# 1 Introduction

## 1.1 *Coronaviridae*

### 1.1.1 Characteristics and taxonomy

*Coronaviridae* belong together with the *Arteriviridae* and *Roniviridae* to the order *Nidovirales*. The family of the *Coronaviridae* comprises two genera, on the one hand the actual Coronaviruses (CoV) and on the other hand the Toroviruses. Until recently, coronaviruses were divided into three groups according to their molecular and serological characteristics. According to current genetic analyses in this thesis five different groups are proclaimed (Fig. 1) (Tang et al., 2006).

**Group I**

Group I coronaviruses, to which the long-known Human coronavirus 229E (HCoV-229E) as well as the Porcine Transmissible Gastroenteritis Virus (TGEV) and Porcine Epidemic Diarrhea Virus (PEDV) belong, are responsible for gastroenteritis and lower respiratory tract infections (Table 1). Especially TGEV can have severe impact on agriculture as gastroenteritis causes great morbidity among pigs with high mortality rates of up to 100% among piglets (Enjuanes et al., 1995; Shibata et al., 2000).

Extensive coronavirus research in the past years has led to the discovery of new group I coronaviruses. In March 2004 a novel human pathogenic coronavirus designated HCoV-NL63 was identified (van der Hoek et al., 2004). The virus was isolated from a seven-month-old baby that had bronchiolitis and conjunctivitis. In further investigations, five babies and three adults (two were immune suppressed) could be identified as virus carriers. HCoV-NL63 was detected, in comparison to HCoV-229E and HCoV-OC43, predominantly among children aged 6 and younger with respiratory infections. In a broad study HCoV-NL63 infection could be associated with croup in children (van der Hoek et al., 2005; Pyrc et al., 2007). Moreover it was discovered that HCoV-NL63 infections are spread globally with high prevalence (Hofmann et al., 2005; Vabret et al., 2005; Han et al., 2007). Up to 10% of all respiratory illnesses are caused by HCoV-NL63 (Arden et al., 2005; Bastien et al., 2005a; Bastien et al., 2005b; Chiu et al., 2005; Ebihara et al., 2005; Kaiser et al., 2005; Suzuki et al., 2005; Vabret et al., 2005; van der Hoek et al., 2005; van der Hoek et al., 2006b).

HCoV-229E and HCoV-NL63 are generally responsible for up to 30% of common colds among humans during wintertime, but can also cause complications i.e. severe diarrhea and

pneumonia especially among immune suppressed humans (Tyrrell et al., 1993; Pene et al., 2003; van der Hoek et al., 2006a; Pyrc et al., 2007).

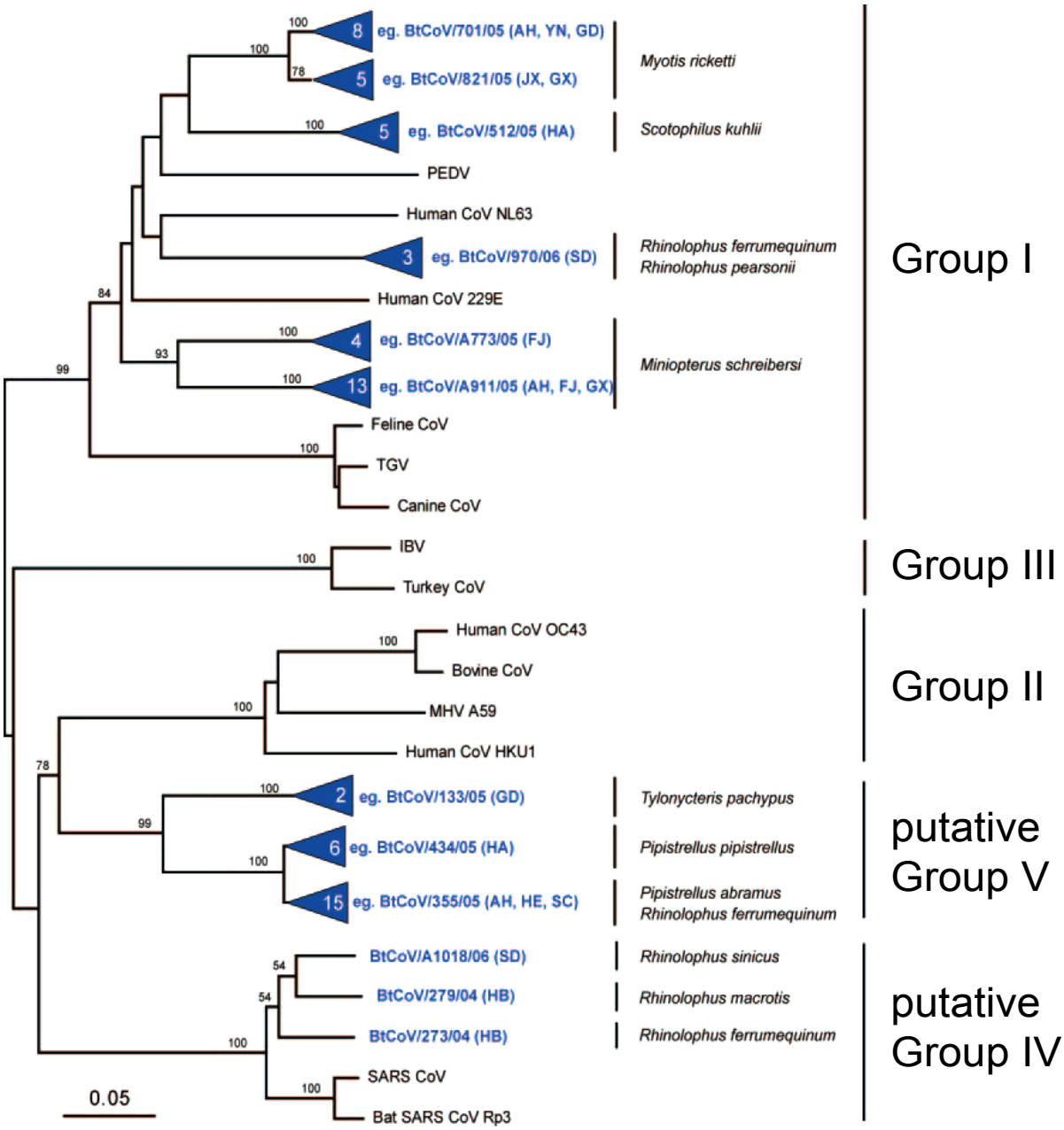


**FIG. 1. Phylogenetic relationships of known human and animal coronaviruses**
The tree was generated based on 440 nucleotides of the RNA-dependent RNA polymerase region by the neighbour-joining method in the MEGA programme. Numbers above branches indicate neighbour-joining bootstrap values (percent) calculated from 1,000 bootstrap replicates. Terminal nodes containing bat coronaviruses isolated in this study are collapsed and represented by a blue triangle with the number of viruses indicated within. The tree was rooted to Breda virus (AY_427798). Scale bar, 0.05 substitutions per site. Abbreviations: AH, Anhui; FJ, Fujian; GD, Guangdong; GX, Guangxi; HA, Hainan; HB, Hubei; HE, Henan; JX, Jiangxi; SC, Sichuan; SD, Shandong; YN, Yunnan. Source: (Tang et al., 2006).

**TABLE 1. Overview of important pathogenic coronaviruses**

| Designation | Group | Disease | References |
|---|---|---|---|
| Human coronavirus (HCoV)-NL63 | I | Common cold, croup | (van der Hoek et al., 2004; van der Hoek et al., 2006a; van der Hoek et al., 2006b) |
| HCoV-229E | I | Common cold, diarrhea, pneumonia | (Tyrrell and Bynoe, 1965; Almeida and Tyrrell, 1967; Bradburne et al., 1967; Tyrrell et al., 1993; Pene et al., 2003) |
| Porcine Transmissible Gastroenteritis Virus (TGEV) | I | Gastroenteritis, LRTI[a] | (Garwes, 1988; Enjuanes et al., 1995; Kim et al., 2000) |
| Porcine Epidemic Diarrhea Virus (PEDV) | I | Gastroenteritis, LRTI[a] | (Kusanagi et al., 1992; Shibata et al., 2000) |
| HCoV-OC43 | II | Common cold, neurological disease, MS[b] associated (?) | (Almeida and Tyrrell, 1967; Bradburne et al., 1967; Hendley et al., 1972; Edwards et al., 2000) |
| HCoV-HKU1 | II | Pneumonia | (Lau et al., 2006; Pyrc et al., 2007) |
| Murine Hepatitis Virus (MHV) | II | Hepatitis, bronchiolitis, encephalitis | (Perlman et al., 1987; Matthews et al., 2002) |
| Avian Infectious Bronchitis Virus (IBV) | III | Bronchiolitis, nephritis | (Ignjatovic and Sapats, 2000; Jia et al., 2002) |
| Human severe acute respiratory syndrome (SARS)-CoV | IV (IIb) | Severe acute respiratory syndrome (ARDS[c], fever, dry cough, dyspnoea, headache and hypoxemia) | (Drosten et al., 2003; Peiris et al., 2003; Rota et al., 2003) |
| Bat-SARS-CoV | IV | Not known | (Lau et al., 2005; Poon et al., 2005; Tang et al., 2006; Woo et al., 2006) |
| Bat-CoV | IV/V | Not known | (Poon et al., 2005; Tang et al., 2006; Woo et al., 2006) |

[a]LRTI: Lower respiratory tract infection
[b]MS: Multiple sclerosis
[c]ARDS: Acute respiratory distress syndrome

### Group II

Other human pathogenic coronaviruses are HCoV-OC43 and recently discovered HCoV-HKU1 that both belong to group II. Similar to HCoV-229E and -NL63 they are responsible

for common colds in humans, but whereas HCoV-OC43 has also been associated with neurological diseases like multiple sclerosis (Edwards et al., 2000) HCoV-HKU1 predominantly causes severe pneumonia (Lau et al., 2006). The Murine Hepatitis Virus (MHV) is also a member of group II, which causes bronchiolitis, infects the liver and brain of mice, and is known for being used as a model organism to study the pathogenesis of coronaviruses (Perlman et al., 1987).

**Group III**

The avian Infectious Bronchitis Virus (IBV), which was the first coronavirus to have been isolated in the 1930s and the Turkey coronavirus are yet the only members of group III (Jia et al., 2002). Today it is one of the foremost causes of economic loss within the poultry industry by causing severe bronchiolitis and nephritis (Cavanagh, 2007).

**Putative group IV**

The most famous putative group IV coronavirus, or distantly related group IIb, was responsible for a global alert and a tremendous challenge for public health authorities. Severe acute respiratory syndrome (SARS) emerged as a newly recognized human disease in the People´s Republic of China late in 2002 and spread globally, causing 8,422 infections with 916 (11%) deaths before being brought under control in 2003. The causative agent was identified as a coronavirus (SARS-CoV), which infected the lower respiratory tract and caused fever, dry cough, dyspnoea, headache and hypoxemia (Chan-Yeung and Xu, 2003; Drosten et al., 2003; Peiris et al., 2003; Pene et al., 2003; Rota et al., 2003). Certainly, the interest for coronaviruses has risen since the SARS-CoV outbreak and details of susceptibility, the cellular receptor, pathogenesis, functions and interactions of different viral proteins have been revealed in the past years. But what distinguishes SARS-CoV from other coronaviruses? Generally, there are significant similarities in properties of structural and non-structural proteins. However, detailed studies have revealed some important differences. The influence of those differences on the severity of clinical symptoms has still to be determined. A special feature of SARS-CoV is the existence of several unique accessory open reading frames (ORFs; 1.1.6) that encode putative proteins with no significant homology to known coronavirus proteins (Tan et al., 2005a; Tan et al., 2006). One of those SARS-CoV genes was thoroughly investigated i.e. ORF3a (1.1.6). The search of a reservoir for SARS-CoV (1.1.2) has led to the discovery of several more coronaviruses, particularly in bats, so that nowadays the existence of an additional putative group V is proclaimed (Tang et al., 2006).

**Putative group V**

In prevalence studies of coronaviruses in bats from China, genetic analysis of an RNA-dependent RNA polymerase gene fragment revealed that bat coronaviruses cluster in three different groups: group I, putative group IV and a novel putative group V (Tang et al., 2006). This group V exclusively contains bat coronaviruses and yet little is known about those bat specific coronaviruses especially as they cannot be propagated in cell cultures. As all of the sampled bats were apparently healthy it seems that persistent or long-term infection without clinical symptoms are likely as already observed for other coronaviruses (Chen and Baric, 1996; Schickli et al., 1997). One member of this group, Bat-CoV/133/05 (BtCoV/133/05), showed special features as it had the longest genome characterized from bats, a large noncoding region at the start of the genome in which a papain-like protease domain was missing and also three ORFs between the spike and the envelope genes instead of one (Tang et al., 2006). What the effects of those features are remain to be resolved.

Coronaviruses, as single-stranded RNA viruses, are fully equipped to adapt rapidly to changing ecological niches as they exhibit high substitution rates i.e. $10^{-4}$ substitutions per year per site (Sanchez et al., 1992; Vijgen et al., 2005). Additionally, homologous recombination has been observed *in vitro* and *in vivo* implying that mutated or recombinant coronavirus variants can emerge anywhere at any time, possibly putting public health at risk as happened in case of SARS-CoV (Makino et al., 1986; Keck et al., 1988; Wang et al., 1993; Jia et al., 1995).

### 1.1.2 Reservoir host of coronaviruses

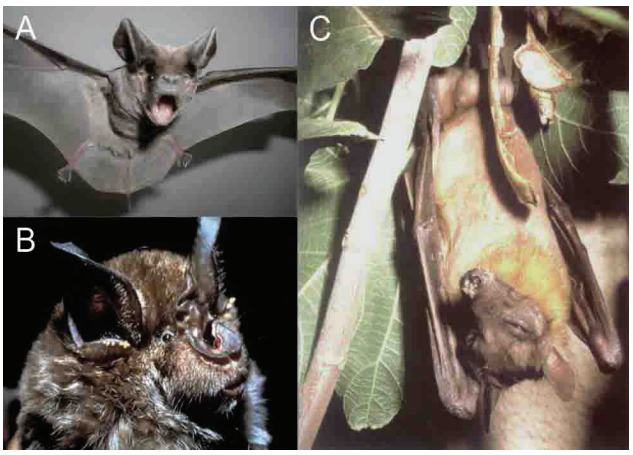


**FIG. 2. Pictures of different bat species**
Shown are the insectivore bats *Mops condylurus* (A) and *Rhinolophus clivosus* (B) as well as the frugivore bat species *Rousettus aegyptiacus* (C).
Sources: www.communitywalk.com; www.mammalogy.org.

Coronaviruses are known to infect rodents, cattle, dogs, pigs, birds and humans, but none of these species seem to be a candidate for hosting coronaviruses as reservoirs (Gorbalenya et al., 2004). After the outbreak of SARS-CoV in 2002 epidemiological studies could show that the earliest cases of SARS were associated with the wildlife meat industry. Investigations trying to find the reservoir of the virus led to the

isolation of related viruses from palm civets (*Paguma larvata*) and raccoon dogs (*Nycereutes procyonoides*) and hog badger (*Arctonyx collaris*) on sale in Chinese food markets. In the marketplace where SARS-CoVs were detected viral RNA from some animals that were seronegative was detected by reverse transcription-PCR, suggesting acute infection, while other animals had antibodies to SARS-CoV but continued to shed virus, suggesting persistent infections (Guan et al., 2003). Although no pathology was associated with SARS-CoV in animals in this market, civets inoculated with human isolates of SARS-CoV had severe lung pathology (Wu et al., 2005). Antibody was found in other species and besides civets a variety of mammals was shown to be susceptible to experimental infection, but none was considered to be natural hosts of the viruses (Guan et al., 2003; Poon et al., 2004).

Further investigation resulted in the identification of SARS-like coronaviruses (SL-CoV) from bats (Fig. 2) in Asia which phylogenetically cluster together with SARS-CoV in a putative group IV including members of several species of Chinese horseshoe bats (suborder *Microchiroptera*, family *Rhinolophidae*, genus *Rhinolophus*) (Lau et al., 2005; Li et al., 2005; Calisher et al., 2006; Tang et al., 2006; Wong et al., 2006; Woo et al., 2006).

The prevalence of antibody to bat SARS-CoV in some species of Chinese horseshoe bats was as high as 84%. Pathology has not been associated with SARS-CoV infection of bats yet. The genomes of SARS-CoV isolates recovered from civets and humans during the 2002 to 2003 outbreak of SARS lay phylogenetically within the broad group of SL-CoV of bats (Li et al., 2005). These data show that the virus responsible for the 2002 outbreak most likely originated from this group of bat-associated viruses. Moreover, antibody against SL-CoV of bats was also detected in *Rousettus leschenaultia*, a cave-dwelling megachiropteran, suggesting that fruit bats also may support infection with SL-CoV. Thus, the natural history of SARS-CoV appears to involve a previously unrecognized SL-CoV of bats being transmitted in markets to amplifying hosts, including masked palm civets, raccoon dogs, and a hog badger, and then spilling over to infect humans in close contact with these intermediate hosts or their tissues. Subsequent human-to-human transmission of the virus was associated with adaptive mutations in the viral genome (Song et al., 2005).

Considering the above mentioned observations with the special feature of this mammal i.e. the high biodiversity, the extraordinary population size, the geographical distribution, migration patterns and the detection of many other emerging viruses in bats, make it reasonable to believe that bats may contain the progenitor of SARS-CoV (Eaton et al., 2005; Leroy et al., 2005; Calisher et al., 2006).

### 1.1.3 Genome structure

A common feature of all coronaviruses is a 27 to 32 kb single-stranded and positive orientated RNA genome making it the biggest known RNA genome. The genome is capped at the 5´end and polyadenylated at the 3´end (Lai, 1990; Brian and Baric, 2005; Masters, 2006).

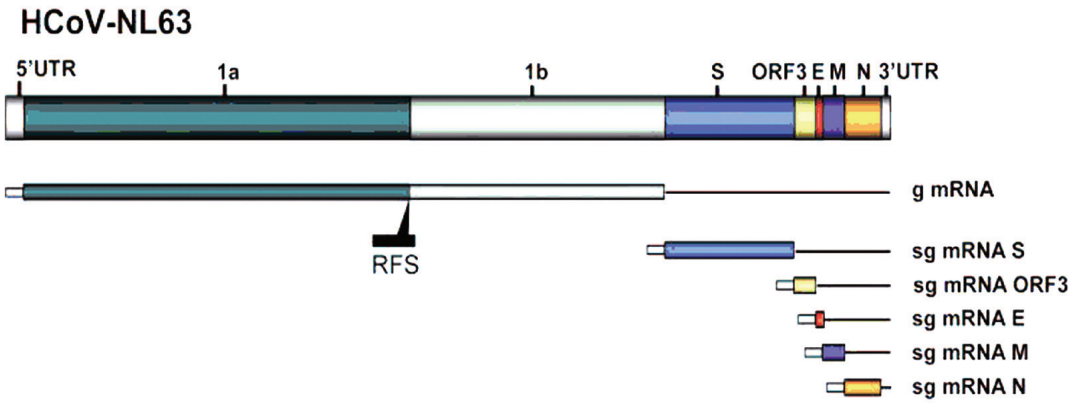


**FIG. 3. Schematic HCoV-NL63 genome organization**
The positive single-stranded RNA genome with all open reading frames (ORFs) as well as the untranslated regions (UTR) of the 5´ and 3´end are shown. Underneath, genomic (g) and subgenomic (sg) mRNAs that are produced by discontinuous transcription during minus-strand synthesis are presented. Only the first ORF of each sg mRNA is translated. The putative ribosomal frameshifting (RFS) element between the 1a and 1b gene is marked. Source: (Pyrc et al., 2007).

The RNA genome has a 5´end and 3´end untranslated region (UTR) and consists of five to six essential genes in the order 1a/b, spike (S), hemagglutinin esterase (HE), envelope (E), membrane (M) and nucleocapsid (N) (Fig. 3).

The two ORFs 1a and 1b comprise two thirds of the genome. Both ORFs share an overlapping region of 40 to 60 nucleotides (nt) and during translation of the polyprotein in 20 to 30% of the cases, a secondary RNA structure pseudoknot initiates a ribosomal frame shift. This causes a failure in stop codon recognition and finally leads to the complete polyprotein 1a/b (approximately 700 kDa). Without this frame shift, only protein 1a is translated. Proteolytic cleavage produces a set of proteins i.e. RNA-dependent RNA polymerase, a helicase, two proteases and a zinc-ion binding protein. Moreover, there is a set of several ORFs encoding non-structural and other structural proteins. Functions are scarcely described and most of the ORFs are non-essential accessory proteins that might have *in vivo* functions (1.1.6).

### 1.1.4 Virus entry and replication

The viral replication cycle starts with the adsorption of the virus at the cell surface with the help of interactions of the viral S protein with a specific cellular receptor (Fig. 4). Different receptors could be identified for coronaviruses i.e. aminopeptidase N for group I coronaviruses TGEV and HCoV-229E (Delmas et al., 1992; Yeager et al., 1992), as well as carcinogenic embryonic antigen (CEA) isoforms for MHV (Dveksler et al., 1991). For the novel SARS-CoV the angiotensin converting enzyme 2 (ACE2), which plays a crucial role in blood pressure regulation, could be identified as the virus receptor (Li et al., 2003). Although HCoV-NL63 is a group I coronavirus that usually apply aminopeptidase N for virus entry into cells, it uses the same cellular receptor as SARS-CoV.

Yet it is still not fully understood how exactly the virus entry takes place. Electron microscopic studies of MHV and IBV infections revealed a possible receptor mediated endocytosis with subsequent fusion of the virus and endosomal membrane (Chasey and Alexander, 1976; Krzystyniak and Dupuy, 1984). Other studies suggest a direct fusion of the virus membrane with the plasma membrane as entry mechanism (Doughri et al., 1976). In case of SARS-CoV the transmembrane S protein binds to ACE2 followed by a conformational change of the S protein by association between two heptad repeat regions (HR1 and HR2). Thereby a six-helix oligomeric complex is formed leading to fusion between the viral and target-cell membrane (Supekar et al., 2004). Accessory proteins like dendritic cell specific ICAM-3-grabbing non-integrin (DC-SIGN) and liver/lymph-node specific (L)-SIGN may also play a role in virus adsorption as both proteins were shown to bind the S protein (Jeffers et al., 2004).

After virus entry, the genomic RNA is directly translated into a polyprotein 1a/b which includes the RNA-dependent RNA polymerase. This polymerase is responsible for the synthesis of minus-stranded RNAs, which serve as templates for the replication of genomic positive-stranded RNA and facilitate the transcription of a nested set of positive-stranded subgenomic (sg) mRNAs (Fig. 3; Fig. 4). The structural features of these mRNAs are unique and particular for the *Nidovirales*. All sg mRNAs have the same 5´end leader sequence consisting of 50 to 90 nt followed by a transcription regulating sequence (TRS) of approximately 7 nt usually a UTAATAA motif. This TRS appears repeatedly in the genomic RNA and is a nicking position for the generation of each sg mRNA. Thus, the first ORF of each sg mRNA that follows the TRS sequence is the one to be translated. The 3´end tail part of each sg mRNA contains the complete genetic information that follows the ORF at the 5´end so that sg mRNAs appear transcriptionally polycistronic. Revealing the mechanism for

the generation of these sg mRNAs has been the aim for many years and nowadays most specialists agree on a discontinuous transcription mechanism during minus-strand synthesis. This means that already during the generation of the genomic minus-stranded template the sg mRNAs are produced by the RNA-dependent RNA polymerase (Jeong and Makino, 1994; Thiel et al., 2001; Masters, 2006).

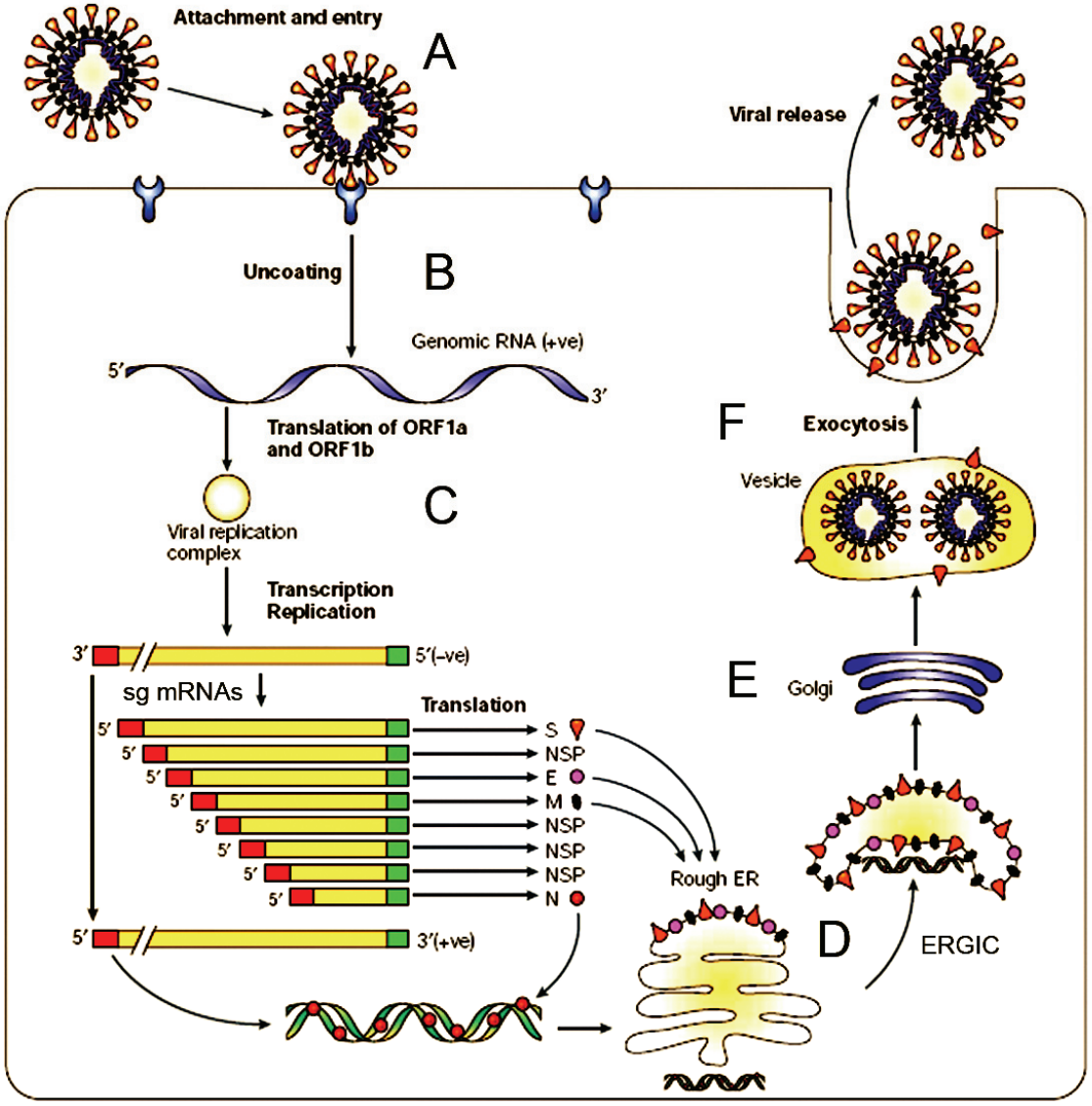


**FIG. 4. The coronavirus life cycle**
After adsorption of viral particles (A) the positive-stranded RNA genome is uncoated (B) and replication starts in the cytosol. A nested set of subgenomic mRNA is produced by discontinuous transcription during minus-strand synthesis (C) resulting in structural (spike (S), envelope (E), membrane (M), nucleocapsid (N)) and non-structural proteins (NSP). Viral proteins assemble and bud at the ER Golgi intermediate compartment (ERGIC; D) and are processed in the Golgi apparatus (E). Newly produced viral particles are incorporated in vesicles and released by exocytosis (F). Source: (Stadler et al., 2003).

A special feature of the HCoV-NL63 genome with its 27,553 nt (Fig. 3), is a relatively low GC content of 34% (usually up to 42%) (Pyrc et al., 2007). One can find conserved TRS core

sequences i.e. AACUAAA throughout the genome responsible for the transcription of five sg mRNA (transcripts encoding S, ORF3, E, M and N). Expression levels of those sg mRNAs increase in 3´end direction of their genomic position. Thus, the sg mRNA for S has the lowest expression level whereas N has the highest level. An exception is the sg mRNA for the E protein which has a lower abundance than the latter sg mRNA for ORF3 protein possibly due to a point mutation in the TRS core sequence. Generally, the E protein is known to have low expression levels in coronaviruses (Pyrc et al., 2004).

### 1.1.5 Structural proteins

As shown in Fig. 5 the RNA genome is associated with the N proteins forming a helical nucleocapsid, which is surrounded by a membrane to form the virion sized 80 to 150 nm (Wang and Zhang, 1999). Although conservation of the N proteins is relatively poor among coronaviruses sequence comparison reveals three domains of which the middle domain has RNA binding potential (Lapps et al., 1987; Parker and Masters, 1990; Nelson and Stohlman, 1993). Within the membrane, there are other structural proteins i.e. the M protein, the E protein and the S protein (Table 2).

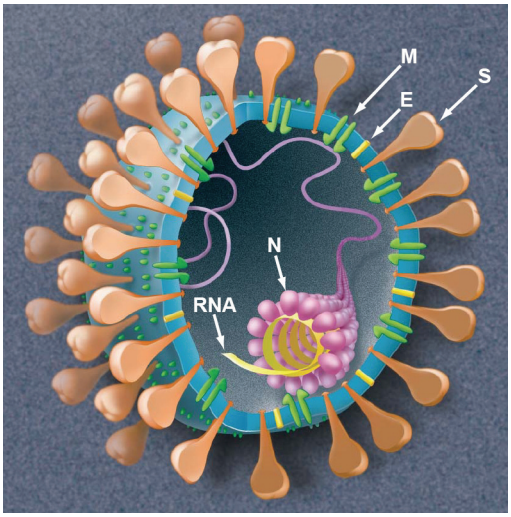


**FIG. 5: The coronavirus virion**
Genomic RNA interacts with the N protein and forms a ribonucleoprotein complex. The virion membrane contains the E, M and S protein. RNA = RNA menome; N = nucleocapsid protein; S = spike protein; M = membrane protein; E = envelope protein. Source: (Holmes and Enjuanes, 2003).

The M protein as a key player of virus assembly and budding is able to interact with the S, E and N protein. It is a triple-spanning membrane protein (20 to 30 kDa) and has a short glycosylated ectodomain and a large carboxyl-terminus (C-terminus) endodomain (de Haan et al., 1999; He et al., 2004b; Voss et al., 2006). This so-called $N_{exo}$-$C_{endo}$ topology was observed for most coronavirus M proteins although in case of TGEV one third of M proteins were shown to have an $N_{exo}$-$C_{exo}$ topology suggesting a quadruple-spanning membrane protein (Escors et al., 2001a). Like for TGEV the M protein of HCoV-NL63 is predicted to have four instead of three transmembrane domains which could mean that the M protein of HCoV-NL63 has also two different conformations but experimental verification is still missing (Pyrc et al., 2005).

The heavily glycosylated S protein has a molecular weight of 150 to 200 kDa and is responsible for the corona-like shape of the virion that can be perceived by electron microscopy. The S protein is essential for virus entry as it interacts with the cellular receptors. Additionally it is presented on the surface of virions and infected cells thereby inducing the production of neutralizing antibodies (Hofmann et al., 2004b; Hofmann and Pohlmann, 2004). During incorporation into the viral envelope and budding, the S protein interacts with the M protein most probably by its transmembrane and C-terminal region (Opstelten et al., 1994; Opstelten et al., 1995; de Haan et al., 1999; Godeke et al., 2000).

**TABLE 2. Characterization of coronavirus structural proteins[a]**

| Designation | Size (kDa) | Motifs | Main function | Localization | Binding partners |
|---|---|---|---|---|---|
| Spike (S) | 150-200 | 1 membrane domain | binding to cellular receptor, mediating fusion | cytoplasm, plasma membrane | ORF3a, M[b] |
| Envelope (E) | ~10 | 1 membrane domain | virus assembly, virus like particles (VLP) with M | ERGIC[c], Golgi | M, ORF3a |
| Membrane (M) | ~26 | 3 membrane domains | virus assembly, VLPs with E | ERGIC[c], Golgi | E, S, N, ORF3a |
| Nucleocapsid (N) | ~42 | self-dimerisation, RNA-binding domain | RNA-binding, induction of apoptosis, actin reorganisation | cytoplasm, nucleolus (also nucleus) | ORF3a, M |
| Open reading frame 3a (ORF3a) | ~26 | 3 membrane domains | S internalisation, potassium ion channel | Golgi, plasma membrane | S, E, M, N |

[a]Data for SARS-CoV; source: (Tan et al., 2005a; Lu et al., 2006; Tan et al., 2006)
[b]Shown for MHV
[c]ERGIC (ER-Golgi intermediate compartment)

Another membrane protein is E (9-12 kDa) which has a low abundance in infected cells. Nonetheless, together with the M protein it plays a crucial role in coronavirus assembly and is responsible for the formation of virus particles (Bos et al., 1996; Vennema et al., 1996). Coronaviral E proteins interact with the M protein by their transmembrane and endodomains and are proposed to act as temporary anchor to relocate M protein into the pre-Golgi compartments (Corse and Machamer, 2002; Corse and Machamer, 2003). In case of SARS-CoV the E protein was moreover found to be a cation-selective ion channel (Wilson et al., 2004).

In some group II coronaviruses, e.g. HCoV-OC43, there is an additional hemagglutinin esterase (HE) protein, which forms dimers and has hemagglutinizing activity (Zhang et al., 1992).

### 1.1.6 Accessory proteins (ORF3 and its homologues)

A variety of genes encoding accessory proteins are interspersed between the structural genes. Their number and location varies within coronavirus genomes. The functions of coronavirus accessory proteins are largely unknown. The emergence of SARS-CoV has led to the characterization of some SARS-CoV specific accessory proteins e.g. ORF3a, 3b, 6, 7a, 7b, 8a, 8b and 9b. Details for thoroughly investigated ORF3a protein are described below. In most cases over-expression of those proteins in cell lines induced apoptosis, stopped cell cycle progression and some accessory proteins were shown to interact with the structural proteins (Tan et al., 2006). Although they could generally be detected in infected cells, deletion of those ORFs in an infectious cDNA clone had no effect on viral replication in cell culture or mouse model experiments (Yount et al., 2005). Reverse genetic analyses of MHV and Feline Infectious Peritonitis Virus (FIPV) also suggest that most accessory proteins are not required for virus replication. Nonetheless, deletion of MHV and FIPV accessory genes results in attenuation in their respective hosts, indicating that accessory genes represent pathogenicity factors and most of them are likely to have an *in vivo* function (Herrewegh et al., 1995; de Haan et al., 2002; Haijema et al., 2004).

Whereas for most coronaviruses structural proteins are thoroughly investigated, only little is known about the ORF3 protein and its homologues. For group I coronaviruses investigations have mainly focused on PEDV and TGEV, that cause severe enteropathogenic diarrhea in swine and incur heavy economic losses (Andries and Pensaert, 1981). It was shown that virulence of those viruses could be reduced by altering the ORF3 gene through cell culture adaptation (Woods, 2001; Song et al., 2003). Interestingly, the ORF3 homologue gene in HCoV-229E named ORF4 undergoes truncation or even splitting into an ORF4a/b by cell culture adaptation. In clinical samples, however, it was shown to be a single conserved ORF4. Although ORF3 seems to be dispensable for cell culture cultivation the attenuation effects of ORF3 deficient strains indicate an important *in vivo* function (Dijkman et al., 2006).
Within group II coronaviruses the MHV accessory proteins have been analyzed by deletion mutants and it could be shown that the ORF3 homologue has no function on virus replication.

Nonetheless it was observed that virus mutants were attenuated in comparison to wild type virus indicating that there might be an *in vivo* function (de Haan et al., 2002).
The highly conserved homologue ORF3a in group III coronavirus IBV could be located in the cytoplasm and putitavely targets a novel domain of the smooth endoplasmatic reticulum. Fractions of ORF3a protein behaved like an integral membrane protein (Pendleton and Machamer, 2005). Like for MHV it was also shown to be dispensable for virus replication but an *in vivo* function cannot be ruled out (Hodgson et al., 2006).
In case of putative group IV coronavirus SARS-CoV the ORF3a protein could be detected in SARS-CoV infected cells and patients showed immune reactivity to ORF3a antigen (Tan et al., 2004a). Additionally the N-terminus ectodomain is able to induce the generation of neutralizing antibodies (Akerstrom et al., 2006). It was shown to be a structural triple-spanning membrane protein with a similar topology as the M protein and it is integrated into virions (Ito et al., 2005; Yuan et al., 2005). Moreover truncated forms were discovered for recombinantly expressed and viral ORF3a which could also be detected in virions (Shen et al., 2005). Unlike the M protein it is not N-glycosylated but O-glycosylated and it was shown to interact with E, M and S protein (Tan et al., 2004b; Nal et al., 2005; Oostra et al., 2006; Voss et al., 2006). It was hypothesized that ORF3a might be able to reduce the expression of S protein on the cell surface (Tan, 2005; Tan et al., 2006).
Subcellular localization of ORF3a could be determined at the Golgi complex, the ER and the plasma membrane where it can be internalized by endocytosis (Tan et al., 2004b; Yuan et al., 2005). Besides, ORF3a was shown to induce apoptosis (Law et al., 2005) and it up-regulates expression of fibrinogen in lung epithelial cells (Tan et al., 2005b). Although small interfering RNAs targeting the ORF3a specific mRNA were able to reduce viral replication (Akerstrom et al., 2007), an ORF3a deletion in an infectious cDNA clone had no effect on viral replication in cell culture and mice (Yount et al., 2005). Recently it was suggested that ORF3a proteins form a homotetramer complex through interprotein disulfide bridges functionally working as a potassium ion channel that modulates virus release and increases membrane permissiveness (Lu et al., 2006).

### 1.1.7 Virus assembly, budding and post-translational modifications

Coronaviruses assemble and bud at membranes of the endoplasmic reticulum (ER)-Golgi intermediate compartment (ERGIC, Fig. 4) (Klumperman et al., 1994; Stertz et al., 2007). The ribonucleoprotein complex buds into the ERGIC where the surface proteins S, M and E need

to be located for virus budding. The N protein interacts with the M and S protein and together with the E protein virions can bud in little smooth long shaped vesicles. Virions accumulate in those vesicles and are released into the extracellular space by exocytosis (Qinfen et al., 2004). While the budding site of several coronaviruses has been localized at the ERGIC, the viral surface proteins can be found in downstream compartments of the secretory pathway. Expressed by the virus or alone the M protein localizes predominantly in the Golgi apparatus (Locker et al., 1995; Escors et al., 2001b) and S is found along the secretory pathway and at the plasma membrane (Opstelten et al., 1995; Lontok et al., 2004), while E is detected in perinuclear regions, the ER and Golgi (Raamsman et al., 2000; Corse and Machamer, 2003; Nal et al., 2005).

Most structural proteins undergo post-translational modification such as phosphorylation and glycosylation. The N protein is known to be phosphorylated in different coronaviruses such as TGEV, MHV and IBV (Wilbur et al., 1986; Calvo et al., 2005; Chen et al., 2005; Jayaram et al., 2005; Shin et al., 2007; White et al., 2007). Phosphorylation results in higher affinity to RNA and influences N protein conformation (Stohlman et al., 1983; Chen et al., 2005). Typically the S and M protein of coronaviruses are glycosylated and it was shown that this glycosylation plays an important role for the generation of their bioactive conformation and influences the fusion activity, receptor binding and antigenic properties of the virus (Alexander and Elder, 1984; Braakman and van Anken, 2000; de Haan et al., 2003; Wissink et al., 2004).

## 1.2 Aims of this thesis

Coronaviruses are, in comparison to other viruses such as Human Immunodeficiency Virus (HIV), relatively scarcely characterized. Although the main features of infection and replication have been investigated in past decades, there are still a lot of open questions. Until 2002 human coronaviruses were known to cause common cold infections and were made responsible for clinical complications among immune suppressed patients. Despite the economic losses that are caused by animal coronaviruses relatively few researchers investigated this field. This changed after the outbreak of SARS-CoV.

The emergence and the identification of a possible reservoir of SARS-CoV was part of the following investigations. By screening African bat species it should be revealed if SARS related coronaviruses are more common in animals and are probably even more widely distributed in the world than expected.

For studying pathogenicity, tissue or organ specificity and replication of viruses it is essential to have cell culture systems for virus propagation. Therefore, the susceptibility of 23 different eukaryotic cell lines for SARS-CoV should be analyzed. Detection of viral nucleic acids should be performed by an established quantitative real-time PCR in order to monitor the progress of viral replication. As a relative quantification by real-time PCR depends on the expression levels of a reference gene it was decided to evaluate the influence of SARS-CoV infection on expression of ten established and widely used reference genes.

Viral proteins of HCoV-NL63 are yet poorly described. Thus, the thesis focuses on the characterization of novel ORF3 protein, a homologous protein to SARS-CoV ORF3a, and the structural proteins M and N. Expression on mRNA and protein level as well as the subcellular localization of those proteins in virus infected and transiently transfected cells should be investigated. To learn more about the features of the ORF3 protein the topology and glycosylation patterns should be analyzed.