


SCIENTIFIC REPORTS

OPEN

Genome Organization of Canada Goose Coronavirus, A Novel Species Identified in a Mass Die-off of Canada Geese

Amber Papineau^{1,2}, Yohannes Berhane¹, Todd N. Wylie^{3,4}, Kristine M. Wylie^{3,4}, Samuel Sharpe⁵ & Oliver Lung^{1,2} 

The complete genome of a novel coronavirus was sequenced directly from the cloacal swab of a Canada goose that perished in a die-off of Canada and Snow geese in Cambridge Bay, Nunavut, Canada. Comparative genomics and phylogenetic analysis indicate it is a new species of *Gammacoronavirus*, as it falls below the threshold of 90% amino acid similarity in the protein domains used to demarcate *Coronaviridae*. Additional features that distinguish the genome of Canada goose coronavirus include 6 novel ORFs, a partial duplication of the 4 gene and a presumptive change in the proteolytic processing of polyproteins 1a and 1ab.

Viruses belonging to the *Coronaviridae* family have a single stranded positive sense RNA genome of 26–31 kb. Members of this family include both human pathogens, such as severe acute respiratory syndrome virus (SARS-CoV)¹, and animal pathogens, such as porcine epidemic diarrhea virus². Currently, the International Committee on the Taxonomy of Viruses (ICTV) recognizes four genera in the *Coronaviridae* family: *Alphacoronavirus*, *Betacoronavirus*, *Gammacoronavirus* and *Deltacoronavirus*. While the reservoirs of the *Alphacoronavirus* and *Betacoronavirus* genera are believed to be bats, the *Gammacoronavirus* and *Deltacoronavirus* genera have been shown to spread primarily through birds³. The first three species of the *Deltacoronavirus* genus were discovered in 2009⁴ and recent work has vastly expanded the *Deltacoronavirus* genus, adding seven additional species³.

By contrast relatively few species within the *Gammacoronavirus* genus have been identified. There are currently two recognized species in the *Gammacoronavirus* genus: avian coronavirus (ACoV) and beluga whale coronavirus SW1 (SW1). ACoVs infect multiple avian hosts and include several important poultry pathogens, such as infectious bronchitis virus (IBV) and turkey coronavirus (TCoV)⁵. IBV was first described in the United States⁶ but has since been described around the globe⁷. Turkey Coronavirus is the cause of acute enteritis in domestic turkeys⁸. The second species in the *Gammacoronavirus* genus SW1 was first discovered in beluga whales⁹ but has since been detected in other cetaceans, such as Indo-Pacific bottlenose dolphins¹⁰. Despite IBV being the first discovered coronavirus and the impact it has on the poultry industry¹¹, the number of identified species within the *Gammacoronavirus* genus remains small in comparison to the other coronavirus genera. Coronaviruses from several other avian hosts for which partial sequences are available suggest relatedness to IBV and TCoV. These viruses, which include goose coronavirus (GCoV), were tentatively classified as part of the ACoV species. An approximately 3 kb region, including the nucleocapsid gene and several accessory genes, of GCoV were previously sequenced from a greylag goose in Norway¹².

Here we present the full genome of Canada goose coronavirus (CGCoV) sequenced directly from the cloacal swab of a Canada goose, which expired in a mass die-off in a remote region near the arctic in Nunavut, Canada. Our analyses demonstrate that it should be classified as a novel species in the *Gammacoronavirus* genus.

¹National Centre for Foreign Animal Disease, Canadian Food Inspection Agency, Winnipeg, MB, Canada.

²Department of Biological Sciences, University of Manitoba, Winnipeg, MB, Canada. ³Department of Pediatrics, Washington University School of Medicine, St. Louis, St. Louis, MO, USA. ⁴McDonnell Genome Institute, Washington University School of Medicine in St. Louis, St. Louis, MO, USA. ⁵Faculty of Veterinary Medicine, University of Calgary, Calgary, AB, Canada. Correspondence and requests for materials should be addressed to O.L. (email: Oliver.Lung@canada.ca)

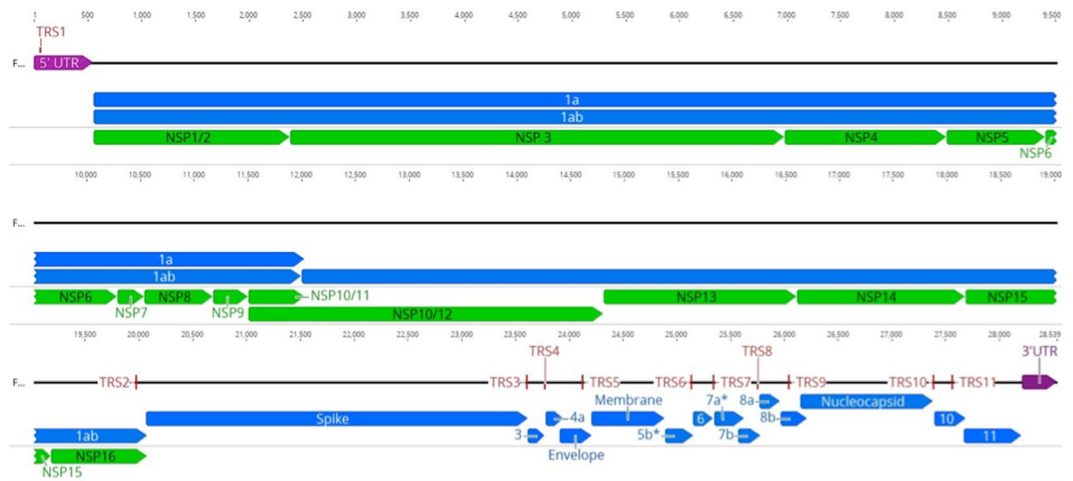


Figure 1. Genome organization of Canada goose coronavirus. Purple indicates untranslated regions, blue indicates putative proteins, green indicates coding region of mature non-structural proteins (NSP) and red indicates transcription regulatory sequences (TRS). The stem loop-like motif and octamer motif are contained within the 3' UTR. Genome organization figure was constructed using Geneious™ (Biomatters, v 9.1.8). *Indicate ACoV 4b homologues. Proteins are named numerically from the 5' end of the genome, with the exception of the structural genes, which are denoted by their common names.

Results and Discussion

Due to the remote location of the die off, samples from the dead birds were not collected immediately and sent to a diagnostic laboratory until severe predation and decomposition had occurred. The poor sample quality, in addition to the difficulty of coronavirus isolation, led to the failure to isolate infectious virus using standard methods. However, the complete genome of a novel gammacoronavirus was assembled from high throughput sequencing reads derived from the cloacal swab of a single Canada goose. The assembled genome of the novel Canada goose coronavirus (CGCoV) is 28,539 nts in length (excluding the poly(A) tail) and has 38.4% GC-content. The genome of CGCoV is approximately 1000 nts longer than the reference genomes for ACoV available in GenBank. The genome organization of CGCoV is presented in Fig. 1. The 5' UTR of CGCoV is 553 nt in length and contains a higher GC content (48.3%) relative to the genome as a whole. The 5' UTR of CGCoV shares only 68% pairwise identity with that of duck coronavirus (DCoV) and 47.5% pairwise identity to that of SW1. Like all coronavirus genomes reported to date, CGCoV's genome is dominated by the coding regions for the large polyproteins 1a and 1ab, followed by the structural and accessory genes. The heptanucleotide slippery sequence UUUAAAC, associated with the ribosomal slippage that produces polyprotein 1ab, was present at nt position 11,995. CGCoV's genome contains genes for all four structural proteins common to coronaviruses; spike (S), envelope (E), membrane (M) and nucleocapsid (N). In addition, CGCoV contains 10 open reading frames (ORFs) predicted to encode accessory proteins. The order of the structural and accessory protein-coding ORFs in CGCoV resembles that of ACoV, but there are notable differences. The general genome organization of ACoV is 1ab-S-3a-3b-E-M-4b-4c-5a-5b-N-6b¹³. However, there is some variance in the genome organization within the ACoV species. For example, Australian IBV strains lack ORFs 4a, 4b and 5b¹⁴. Overall, CGCoV contains a larger number ($n = 14$) of ORFs coding for predicted accessory and structural proteins downstream of the polyprotein 1ab coding region. Two additional ORFs (7a and 7b) are found between the CGCoV M and N ORFs. There are also two additional ORFs (10 and 11) following the N gene. While some ACOVs do have ORFs following the N gene, ORFs 10 and 11 in CGCoV do not share obvious homology to those of IBV and TCoV. The 3' UTR of CGCoV is 301 nucleotides in length and contains the stem loop-like motif 113 bp upstream from the poly(A) tail. This stem loop-like motif was first identified in astroviruses¹⁵ but is also present in ACOVs and SARS-CoV¹³. Further downstream in the 3' UTR, the octanucleotide motif (GGAAGAGC) is found 71 bp upstream of the poly(A) tail. The 3' UTR of CGCoV shares 98% pairwise identity to the partially sequenced GCoV and 84% pairwise identity to IBV.

A trait suggesting common ancestry between CGCoV and ACoV is the canonical ACoV transcription regulatory sequence (TRS) found at the end of the leader sequence in CGCoV. The TRS of CGCoV is identical to that identified by Cao *et al.* (2008) as the TRS of TCoV (CTTAACAAA). Body TRSs regulate viral gene expression by forming a complex with the leader TRS, causing discontinuous transcription of mRNA¹⁶. Ten putative body TRSs were found in the 3' end of the CGCoV genome (Fig. 1). Four of the ten putative TRSs (4, 6, 8, 9) were exact matches to the canonical leader TRS. Three TRSs (2, 7, 11) contained one mismatch and the remaining three TRSs (3, 5, 10) contained two mismatches to the leader TRS. The functionality of these TRSs would need to be experimentally determined; however, previous studies have shown that TRSs of ACOVs are subject to some variation^{13,17}. CGCoV contains twice the number of TRSs as ACOVs and a similar number compared to the nine contained in SW1⁹. Table 1 demonstrates the nucleotide distances between the TRS and the start codon of ORFs found in CGCoV's, which are comparable to those of TCoV¹³.

The start codon of CGCoV's polyprotein 1ab is located 567 nucleotides downstream of the leader TRS. The coronavirus polyprotein 1ab is cleaved into 15–16 non-structural proteins (NSPs) by two viral proteases¹⁸.

Protein	Top Match in NCBI	Top match - aa % identity*	Size (aa)	Distance between TRS and start codon (nt)
1a	1a-Infectious bronchitis virus strain B1648	43	3825	480
1ab	1ab-Infectious bronchitis virus strain ck/CH/LJL/051	57	6510	480
Spike	Spike-Infectious bronchitis virus strain N2-75	53	1184	82
3	n/a	n/a	53	0
4a	n/a	n/a	55	3
Envelope	Envelope-Infectious bronchitis virus strain IS-1494	69	100	n/a
Membrane	Membrane-Duck Coronavirus isolate DK/GD/2014	72	235	74
5b	4b-Infectious bronchitis virus strain Georgia 1998 Vaccine	41	88	n/a
6	n/a	n/a	63	5
7a	4b-Duck Coronavirus isolate DK/GD/2014	23	92	3
7b	n/a	n/a	69	n/a
8a	5a-Duck Coronavirus isolate DK/GD/2014	37	65	4
8b	5b-Duck Coronavirus isolate DK/GD/2014	46	85	n/a
Nucleocapsid	Nucleocapsid-Goose Coronavirus	94	414	94
10	ORF _{Xg} -Goose Coronavirus	92	97	0
11	ORF _{Yg} -Goose Coronavirus	81	180	91

Table 1. Putative viral proteins of Canada goose coronavirus. *Matches below 20% coverage not shown.

Protein	CGCoV		TCoV		IBV		SW1	
	Cleavage site	Size aa	Cleavage site	Size aa	Cleavage site	Size aa	Cleavage site	Size aa
NSP1/2	AG [^] GH	609	AG [^] GK	673	AG [^] GK	673	VD [^] GD	636
NSP3	AG [^] GV	1532	AG [^] GV	1594	AG [^] GI	1592	LG [^] GV	1586
NSP4	LQ [^] AG	503	LQ [^] AG	514	LQ [^] SG	514	LQ [^] AG	537
NSP5	LQ [^] SN	307	LQ [^] SS	307	LQ [^] SS	307	LQ [^] SN	303
NSP6	VQ [^] SK	295	VQ [^] SK	297	VQ [^] AK	293	VQ [^] SK	303
NSP7	LQ [^] AV	83	LQ [^] SV	83	LQ [^] SV	83	LQ [^] AV	83
NSP8	LQ [^] NN	212	LQ [^] NN	210	LQ [^] NN	210	LQ [^] NN	198
NSP9	LQ [^] GK	111	LQ [^] SK	111	LQ [^] SK	111	LQ [^] HG	112
NSP10	SRFV*	173	VQ[^]SA	145	VQ[^]SV	145	LQ[^]SV	189
NSP11	—	—	—	23	—	23	—	17
NSP12	SRFV*	1101	VQ[^]SA	941	VQ[^]SV	940	LQ[^]SV	926
NSP13	LQ [^] SC	599	LQ [^] SC	601	LQ [^] SC	600	LQ [^] AS	601
NSP14	LQ [^] SN	522	LQ [^] GT	521	LQ [^] GT	514	LQ [^] SQ	528
NSP15	LQ [^] SI	338	LQ [^] SI	338	LQ [^] SI	338	LQ [^] SL	349
NSP16	LQ [^] SG	298	LQ [^] SA	302	LQ [^] SA	302	LQ [^] SD	312

Table 2. Non-structural proteins size and cleavage site of gammacoronaviruses. *Amino acids present in CGCoV where putative protease cleavage sites were observed in TCoV, IBV and SW1.

Putative cleavage sites for these proteases are present in CGCoV's 1a and 1ab polyproteins, with the exception of the NSP 10/11 (polyprotein 1a) and NSP 10/12 (polyprotein 1ab) cleavage sites. The missing cleavage site would be located near the end of polyprotein 1a, producing the NSPs 10 and 11, and also in the alternatively transcribed polyprotein 1ab, producing NSPs 10 and 12. The absence of the NSP10/11 and 10/12 protease recognition site was confirmed with Sanger sequencing. With the exception of the missing cleavage sites, the putative cleavage sites would produce NSPs of sizes congruent with other *Gammacoronavirus* species (Table 2). No *Gammacoronavirus* species to date, including CGCoV, have a papain-like protease cleavage site between NSP 1-2¹⁹.

While the genome structure of CGCoV resembles that of ACoV, there are some notable differences. For example, there are no homologues to ACoV's 3a or 3b accessory proteins in CGCoV, a trait shared with SW1. Furthermore, CGCoV has a number of ORFs that do not appear to have homologues in other sequenced *Gammacoronavirus* species, such as the ORFs for putative proteins 3 and 4a (Fig. 1). These two ORFs are found in CGCoV in the corresponding location of ACoV's 3a and 3b ORFs (between the S and E ORFs) and are also similar in size to ACoV's 3a and 3b proteins. However, they share no obvious sequence similarity with any 3a or 3b gene, or any other entry in NCBI (Table 1). ACoV's 3a and 3b proteins have been shown to be unnecessary for replication²⁰, however knock-out mutants for these accessory genes are attenuated²¹. The IBV's 3 gene is functionally tricistronic, meaning the 3a, 3b and E proteins are under the control of a single TRS^{22,23}. This is not the case

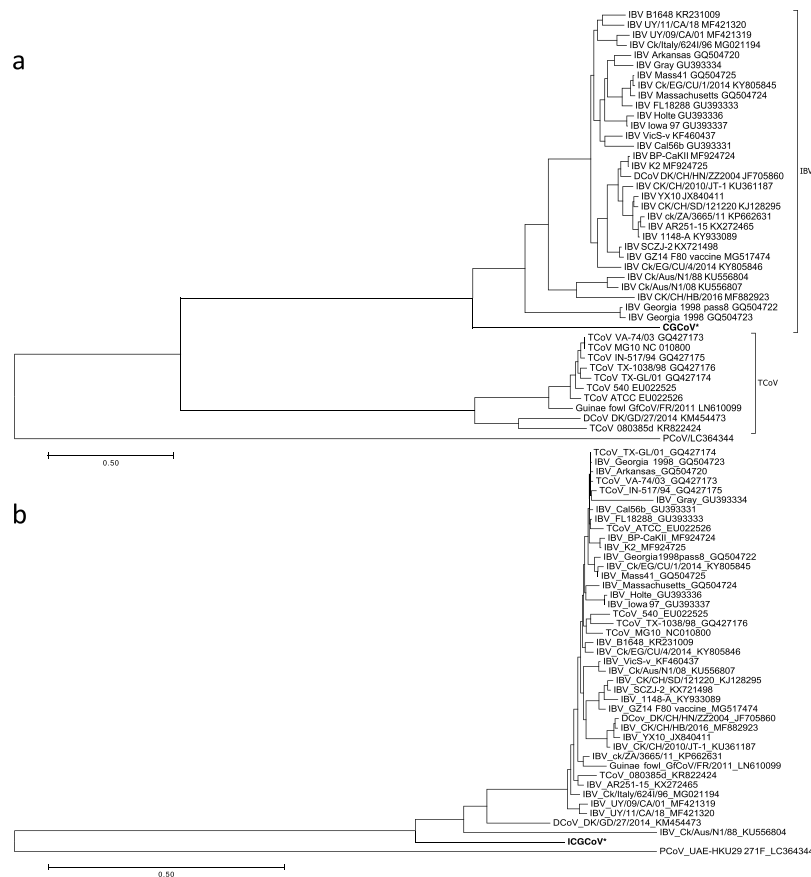


Figure 2. The phylogeny of gamma-coronavirus spike and nucleocapsid proteins. A maximum likelihood tree built, using the amino acid sequences of the spike protein (a) and nucleocapsid protein (b) domains aligned with ClustalW³¹, in MEGA X using the Jones-Taylor-Thornton (JTT) substitution model and 1000 bootstraps³². IBV Infectious Bronchitis virus, TCoV Turkey Coronavirus, PCoV Pigeon Coronavirus, DCoV Duck Coronavirus.

in CGCoV, as the E ORF of CGCoV shares a TRS with only the 4a ORF in CGCoV and 3 ORF is preceded by a separate TRS (Fig. 1).

An additional TRS is also found in between CGCoV's M and N ORFs, preceding the proteins 7a and 7b (Fig. 1). Commonly ACoV's have two ORFs between the M and 5 genes, coding for the 4b and 4c accessory proteins. CGCoV contains 4 ORFs between the M and 8 gene (ACoV 5 gene homologue). Two of these ORFs (5b and 7a) are ACoV 4b homologues, likely the result of gene duplication. This area in IBV has been identified as a hotspot for recombination²⁴. The region between the ACoV M and 5 gene was formally called the intergenic region because of the lack of a TRS. However, it was later shown that gene 4 is expressed using an alternative TRS in IBV¹⁷. Notably, one of the 4b homologs (i.e. 5b) in CGCoV does have a TRS (Fig. 1). The use of template switching at TRSs is thought to lend to recombination in coronaviruses²⁵. The two CGCoV 4b homologs are not identical to each other (Table 1). Amino acid sequence identity to other 4b proteins is low for both CGCoV 4b homologues, 41% to IBV and 23% to DCoV respectively. The gene 4 duplication was also confirmed by Sanger sequencing of the genomic region between the M ORF to the 8 gene.

The ACoV 5a and 5b accessory proteins (8a and 8b in CGCoV) appear to be the only accessory proteins conserved in all 3 *Gammacoronavirus* species, although gene order differs. ORFs encoding putative proteins 5a and 5b belong to the bicistronic gene 5 of ACoVs and are also unnecessary for replication²¹. To date, all publically available sequence information suggest that *Gammacoronavirus* species have lost the NSP1 cleavage site. The function of NSP1 in alphacoronaviruses and betacoronaviruses is the inhibition of host protein production. Accessory protein 5a is shown to have adopted this function in place of NSP1 in IBV¹⁹.

The majority of structural proteins of CGCoV also share low amino acid sequence identity (53–72%) with IBV and DCoV. Phylogenetic analysis of the spike gene show that the CGCoV spike gene clusters with the IBV spike gene, separate from the TCoV cluster (Fig. 2a). Figure 2b also demonstrates the nucleocapsid gene of CGCoV is distantly related to those of ACoVs. However the CGCoV nucleocapsid protein does share 94% amino acid sequence identity with the nucleocapsid protein encoded in the partially sequenced graylag GCoV genome¹³. In addition, ORFs 10 and 11, which are preceded by the nucleocapsid gene, also share high amino acid identity with graylag GCoV proteins, 92% and 81% respectively. It should be noted that, among full and partial genomes of gammacoronaviruses sequenced to date, ORFs 10 and 11 seem to be unique to CGCoV and GCoV and are both preceded by a TRS, suggesting that these ORFs are very likely expressed. The fact that some CGCoV proteins

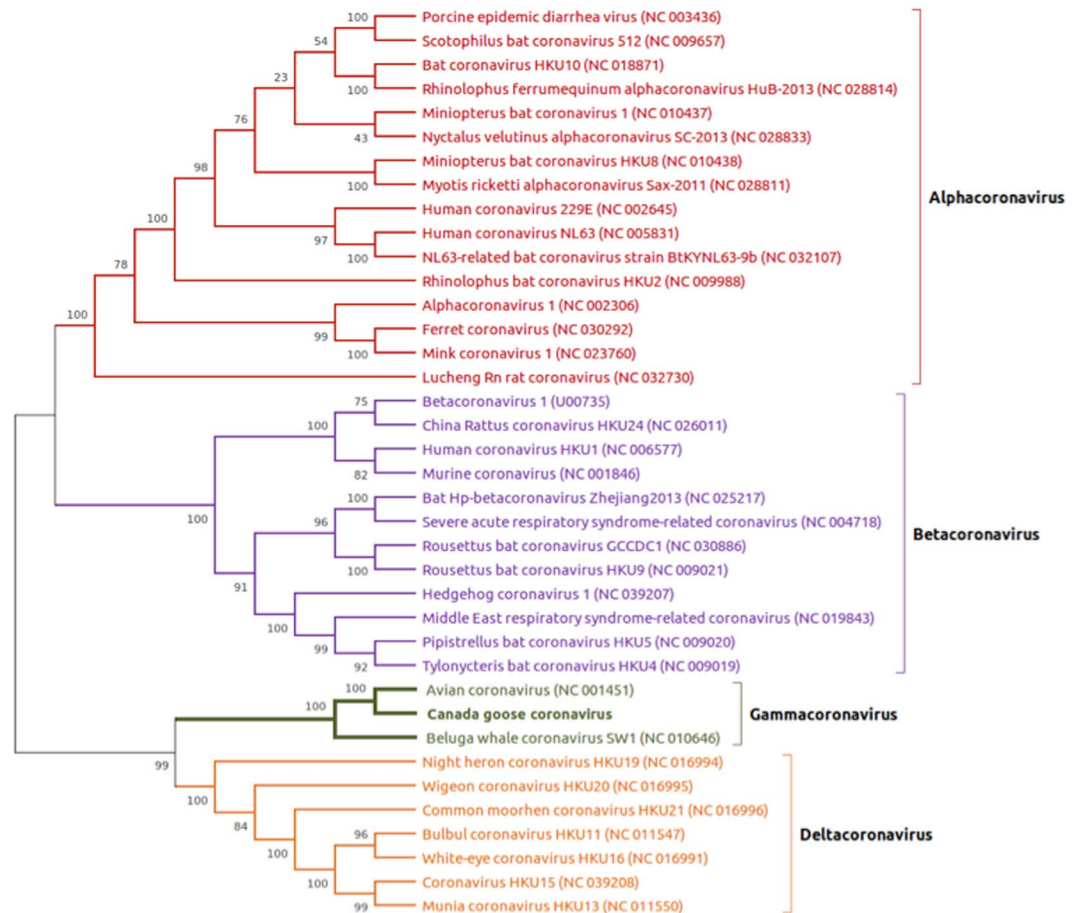


Figure 3. The phylogeny of Canada goose coronavirus. A maximum likelihood tree built, using the concatenated amino acid sequences of the replicase and helicase protein domains aligned with ClustalW³¹, in MEGA X using the Jones-Taylor-Thornton (JTT) substitution model and 1000 bootstraps³². Numbers at nodes indicate the bootstrap value.

share higher amino acid sequence similarity with the partial GCoV sequences available suggest these two viruses are more closely related to each other than to other gammacoronaviruses known to date.

The phylogenetic tree built using the coding regions for the conserved replicase and helicase domains demonstrates that CGCoV clusters with gammacoronaviruses and shares a more recent common ancestor with ACoV than with the cetacean gammacoronaviruses (Fig. 3). Further comparisons suggest that CGCoV is a separate species from ACoV. Current taxonomy of *Coronaviridae* is determined using pairwise comparisons of the amino acid sequence of seven conserved domains in the 1ab polyprotein. Members of the same species share over 90% amino acid identity in these seven conserved domains⁵. Percent identity of CGCoV falls well below the 90% threshold set by ICTV with ACoV and SW1, suggesting CGCoV is a separate species (Table 3). Within *Coronaviridae*, CGCoV shares the highest homology (68%) in the 7 conserved domains to the gammacoronaviruses TCoV and DCoV.

As the full genome was sequenced from only the cloacal swab of a single Canada goose, a screening PCR was designed based on the 4b duplication region unique to CGCoV and performed on all samples. The Sanger sequencing primers of the region between the M and 8 gene were used, as this area of the genome is specific to CGCoV. All samples were found to be positive, with the exception of the pharyngeal swab of the snow goose and the lung tissue of the second Canada goose which could not be tested as the sample was exhausted. Amplicons were Sanger sequenced and confirmed to match the CGCoV genome. High throughput sequencing conducted on RNA extracted from cloacal swabs from the second Canada goose and the snow goose also resulted in partial (64 and 18%) genomes of the CGCoV. While this does not confirm the virus's presence in all animals that perished in the die off, this shows CGCoV was present in all birds that were available for testing. Further studies will require the availability of an infectious virus to determine the pathogenicity of CGCoV and its ability to cause mortality in Canada geese and snow geese.

To summarize, the complete genome of CGCoV, a novel *Gammacoronavirus* species was sequenced directly from the cloacal swab of a Canada goose associated with a mass die-off. The CGCoV genome was also detected in samples derived from a second Canada goose and a snow goose that perished in the die-off, using PCR, Sanger and high throughput sequencing. Comparative genomics and phylogenetic analysis indicate CGCoV clusters with ACoV but is a distinct *Gammacoronavirus* species. Interesting features of this new species include the

Domain	aa % identity to IBV	aa % identity to TCoV	aa % identity to DCoV	aa % Identity to SW1
ADP-ribose-1''-phosphatase	42	43	38	23
3C-like Protease	56	58	57	49
RdRp	80	80	83	69
Helicase 1	89	90	92	78
Exonuclease	78	72	77	56
Endoribonuclease	53	53	54	41
Ribose-2'-O methyltransferase	74	77	76	65
Average	67	68	68	54

Table 3. Comparison of the amino acid pairwise identity of 7 conserved coronavirus domains in the poly1ab protein of Canada goose coronavirus to other gammacoronaviruses.

presence of two 4b homologues, a putative change in the proteolytic processing of the polyproteins 1a and 1ab, and six novel accessory genes.

Methods

Source of samples. A large die off of Canada and snow geese occurred in the fall of 2017 near the arctic in Cambridge Bay, Nunavut, Canada. Due to poor carcass quality and remote location, samples were only collected from two dead Canada geese and one Snow goose, all of which had undergone predation and decomposition. Cloacal and pharyngeal swabs were collected from all three birds, lung tissue was collected from one Canada goose. Other organs were not present or were in extremely poor condition. Detection of both common avian pathogens, such as avian influenza and avian paramyxovirus by the National Reference Laboratory, by routine laboratory testing gave negative results. Virus isolation was performed by two serial passages in SPF chicken eggs using protocols prescribed by the World Organization for Animal Health (OIE) for the most closely related gammacoronavirus, infectious bronchitis virus (IBV). Samples were then subjected to targeted sequence enrichment²⁶ and next-generation sequencing on an Illumina MiSeq platform.

Sample pre-treatment. Tissues were homogenized using a Precellys Evolution homogenizer (Bertin Instruments) according to the manufacturer's instructions. Following a clarification by centrifugation at 3000 rpm for 10 minutes, nucleic acids were extracted using the MagMAX Pathogen RNA/DNA Kit (Ambion) according to the manufacturer's instructions.

cDNA synthesis was then performed using SuperScriptTM IV First-Strand Synthesis System (SSIV) (ThermoFisher) according to the manufacturer's recommendation. A total of 11 μ L of extracted total nucleic acid was mixed with dNTPS (10 mM) and a tagged random nonamer primer (40 μ M) (GTT TCC CAG TCA CGA TAN NNN NNN NN). Samples were incubated at 65 °C for 5 minutes, and then placed on ice for 1 minute. A reagent mixture of 5x SSIV Buffer, Ribonuclease Inhibitor (40 U/ μ L), DTT (100 mM) and SuperScriptTM IV Reverse Transcriptase was then added. The samples were incubated for 10 minutes at 23 °C, 10 minutes at 50 °C and 10 minutes at 80 °C.

Second strand synthesis was performed using Sequenase Version 2.0 DNA Polymerase (ThermoFisher) according to the manufacturer's recommendation. The first strand synthesis product was incubated with 10 μ L of Sequenase Version 2.0 DNA Polymerase diluted in 5x reaction buffer and nuclease free water. Samples were then heated to 37 °C over five minutes and incubated at 37 °C for 12 minutes, followed by 2 minutes at 95 °C. Samples were then cooled to 10 °C and 1.2 μ L of Sequenase DNA polymerase in dilution buffer was added. Samples were again ramped to 37 °C over five minutes and incubated at 37 °C for 12 minutes, followed by 8 minutes at 95 °C. A total of 6 μ L of the second strand synthesis product was then used as template for amplification. AccuPrimeTM Taq DNA Polymerase (ThermoFisher) was mixed with 10X AccuPrimeTM PCR Buffer I, nuclease free water and a primer for the nonomer's tag (100 μ M). 30 cycles of PCR were then performed with the following parameters: 30 seconds at 94 °C, 30 seconds at 40 °C, 30 seconds at 50 °C and 1 minute at 72 °C. cDNA/DNA mixtures were then cleaned with Genomic DNA Clean & Concentrator columns (Zymo Research) and eluted in 20 mM Tris (ThermoFisher).

Library preparation and sequencing. Sequence libraries were prepared with the KAPA HyperPlus library kit (Roche). Sequence library construction and capture were carried out according to Nimblegen's SeqCap EZ HyperCap Workflow User's Guide v1. Samples were pooled in equal amounts by weight prior to capture. Sequencing was performed on an Illumina Miseq instrument in the National Centre for Foreign Animal Disease biocontainment level 3 sequencing facility. A V2 flow cell was used with a 500 cycle reagent cartridge (Illumina).

5' Race and Sanger sequencing. 5' RACE was used to obtain the missing leader sequence (52 bp). The SMARTer 5' RACE and 3' RACE kit (TakaraBio) was used according to the kit instructions. The gene specific primer used for 5' RACE was TCAGCTACAGTAGAGGGAGATGTCATAGGTGC. For Sanger sequencing, amplicons was performed using KAPA HiFi HotStart ReadyMixPCR Kit (KAPABiosystems). The primers CTAAAGAGAAGGTGGACACTGGT and CTAAGAATGCGAACTTCACAGAGC were used to amplify the gene 4b homologue region. The primers GTTGTTGTGTTACAAGGCAAGGG and GGATTATGATCAAACCATGAACCTGG were used to amplify the NSP 10/12 region. Cycling conditions

used to generate amplicon for Sanger sequencing were: 1 cycle: 95 °C for 3 minutes, 40 cycles: 98 °C for 20 seconds, 65 °C for 15 seconds, 72 °C for 2.5 minutes, and 1 cycle: 72 °C for 3 minutes. Amplicons were cleaned using AMPure XP beads (Beckman Coulter) according to the manufacturer's directions. Sanger sequencing was performed on the ABI Genetic Analyzer 3130XL platform using the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) according to the user manual.

Bioinformatics. Read quality was assessed using FastQC and trimmed using Trimmomatic²⁷ (Version 0.36). Host reads were then filtered with RAMBO-K, using the only complete genome of a goose species (*Anser cygnoides*) currently available and DCoV²⁸. The near complete genome sequence of CGCoV was assembled from NGS derived sequences from a cloacal swab of one Canada goose using SPAdes²⁹. Sanger reads were aligned to the draft genome in GeneiousTM (Biomatters, v 9.1.8). Annotations were performed using Geneious and protein domains were identified using PFAM³⁰. The Canada goose coronavirus genome is available under accession number MK359255 on NCBI.

References

1. Tsang, K. W. *et al.* A cluster of cases of severe acute respiratory syndrome in Hong Kong. *N. Engl. J. Med.* **348**, 1977–1985 (2003).
2. Wood, E. N. An apparently new syndrome of porcine epidemic diarrhoea. *Vet. Rec.* **100**, 243–244 (1977).
3. Woo, P. C. Y. *et al.* Discovery of seven novel Mammalian and avian coronaviruses in the genus deltacoronavirus supports bat coronaviruses as the gene source of alphacoronavirus and betacoronavirus and avian coronaviruses as the gene source of gammacoronavirus and deltacoronavirus. *J. Virol.* **86**, 3995–4008 (2012).
4. Woo, P. C. Y. *et al.* Comparative analysis of complete genome sequences of three avian coronaviruses reveals a novel group 3c coronavirus. *J. Virol.* **83**, 908–917 (2009).
5. de Groot, R. J. *et al.* Revision of the family Coronaviridae. *Taxonomic proposal to the ICTV Executive Committee*, 1–37 (2008).
6. Fabricant, J. The early history of infectious bronchitis. *Avian Dis.* **42**, 648–650 (1998).
7. Bande, F. *et al.* Global distributions and strain diversity of avian infectious bronchitis virus: a review. *Anim. Health Res. Rev.* **18**, 70–83 (2017).
8. Lin, T. L. *et al.* Characterization of turkey coronavirus from turkey poults with acute enteritis. *Vet. Microbiol.* **84**, 179–186 (2002).
9. Mihindukulasuriya, K. A., Wu, G., St Leger, J., Nordhausen, R. W. & Wang, D. Identification of a novel coronavirus from a beluga whale by using a panviral microarray. *J. Virol.* **82**, 5084–5088 (2008).
10. Woo, P. C. Y. *et al.* Discovery of a novel bottlenose dolphin coronavirus reveals a distinct species of marine mammal coronavirus in Gammacoronavirus. *J. Virol.* **88**, 1318–1331 (2014).
11. Jones, R. C. Viral respiratory diseases (ILT, aMPV infections, IB): are they ever under control? *Br. Poult. Sci.* **51**, 1–11 (2010).
12. Jonassen, C. M. *et al.* Molecular identification and characterization of novel coronaviruses infecting graylag geese (*Anser anser*), feral pigeons (*Columba livia*) and mallards (*Anas platyrhynchos*). *J. Gen. Virol.* **86**, 1597–1607 (2005).
13. Cao, J., Wu, C.-C. & Lin, T. L. Complete nucleotide sequence of polyprotein gene 1 and genome organization of turkey coronavirus. *Virus Res.* **136**, 43–49 (2008).
14. Mardani, K., Noormohammadi, A. H., Hooper, P., Ignjatovic, J. & Browning, G. F. Infectious bronchitis viruses with a novel genomic organization. *J. Virol.* **82**, 2013–2024 (2008).
15. Jonassen, C. M., Jonassen, T. O. & Grinde, B. 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**(Pt 4), 715–718 (1998).
16. Dufour, D., Mateos-Gomez, P. A., Enjuanes, L., Gallego, J. & Sola, I. Structure and functional relevance of a transcription-regulating sequence involved in coronavirus discontinuous RNA synthesis. *J. Virol.* **85**, 4963–4973 (2011).
17. Bentley, K., Keep, S. M., Armesto, M. & Britton, P. Identification of a noncanonically transcribed subgenomic mRNA of infectious bronchitis virus and other gammacoronaviruses. *J. Virol.* **87**, 2128–2136 (2013).
18. Ziebuhr, J., Snijder, E. J. & Gorbalenya, A. E. Virus-encoded proteinases and proteolytic processing in the Nidovirales. *J. Gen. Virol.* **81**, 853–879 (2000).
19. Kint, J. *et al.* Infectious bronchitis coronavirus limits interferon production by inducing a host shutoff that requires accessory protein 5b. *J. Virol.* **90**, 7519–7528 (2016).
20. Hodgson, T., Britton, P. & Cavanagh, D. Neither the RNA nor the proteins of open reading frames 3a and 3b of the coronavirus infectious bronchitis virus are essential for replication. *J. Virol.* **80**, 296–305 (2006).
21. Laconi, A. *et al.* Deletion of accessory genes 3a, 3b, 5a or 5b from avian coronavirus infectious bronchitis virus induces an attenuated phenotype both *in vitro* and *in vivo*. *J. Gen. Virol.* **99**, 1381–1390 (2018).
22. Liu, D. X., Cavanagh, D., Green, P. & Inglis, S. C. A polycistronic mRNA specified by the coronavirus infectious bronchitis virus. *Virology* **184**, 531–544 (1991).
23. Brooks, J. E. *et al.* Comparisons of envelope through 5B sequences of infectious bronchitis coronaviruses indicates recombination occurs in the envelope and membrane genes. *Virus Res.* **100**, 191–198 (2004).
24. Hewson, K. A., Ignjatovic, J., Browning, G. F., Devlin, J. M. & Noormohammadi, A. H. Infectious bronchitis viruses with naturally occurring genomic rearrangement and gene deletion. *Arch. Virol.* **156**, 245–252 (2011).
25. Simon-Loriere, E. & Holmes, E. C. Why do RNA viruses recombine? *Nat. Rev. Microbiol.* **9**, 617–626 (2011).
26. Wylie, T. N., Wylie, K. M., Herter, B. N. & Storch, G. A. Enhanced virome sequencing using targeted sequence capture. *Genome Res.* **25**, 1910–1920 (2015).
27. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
28. Tausch, S. H., Renard, B. Y., Nitsche, A. & Dabrowski, P. W. RAMBO-K: Rapid and Sensitive Removal of Background Sequences from Next Generation Sequencing Data. *PLoS ONE* **10**, e0137896 (2015).
29. Bankevich, A. *et al.* SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).
30. El-Gebali, S. *et al.* The Pfam protein families database in 2019. *Nucleic Acids Res.*, <https://doi.org/10.1093/nar/gky995> (2018).
31. Larkin, M. A. *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948 (2007).
32. Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol. Biol. Evol.* **35**, 1547–1549 (2018).

Acknowledgements

The authors acknowledge funding from Canadian Food Inspection Agency (CFIA) project WIN-A-1408 and Canadian Safety and Security Program project TI-2222 for the student stipend of A.P. The authors would also like to acknowledge Michelle Nebroski and Mathew Fisher for review of the manuscript and technical assistance.

Author Contributions

A.P. and O.L. designed the experiment and wrote the main manuscript text. Y.B. and S.S. performed sample collection and routine testing. T.W. and K.W. designed the targeted sequence capture method used for enrichment of viral sequences. A.P. performed the experimental work and performed the analysis. All authors reviewed and approved the manuscript.

Additional Information

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019