CHAPTER EIGHT

Hosts and Sources of Endemic Human Coronaviruses

Victor M. Corman*,†, Doreen Muth*,†, Daniela Niemeyer*, Christian Drosten*,†,1
*Charité–Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Institute of Virology, Berlin, Germany
†German Center for Infection Research (DZIF), Berlin, Germany
1Corresponding author: e-mail address: christian.drosten@charite.de

Contents
1. Introduction 164
   1.1 Endemic Human Coronavirus Disease 165
   1.2 Definitions and Concepts 166
   1.3 Natural Hosts and Zoonotic Sources of Human Coronaviruses 170
2. HCoV-NL63 170
3. HCoV-229E 173
4. HCoV-OC43 175
5. HCoV-HKU1 177
   5.1 Perspectives 178
Acknowledgment 179
References 179

Abstract

The four endemic human coronaviruses HCoV-229E, -NL63, -OC43, and -HKU1 contribute a considerable share of upper and lower respiratory tract infections in adults and children. While their clinical representation resembles that of many other agents of the common cold, their evolutionary histories, and host associations could provide important insights into the natural history of past human pandemics. For two of these viruses, we have strong evidence suggesting an origin in major livestock species while primordial associations for all four viruses may have existed with bats and rodents. HCoV-NL63 and -229E may originate from bat reservoirs as assumed for many other coronaviruses, but HCoV-OC43 and -HKU1 seem more likely to have speciated from rodent-associated viruses. HCoV-OC43 is thought to have emerged from ancestors in domestic animals such as cattle or swine. The bovine coronavirus has been suggested to be a possible ancestor, from which HCoV-OC43 may have emerged in the context of a pandemic recorded historically at the end of the 19th century. New data suggest that HCoV-229E may actually be transferred from dromedary camels similar to Middle East respiratory syndrome (MERS) coronavirus. This scenario provides important ecological parallels to the present prepandemic pattern of host associations of the MERS coronavirus.
1. INTRODUCTION

Coronaviruses (CoV) (order Nidovirales, family Coronaviridae, subfamily Coronavirinae) are enveloped, positive stranded RNA viruses. The subfamily Coronavirinae contains the four genera Alpha-, Beta-, Gamma-, and Deltacoronavirus. Coronaviruses infect birds (gamma- and deltacoronaviruses) and several mammalian species (mainly alpha- and betacoronaviruses), including humans. Animal CoVs, which include important livestock pathogens such as transmissible gastroenteritis virus (TGEV) of swine, bovine CoV (BCoV), and feline coronavirus (FCoV) have been known for more than 80 years (Saif, 2004). Six different CoVs have been identified in humans. The earliest reports of endemic human CoV (HCoV) date back to the 1960s, when HCoV-OC43 and -229E were described (Hamre and Procknow, 1966; McIntosh et al., 1967). HCoV-NL63 and -HKU1 were discovered only in 2004 and 2005, respectively (van der Hoek et al., 2004; Woo et al., 2005). In addition to these four endemic HCoVs, two epidemic CoVs have emerged in humans in the last 2 decades, severe acute respiratory syndrome (SARS)-CoV and the Middle East respiratory syndrome (MERS)-CoV discovered in 2003 and 2012, respectively (Drosten et al., 2003; Zaki et al., 2012). Both viruses belong to the genus Betacoronavirus and were responsible for outbreaks involving high case fatality rates. SARS-CoV was responsible for an outbreak of viral pneumonia in 2002/2003. This outbreak affected at least 8000 individuals and was characterized by a case fatality rate of approximately 10% (Cheng et al., 2007; Drosten et al., 2003). The epidemic lineage of SARS-CoV is believed to have been acquired by humans from carnivorous wild game such as civet cats, which in turn are thought to have acquired the virus from rhinolophid bats (Drexler et al., 2010; Drosten et al., 2003; Ge et al., 2013; Peiris et al., 2003; Poon et al., 2005; Yang et al., 2015).

Since 2004, no human SARS-CoV cases have been reported, but SARS-CoV or closely related viruses carried by bats may still be able to cause human disease after spillover infection (Ge et al., 2012, 2013).

The other highly pathogenic CoV infecting humans—the MERS-CoV—was incidentally discovered in a fatal human case of pneumonia in Saudi Arabia in 2012 (Zaki et al., 2012). A large body of subsequent work suggests that humans regularly and frequently acquire MERS-CoV as a zoonotic infection from dromedary camels, a major livestock species in the Middle East. Dromedary camels across their habitats in Africa, the Middle East, and Asia are now known to be seropositive for MERS-CoV at very high
proportions (Chu et al., 2014, 2015; Corman et al., 2014b; Muller et al., 2014; Reusken et al., 2013; Saqib et al., 2017). Conspecific viruses in coancestral relationship to dromedary-associated viruses were found in bats. However, these bat-associated MERS-related CoVs are far more genetically distant from their human-infecting counterpart than bat-associated viruses related to SARS-CoV (Annan et al., 2013; Corman et al., 2014a; de Groot et al., 2013; Ithete et al., 2013; van Boheemen et al., 2012). Since the discovery of MERS-CoV, more than 2000 human cases were reported. The case fatality rate in hospital-associated clusters ranged around 35% (WHO, 2017). Hospital-based outbreaks are known to have involved up to four consecutive steps of human-to-human transmission before further spread could be halted by intensified measures of infection control (Drosten et al., 2015; Kim et al., 2017; Oboho et al., 2015). Owing to the role of dromedaries as a major livestock species in the Middle East, MERS-CoV represents a serious zoonotic threat involving an unknown epidemic and pandemic potential.

In extension of our knowledge on origins of MERS- and SARS-CoV in bats, it has been proposed that all HCoVs may be of zoonotic origin, and may indeed originate from bats (Drexler et al., 2014; Vijaykrishna et al., 2007; Woo et al., 2009a, b). The common scenario of CoV evolution then involves past transitions into intermediate hosts such as livestock that have closer interaction with humans, and that may carry a diversity of viruses including variants directly related to ancestral strains. Discovering intermediary viruses may enable comparisons between original and current viral characteristics in humans, elucidating the process of human adaptation. However, there is still a gross lack of comprehensive data on the evolutionary history of most HCoVs. Only for HCoV-OC43, which belongs to the species Betacoronavirus 1 (BetaCoV 1), a zoonotic acquisition from ungulate livestock is widely accepted (de Groot et al., 2012a, b; Vijgen et al., 2005, 2006). Recently, a number of studies of CoV in wildlife and livestock have advanced our knowledge of origins of the other HCoVs. In this text, we will provide definitions that are important to correctly describe the process of host transition during emergence of HCoVs, and continue to summarize what is known and thought about the natural histories of emergence of HCoVs.

### 1.1 Endemic Human Coronavirus Disease

The alphacoronaviruses HCoV-NL63 and -229E and the betacoronaviruses HCoV-OC43 and -HKU1 are established human pathogens. They are
responsible for episodes of common cold in humans worldwide (Annan et al., 2016; Graat et al., 2003; Mackay et al., 2012; Owusu et al., 2014; van Elden et al., 2004). Depending on the study setting, up to 20% of tests in individuals with respiratory disease yielded evidence of acute infection with these viruses (Annan et al., 2016; Arden et al., 2005; Bastien et al., 2005; Berkley et al., 2010; Dijkman et al., 2012; Fielding, 2011; Gaunt et al., 2010; Larson et al., 1980; Walsh et al., 2013). HCoV-229E was first isolated in 1967 and shares only 65% nucleotide identity with the other human alphacoronavirus, HCoV-NL63. The latter was first isolated in 2003 from a 7-months-old child suffering from bronchiolitis and conjunctivitis (Hamre and Procknow, 1966; van der Hoek et al., 2004). HCoV-OC43 is already known since the 1960s, whereas HCoV-HKU1 was discovered only in 2005 in a 71-year-old man with pneumonia treated in Hong Kong (McIntosh et al., 1967; Woo et al., 2005).

Although the majority of infections with HCoVs cause only mild respiratory tract illness, all HCoVs can also induce fulminant courses of disease, especially but not exclusively in immunosuppressed patients and infants (Konca et al., 2017; Mayer et al., 2016; Oosterhof et al., 2010; van der Hoek, 2007). Beside the occurrence in the respiratory tract, all endemic CoVs can also be detected in stool samples but they do not seem to be a major cause of gastroenteritis (Esper et al., 2010; Paloniemi et al., 2015; Risku et al., 2010). In particular, HCoV-OC43 has been suspected to play a role in neurological diseases such as chronic demyelinating disease and acute encephalomyelitis (Morfopoulou et al., 2016; Murray et al., 1992; Yeh et al., 2004).

1.2 Definitions and Concepts

1.2.1 Virus Species

The International Committee for the Taxonomy of Viruses (ICTV) endorsed the following definition for a virus species in 1991: “A virus species is a polythetic class of viruses that constitute a replicating lineage and occupy a particular ecological niche”. In spite of this comprehensive definition, the delineation of specific viral species is often not well defined. For CoVs, the ICTV coronavirus study group has suggested a species criterion based on rooted phylogenies and pair wise amino acid distances in seven concatenated domains of the nonstructural part of the CoV genome (de Groot et al., 2013). While the group is presently working on a comprehensive revision of the present CoV classification, we use the above-described classification and current taxonomic virus designations for this overview.
ICTV classifies HCoV-NL63, -229E, and -HKU1 as independent viral species, whereas HCoV-OC43 is part of a virus species named *Betacoronavirus* 1 (BetaCoV 1). Beside HCoV-OC43, BetaCoV1 comprises BCoV and several closely related viruses found in odd- and even-toed ungulates, carnivores, and lagomorphs (Alekseev et al., 2008; Guy et al., 2000; Hasoksuz et al., 2007; Lau et al., 2012; Lim et al., 2013; Majhdi et al., 1997).

### 1.2.2 Application of Niche- vs Genetic Distance Criteria in Species Classification

The ICTV species definition includes a general recognition of the role of habitat in the process of speciation. It also implicates genetic restrictions separating species by referring to a species as “a replicating lineage.” However, only for some virus groups have the genetic basis of speciation—limited fitness of recombinants—been used to systematically delimit species. Habitat or niche separation leads to physical, but not necessarily genetic reproductive isolation. In classical approaches to species definition in animals, populations living in disconnected habitats continue to belong to one same species as long as they can generate fit and fertile progeny. In some viral taxa including the CoVs, genetic distance criteria have been established, using data that at least partly consider the empirical capability of viral lineages to form viable recombinants (de Groot et al., 2012a, b). Genetic species delimitation criteria enable us to discover cases in which viral species exist over several host systems. We can then identify those host species that contain a higher genetic viral diversity, and thereupon derive source attributions in zoonotic or evolutionary scenarios. Adequate comparisons of genetic diversity can only be conducted as long as the compared viruses belong to one same species. This is because the biology of separated species—even if they belong to closely related phylogenetic clades—may be so different that genetic diversity could have developed at considerably different pace. Consequently, information on ecological niche—often a particular host species—should not be used as a leading criterion for the classification of viral species.

### 1.2.3 Natural Hosts

As ICTV’s general view on species involves the concept of niche, we have to reflect on the implications for multihost species of viruses. For the purpose of the present text, we will define a *natural host* as the long-term ecological niche (resembling a habitat) of a viral population or metapopulation, whereas *dead-end host* is a niche in which the maintenance of populations
is routinely unsuccessful (and success may lead to a pandemic). It should be noted that “host” in this context does not necessarily mean a species of animals but can extend over wider taxonomic groups (such as a genus). The host niche can also be restricted by nontaxonomic properties such as geographic range.

Natural hosts are expected to show typical common characteristics related to infection patterns: First, they should contain a higher genetic virus diversity than other host species. Second, they should harbor the virus continuously, at least on the level of social groups. And third, they should be naturally infected beyond the geographic range of present social groups (Drexler et al., 2012; Greger, 2007; Haydon et al., 2002). For some viral species all of these criteria are met, and consensus regarding natural host associations can be reached. For instance, virologists agree on the association of rabies virus with dogs and related carnivores, as well as influenza A virus with several species of waterfowl.

In many studies, the concept of natural host is used with an evolutionary connotation. To discriminate evolutionary concepts from epidemiological concepts, we will hereafter use the term primordial host when we imply animal taxa in (or from which) the virus of interest is thought to have speciated. Primordial hosts are crown group taxa, meaning that they can contain extinct species.

1.2.4 Zoonotic Sources
Several zoonotic spillover infections or epidemics in humans are explained by the involvement of intermediate hosts, creating additional complexity when analyzing multihost species of viruses. In cases in which the virus establishes long-term endemicity in intermediate hosts, intermediate hosts can become natural hosts according to the above definition. The viruses existing in the primordial host and the intermediate host can both belong to the same species, but the viral source population involved in zoonotic transmission may be disconnected from that in the primordial host. For clarity of source implications, we will therefore refer to source species in zoonotic transmission processes as zoonotic sources, irrespective of whether they also fulfill all criteria of natural hosts. A prominent example is rabies virus that has highly diversified conspecifics in bats but is enzootic in dogs to such a degree that dogs are considered zoonotic sources and natural hosts at the same time, but not primordial hosts. The transmission of Nipah and Hendra virus (belonging to the family Paramyxoviridae) from bats to humans involves swine or horses that act as zoonotic sources but are not considered natural
hosts because they are only accidentally infected (Eaton et al., 2006). Intermediary hosts may act as zoonotic sources for dead-end hosts not only because they close gaps of contact between species, but also because they could make the virus transmissible by intermediary adaptation (Caron et al., 2015; Plowright et al., 2015). Intermediary adaptation was strongly suspected for the SARS-CoV whose zoonotic source is in carnivores, in which the virus would have evolved human-compatible receptor tropism (Graham and Baric, 2010; Guan et al., 2003; Song et al., 2005; Wang et al., 2005; Xu et al., 2004). However, viruses directly infecting human cells have later been found in rhinolophid bats, the primordial natural host of the species SARS-related CoV (Drexler et al., 2014; Ge et al., 2013; Yang et al., 2015). Of note, these viruses are conspecific with SARS-CoV but do not fall into the viral clade that was transferred from carnivores to humans and initiated the epidemic. The actual ancestors of the SARS-CoV may continue to exist in the natural reservoir, as new virus variants are continuously formed in the reservoir. If the natural host (e.g., a bat species) harbors different virus populations of the same species, recombination contributes new variants. The proofreading and repair mechanisms typical in CoVs may aid the survival and selection of recombinant viruses in populations (Eckerle et al., 2010; Hanada et al., 2004; Minskaia et al., 2006). Indeed, recombination between the S1 and S2 subunits of the spike gene has been discussed as one of the major mechanism involved in the emergence of human SARS-CoV strains from bat and civet ancestors (Eckerle et al., 2010). On the other hand, ongoing recombination maintains stability of the gene pool of a defined virus species by shuffling sequences and thus limiting divergence processes (Lukashev, 2010).

The ubiquitous occurrence and the vast diversity of bat-associated CoVs have led to the assumption that bats are the primordial hosts of CoVs (Vijaykrishna et al., 2007). Whereas it seems possible that all mammalian CoVs may have originated in bats (Drexler et al., 2014; Woo et al., 2009a, b), this hypothesis should be reevaluated after more complete data on the diversity in other, similarly complex groups of mammalian hosts are available.

### 1.2.5 Tropism Changes During Emergence

After host transition, fidelity-associated mutation and selection in the novel host environment may gradually optimize virus–host interactions maintained from the former host environment, for instance regarding interactions between spike protein and receptor (Eckerle et al., 2010). Provided
that receptor distribution is similar in donor and recipient host, these gradual changes are not expected to cause general changes of disease pattern. Nevertheless, the emergence of coronaviral diseases has also involved drastic changes including switches in tissue tropism by combinations of gene deletion and recombination. For instance, in TEGV of swine, deletions in the spike protein have been associated with changed tropism. TEGV with a full-length spike gene has tropism for both the respiratory and the enteric tracts, whereas strains with deletions (termed PRCV for porcine respiratory coronavirus) mainly replicate in and are transmitted via the respiratory tract (Kim et al., 2000; Sanchez et al., 1992). PRCV outcompeted TGEV in pig populations in spite of its virtual identical antigen composition (Openshaw, 2009). As respiratory viruses do not require direct contact for transmission, they are naturally more contagious irrespective of antigen variation and escape of population immunity. Because of increased prevalence in zoonotic sources, such viruses may then undergo onward host transition.

In humans, infections with CoVs are thought to cause mainly respiratory tract infections, while many (but not all) livestock CoVs cause infections of the gastrointestinal tract. Although many studies examining human intestinal specimens reported that coronaviral RNA can be detected in stool samples (Risku et al., 2010; Zhang et al., 1994), it seems that this detection is most likely explained by the presence of ingested virus particles from the respiratory tract, than resulting from productive replication in intestinal tissue (Jevsnik et al., 2013). It is worth mentioning that viruses in coancestral relationship to human CoVs are well known in bats. As these viruses are almost universally detected in feces, this suggests a primordial tropism for the intestinal tract before emergence as human respiratory pathogens (Smith et al., 2016).

1.3 Natural Hosts and Zoonotic Sources of Human Coronaviruses

To elucidate natural hosts and potential zoonotic sources of HCoVs, it is helpful to review the virus diversity and genome characteristics of related viruses. Fig. 1 provides a schematic overview of animal groups that may have played a role in the evolution and emergence of HCoVs.

2. HCoV-NL63

HCoV-NL63 was discovered by Dutch researchers in the supernatant of tertiary monkey kidney cells used to screen patients with respiratory
disease (van der Hoek et al., 2004). An independent study, also from the Netherlands, found the same virus in blind cell culture isolate that had been stored for many years before final characterization (Fouchier et al., 2004). American researchers identified the same virus, then termed HCoV-NH (for New Haven), by RT-PCR (Esper et al., 2005). The first bat CoVs related to HCoV-NL63 were found in feces of European and African bats belonging to the family Vespertilionidae (Drexler et al., 2010; Gloza-Rausch et al., 2008; Pfefferle et al., 2009). While these viruses are related to HCoV-NL63, they are not conspecific. The same applies to another virus found in

Fig. 1 Summary diagram of the animal groups representing natural hosts and the putative intermediate hosts for the six CoVs found in humans.
the American tricolored bat (*Perimyotis subflavus*, also a vespertilionid species), termed ARCoV.2, which is also not conspecific with HCoV-NL63 by definition (Donaldson et al., 2010). Nevertheless, this study provides functional data that bat cells can support HCoV-NL63 replication by conducting infection experiments in an immortalized lung cell line from the American tricolored bat (Donaldson et al., 2010; Huynh et al., 2012). The cell line was infected with HCoV-NL63 and virus replication was confirmed by the detection of subgenomic RNA and the production of nucleocapsid protein. Foci of infected cells appeared to increase when observed under a fluorescence microscope. Interestingly, the number of infectious particles as determined by plaque assay remained low, which was explained by a potential block in viral egress (Huynh et al., 2012). It was proposed that the functional and genetic findings for American tricolored bats and ARCoV.2 have implications on origins of HCoV-NL63 (Huynh et al., 2012).

More recently, other researchers presented three different sequences of CoVs in *Triaenops affer* (family *Hipposideridae*) from Kenya that are more closely related to HCoV-NL63 (Tao et al., 2017). These bats are not closely related to vespertilionids. The report by Tao et al. included one strain (BtKYNL63-9a, GenBank ACC no. KY073744) that exceeded the 90% amino acid sequence identity threshold applicable for species typing in three of the seven conserved gene domains (Tao et al., 2017). The genome organization of the NL63-related bat CoV was similar to that of HCoV-NL63 (ORF1ab-S-ORF3-E-M-N) with the exception that the bat CoV genomes coded for an additional open reading frame (ORF X) downstream of the N gene. Even if distance-based typing did not allow to formally classify the virus as conspecific with HCoV-NL63, there was evidence for recombination between ancestral lineages of BtKYNL63-9a and bat-associated HCoV-229E-related viruses (Tao et al., 2017). This is important as HCoV-NL63 and -229E form sister species. The recombination processes involved the spike gene via two breakpoints, one located near the gene’s 5’-end and the second around 200 nucleotides upstream of the 3’-end (Tao et al., 2017). Similar recombination breakpoints were also reported for SARS- and SARS-like CoV and HCoV-OC43 (Hon et al., 2008; Lau et al., 2010, 2011). Because the existence of viable recombinants could indicate incomplete speciation, the existence of BtKYNL63-9a should be regarded as a hint toward the primordial host of HCoV-NL63. Its ancestors may well have existed in hipposiderid or rhinolophid bats.

We have no knowledge about zoonotic sources of HCoV-NL63.
3. HCoV-229E

The human respiratory agent HCoV-229E was first described in 1966 (Hamre and Procknow, 1966). It was isolated in cell cultures inoculated with samples from diseased student volunteers. In 2007, sequences related to HCoV-229E were detected in captive Alpacas (*Vicugna pacos*) during a trade fair in California. The single virus strain, termed Alpaca coronavirus (ACoV), was fully sequenced and found to be highly related to known HCoV-229E strains. Epidemiological records suggested that the virus had caused a small outbreak of respiratory disease in exposed alpacas (Crossley et al., 2010, 2012), but it was never observed in feral animals. As no knowledge existed on related viruses other than HCoV-229E at that time, the finding was difficult to interpret as the strain could simply have been acquired from humans.

Around the same time, studies of CoV diversity in bats revealed viruses in close and conspecific relationship to HCoV-229E in hipposiderid bats in Africa (Corman et al., 2015; Pfefferle et al., 2009). Phylogenetic analysis identified lineages closely related to HCoV-229E and Alpaca-CoV, as well as novel sister and cousin groups that altogether showed a much larger viral diversity in hipposiderid bats, than in humans and alpaca together. Epidemiological findings and physiological data suggested long-term endemicity and absence of disease in infected bats, which provides further support for a natural host relationship (Corman et al., 2015; Pfefferle et al., 2009). While humans can come in contact with hipposiderid bats in their natural habitats, the existence of the alpaca virus remained difficult to explain except by human-derived infection; hipposiderid bats and alpaca do not share habitats (Old World vs New World species) (Corman et al., 2015).

In studies on MERS-CoV in dromedary camels, we and others found 229E-related CoVs in dromedary camels in the Arabian Peninsula and Africa (Corman et al., 2016; Sabir et al., 2016). Phylogenetic analyses of complete genomes placed the dromedary-associated viruses in sister relationship to HCoV-229E, while the Alpaca-CoV clustered with dromedary camel viruses. Alpaca is a species in the genus *Vicugna* of the family *Camelidae*. The monophyletic clade formed by human and camelid viruses has an immediate sister lineage within bat-associated viruses (Fig. 2A). Whereas the novel finding does not challenge the assumption of bats acting as natural hosts of HCoV-229E, the virus could have been acquired by humans or camels in either direction of transmission, and could also have been acquired
Fig. 2 Genome characteristics and mutation patterns of HCoV-229E and related viruses from animals. (A) Maximum-likelihood phylogeny of complete genomes of representative HCoV-229E-related coronaviruses from humans, camels, and bats. Filled circles at nodes indicate bootstrap supports of 100% (200 replicates). HCoV-NL63 GenBank (accession no. NC_005831) was used as an outgroup (branch truncated). (B) Genomic organization of 229E-related coronaviruses. Regions with deletions are given in red. (C and D) Deletion patterns in ORF 8 homologs of 229E-related coronaviruses regions with deletions are given in red (numbered I–IV). Asterisks represent triplets that would act as in-frame stop codons; arrows represent possible start codons. Figure was adapted from Corman, V.M., Eckerle, I., Memish, Z.A., Liljander, A.M., Dijkman, R., Jonsdottir, H., Juma Ngeiywa, K.J., Kamau, E., Younan, M., Al Masri, M., Assiri, A., Gluecks, I., Musa, B.E., Meyer, B., Muller, M.A., Hilali, M., Bornstein, S., Wernery, U., Thiel, V., Jores, J., Drexler, J.F., Drosten, C. 2016. Link of a ubiquitous human coronavirus to dromedary camels, Proc. Natl. Acad. Sci. U. S. A., 113, 9864–9869.
by both from an unknown source including bats. However, more detailed analyses suggest HCoV-229E to have evolved toward the human genotype in camelids, thereby identifying the latter as a zoonotic source of human infection, and creating a scenario of emergence that is congruent with the present prepandemic pattern of MERS-CoV infection (Corman et al., 2015, 2016).

In particular, the spike genes of all HCoV-229E and camelid-associated 229E viruses contain deletions as compared to bat-associated viruses (Fig. 2B). Moreover, all 229E viruses from camels and all strains from bats showed an additional gene (termed ORF 8) downstream of the nucleocapsid gene (Fig. 2C). HCoV-229E has retained the transcription regulatory sequence upstream of a remnant ORF 8 (Corman et al., 2016). Interestingly, a process of gradual deletion of ORF 8 sequence can be traced from the camelid-associated viruses via early human isolates to contemporary human strains, defining hallmarks in genome evolution that provide additional topological information to the phylogenetic tree (Fig. 2D).

The deletions in the spike gene of bat-associated 229E viruses could have played a role in host switching by enabling an altered organ tropism in recipient hosts (Corman et al., 2015). This is because camelid- and human strains, in contrast to bat-associated viruses, replicate primarily in the respiratory tract (Corman et al., 2016; Crossley et al., 2010). Loss of spike gene content could have occurred by recombination, as recombination breakpoints were identified in bat-associated viruses within the ORF1ab gene upstream of the spike gene, as well as at the border of the S1 and S2 domains. This is consistent with previous evidence of major recombination events which in part lead to changes in tropism or host association in other CoVs (Corman et al., 2014a; Eckerle et al., 2010; Sabir et al., 2016; Tao et al., 2017).

It should be mentioned that the identification of a wide diversity of 229E-related viruses in dromedary camels provides an interesting possibility to explain the occurrence of Alpaca-CoV; dromedaries are often displayed in alpaca trade fairs and are sometimes kept along with alpacas in husbandries and zoological gardens. Routine testing of dromedaries for 229E-related viruses may be an option to prevent similar outbreaks.

4. **HCoV-OC43**

The human endemic pathogen HCoV-OC43 is part of the virus species BetaCoV1. It was first isolated in the 1960s based on human tracheal explants kept in organ culture, hence the “OC” in the virus name. Other
Bona fide human CoVs were isolated the same way but were not followed up upon and forgotten (McIntosh, 2005; McIntosh et al., 1967; Tyrrell and Bynoe, 1965).

BetaCoV 1 differs from many other CoV species in that it comprises strains detected in highly divergent host species. The bovine CoV (BCoV) is the best-studied representative, whereas essentially the same virus is present in other ungulates whose common names are often identified in virus designations (e.g., Giraffe-CoV; Hasoksuz et al., 2007). The recently described camel CoV HKU23 is a BCoV variant previously known to exist in camels (Corman et al., 2016; Reusken et al., 2013; Sabir et al., 2016; Woo et al., 2016; Wunschmann et al., 2002). All of these viruses belong to BetaCoV1, whose hosts comprise primates, lagomorphs, artiodactyls, and perissodactyls, as well as carnivores (Alekseev et al., 2008; Drexler et al., 2014; Erles et al., 2003; Guy et al., 2000; Hasoksuz et al., 2007; Lau et al., 2012; Lim et al., 2013; Majhdi et al., 1997; Tsunemitsu et al., 1995; Woo et al., 2014) (Fig. 3).

Fig. 3 Maximum-likelihood phylogeny of complete genomes of representatives of Betacoronavirus 1, Mouse Hepatitis virus (MHV), and HCoV-HKU1 strains from humans and animals. Virus designations include GenBank accession numbers and host information. ICTV species are given to the right of clade designations. Filled circles at nodes indicate bootstrap supports of 100% (200 replicates). SARS-CoV (accession no. NC_004718) was used as an outgroup (branch truncated). PHEV; porcine hemagglutinating encephalomyelitis virus.
The emergence of HCoV-OC43 in humans was proposed to be linked to a host-switching event around the year 1890, a time that coincides with a pandemic of respiratory disease recorded in humans (Vijgen et al., 2005, 2006). Whether the zoonotic sources for this host transition were cattle, pigs, or other animals are not definitely resolved. However, the high diversity of related BetaCoV1 strains recently described in livestock species support their role as zoonotic sources in the emergence of HCoV-OC43 in humans. Why BetaCoV1 strains are found in many host species is not understood, but like for other CoVs recombination and deletion events likely played a role in these host-transition or adaptation processes. A deletion downstream of the spike gene of HCoV-OC43 was proposed to reflect adaptation to humans (Vijgen et al., 2005, 2006). Recombination within the species seems to promote the emergence of variants. For instance, Lu et al. recently described a novel canine respiratory coronavirus strain, which likely resulted from recombination between a Chinese canine respiratory coronavirus and BCoV (Lu et al., 2017).

Unlike other alpha- and betacoronavirus species, BetaCoV1 and the related mouse hepatitis virus (MHV) and HCoV-HKU1 do not seem to have ancestral links to bats. Several recent descriptions of novel BetaCoV1 members or relatives in rodents suggest them as primordial hosts (Hu et al., 2017; Lau et al., 2015; Wang et al., 2015). HCoV-OC43 was found to have tropism for neural cells in vitro and in experimentally infected mice, resembling MHV (Arbour et al., 1999; Jacomy et al., 2006). Presence of virus in the central nervous system was reported in patients with chronic demyelinating disease and acute encephalomyelitis (Morfopoulou et al., 2016; Murray et al., 1992; Yeh et al., 2004). The mechanisms that enable neurotropism are not fully understood, but once again mutations within the spike gene may play a role (St-Jean et al., 2004).

There is a single study reporting the isolation of BCoV from a 6-year-old child with watery diarrhea (Zhang et al., 1994). Although the report does not provide information on the patient’s disease and immune status, it reminds us of the potential of BetaCoV1 members to cause spillover infections upon contact with zoonotic sources.

5. HCoV-HKU1

Human coronavirus HKU1 (for Hong Kong University) was detected in 2004 in a patient with viral pneumonia (Woo et al., 2005). The virus has subsequently been isolated in tissue cultures and is now recognized as a
human respiratory agent (Pyrc et al., 2010; Woo et al., 2006, 2009b). There are no conspecific virus sequences from any other animal species. However, as HCoV-HKU1 is a sister taxon to MHV and rat sialodacryoadenitis virus, and together with these stands in sister relationship to BetaCoV1 with its basal rodent-associated viruses, a primordial association with rodent hosts may be considered (Hu et al., 2017; Lau et al., 2015; Wang et al., 2015; Woo et al., 2005).

5.1 Perspectives

There is a growing trend in the field of molecular epidemiology to conduct studies of host associations guided by taxonomy. Comprehensive studies suggest that the large lineages in the mammalian CoV phylogeny have primordial links to small mammals, focused but not limited to bats (Drexler et al., 2014; Vijaykrishna et al., 2007; Woo et al., 2009a, b). The enormous effort behind field studies can cause biases in the spectrum of potential host associations studied. The increased attention toward bats triggered by the discovery of natural host relationships for the SARS-CoV will continue to influence the current interpretations of CoV host associations (Guan et al., 2003). Already we sense that rodents, the largest group of mammals, seem to be underrepresented in studies of CoV hosts. The important role of rodents is highlighted by the recent findings of rodent CoVs that hint at primordial host relationships for the BetaCoV1 species. Data bias is further exemplified by the very few studies on CoVs in shrews and hedgehogs, belonging to the order Eulipotyphla. The 450 species in the order suggest a smaller genetic diversity as compared to the orders Rodentia and Chiroptera with 2269 and 1241 species, respectively. Nevertheless, surprisingly diverse alpha- and betacoronaviruses were detected in Eulipotyphla, including viruses clustering with important CoV clades such as MERS-CoV (Corman et al., 2014c; Tsoleridis et al., 2016; Wang et al., 2017). Other CoVs in shrews and hedgehogs are genetically equidistant and represent relatively deep lineages in the phylogenetic tree. It seems possible that Eulipotyphla may provide missing links in the evolutionary history of some major CoV groups.

Despite a clear role for intermediate hosts in the transmission of viruses from their natural hosts to humans, we have little information on adaptive processes that contributed by passage through these intermediate hosts. For instance, while it has long been assumed that adaptive changes happened in the SARS-CoV spike protein during passage in civets, there are now
reports suggesting direct infectivity, for humans, of some but not all virus variants existing in the natural host (Eckerle et al., 2010; Guan et al., 2003; Wang et al., 2005). Has the virus evolved in civets to become more compatible to humans? Or was it selected from a wider spectrum of variants that have different fitness upon host switching? Understanding (and eventually predicting) the process of host switching will have to be based on functional comparison of natural host-associated viruses with their epidemic counterparts.

Similarly, the ubiquitous presence of MERS–CoV and the high diversity of HCoV–229E-related strains in camels suggests that these animals may be an important zoonotic source for pandemic CoVs in humans—including the prepandemic MERS–CoV (Anthony et al., 2017; Corman et al., 2014a, 2015, 2016). Understanding the past natural history of endemic HCoVs may help to recognize early signs of emergence from highly exposed hosts such as livestock species. We have ideas regarding the emergence of HCoV–OC43 that can potentially be refined (Vijgen et al., 2006). It seems peculiar that there is a complete absence of host-specific CoVs in great apes and other primates. This absence provides further support to the suspicion that contact with domestic animals may have been essential in human acquisition of most or all endemic CoV. For HCoV–HKU1 and HCoV–NL63, we therefore should study more of the virus diversity in livestock including closely related wildlife species to find footprints of their evolution and emergence. Finally, humans and domestic animals may have carried evolutionary intermediates that explain the emergence of major viral taxa, but that are extinct today.

ACKNOWLEDGMENT
This contribution was supported by a grant from Deutsche Forschungsgemeinschaft (DR 772/7–2).

REFERENCES


characterization of viruses related to the SARS coronavirus from animals in southern China. Science 302, 276–278.


Hosts and Sources of Endemic Human Coronaviruses


