



Sveriges lantbruksuniversitet
Swedish University of Agricultural Sciences

Fakulteten för veterinärmedicin och
husdjursvetenskap
Institutionen för biomedicin och veterinär
folkhälsovetenskap

Genetic characterization of canine respiratory coronavirus

Simon Larsson

Uppsala 2015

Kandidatarbete 15HP i biologi

Genetisk karakterisering av hundens respiratoriska coronavirus

Genetic characterization of canine respiratory coronavirus

Simon Larsson

Handledare: Maja Malmberg, institutionen för biomedicin och veterinär folkhälsovetenskap

Andra handledare: Jonas Johansson Wensman, institutionen för kliniska vetenskaper

Examinator: Mikael Berg, institutionen för biomedicin och veterinär folkhälsovetenskap

Kandidatarbete i Biologi

Omfattning: 15 hp

Nivå och fördjupning: grundnivå, G2E

Kurskod: EX0689

Utgivningsort: Uppsala

Utgivningsår: 2015

publicering: <http://stud.epsilon.slu.se>

Nyckelord: Coronavirus, hund, kennelhosta, spikprotein, hemagglutinin-esteras, virus,

Key words: Coronavirus, Canine, Kennel Cough, Canine Infectious Respiratory Disease, Spike Protein, Virus, Hemagglutinin-e

Table of contents

Sammanfattning på svenska	1
Abstract	2
Introduction.....	3
Canine respiratory coronavirus and its epidemiology.....	6
Molecular virology and mechanisms in the replication cycle.....	6
The coronavirus spike protein (S).....	7
The coronavirus membrane protein (M).....	8
The coronavirus nucleoprotein (N).....	8
The coronavirus envelope protein (E).....	9
Accessory proteins.....	9
Laboratory procedure	8
Results	11
Discussion	16
Acknowledgements	19
References.....	20

Sammanfattning på svenska

Kennelhosta är ett samlingsnamn för en grupp virus- och bakterieinfektioner som ofta drabbar hundar som är tillsammans med större grupper av andra hundar. Symptomen är oftast milda, vanligtvis med en torr, hackande hosta, men kan i vissa fall ta sig uttryck som svårare symptom som lunginflammation. De vanligaste infektionsämnena är hundens parainfluenzavirus (CPIV), hundens respiratoriska coronavirus (CRCoV), hundens adenovirus typ 2 (CAV-2) och bakterien *Bordetella bronchiseptica*. Hundens respiratoriska coronavirus är det virus som undersöktes i studien som ligger till grund för den här rapporten. Familjen coronavirus innehåller de största membranomslutna virusen med RNA som arvs massa med genomstorlekar på upp till 32000 baser. Dess medlemmar har en gemensam genomstruktur med två stora öppna läsramar (ORF:s) i sin 5'-ände och gener som kodar för strukturella protein, det vill säga protein som bygger upp viruskapsiden, i 3'-änden i ordningen S – E – M – N där S är ett spikprotein, E är ett höljeprotein, M är ett membranprotein och N är ett nukleoprotein. Mellan dessa proteiner finns ofta olika mindre protein som inte är gemensamma för hela familjen, varav ett är ett influensa-liknande hemagglutinin-esteras.

I studien som ligger till grund för den här rapporten undersöktes sekvensen som kodar för tre strukturella protein hos hundens respiratoriska coronavirus; spikprotein, membranprotein och hemagglutinin-esteraset. Hundens respiratoriska coronavirus (CRCoV) tillhör underfamiljen betacoronavirus som också innehåller bland annat nötboskapens coronavirus, människans coronavirus OC43, musens hepatitvirus och hästens coronavirus. Fylogenetiska undersökningar gjordes där CRCoV jämfördes med andra betacoronavirus, samt de tio specifika prover som tagits i Sverige. Ytterligare jämförelser gjordes även mellan andra sekvenser från tidigare studier av CRCoV runt om i världen (Korea, Storbritannien, Italien och Japan).

Det laborativa arbetet bestod i att till en början syntetisera cDNA från RNA som extraherats från nos- och svalgsvabbar från hundar som besökt utvalda veterinärkliniker runt om i landet och som visats vara PCR-positiva för CRCoV. cDNA:t användes sedan som templat för PCR-amplifikation av den virala arvs massan och efter gelelektrofores, antingen enbart för att fastslå att PCR-reaktionen var lyckad eller för att också i vissa fall gelrena PCR-produkten, skickades PCR-produkterna för sekvensering. Primerpar användes för totalt sex genfragment, hemagglutinin-esterasgenen, membranproteingenen och fyra fragment för spikproteingenen. Antalet sekvenser per prov blev varierat; i vissa prover blev samtliga fragment sekvenserade och i andra bara några få fragment. Två prover gav aldrig några PCR-produkter över huvudtaget.

De CRCoV-genom som undersöktes verkade intressant nog ligga närmast ett isolat från en hund i Japan. Det är intressant eftersom sekvenser från virus isolat i England, som geografiskt ligger betydligt närmare, var förhållandevis långt bort från de svenska proverna rent fylogenetiskt. Även sekvenser tagna från coronavirus från svenska kor låg långt bort i trädets vilken borde betyda att det är osannolikt att det CRCoV som finns endemiskt i svenska hundar kommer från en separat överföring från svenska kor utan snarare antingen från andra hundar som besökt Sverige eller från utländska hundar som importerats till Sverige.

Abstract

Kennel cough is a disease complex caused by several different pathogens, both viruses and bacteria. Most commonly, kennel cough presents with very mild clinical signs but occasionally more severe signs like bronchopneumonia can erupt. Kennel cough is caused by one of five pathogens; canine parainfluenzavirus (CPIV), canine respiratory coronavirus (CRCoV), canine adenovirus type-2 (CAV-2), canine influenza virus (CIV) and the bacterial pathogen *Bordetella bronchiseptica*. This study focused on the genetic characterization of field isolates of CRCoV.

The coronavirus family contains the largest known enveloped RNA viruses with genome sizes ranging from 28kb to 32kb. Its members share a common genome structure with two large open reading frames in the 5'-end and the genes coding for structural proteins lumped together in the 3'-end. There is a conserved order in the four shared structural proteins; 5' – S – E – M – N – 3', where S is the spike protein, E is an envelope protein, M a small membrane protein and N is a nucleoprotein. A large and diverse group of smaller proteins have been observed in many of the coronaviruses and these differ from species to species. An example of this is the influenza-like hemagglutinin-esterase found on the surface of coronaviruses of the betacoronavirus subfamily. In this study, the nucleotide sequence of three of the structural protein genes found in the CRCoV genome was examined. CRCoV belongs to the betacoronavirus subfamily which also contains bovine coronavirus, human coronavirus OC43, murine hepatitis virus (MHV) and equine coronavirus, among other species. Phylogenetic analyses were conducted, both on CRCoV sequences in relation to the other coronaviruses and in relation to available CRCoV sequences from different countries (Korea, The United Kingdom, Italy and Japan).

The laboratory work consisted of four major steps. First cDNA was synthesized from RNA extracted from samples taken from nasopharyngeal swabs from dogs admitted at veterinary clinics from around Sweden and proven to be PCR-positive for CRCoV. Then the cDNA was used as template for PCR-amplification and the PCR products were analysed using gel electrophoresis. Some samples needed to be purified and these were purified by running all the PCR-product for that sample on a separate gel. The bands were then cut out with a scalpel and DNA was extracted using a gel purification kit. Sanger sequencing was done by MacroGen at their facility in Amsterdam. The sequences were quality controlled, trimmed and assembled using CodonCode Aligner and the phylogenetic trees were constructed using the PHYLIP neighbour joining algorithms with 1000x Bootstrap in the Ugene (Unipro) software. All phylogenetic trees were constructed using FigTree.

The CRCoV sequences that were analysed for phylogeny with the rest of the betacoronavirus family were most closely related to an isolate from a Japanese dog, which is interesting considering that samples taken geographically closer to Sweden were included. Sequences taken from Swedish cattle (bovine coronavirus) were also considerably distant, genetically, even compared to other BCoV sequences. The most reasonable conclusion is that the CRCoV strains found in Sweden most likely entered the Swedish dog population either through dogs traveling in and out of Sweden or through international pet trade as opposed to from a separate cross-over event from bovine coronavirus found in Swedish cattle.

Introduction

Kennel cough (also called canine infectious respiratory disease, or CIRDC), like the common cold, can be caused by different pathogens. It is a respiratory disease that affects dogs and while most dogs go through the infection asymptotically others might suffer clinical signs ranging from mild nasal congestion to severe bronchopneumonia. The disease is very recognizable due to the common dry, hacking cough. In most cases infected dogs will have cleared the infection on their own within a period of one to three weeks but if bronchopneumonia occurs, veterinary care might be necessary since in a few cases it has led to death or euthanasia (Appel et al, 1987). Dogs that are at risk for infection are usually dogs that spend time in places where close contact with other dogs is common, such as rehoming kennels, dog parks, dog shows or dog day cares.

As previously mentioned, the term kennel cough describes a collection of infectious agents consisting of viruses like canine parainfluenzavirus (CPIV), canine respiratory coronavirus (CRCoV) canine adenovirus type-2 (CAV-2) and canine influenza virus (CIV) as well as the bacterium *Bordetella bronchiseptica*. Canine herpesvirus and canine reovirus have also been associated with the disease but only tangentially (during co-infection with one of the other, more common pathogens) and they are not believed to have an important role in the disease complex (Buonavoglia et al, 2006). However, it has been shown that the pathology of kennel cough is generally mild for any of the infectious agents as long as there is only one infection raging at any given time. The few times where more severe signs have been found, illness has been shown to be due to co-infection from more than one pathogen. (Erles et al, 2003).

Prevalence of the pathogens involved in kennel cough have been shown to vary based on geography but the general trend is that the most common infection is canine parainfluenzavirus followed by either *B. bronchiseptica* or by canine respiratory coronavirus (Schulz et al, 2014). The closest location to Sweden where these ratios have been examined is Germany where the following numbers were reported (B.S. Shultz et al, 2014): canine parainfluenzavirus (23/61 dogs were PCR-positive), canine adenovirus type-2 (0/61 dogs were PCR positive), canine influenza virus (0/61 dogs were PCR-positive), canine respiratory coronavirus (6/61 dogs were PCR-positive), canine herpesvirus (0/61 dogs were PCR-positive) and *B. bronchiseptica* (48/61 dogs were PCR-positive).

There are vaccines for both canine parainfluenzavirus and for *B. bronchiseptica*, but despite the high seroprevalence there is no vaccine for canine respiratory coronavirus. The reason for this could be that the virus was discovered relatively recently and has not been properly studied. An outbreak of CRCoV could potentially cause economic losses for people involved in the dog industry, such as breeders or dog day care providers, or for people who rely on their dogs for specific work, such as people who compete with their dogs or use them for sledding. Therefore, there is arguably a need for a vaccine against CRCoV and since there is almost no research done on the virus, a good start is sequencing genes coding for important proteins for vaccine development. This, as well as finding out if there are any differences in CRCoV over time or across geography in Sweden, is the purpose of this study.

Canine Respiratory Coronavirus and its epidemiology

The coronavirus family consists of the largest known enveloped RNA-viruses with genome sizes ranging from 27kb to 32kb, with genomes rolled up in a helical conformation together with nucleoproteins and envelope proteins (Acheson, 2007). The coronavirus family is generally divided into four distinct genera aptly named alpha-, beta-, delta- and gammacoronavirus where the division depends on which receptor is being used to enter the host cell (King et al, 2012). A complete coronavirus virion is between 80nm and 120nm in diameter and the surface spike proteins protrude an extra 10nm. In betacoronaviruses there is an “underbrush” of significantly shorter Hemagglutinin-Esterase proteins. There are several coronaviruses infecting humans and some notorious members have been known to cause very dangerous syndromes, such as SARS (Severe Acute Respiratory Syndrome) which has a mortality rate of just under 10% (WHO 2015) (Bermingham, 2012). On the other hand, coronaviruses are also known to make up about 30% of common colds so mild infections can occur. The two main transmission mechanisms are through direct contact and through aerosol but other ways have been shown to cause disease, such as faecal-oral transmission and blood transfusion (Acheson, 2007).

Coronaviruses are wide-spread among animals from different branches of the tree of life, such as cattle, dogs and giraffes from the mammal branch and geese, ducks and puffins from the bird branch. Almost all alpha- and betacoronaviruses infect mammals whereas all deltacoronaviruses infect birds. In dogs there are two different coronaviruses, one causing gastrointestinal illness (CCOV) and one causing respiratory illness (CRCoV). CCOV, unlike CRCoV, belongs to the Alphacoronavirus subfamily and has very little in common with CRCoV other than belonging to the same virus family and infecting the same host. The CRCoV has previously been reported to be very similar to bovine coronavirus (BCoV) and one of the human coronaviruses (OC43) with over 90% similarity in both nucleotide sequence and amino acid sequence, however this was determined through alignment of the respective replicase proteins (Masters et al, 2006). Traditionally, the replicase proteins have been used to compare different viruses of the same subfamily because they are more conserved than the structural proteins (Erles et al, 2003).

Different seroprevalences of CRCoV have been reported from geographically separated populations; 54.7% in the United States, 36.6% in Great Britain, 17.8% in Japan, 9.8% in Germany and 32% in Italy. Moreover, the seroprevalence increased as the dogs aged so it is probable that the infection is endemic in these geographical areas and not introduced in short burst epidemics (Priestnall et al, 2006) (Schulz et al, 2014). Another epidemiological fact is that seronegative dogs that were introduced to an infectious environment (i.e. a kennel with infected dogs) turned seropositive usually within a week after introduction (Brownlie et al, 2004). Another study suggested that upon introduction to a rehoming kennel 30% of dogs were already seropositive and after three weeks 99% were seropositive (Erles et al, 2003). Together, this shows that CRCoV is highly contagious in dog populations living in close proximity with one another.

The virus was first discovered in 2003 by Erles et al., but in retrospective studies of older, preserved samples from dogs in Canada it has been shown that CRCoV were already circulating in dogs in 1996 (Ellis et al, 2005). Evidence from sequence identity data points to the virus being transmitted to dogs from cattle and that would imply that BCoV is the ancestral virus to CRCoV (Erles et al, 2007).

Molecular virology

A coronavirus genome essentially looks like any given mRNA, except that it is several times larger than the average mRNA. It has a 5'-cap as well as a 3'-poly(A)-tail attached to it (Lai et al, 1981). Despite the fact that both positive-sense RNA and negative-sense RNA is synthesized, positive-sense RNA is selected for and strongly favoured to form the new virion. Some coronaviruses are able to select for exclusively positive-sense RNA but others have significant portions of negative sense RNA in new virions (Makino et al, 1990) (Hofmann et al, 1990).

All coronaviruses have a conserved genome organization of their essential proteins, 5' – (Pol) – S – E – M – N – 3', often with other proteins interspersed between them. This order is maintained in the CRCoV genome as well (**Fig. 1**).



Figure 1: Genome structure of CRCoV, where the first two ORF:s (orange and green) code for the replicase proteins. The other genes are, from left to right, a 32kD non-structural protein gene, a hemagglutinin-esterase gene, the spike protein gene, a 4,8kD non-structural protein gene, an envelope protein gene, a 4,9kD non-structural protein gene, a membrane protein and a nucleoprotein gene. In a frame shifted ORF within the nucleoprotein there is also a structural protein named internal protein.

Interestingly, it has been shown that while this order is highly conserved in nature, it is not essential for the virus to be viable (de Haan et al, 2002). Since the RNA is part of the virion, one idea for the function of the small “non-structural protein genes” could have been to ensure appropriate length of the RNA molecule for a symmetric virion to form. However, this conclusion is probably not true since the sequences of these genes are conserved. Their protein-binding functionality would have to be sequence-specific and if it was the structure would fall apart when the ORFs were shuffled around.

The coronavirus spike protein (S)

The spike protein is the third largest protein in coronaviruses, after the replicase and a non-structural protein encoded in one of the polyproteins in the first two-thirds of the genome. Pre-glycosylation it has a molecular weight of between 128kDa and 160kDa depending on species of coronavirus and post-glycosylation it weighs between 150kDa and 200kDa (Yamada et al, 2000). This means that the protein is heavily glycosylated and while it is not known how many of- and which of the 23 candidate glycosylation sites are glycosylated, it is probably pretty close to SARS which has 12 glycosylated amino acids (Krokhin et al, 2003). Most glycosylation in coronavirus proteins is O-linked but there is sometimes N-linked glycosylation and depending on how many N-linked glycosylation sites the CRCoV spike protein has, there could be more than 23 candidate sites (Yamada et al, 2000).

The exact mechanism of viral entry of coronaviruses is not known but among the coronaviruses it has been studied in detail only for SARS coronavirus. The spike protein is believed to interact with the

host receptor through a stretch of around 500 amino acids that are believed to be part of a unique domain. This interaction then leads to a conformational change in the spike protein which in turn, by some currently unknown mechanism, leads to the fusion of the viral envelope with the host membrane (Babcock et al, 2004).

No direct evidence for the specific receptor that the spike protein of CRCoV binds to exists to date but because the receptor for BCoV is known to be 9-O-acetosialic acid it is likely that this is the receptor for CRCoV too. The fusion mechanism appears to be roughly similar to those outlined for the HA-protein of influenza and the gp41 from HIV, but the details have never been fully worked out.

An interesting feature of the coronavirus spike protein is that occasionally some of them do not make it into a forming virion before it buds but still manages to get to the cell membrane. Once there it can still be active and fuse an infected cells with an uninfected cell, generating a so-called multinucleate syncytium, which is a cell consisting of the cell membrane of both cells with two nuclei. Since this mechanism of spread does not require an extracellular virion that can be detected by immune cells, the infection can occur without an immune response until complete virions starts forming again.

The coronavirus membrane protein (M)

The M-protein is the most abundant protein in coronavirus virions with regards to the number of protein units and even though it is a fairly small protein (25kDa to about 30kDa) (Sturman, 1977), it is heavily glycosylated which increased the overall weight of the protein. The level of glycosylation often differs from one individual protein to another (Krijnse Locker et al, 1992). One important task for the membrane protein is to shepherd the spike protein to the budding site of new virions. The coronavirus spike proteins do not have a high affinity to the budding site. Membrane proteins attach themselves to the base of the spike protein and use their affinity for the budding site to move the spike proteins there (Opstelten et al, 1995).

The coronavirus nucleoprotein (N)

Coronavirus nucleoproteins range in size from 43kDa to 50kDa and mainly form the protein component of the nucleocapsid (Laude et al, 1995). The protein is phosphorylated quickly after translation and while it has several residues that could be phosphorylated, only a small subset actually is (Siddell et al, 1981) (Wilbur et al, 1986). It is not known why the protein is phosphorylated but some theories are that it increases the affinity to intercellular membranes, like Golgi, and that it helps distinguish between viral RNA substrates and host RNA substrates. Another suggestion is that it is important for virion formation and assembly, but this theory lacks supporting evidence. (Chen et al, 2005)

It is possible that the coronavirus nucleoprotein is somehow involved in the viral replication on a more complex level than simply forming the nucleocapsid together with the viral RNA. It has been shown experimentally that all essential activity for the replication occurs outside the nucleus. In fact, it has even been shown that the replication can occur in the absence of a nucleus as long as other,

more important cellular functions are still intact. However, the nucleoprotein is transported into the nucleus, more specifically to the nucleolus, presumably to fill a modulating function such as inhibiting the cell cycle. This function is, however, not essential for virus replication and could probably be considered a non-essential function to make the environment more beneficial for the virus (Wurm et al, 2001) (Hiscox et al, 2001).

Envelope protein (E)

The smallest of the structural proteins of coronaviruses is the envelope protein. It weighs 8.4kDa to 12kDa, which translates to 76 to 109 amino acids, and there are not many of them in each virion. It differs greatly between coronavirus groups and even between members of the same group. Its tasks in the virion include resistance to alkalinity as well as cleaving the budding virion out of the host cell (Coarse et al, 2000). Like the other structural proteins in the coronavirus virion it goes through post-translational modification although instead of glycosylation or phosphorylation the E-protein is palmitoylated, meaning that it is a lipoprotein (Corse et al, 2002).

When doing mutational analyses on the E-protein it was found that while it is a very important protein for budding, it is not essential. When the envelope protein was missing or damaged, budding still took place but at a much slower rate and often with defective virions as a result (Fischer et al, 1992).

Accessory proteins

Except for the structural proteins mentioned above, many coronaviruses have extra proteins that may differ slightly from one coronavirus to another, even within single subfamilies, such as with the “non-structural proteins” of the betacoronaviruses (**Fig. 2**). Historically, these were called non-structural proteins plus their kDa but it has been shown that some of these proteins actually are part of the virion. Most of them are presumed to have been acquired horizontally, either from host cells or from other co-infecting viruses. It is thought that many of these proteins encode so-called luxury functions that are not necessarily essential to the replication of the virus but rather make it easier for the virus to replicate by different mechanisms, such as regulating the immune system or stopping cell division (Mazumder et al, 2002).

	Bovine (cow)	Giraffe	Dromedary camel	Equine (horse)	Porcine (pig)	Sable antelope	Canine (dog)
Spike protein	Green	Green	Green	Green	Green	Green	Green
Envelope protein	Green	Green	Green	Green	Green	Green	Green
Membrane protein	Green	Green	Green	Green	Green	Green	Green
Nucleoprotein	Green	Green	Green	Green	Green	Green	Green
Hemagglutinin-esterase	Green	Green	Green	Green	Green	Green	Green
32kDa non-structural protein	Green	Green	Green	Green	Green	Green	Green
12.7kDa non-structural protein	Green	Green	Green	Green	Green	Green	Green
4.9kDa non-structural protein	Green	Green	Red	Red	Green	Green	Green
4.8kDa non-structural protein	Green	Green	Red	Green	Red	Green	Green
Internal protein	Red	Red	Green	Green	Green	Red	Green

Figure 2: Protein set-up of some closely related betacoronaviruses where green means that the virus has the gene for the protein and red means that it does not. The internal protein is a different frameshift of the nucleoprotein.

An example of an accessory protein is the hemagglutinin-esterase protein (HE) that is exclusive to the betacoronavirus genus. It is related to the hemagglutinin-esterase named HA1 in influenza C. Interestingly, some betacoronaviruses have HE-genes that encode sialate-9-O-acetylerases while others have HE-genes that encode sialate-4-O-acetylerases (Mazumder et al, 2002). This would imply that two proteins of roughly the same function have been acquired horizontally twice, independently of one another and the chances of this happening without it granting the virus some selective advantages are astronomical. However, a conclusive finding that illuminates the function of the HE-gene has not yet been made; replication seems to go on as usual without it and an experiment with three mutants, one with a functional HE-protein, one with a deactivated HE-protein and one completely lacking the HE-gene, showed that a virus lacking the HE-gene all together had a higher growth rate than counterparts that did not have the gene. This implies that having the HE-gene is a burden in the viral replication, which if it did not have a function that gave the virus selective advantage would be quickly removed. The proposed function involves being able to infect more cell types and/or hosts but apart from a hard-to-reproduce study suggesting that it gave the virus higher neurovirulence, this has yet to be proved (Cornelissen et al, 1997) (Kazi et al, 2005) (Lissenberg et al, 2005).

Laboratory procedure

Sample preparation

The samples used for the analysis were collected in different veterinary clinics in Sweden as part of a clinical study of the CRID complex (Kennelhosta hos hund). RNA was extracted at SVA (Statens Veterinärmedicinska Anstalt) using a Magnatrix 8000+ 96-well extraction robot. Extracted RNA was kept at -80°C until it was used. A total of twelve samples were used initially but two of the samples never gave any PCR-products so for some of the PCR-amplifications only ten of the samples were used (**Table 1**).

Table 1: Samples used in the study. The two samples marked with red never gave any PCR-products.

Sample	Sampling site	Sampling date
1	Helsingborg	April 12th 2013
2	Helsingborg	September 25th 2013
3	Stockholm (Albano)	December 4th 2013
4	Stockholm (Albano)	December 30th 2013
5	Stockholm (Albano)	January 30th 2014
6	Stockholm (Albano)	January 30th 2014
7	Västerås	February 11th 2014
8	Västerås	February 20th 2014
9	Västerås	February 20th 2014
10	Helsingborg	March 7th 2014
11	Härnösand	March 19th 2014
12	Stockholm (Södra djursjukhuset)	December 19th 2014

cDNA synthesis

cDNA was synthesized from 3µl of RNA using the SuperScript™ III Reverse Transcriptase (Invitrogen, 18080-093). The cDNA was made double-stranded using, Klenow Fragment™ 3'->5' exo- (New England Biolabs, M0212S). Both were used according to protocol except for the Klenow Fragment™ being used without the allocated dilution buffer during the synthesis of the first batch. The cDNA was stored in at -20°C and thawed for use. When the first batch ran out, a second one was made using the same protocol except this time 5µl was used for two of the samples to get better yields and the allocated buffer for Klenow Fragment™ was used.

PCR amplification

All PCR-amplifications were run on a ProFlex™ 3 x 32-well PCR System (Life Technologies, 4484073) and all sample volumes were 25µl. Primer sequences are shown in Table 2.

Table 2: Primers used for PCR-amplification. The S1-S4-pairs cover the four fragments of the spike protein gene, the M1-pair covers the membrane protein gene and the HE1-pair covers the hemagglutinin-esterase gene.

Primer	Sequence	Direction (5'-3' is "forward)	Product size	Annealing temp. (°C)	Extension time
S1F ¹	GCTGCATGATGCTTAGACCA	Forward	1067	60	15
S1R ¹	TTAATGGAGAAGGCACCGAC	Reverse			
S2F ¹	AACGGTTACTGTTTCAGCC	Forward	1376	65	15
S2R ¹	TCGATCTACGACTTCGTCTT	Reverse			
S3F ¹	TTCACGACAGCTGCAACCTA	Forward	1107	60	15
S3R ¹	CTGAGCTTGCGCTTCAAGAG	Reverse			
S4F ¹	GCAGCAGCAGGTGTACCATT	Forward	1133	60	15
S4R ¹	GTCGTCATGTAAGGTTTTAATTAC	Reverse			
M1 ²	AGAGTTCCTAGTCATGCTTGGTG	Forward	1311	65	30
M1R ¹	TTTGCTTGGGTTGAGCTCTT	Reverse			
HE1F ¹	CATCACCGGCTAGACTTGAA	Forward	1257	67	15
HE1R ¹	AGACTGCCTGGCATTGTCC	Reverse			

¹ Genetic analysis of canine group 2 coronavirus in Korean dogs, An et al, *Veterinary Microbiology*, 46- 52, 2010

² Isolation and sequence analysis of canine respiratory coronavirus, Erles et al, *Virus Research*, 78-87, 2007

The PCR-reactions were run on different temperature programs depending on annealing temperatures and sometimes based on optimization efforts after failed attempts. The basic temperature program was 95°C for 3 minutes and then 40 cycles of 95°C for 15 seconds, a reaction specific annealing temperature for 15 seconds (Table 1), 72°C for a reaction-specific amount of time and finally 72°C for 1 minute (Table 1). All reactions were run with a positive control, which was cDNA from BCoV, and a negative control, which was nuclease-free water. At first the enzyme Platinum® Taq DNA Polymerase High Fidelity (Life Technologies, 11304-011) was used but it did not give any PCR-products so instead KAPA2G Robust HotStart ReadyMix PCR Kit (KAPA Biosystems, KK5701) was used. It was always used according to protocol but without added MgCl₂ (Protocol: KR0381).

Gel electrophoresis and purification

Gels were made using TBE-buffer (Tris/Borate/EDTA buffer) with between 1% and 1.5% agarose concentration and run at 80V for between 2 and 3 hours depending on whether or not optimal separation was necessary. When loading the gel, 5µl of sample was mixed with 1µl DNA Gel Loading Dye (6X) (Life Technologies, R0611) and added into the well. When gel filtration was necessary, larger wells were made and 20µl sample and 4µl DNA Gel Loading Dye (6X) was pipetted into them. The gels were also allowed to run for 3 hours instead of the normal 2 hours. Bands were cut out using a scalpel and gel purified using a GeneJet Gel Extraction Kit (Life Technologies, K0691).

DNA Sequencing

PCR-products were sent to MacroGen Inc. in Amsterdam for sequencing. Before anything was done with the sequences, ends were trimmed with CodonCode Aligner using the quality information that came with the finished sequences from the company. The end-clipping was done with 0.8 error rate instead of the default 1.0. This stricter parameter was chosen because the default only led to severe mismatches in the extended overlap regions on the spike protein contig and it led to several sequences not being recognized by the program.

Phylogenetic analysis

All phylogenetic trees were generated using MUSCLE alignment with the settings Whole Alignment and MUSCLE default mode. Sequences were retrieved from the nucleotide database at NCBI (for nucleotide sequences) and the amino acid sequence database also at NCBI (for amino acid sequences). The alignments were then turned into trees using the PHYLIP neighbour joining algorithm with 1000x Bootstrap in the build-in tool in the Ugene software (Unipro).

Results

Six of the samples plus the bovine coronavirus positive control were successfully PCR amplified for all six fragments, two of the samples for all but one fragment, one for all but two fragments and the last sample only had two fragments sequenced. Some sequences were of such low quality that they were deemed unusable when the assembly was done. Five samples had all six fragments sequenced with good enough overlap in the spike protein gene. Four out of these five samples were used in the phylogenetic analyses of the individual genes (the fifth had some questionable quality in one overlap). The other five samples were used wherever it was possible to use them, as for instance when there was a fragment for the hemagglutinin-esterase gene or the membrane protein gene since these only consisted of one fragment anyways and overlap was not necessary to generate a full gene for comparison.

Gel electrophoresis of samples

The gel electrophoreses had varying results depending on temperature programs and other optimization efforts. Some gels had clear bands with almost no smearing and others had obvious contaminations. An example of a good gel can be seen in **figure 3** where most lanes have strong, defined bands and no background smearing.

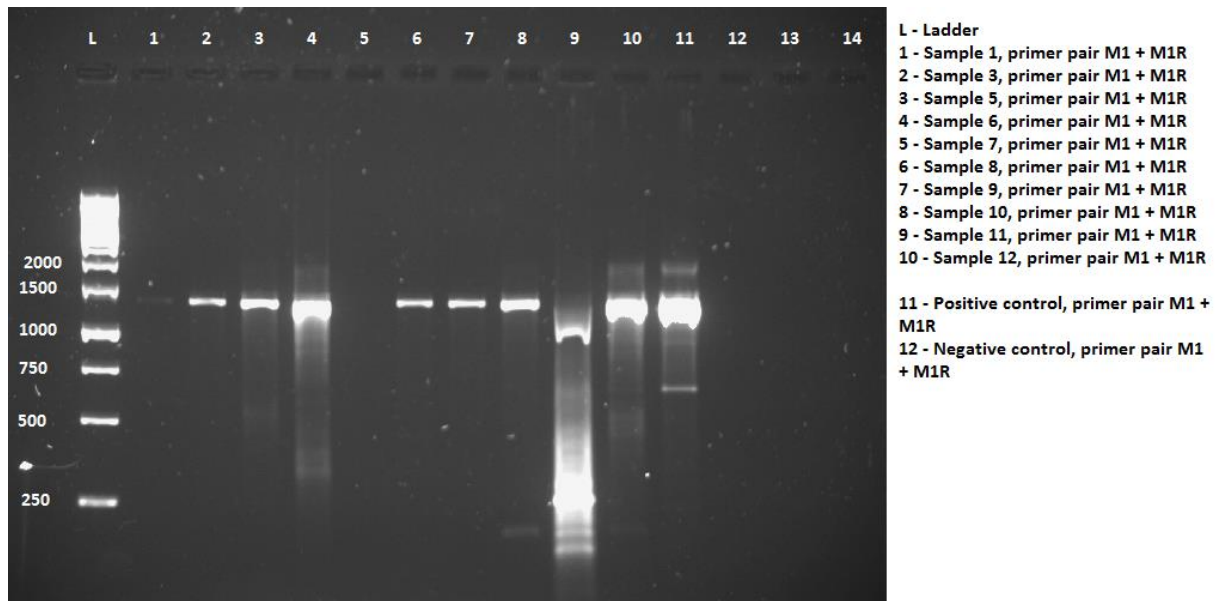


Figure 3: An example of a gel electrophoresis that went well. The gel had an agarose concentration of 1% and was run for 150 minutes. The target fragment is 1312 base pairs long and the DNA ladder was a GeneRuler™ 1kb DNA Ladder.

Single nucleotide polymorphisms (SNPs)

All sequences showed some SNPs compared to the reference CRCoV genome used, isolated from a Korean dog (Accession: JX860640.1). A clear majority of the SNPs were cytosine-to-thymine mutations or thymine-to-cytosine mutations, so called transition mutations, and there were more mutations in the hemagglutinin-esterase gene and the membrane protein genes than in the spike protein genes (**Table 3**) (**Table 4**). At one of the junctions in the spike protein sequence, between fragments two and three, there were a lot of mutations but they are presumed to be due to the length of the sequences and bad overlap and were thus not counted. The spike protein has two subunits and there was no difference between them in terms of how many SNPs they had per base pair.

Table 3: List of SNPs found in the spike protein gene. Only mutations that could be confirmed by looking at the quality readings from the Sanger sequencing have been included. Where it says “all of the samples”, this means that the SNP existed in all the Swedish field isolates but not in the CRCoV reference genome used (Accession: JX860640.1).

Position in the genome	Mutation	Comments
25241	Transition (T to C)	All of the samples
25243	Transition (T to C)	All of the samples
26589	Transition (C to T)	All of the samples
27085	Transversion (C to A)	All of the samples
27468	Transversion (C to A)	All of the samples
27555	Transversion (T to A)	All of the samples
27572	Transition (T to C)	All of the samples
28505	Transversion (T to A)	All of the samples
29078	Transition (G to A)	All of the samples

Table 4: List of SNPs found in the hemagglutinin-esterase gene. Where it says “all of the samples”, this means that the SNP existed in all the Swedish field isolates but not in the CRCoV reference genome used (Accession: JX860640.1). The last hundred bases had a lot of irregularities and the number of mutations were too many to count.

Position in the genome	Mutation	Comments
22832	Transition (C to T)	On all samples
22955	Transition (C to T)	On all samples
23035	Transition (C to T)	Only on one of the samples
23053	Transversion (T to G)	Only on one of the samples
23070	Deletion	Only on one of the samples (8)
23080	Transversion (C to G)	Only on one of the samples
Last hundred bases	???	A lot of irregularities

Phylogenetic analysis of Swedish CRCoV field isolates and similar betacoronaviruses

The first phylogenetic analysis was performed to elucidate the relationship between the field isolates examined during the experimental part of this project and other, similar sequences. It had previously been reported that the two closest relatives to CRCoV were bovine coronavirus (BCoV) and the human coronavirus OC43 so a few sequences of these viruses were included (Fig 4).



Figure 4: Phylogenetic analysis of the spike protein gene of four of the samples as well as some related sequences. The samples from the Swedish dogs were closest to one another than to any other sequence, which is to be expected. The closest relative seems to be an isolate from a Japanese dog.

A second phylogenetic analysis was made using other members of the betacoronavirus subfamily, showing that the samples were most closely related to the other canine respiratory coronavirus sequence and then the bovine coronavirus, the human coronavirus OC43, the equine coronavirus and the murine hepatitis virus in that order. These are all closely related betacoronaviruses and as a control of the veracity of the tree, one more distantly related coronavirus, the MERS-CoV (Middle-East Respiratory Syndrome Coronavirus) was included. It was the least related to CRCoV (Fig 5). All the known related sequences are bundled together, for instance all the Swedish bovine coronavirus sequences are bundled together and all the human coronavirus OC43 sequences were bundled together.



Figure 5: Phylogenetic analysis of the spike protein gene sequence from four of the samples and from other betacoronaviruses. Sample 1,2,3 and 5 are the four of the five spike protein sequences with the best quality.

Lastly, a phylogenetic analysis of all ten samples collected from different veterinary clinics around Sweden was made in order to see if it was possible to elucidate any information about virus transmission over time and sampling sites (Fig. 6).

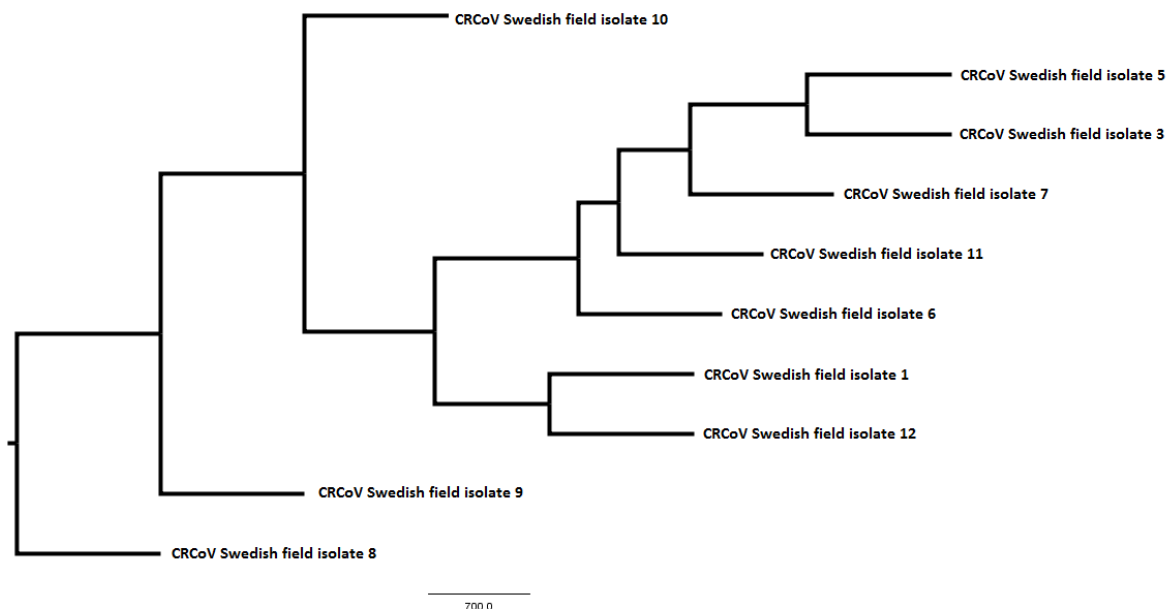


Figure 6: Phylogenetic tree of all the field isolates examined in the study.

A phylogenetic analysis of the hemagglutinin-esterase gene was also performed using the sequences from the Swedish CRCoV isolates where everything but the hemagglutinin-esterase gene had been truncated out. It was run with four hemagglutinin-esterase gene sequences found on GenBank (**Fig. 7**).

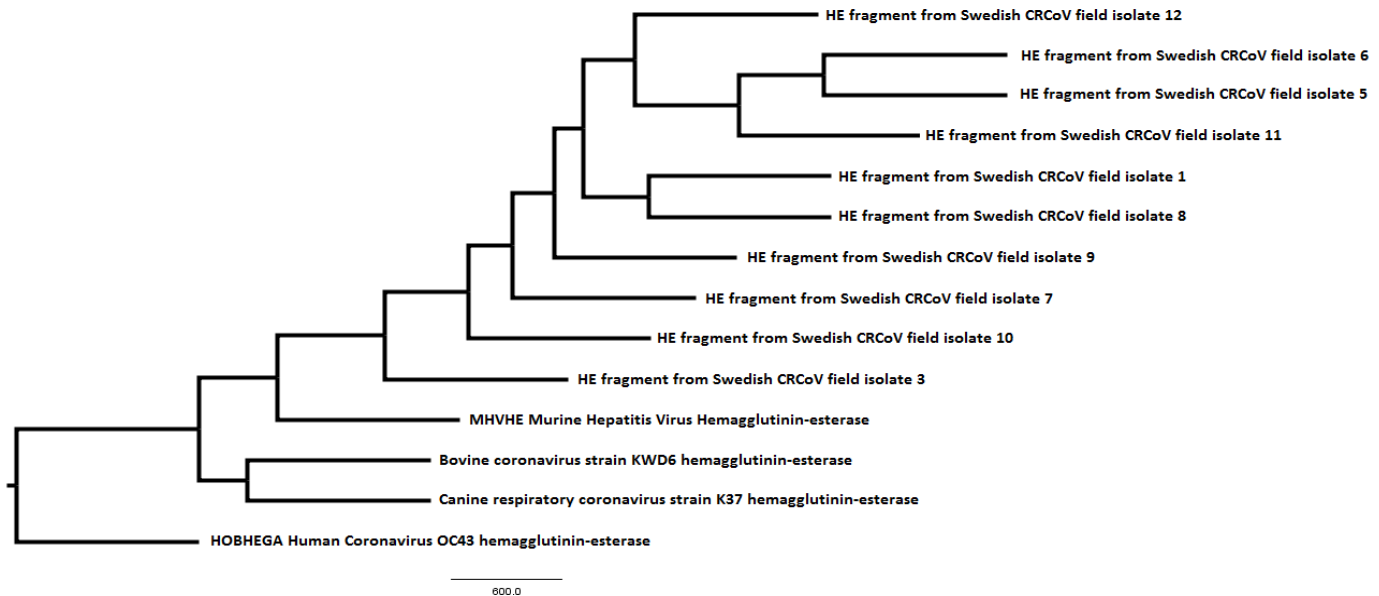


Figure 7: Phylogenetic tree of the parts of the hemagglutinin-esterase gene from the Swedish field isolates and different GenBank-entries of hemagglutinin-esterase genes in the betacoronavirus family.

Discussion

There were some initial doubts whether or not the bovine coronavirus and the human coronavirus OC43 were the closest relatives to the CRCoV due to some differences in sequence between a known CRCoV genome and some BCoV/OC43 sequences. However, since all the primers that worked for the CRCoV samples also worked for the BCoV positive control sample this doubt might have been unfounded. The phylogenetic trees also seem to show that both BCoV and human coronavirus OC43 were very close to CRCoV. Most of the phylogenetic trees are in accordance with what is previously known about the canine respiratory coronavirus and the betacoronavirus family. As expected the sequences of the Swedish CRCoV field isolates were closer to each other than they were to the other sequences used in the analyses.

One interesting aspect is that the closest related non-Swedish sequence was from an isolate taken from a Japanese dog. It might be the case that these sequences are so similar to each other that random chance made it so that one is closer but since all the other sequences seem to be so neatly grouped, this seems unlikely. It is not known whether this resemblance is due to a Japanese dog visiting Sweden, a Swedish dog visiting Japan or perhaps even more likely; that Swedish pet importers and Japanese pet importers ordered dogs from the same source and both got the infection from the same source.

SNPs

On first inspection of tables 3 and 4 it might seem that there are roughly the same number of SNPs on the hemagglutinin-esterase gene as there is on the spike protein gene but the spike protein gene is four times as large as the hemagglutinin-esterase gene. The fact that there were more SNPs in the hemagglutinin-esterase gene than in the spike protein gene is very interesting and a couple of reasons for this will be proposed. First, it is important to think about what an SNP really represents. If the mutation rate of any given length of DNA or RNA is considered constant, given the same polymerase fidelity and repair or lack thereof, the deciding factor on the number of mutations that show up in a population should be which mutations stay in the genome. Not all mutations will confer an evolutionary advantage to the organism but the ones that do will be conserved. Mutations that do not confer any advantage but also are not a disadvantage to the organism could presumably remain too. When one virus crosses from one species to another, such as when some ancestral bovine coronavirus infected a dog as the theory suggests with the canine respiratory coronavirus (Ellis et al, 2005), the virus will most likely be less adapted to the new species and there will thus be a higher selection pressure. This should imply that the rate of SNPs that remain, that is mutations that confer an advantage and those that confer neither an advantage nor a disadvantage, should increase (Fig. 8)

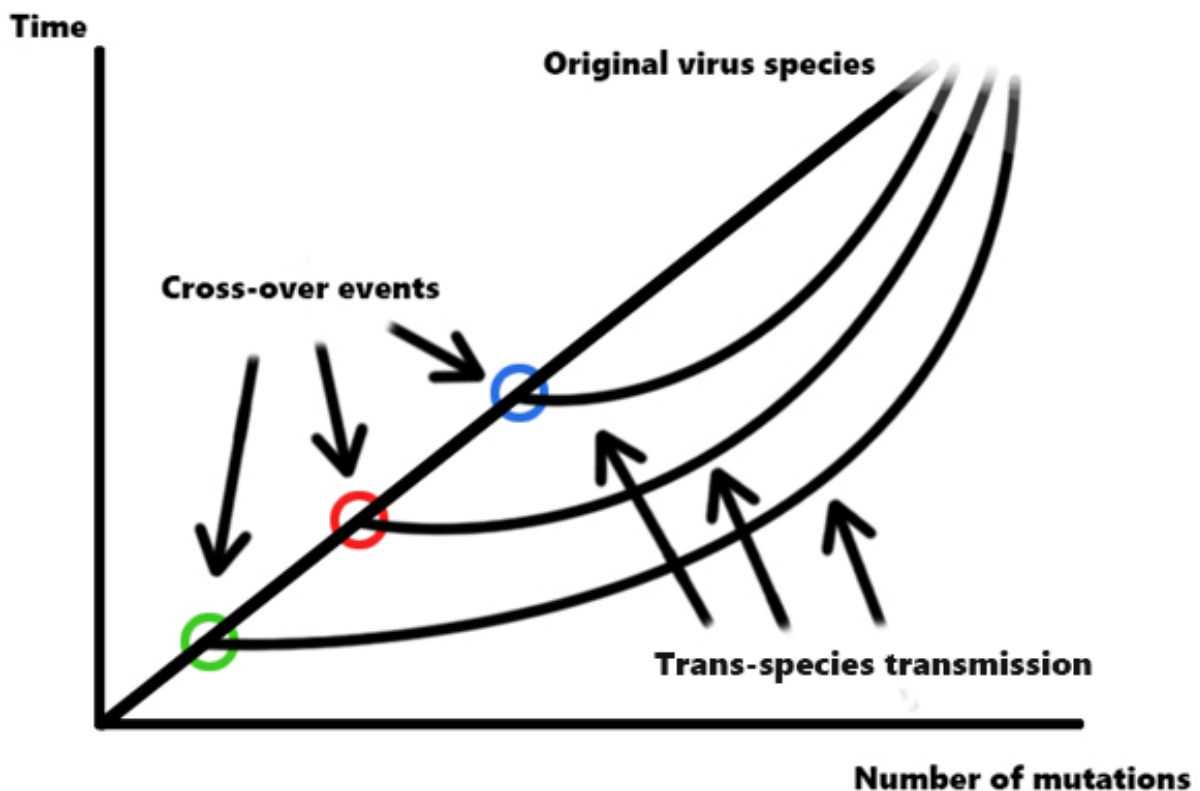


Figure 8: The number of accumulated mutations in a population plotted against time. The coloured circles point out where cross-over events take place and directly after these events a separate virus strain is born where the mutations per unit of time, that is the mutation rate, is higher. In the graph the inverse of the slope is the rate of mutations.

The mutation rate of the new virus species depends on the similarity between the two hosts, such as codon preferences and structure of the receptor protein. Some proteins will be under higher or lower adaptive pressure depending on to which degree their specific function is impaired by the introduction to the new host. This is one possible explanation why some proteins accumulate more mutations than others.

Another explanation could be that since the hemagglutinin-esterase is not vital for the viral replication cycle (Kazi et al, 2005), more mutations are tolerated with the virus still maintaining biological viability. In other words, random mutations that are not detrimental to the viral life cycle are more common and since these mutations sometimes stick, the mutation can become more common in the viral population.

A third explanation is derived from the very intricate structure-function relationship found in the spike protein. The spike proteins exist as trimers on the viral surface with protein-protein interactions keeping the different monomers together which could potentially require highly specific amino acid interactions. When the protein is activated, a very intricate rearrangement of the head part of the spike is what leads to the eventual viral entry and this is probably also governed by intricate interactions on the amino acid level. All this could mean that there is simply less tolerance for point mutations in the spike protein before the virus loses its viability. The spike protein also consists of two major subunits, S1 and S2, where S1 covers roughly the first two thirds of the gene and S2 covers the latter third. Since only five complete sequences of the spike protein were generated, the data is a little too weak to support anything definitive but there is seemingly no difference in the number of mutations in the two subunits. Only one of the sequences had a polymorphism that differed from the other four and it was in the S1 subunit.

The membrane protein has been shown to vary more in its sequence than all other proteins in the betacoronavirus subfamily. This could be for many different reasons, one of which is that it has a very small task in the viral assembly which is to bring the spike protein to the budding site and that this interaction might not require a lot of very specific structural motifs.

Phylogenetic tree of all the contigs from samples

Luckily, the phylogenetic analyses seem to confirm that most of the laboratory work was at least reasonably successful (**Fig. 6**). Samples 8 and 9, which are at the bottom and close to one another, were collected from the same clinic and from two dogs that are friends. It is likely that one of those dogs contracted the infection from the other and that should mean that the viruses should be closely related, and indeed they are. Samples 3 and 5, which seem to be the two most closely related samples, were both collected at the Albano animal clinic, which makes epidemiological sense because that should mean that the dogs lived in close proximity to each other. However, sample 7 was collected at the same animal clinic as samples 8 and 9 but seems to be quite dissimilar from them. The owner of the dog from which sample 7 was taken told the veterinarian that many other dogs lived in the area where he or she lived when asked about a possible source of contamination. If this dog and the dogs that produced samples 8 and 9 do not live in the same part of the city or were not in contact with each other for some other reason, this could mean that there are two very similar

but different strains circulating in that city. It is also interesting that samples 1 and 12 are very similar despite being taken with 20 months in between them and in different cities. In both cases there were no apparent crossed paths and the dogs in question do not partake in any similar activities that could explain how they could be infected by the same strain. Since such a long time had passed between the two sampling occasions, it is strange that there is not at least some genetic drift that causes difference between the two sequences. The tree from the hemagglutinin-esterase sequences showed a completely different result but in a way, the information that it provides could explain the strangeness of the first one. If the difference between the samples is so small that background mutations are what determines the phylogeny in these trees that would explain the seemingly random positioning.

Conclusion

All the Swedish field isolates seem to be similar to one another and no difference over time or across geography seems to be discernible. The few mutations that were found were either communal, that is they existed in all Swedish samples, or they were easily explained through errors of methodology or could be considered background mutations rather than mutations to achieve adaptation. There were more mutations in the hemagglutinin-esterase than there were in the spike protein roughly by a factor of three which could be explained by the dependence of the intricate structure of the spike protein for the virus to function. Just like previous research has suggested, bovine coronavirus seems to be the closest relative of CRCoV, followed by the human coronavirus OC43.

Acknowledgements

This study was funded by Agria Animal Insurances and the Swedish Kennel Club Research Foundation, Jan Skogborgs Foundation, Sveland Foundation for Animal Health and Welfare and MSD Animal Health. RNA extraction was done by Anna-Karin Theelke at SVA (Statens Veterinärmedicinska Anstalt). Supervision was in part paid for via the project "Formas (221-2012-586.)".

References

- Acheson, Nicholas H. 2007. *Fundamentals of Molecular Virology*. Hoboken, NJ: Wiley Press.
- Appel, Max J. G. 1987. "Canine infectious tracheobronchitis short review: kennel cough." In *Virus Infections of Carnivore*, by Max J. G. Appel, 201-211. Amsterdam: Elsevier Science Publishing.
- B.S. Shultz, S. Kurz, K. Weber, H.-J. Balzer, K. Hartmann. 2014. "Detection of respiratory viruses and Bordetella Bronchiseptica in dogs with acute respiratory tract infections." *The Veterinary Journal* 365-369.
- Babcock, G., Eshshaki, D., Thomas, W., Ambrosino, D. 2004. "Amino Acids 270 to 510 of the Severe Acute Respiratory Syndrome Coronavirus Spike Protein Are Required for Interaction with Receptor." *Journal of Virology* 4552-4560.
- Bermingham, A. 2012. "Severe respiratory illness caused by a novel coronavirus, in a patient transferred to the United Kingdom from the Middle East, September 2012." *Euro Surveillance* 17.
- Brownlie, Joe. 2004. "Longitudinal study of Viruses Associated with Canine Infectious Respiratory Disease." *Journal of clinical microbiology* 4524-4529.
- Buonavoglia, Canio. 2006. "Canine Respiratory Viruses." *Veterinary Research* 355-373.
- Chen, H., Gill, A. Dove, B., Emmett, S. R., Kemp, C. F., Ritchie, M. A., Dee, M., Hiscox, J. A. 2005. "Mass spectroscopic characterization of the coronavirus infectius bronchitis virus nucleocapsid and elucidation of the role of phosphorylation in RNA binding by using surface plasmon resonance." *Journal of Virology* 1164-1179.
- Coarse, E., Machamer, C. E. 2000. "Infectious bronchitis virus E protein is targeted to the Golgi complex and directs release of virus-like particles." *Journal of Virology* 4319-4326.
- Cornelis A.M. de Haan, Haukeline Voldets, Cheri A. Koetznet, Paul S. Masters, Peter J. M. Rottier. 2002. "Coronavirus Maintain Viability despite Dramatic Rearrangements of the Strictly Conserved Genome Organization." *Journal of Virology* 12491-12502.
- Cornelissen, L.A., Wierda, C. M., van der Meer, F. J., Herrewegh, A. A., Horzinek, M. C., Egberink, H. F., de Groot, R. J. 1997. "Hemagglutinin-esterase, a novel structural protein of torovirus." *Journal of Virology* 5277-5286.
- Corse, E., Machamer, C. E. 2002. "The cytoplasmic tail of infectious bronchitis virus E protein directs Golgi targeting." *Journal of Virology* 1273-1284.
- Dong-Jun An, Wooseog Jeong, Sook-Hee Yoon, Hye-Young Jeoung, Hyun-Jeong Kim, Bong-Kyun Park. 2010. "Genetic Analysis of Canine Type 2 coronavirus in Korean dogs." *Veterinary Microbiology* 46-52.
- Erles, K. 2007. "Isolation and sequence analysis of canine respiratory coronavirus." *Virus Research* 669-702.
- Erles, K. 2003. "Detection of a group 2 coronavirus in dogs with canine infectious respiratory disease." *Virology* 216-223.

- Fischer, F., Stegen, C. F., Masters, P. S., Samsonoff, W. A. 1992. "Analysis of constructed E gene mutants of mouse hepatitis virus confirms a pivotal role for E protein in coronavirus assembly." *Journal of Virology* 7885-7894.
- Gerard, G.F., D'Alessio, J.M, Driftmier, K.M., Blodgett, K.P., and Noon, M.C. 1986. "Cloning and overexpression of Moloney murine leukemia virus reverse transcriptase in Escherichia Coli." *DNA* 249-258.
- Hiscox, J. A., Wurm, T., Wilson, L., Britton, P., Cavanagh, D., Brooks, G. 2001. "The coronavirus infectious bronchitis virus nucleoprotein localizes in the nucleolus." *Journal of Virology* 506-512.
- Hofmann, M. A., Sethna, P. B., Brian, D. A. 1990. "Bovine coronavirus mRNA replication continues throughout persistent infection in cell culture." *Journal of Virology* 259-263.
- J.A. Ellis, N. McLean, R. Hupaelo, D.M. Haines. 2005. "Detection of coronavirus in cases of tracheobronchitis in dogs: A retrospective study from 1971 to 2003." *Canadian Veterinary Journal* 447-448.
- K. Erles, C. Toomey, H.W. Brooks, J. Brownlie. 2003. "Detection of a group 2 coronavirus in dogs with canine infectious respiratory disease." *Virology* 216-223.
- Kazi, L., Lissenberg, A., Watson, R., de Groot, R.J, Weiss, S. R. 2005. "Expression of hemagglutinin-esterase protein from recombinant mouse hepatitis virus enhances neurovirulence." *Journal of Virology* 15064-15073.
- King, A., Adams, M.J., Carstens, E.B. and Lefkowitz, E.J. 2012. *Virus Taxonomy: Classification and Nomenclature of viruses: Ninth Report of the International Committee on Taxonomy of Virus*. San Diego: Elsevier Academic Press.
- Krijnse Locker, J., Rose, J. K., Griffiths, G., Horzinek, M. C., Rottier, P. J. M. 1992. "O-glycosylation of the coronavirus M-protein. Differential localization of sialyltransferases in N- and O-linked glycosylation." *Journal of Biological Chemistry* 14094-14101.
- Krokhin, O., Li, Y., Andonov, A., Feldmann, H., Flick, R., Jones, S., Stroehel, U., Bastien, N., Dasuri, K. V., Cheng, K, Simonsen, J. N., Perreault, H. 2003. "Mass spectrometric characterization of proteins from the SARS virus: A preliminary report." *Molecular Cell Proteomics* 346-356.
- Lai, M. M. C., Stohlman, S. A. 1981. "Comparative analysis of RNA genomes of mouse hepatitis viruses." *Journal of Virology* 661-670.
- Laude, H., Masters, P. S. 1995. *The Coronaviridae*. New York: Plenum.
- Lissenberg, A., Vrolijk, M. M., van Vliet, A. L. W., Langereis, M.A., de Groot-Mijnes, J. D. F., Rottier, P. J. M., de Groot, R. J. 2005. "Luxury at a cost? Recombinant mouse hepatitis viruses expressing the accessory hemagglutinin-esterase protein display reduced fitness in vitro." *Journal of Virology* 15054-15063.
- Makino, S., Yokomori, K., Lai, M. M. C. 1990. "Analysis of efficiently packaged defective interfering RNAs of murine coronavirus: Localization of a possible RNA packaging signal." *Journal of Virology* 6045-6041.
- Masters, Paul S. 2006. "The Molecular Biology of Coronaviruses." *Advances in Virus Research, Vol 66* 193-292.

- Mazumder, R., Iyer, L., Vasudevan, S., Aravind, L. 2002. "Detection of novel members, structure-function analysis and evolutionary classification of the 2H phosphoesterase superfamily." *Nucleic Acids Research* 5229-5243.
- Opstelten, D. J. E., Raamsman, M. J. B., Wolfs, K., Horzinek, M.C., Rottier, P. J. M. 1995. "Envelope glycoprotein interactions in coronavirus assembly." *Journal of Cell Biology* 339-349.
- Priestnall, Simon. 2006. "Serological prevalence of canine respiratory coronavirus." *Veterinary Microbiology* 43-53.
- S. Priestnall, J. Brownlie, E.J. Dubovi, K. Erles. 2006. "Serological prevalence of canine respiratory coronavirus." *Veterinary Microbiology* 43-55.
- Schulz, B.A., Kurz, S., Weber, K., Balzer, H. J., Hartmann, K. 2014. "Detection of respiratory viruses and Bordetella Bronchiseptica in dogs with acute respiratory tract infections." *The Veterinary Journal* 365-369.
- Seong-in Lim, Sarah Choi, Ji-Ae Lim, Hye-Young Jeoung, Jae-Young Song, R.C. Dela Pena. 2013. "Complete Genome Analysis of Canine Respiratory Coronavirus." *Genome Announcements* 1-3.
- Siddell, S. G., Barthel, A., Ter Meulen, V. 1981. "Coronavirus JHM: A virion-associated protein kinase." *Journal of General Virology* 235-243.
- Sturman, L. S. 1977. "Characterization of a coronavirus: I. Structural proteins: Effect of preparative conditions on the migration of protein in polyacrylamide gel." *Virology* 637-649.
- WHO. 2015. *WHO.int*. 27th April. http://www.who.int/csr/sars/country/table2004_04_21/en/.
- Wilbur, S. M., Nelson, G. W., Lai, M. M. C., McMillan, M., Stohlman, S. A. 1986. "Phosphorylation of the mouse hepatitis virus nucleocapsid protein." *Biochemistry, Biophysics, Residual communications* 7-12 Erratum 884.
- 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 cell division." *Journal of Virology* 9345-9356.
- Yamada, Y. K., Yabe, M., Ohtsuki, T., Taguchi, F. 2000. "Unique N-linked glycosylation of murine coronavirus (MHV-2) membrane protein at the conserved O-linked glycosylation site." *Virus Research*. 149-154.