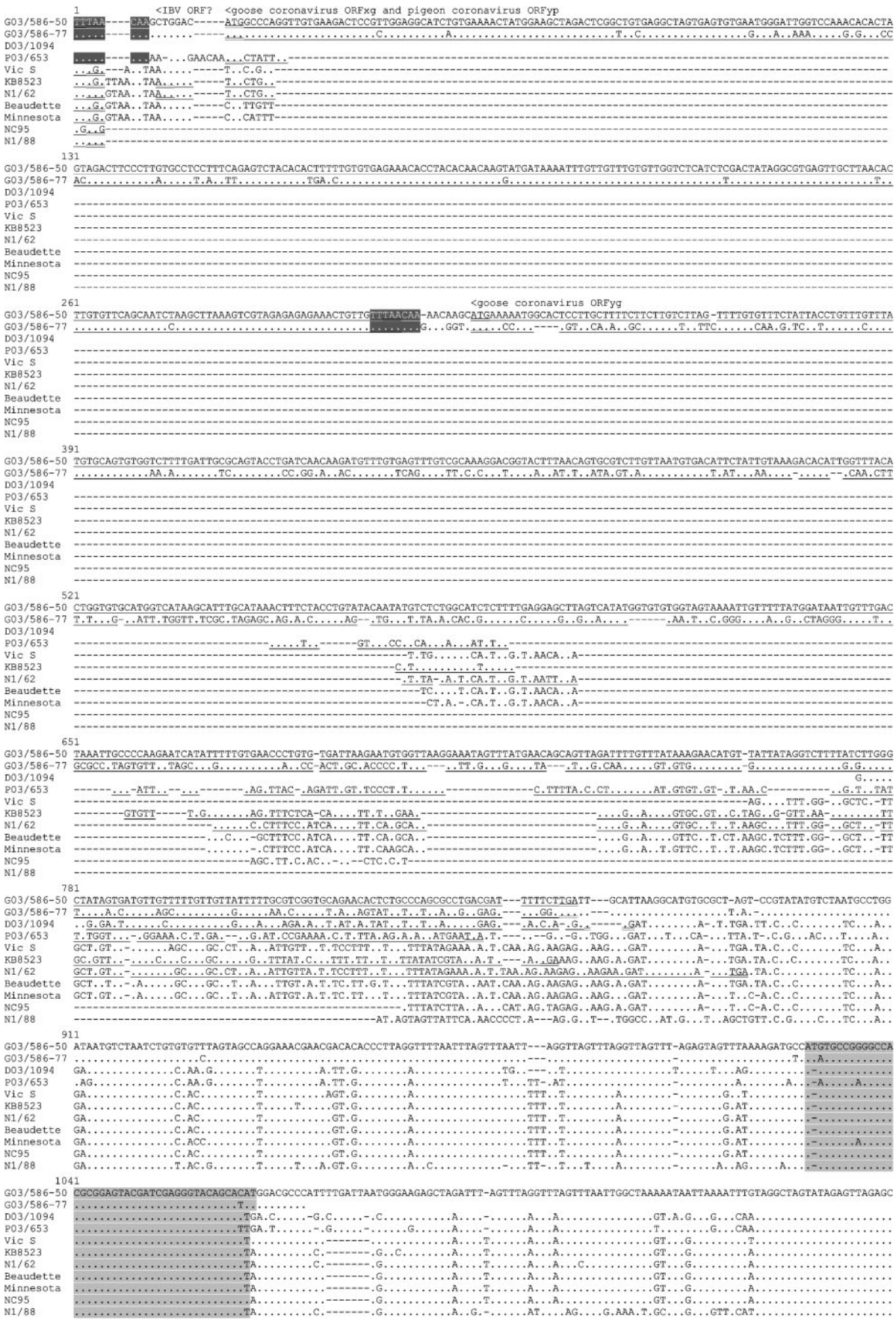


Fig. 1. Phylogenetic analysis on part of the replicase gene (a) of the coronavirus family, including goose (GCoV), duck (DCoV) and pigeon (PCoV) coronaviruses, and (b) of all the goose coronavirus isolates sequenced, using duck coronavirus (DCoV) as an outgroup. Samples 03/586-n were sampled in 2003 from Smøla, while the remaining 03/samples were the fresh bird droppings from Rogaland. The samples from 2004 were sampled from Smøla (04/Sn) or Vega (04/Vn). (c) Phylogenetic analysis on the complete nucleocapsid protein of the coronavirus family, including GCoV and PCoV. The accession numbers of the coronaviruses used in the analyses are indicated on the trees. *Murine hepatitis virus* (MHV); *Bovine coronavirus* (BCoV); *Porcine haemagglutinating encephalomyelitis virus* (PHEV); *Human coronavirus* (HCoV) *OC43*; Equine coronavirus (ECoV); HCoV HUK1; SARS CoV; *Porcine epidemic diarrhoea virus* (PEDV); HCoV 229E; Porcine transmissible gastroenteritis virus (TGEV); *Feline coronavirus* (FCoV); HCoV NL63; *Infectious bronchitis virus* (IBV) Beaudette; IBV LX4; IBV KB8523; IBV Vic S; IBV N1/88; IBV V18/91; *Turkey coronavirus* (TCoV) VR-911; TCoV NC 95. The trees were constructed by the neighbour-joining method using the F84 model of nucleotide substitution (a and b) or the Jones–Taylor–Thornton model of amino acid substitution (c), and bootstrap values calculated from 1000 replicates (a and c).


1
<IBV ORF?
<goose coronavirus ORFxg and pigeon coronavirus ORFyp
G03/586-50
G03/586-77
D03/1094
P03/653
Vic S
KB8523
N1/62
Beaudette
Minnesota
NC95
N1/88
131
261
<goose coronavirus ORFyg
391
521
651
781
911
1041

the part of the replicase gene that was sequenced in this study, and Fig. 1(b) illustrates the variation in the goose coronavirus isolates in that same gene part.

The nucleocapsid genes from both pigeon and goose coronaviruses were sequenced, as well as the rest of the 3′ end of the genomes, while only a short part of the 3′ end was sequenced from the duck coronavirus, due to the limited quantity of the starting material. Fig. 1(c) shows a phylogenetic tree of the coronavirus family based on the nucleocapsid protein, and Fig. 2 shows the nucleotide alignment of the avian coronaviruses downstream of the nucleocapsid gene.

The goose coronavirus was found to have two additional open reading frames (ORF) between the nucleocapsid gene and the 3′-UTR, named ORFxg and ORFyg, potentially encoding two proteins, 95 and 169–176 aa long, respectively, with no sequence homology to any other protein in GenBank. Several goose coronavirus isolates were sequenced in the 3′ region, all displaying the two additional ORFs, even if nucleic acid heterogeneity, including insertions and deletions, was observed, especially in ORFyg. Amino acid sequence analysis predicted two transmembrane domains in the protein encoded by ORFyg for all analysed goose coronavirus samples (Fig. 3). No characteristic protein domain was predicted for ORFxg.

Similarly, downstream of the nucleocapsid gene, the pigeon coronavirus was found to have one additional ORF, named ORFyp, potentially encoding a 63 aa long protein with a predicted transmembrane domain (Fig. 3). ORFyp showed no sequence homology to any other proteins in GenBank, nor to the additional ORFs of the goose coronavirus, at the amino acid level, but showed sequence similarity at the nucleic acid level with the 3′ part of ORFyg, and the 5-terminal part of the 3′-UTR of some IBV and TCoV, as shown in Fig. 2.

DISCUSSION

Coronaviruses could be detected by RT-PCR in all the three avian species that were included in this survey. Although virus could not be isolated in embryonated chicken eggs by methods used in this study, sequence homology analysis of the sequenced PCR products obtained both in the replicase region and in the 3′ part of the viral genome, indicates that virus is present in these animals. Furthermore, the present coronavirus screening showed that RT-PCR based on primers aimed at detection of members of the different groups of coronaviruses (Stephensen *et al.*, 1999) could be used for detection of novel coronaviruses. Coronavirus infection seemed to be widespread in the graylag goose populations sampled, as every fourth goose harboured coronavirus. The prevalence seemed to vary between geographical locations, years and age groups, and was higher in 2004 (38 %) than in 2003 (16 %). This could be explained by different phases in the infection dynamics of the different populations at the time of sampling. The observed variation in the sequences of the goose coronavirus replicase gene further suggests that the virus is actively circulating in the geese. As juvenile and adult birds were found to be infected in approximately equal proportions, the virus does not seem to be the cause of a disease particularly affecting juveniles.

The result of the regression analysis demonstrated that being virus positive had a significant influence on body weight, indicating that the infection has a clinical significance for the birds. From the tissue samples collected, virus could be detected in the kidneys of several geese, similarly to some IBV strains that have been shown to display a tropism toward kidneys (Albassam *et al.*, 1986). As no tracheal swabs were sampled from the geese, the respiratory route of transmission was not investigated. Whether coronavirus infection in geese is predominantly an enteric, renal, respiratory or systemic disease therefore remains to be clarified. However, the large proportion of positive cloacal swab samples suggests that the faecal–oral transmission route is important, with the intestinal tract of the geese as a potential primary target organ for coronavirus replication. The presence of virus both in fresh droppings collected upon the arrival of the geese to Norway in springtime, and during the autumn hunting, suggests that a great proportion of the birds carry coronavirus throughout large parts of the year.

One of the pooled goose droppings sample (03/436 7-12) displayed several ‘wobbled’ nucleotides suggesting it could be a mix of goose and duck coronaviruses, even if care was taken to pick up only freshly laid faecal samples upon gathering. The area where the droppings were gathered hosts several migrating bird species in the springtime, and some duck faeces might have been picked up together with goose faeces. Further investigations are, however, needed

Fig. 2. Nucleotide alignment of the 3′ part of avian coronavirus genomes, showing two isolates of goose coronavirus (G03/586-50 and G03/586-77), limited sequence data of the duck coronavirus (D03/1094), and one pigeon coronavirus isolate (P03/653), compared with previously published IBV (Vic S, KB8523, N1/62, Beaudette, N1/88) and TCoV strains (Minnesota and NC95). The alignment starts 2 nt upstream of the stop codon of the nucleocapsid gene, and ends immediately upstream of the poly(A) tail. Dots indicate nucleotides identical to the G03/586-50 sequence, gaps are shown as dashes, and missing data are shown blank. Predicted ORFs are underlined, with double lines under start and stop codons. Start of ORFs are also indicated by < above the sequences. For goose and pigeon coronaviruses, the putative TRS are shown in white on a grey background. s2m is shown as a shaded box.

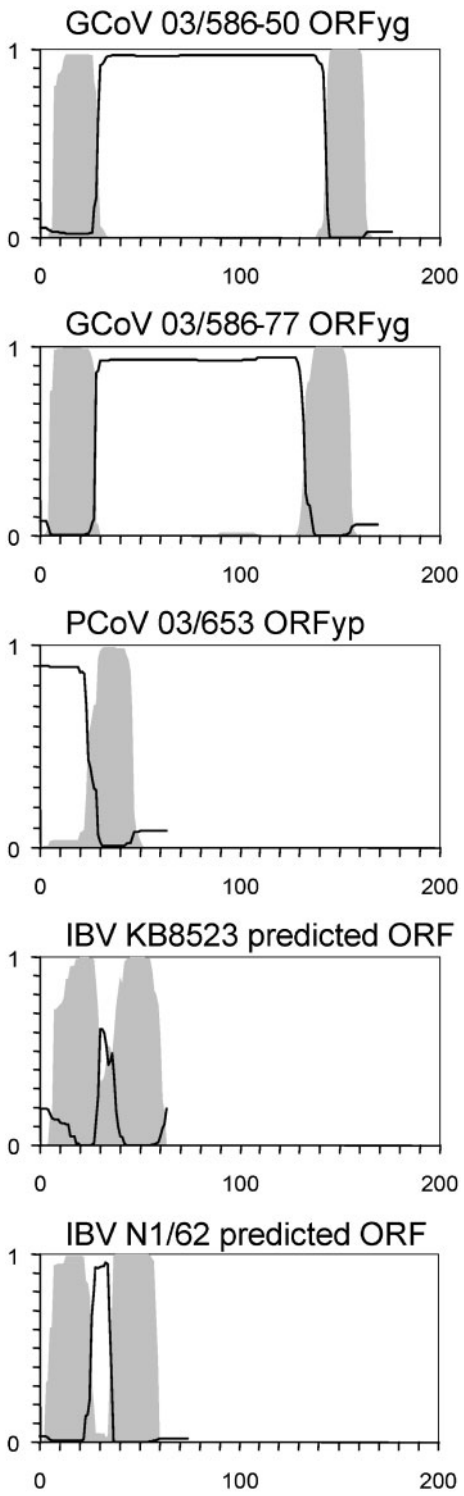


Fig. 3. Probability of transmembrane helices (shaded curves) and non-cytoplasmic localization (open curves), as predicted by TMHMM, for the proteins encoded by the ORFyg of two goose coronavirus (GCoV) isolates, the ORFyp of one pigeon coronavirus (PCoV) isolate and for the ORFs downstream of the nucleocapsid gene predicted in some IBV strains. Amino acid numbering is shown on the horizontal axis and probability on the vertical axis.

to shed light on how widespread coronavirus infection is in mallards, as only five individuals were sampled.

Coronavirus infecting pigeons was found in only two of 100 analysed cloacal swabs, indicating a lower prevalence or shorter duration of the infection, as compared with graylag goose. Virus was detected in the liver and spleen of one of the swab-positive pigeons, suggesting a viraemic stage of infection in this species. Virus was detected in the tracheal swab sample but not in the lungs of the pigeon, leaving the mode of transmission of pigeon coronavirus uncertain. Necropsy did not reveal enlarged spleen, or other significant pathological findings in the coronavirus infected pigeons.

Most IBV strains, as well as pheasant coronavirus, have been propagated in embryonated chicken eggs by inoculation of the allantoic cavity (Gough *et al.*, 1996; Jordan & Nassar, 1973), while TCoV has been shown to replicate in embryonated chicken eggs by inoculation of the amniotic cavity (Adams & Hifstad, 1971). Propagation of the goose coronavirus was attempted by both allantoic cavity and amniotic cavity inoculation of embryonated chicken eggs, to include the enteric route of transmission, while propagation of the pigeon coronavirus was attempted by allantoic cavity inoculation alone. No replication in embryonated chicken eggs could be achieved for the pigeon or goose coronavirus in the present study, suggesting that these coronaviruses do not readily infect chicken embryo. Further attempts will be made to isolate these viruses either in embryonated eggs from goose or in cell culture.

Phylogenetic analyses performed both in the replicase gene, and on the nucleocapsid protein, indicate that all the coronaviruses found in this study are novel, as opposed to the newly reported IBV infections in a domestic peafowl and in a teal from domestic bird flocks (Liu *et al.*, 2005). IBV infection has also been reported in a flock of racing pigeons in Australia, and confirmed by serological tests to be of the same serotype as the widely used Australian IBV vaccine strains, but no sequence information is available for this isolate (Barr *et al.*, 1988). All the novel bird coronaviruses branch off from group III coronaviruses, indicating a monophyletic origin with other bird coronaviruses. Several IBV strains isolated in Australia are phylogenetically distant from most other TCoV and IBV strains, in the spike and nucleocapsid proteins (Sapats *et al.*, 1996a, b), and even if all the IBV/TCoV strains clustered together in a monophyletic group as opposed to goose and pigeon coronaviruses in the nucleocapsid protein, this branching is not strongly supported by bootstrap analysis (Fig. 1c).

One of the unique features of the goose coronavirus, as compared with other bird coronaviruses is the presence of two additional ORFs at the 3′ end of its genome, with no amino acid sequence homology to other known proteins. While no specific domains were predicted for the protein encoded by ORFxg, two transmembrane domains

were predicted for ORFyg. However, the putative RNA transcript of ORFyg harbours two very closely located methionine codons in different frames. The first methionine is the start of the longer frame, ORFyg, while the second is the start of a shorter ORF encoding a short protein (45 aa). The second methionine codon is, however, in a better context of translation start in some of the goose coronavirus isolates, according to the rules of translation initiation proposed by Kozak (1995, 1997), with a G as the fourth base, immediately downstream of the ATG, and it cannot be ruled out that a protein encoding the shorter ORF is also synthesized. The protein encoded by this short ORF contains a transmembrane domain as well (data not shown). The group I coronaviruses transmissible gastroenteritis virus (TGEV) and feline coronavirus (FCoV) have also been shown to encode one and two proteins downstream of the nucleocapsid gene, respectively (Herrewegh *et al.*, 1995; Tung *et al.*, 1992). The additional ORF of TGEV and one of the additional ORFs of FCoV also encode proteins with transmembrane domains. However, in FCoV, as opposed to goose coronavirus, it is the most upstream additional ORF that encodes a protein with transmembrane domains.

The short additional ORF found downstream of the nucleocapsid gene of the pigeon coronavirus is similar in length to the predicted ORF located in the hypervariable region (HVR) of the 3′-UTR in some strains of IBV (Breslin *et al.*, 1999; Sapats *et al.*, 1996a; Williams *et al.*, 1993). Both goose and pigeon coronaviruses have a sequence (CTTAACAA) upstream of their nucleocapsid gene, similar to the transcription regulatory sequence (TRS) of IBV and TCoV [CT(T/G)AACAA]. A similar putative TRS (TTTAACAA) was found, upstream of both additional ORFs of the goose coronavirus, and of the single additional ORF of the pigeon coronavirus (Fig. 2), suggesting that mRNA specific for these ORFs are synthesized.

The pigeon coronavirus ORFyp and the predicted short ORFs of some IBV strains, all display some nucleotide sequence similarity with the 3′ part of the goose coronavirus ORFyg, suggesting a possible homologous origin of these ORFs in bird coronaviruses (Fig. 2). The short predicted ORFs of IBVs, ORFyp and ORFyg all potentially encode proteins with transmembrane domains (Fig. 3), predicted by PSORTII to have a membrane topology of the 3a type (Hartmann *et al.*, 1989; Singer, 1990), which locate toward the endoplasmic reticulum. This could suggest a similar function of these proteins to the short protein synthesized from the additional ORF at the 3′ end of the TGEV genome, that has been found to be associated with the endoplasmic reticulum and cell surface membranes of infected cells (Tung *et al.*, 1992). However, in contrast to the goose coronavirus ORFyg and pigeon coronavirus ORFyp, the additional ORF in some IBV strains, is not preceded by the TRS specific for IBV, suggesting that no mRNA specific for this ORF is synthesized. Even if internal ribosome entry has been suggested for the translation of some of the IBV proteins (Liu & Inglis, 1992), the fact that this ORF is mostly located in a part of the genome that is highly variable in IBV and TCoV, with loss of coding potential, or even largely deleted in some strains, indicates that it is not subject to strong conservative evolutionary constraints. Besides, the HVR has been shown experimentally not to be involved in virus replication and packaging, nor to be a part of the 3′-UTR that binds to nucleocapsid (Dalton *et al.*, 2001; Zhou & Collisson, 2000). This could suggest that the HVR in IBV and TCoV is a truncated, presumably non-functional, remnant of a longer gene of a common ancestor of bird coronaviruses that is still present in the goose coronavirus, encoding two potential proteins. The pigeon coronavirus also harbours a truncated gene version, but the TRS upstream of it suggests that a specific mRNA is synthesized and potentially translated into a functional protein. Further studies are required to verify the presence and the function in virus replication of the additional proteins in goose and pigeon coronaviruses. The part of the 3′ end sequence obtained from the duck coronavirus was too short to assess the presence of an entire additional ORF, with a TRS and a start codon, but the extensive sequence similarity, both in nucleotide and predicted amino acid sequence, with goose coronavirus in the 3′ terminal part of ORFyg (Fig. 2), suggests that duck coronavirus also encodes a membrane-associated protein at its 3′ end.

The sequences encoding the nucleocapsid gene in both pigeon and goose coronaviruses contain some internal ORFs, 40–70 codons long. No function was predicted for the proteins encoded by these ORFs, except for a nuclear localizing signal (NLS) in one of the goose and one of the pigeon coronavirus ORFs (Hicks & Raikhel, 1995; Robbins *et al.*, 1991). Some other coronaviruses have been shown to synthesize proteins from internal ORFs in the nucleocapsid genes (Fischer *et al.*, 1997; Singh, 1999), with no known function. Two internal ORFs were reported within the nucleocapsid gene of the SARS coronavirus, as well (Marra *et al.*, 2003), and it cannot be ruled out that some additional proteins are synthesized in the pigeon and goose coronaviruses from the internal ORFs. Besides, the nucleocapsid gene seemed to be much more conserved than the part of the replicase gene in the five goose coronavirus isolates that were sequenced in both regions (data not shown), suggesting some evolutionary constraints in the nucleocapsid gene compatible with the presence of translated overlapping reading frames.

The nucleocapsid protein of several coronaviruses, including IBV, has been reported to be localized in the nucleolus (Hiscox *et al.*, 2001; Ren *et al.*, 2004; Wurm *et al.*, 2001). The nucleolus is the site of ribosome biogenesis, and is involved in cell growth regulation. Cells expressing IBV-N protein have been shown to have retarded cell growth, delaying the cell cycle at interphase where maximum translation of viral mRNAs can occur (Chen *et al.*, 2002). The

nucleocapsid proteins of both pigeon and goose coronaviruses contained several NLSs (Hicks & Raikhel, 1995; Robbins *et al.*, 1991), suggesting that their nucleocapsid proteins locate toward the nucleus as well.

In conclusion, the present identification of novel coronaviruses indicates that coronaviruses also infect wild-bird species, and confirms the grouping of all known avian coronaviruses in group III. The findings suggest a monophyletic origin of the avian coronaviruses, as compared with the coronaviruses infecting mammals, and thus could not confirm a bird coronavirus involvement in the history of the SARS coronavirus. The relationship of the SARS coronavirus to group III coronaviruses was first postulated, since the presence of s2m at the 3′ end of the SARS coronavirus genome was one of the first evidenced sequences in the original search for the aetiological cause of SARS (Ksiazek *et al.*, 2003; Wang *et al.*, 2003). s2m was found in all of the novel bird coronaviruses, and seems to be a common feature of all group III coronaviruses. The SARS coronavirus could have acquired s2m in an event independent of its presence in the bird coronaviruses, as s2m seems to be readily transferred between viral RNA genomes. Besides, the 3′-UTR of the SARS coronavirus has been found to replace functionally its counterpart in the group II mouse hepatitis virus (Goebel *et al.*, 2004). s2m has recently been shown to have a rigorously conserved three-dimensional structure (Robertson *et al.*, 2005), and further studies are required to shed light on its function.

ACKNOWLEDGEMENTS

The study was supported by grant 158565 from the Norwegian Research Council. We would like to thank Lone Thiel Engerdahl for excellent technical assistance, Martin Dagsland for collecting the goose droppings, and Inger Marit Brunborg and Tom Øystein Jonassen for revising the manuscript. Permission to capture mallards and euthanize pigeons for sampling was given by the Directorate for Nature Management (ref: 2003/3992 ARTS-VI-ID).

REFERENCES

Adams, N. R. & Hofstad, M. S. (1971). Isolation of transmissible enteritis agent of turkeys in avian embryos. *Avian Dis* **15**, 426–433.

Albassam, M. A., Winterfield, R. W. & Thacker, H. L. (1986). Comparison of the nephropathogenicity of four strains of infectious bronchitis virus. *Avian Dis* **30**, 468–476.

Barr, D. A., Reece, R. L., O'Rourke, D., Button, C. & Faragher, J. T. (1988). Isolation of infectious bronchitis virus from a flock of racing pigeons. *Aust Vet J* **65**, 228.

Bell, D., Roberton, S. & Hunter, P. R. (2004). Animal origins of SARS coronavirus: possible links with the international trade in small carnivores. *Philos Trans R Soc Lond B Biol Sci* **359**, 1107–1114.

Breslin, J. J., Smith, L. G., Fuller, F. J. & Guy, J. S. (1999). Sequence analysis of the turkey coronavirus nucleocapsid protein gene and 3′ untranslated region identifies the virus as a close relative of infectious bronchitis virus. *Virus Res* **65**, 187–193.

Cavanagh, D., Mawditt, K., Sharma, M., Drury, S. E., Ainsworth, H. L., Britton, P. & Gough, R. E. (2001). Detection of a coronavirus from turkey poults in Europe genetically related to infectious bronchitis virus of chicken. *Avian Pathol* **30**, 355–368.

Cavanagh, D., Mawditt, K., Welchman, D. B., Britton, P. & Gough, R. E. (2002). Coronaviruses from pheasants (*Phasianus colchicus*) are genetically closely related to coronaviruses of domestic fowl (infectious bronchitis virus) and turkeys. *Avian Pathol* **31**, 81–93.

Chen, H., Wurm, T., Britton, P., Brooks, G. & Hiscox, J. A. (2002). Interaction of the coronavirus nucleoprotein with nucleolar antigens and the host cell. *J Virol* **76**, 5233–5250.

Dalton, K., Casais, R., Shaw, K., Stirrups, K., Evans, S., Britton, P., Brown, T. D. & Cavanagh, D. (2001). *cis*-acting sequences required for coronavirus infectious bronchitis virus defective-RNA replication and packaging. *J Virol* **75**, 125–133.

Drosten, C., Gunther, S., Preiser, W. & 23 other authors (2003). Identification of a novel coronavirus in patients with severe acute respiratory syndrome. *N Engl J Med* **348**, 1967–1976.

Enjuanes, L., Brian, D., Cavanagh, D. & 9 other authors (2000). *Coronaviridae*. In *Virus Taxonomy. Classification and Nomenclature of Viruses*. Edited by M. H. V. van Regenmortel, C. M. Fauquet, D. H. L. Bishop, E. B. Carsten, M. K. Estes, S. M. Lemon, D. J. McGeoch, J. Maniloff, M. A. Mayo, C. R. Pringle & R. B. Wickner. San Diego: Academic Press.

Fischer, F., Peng, D., Hingley, S. T., Weiss, S. R. & Masters, P. S. (1997). The internal open reading frame within the nucleocapsid gene of mouse hepatitis virus encodes a structural protein that is not essential for viral replication. *J Virol* **71**, 996–1003.

Goebel, S. J., Taylor, J. & Masters, P. S. (2004). The 3′ *cis*-acting genomic replication element of the severe acute respiratory syndrome coronavirus can function in the murine coronavirus genome. *J Virol* **78**, 7846–7851.

Gonzalez, 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**, 2207–2235.

Gough, R. E., Cox, W. J., Winkler, C. E., Sharp, M. W. & Spackman, D. (1996). Isolation and identification of infectious bronchitis virus from pheasants. *Vet Rec* **138**, 208–209.

Guan, Y., Zheng, B. J., He, Y. Q. & 15 other authors (2003). Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China. *Science* **302**, 276–278.

Hartmann, E., Rapoport, T. A. & Lodish, H. F. (1989). Predicting the orientation of eukaryotic membrane-spanning proteins. *Proc Natl Acad Sci U S A* **86**, 5786–5790.

Herrewegh, A. A., Vennema, H., Horzinek, M. C., Rottier, P. J. & de Groot, R. J. (1995). The molecular genetics of feline coronaviruses: comparative sequence analysis of the ORF7a/7b transcription unit of different biotypes. *Virology* **212**, 622–631.

Hicks, G. R. & Raikhel, N. V. (1995). Protein import into the nucleus: an integrated view. *Annu Rev Cell Dev Biol* **11**, 155–188.

Hiscox, J. A., Wurm, T., Wilson, L., Britton, P., Cavanagh, D. & Brooks, G. (2001). The coronavirus infectious bronchitis virus nucleoprotein localizes to the nucleolus. *J Virol* **75**, 506–512.

Jonassen, C. M., Jonassen, T. O. & Grinde, B. (1998). A common RNA motif in the 3′ end of the genomes of astroviruses, avian infectious bronchitis virus and an equine rhinovirus. *J Gen Virol* **79**, 715–718.

Jordan, F. T. & Nassar, T. J. (1973). The combined influence of age of embryo and temperature and duration of incubation on the replication and yield of avian infectious bronchitis (IB) virus in the developing chick embryo. *Avian Pathol* **2**, 279–294.

Klempner, M. S. & Shapiro, D. S. (2004). Crossing the species barrier – one small step to man, one giant leap to mankind. *N Engl J Med* **350**, 1171–1172.

Kozak, M. (1995). Adherence to the first-AUG rule when a second AUG codon follows closely upon the first. *Proc Natl Acad Sci U S A* **92**, 2662–2666.

Kozak, M. (1997). Recognition of AUG and alternative initiator codons is augmented by G in position +4 but is not generally affected by the nucleotides in positions +5 and +6. *EMBO J* **16**, 2482–2492.

Ksiazek, T. G., Erdman, D., Goldsmith, C. S. & 24 other authors (2003). A novel coronavirus associated with severe acute respiratory syndrome. *N Engl J Med* **348**, 1953–1966.

Kuiken, T., Fouchier, R. A., Schutten, M. & 19 other authors (2003). Newly discovered coronavirus as the primary cause of severe acute respiratory syndrome. *Lancet* **362**, 263–270.

Liu, D. X. & Inglis, S. C. (1992). Internal entry of ribosomes on a tricistronic mRNA encoded by infectious bronchitis virus. *J Virol* **66**, 6143–6154.

Liu, S., Chen, J., Chen, J. & 7 other authors (2005). Isolation of avian infectious bronchitis coronavirus from domestic peafowl (*Pavo cristatus*) and teal (*Anas*). *J Gen Virol* **86**, 719–725.

Marra, M. A., Jones, S. J., Astell, C. R. & 56 other authors (2003). The genome sequence of the SARS-associated coronavirus. *Science* **300**, 1399–1404.

Poon, L. L., Guan, Y., Nicholls, J. M., Yuen, K. Y. & Peiris, J. S. (2004). The aetiology, origins, and diagnosis of severe acute respiratory syndrome. *Lancet Infect Dis* **4**, 663–671.

Ren, A. X., Xie, Y. H., Kong, Y. Y., Yang, G. Z., Zhang, Y. Z., Wang, Y. & Wu, X. F. (2004). Expression, purification and sublocalization of SARS-CoV nucleocapsid protein in insect cells. *Acta Biochim Biophys Sin* **36**, 754–758.

Rest, J. S. & Mindell, D. P. (2003). SARS associated coronavirus has a recombinant polymerase and coronaviruses have a history of host-shifting. *Infect Genet Evol* **3**, 219–225.

Robbins, J., Dilworth, S. M., Laskey, R. A. & Dingwall, C. (1991). Two interdependent basic domains in nucleoplasmin nuclear targeting sequence: identification of a class of bipartite nuclear targeting sequence. *Cell* **64**, 615–623.

Robertson, M. P., Igel, H., Baertsch, R., Haussler, D., Ares, M., Jr & Scott, W. G. (2005). The structure of a rigorously conserved RNA element within the SARS virus genome. *PLoS Biol* **3**, E5.

Sapats, S. I., Ashton, F., Wright, P. J. & Ignjatovic, J. (1996a). Novel variation in the N protein of avian infectious bronchitis virus. *Virology* **226**, 412–417.

Sapats, S. I., Ashton, F., Wright, P. J. & Ignjatovic, J. (1996b). Sequence analysis of the S1 glycoprotein of infectious bronchitis viruses: identification of a novel genotypic group in Australia. *J Gen Virol* **77**, 413–418.

Singer, S. J. (1990). The structure and insertion of integral proteins in membranes. *Annu Rev Cell Biol* **6**, 247–296.

Singh, M. (1999). A novel internal open reading frame product expressed from a polycistronic mRNA of porcine epidemic diarrhoea virus may not contribute to virus attenuation. *J Gen Virol* **80**, 1959–1963.

Snijder, E. J., Bredenbeek, P. J., Dobbe, J. C. & 7 other authors (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**, 991–1004.

Stanhope, M. J., Brown, J. R. & Amrine-Madsen, H. (2004). Evidence from the evolutionary analysis of nucleotide sequences for a recombinant history of SARS-CoV. *Infect Genet Evol* **4**, 15–19.

Stavrinides, J. & Guttman, D. S. (2004). Mosaic evolution of the severe acute respiratory syndrome coronavirus. *J Virol* **78**, 76–82.

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.

Sun, Z. F. & Meng, X. J. (2004). Antigenic cross-reactivity between the nucleocapsid protein of severe acute respiratory syndrome (SARS) coronavirus and polyclonal antisera of antigenic group I animal coronaviruses: implication for SARS diagnosis. *J Clin Microbiol* **42**, 2351–2352.

Traavik, T., Mehl, R. & Kjeldsberg, E. (1977). "Runde" virus, a coronavirus-like agent associated with seabirds and ticks. *Arch Virol* **55**, 25–38.

Tung, F. Y., Abraham, S., Sethna, M., Hung, S. L., Sethna, P., Hogue, B. G. & Brian, D. A. (1992). The 9-kDa hydrophobic protein encoded at the 3′ end of the porcine transmissible gastroenteritis coronavirus genome is membrane-associated. *Virology* **186**, 676–683.

Wang, D., Urisman, A., Liu, Y. T. & 11 other authors (2003). Viral discovery and sequence recovery using DNA microarrays. *PLoS Biol* **1**, E2.

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.

Wurm, T., Chen, H., Hodgson, T., Britton, P., Brooks, G. & Hiscox, J. A. (2001). Localization to the nucleolus is a common feature of coronavirus nucleoproteins, and the protein may disrupt host cell division. *J Virol* **75**, 9345–9356.

Zhou, M. & Collisson, E. W. (2000). The amino and carboxyl domains of the infectious bronchitis virus nucleocapsid protein interact with 3′ genomic RNA. *Virus Res* **67**, 31–39.