Review

Ecology, evolution and classification of bat coronaviruses in the aftermath of SARS

Jan Felix Drexler⁎, Victor Max Corman, Christian Drosten

Institute of Virology, University of Bonn Medical Centre, Germany

A R T I C L E   I N F O

Article history:
Received 22 July 2013
Revised 30 September 2013
Accepted 21 October 2013
Available online 31 October 2013

Keywords:
Coronaviridae
Alphacoronavirus
Betacoronavirus
Bats
Zoonosis
Taxonomy

A B S T R A C T

In 2002/2003, a novel coronavirus (CoV) caused a pandemic, infecting more than 8000 people, of whom nearly 10% died. This virus, termed severe acute respiratory syndrome-CoV was linked to a zoonotic origin from rhinolophid bats in 2005. Since then, numerous studies have described novel bat CoVs, including close relatives of the newly emerging Middle East respiratory syndrome (MERS)-CoV. In this paper we discuss CoV genomic properties and compare different taxonomic approaches in light of the technical difficulties of obtaining full genomic sequences directly from bat specimens. We first present an overview of the available studies on bat CoVs, with details on their chiropteran hosts, then comparatively analyze the increase in bat CoV studies and novel genomic sequences obtained since the SARS pandemic. We then conduct a comprehensive phylogenetic analysis of the genera Alpha- and Betacoronavirus, to show that bats harbour more CoV diversity than other mammalian hosts and are widely represented in most, but not all parts of the tree of mammalian CoVs. We next discuss preliminary evidence for phylogenetic co-segregation of CoVs and bat hosts encompassing the Betacoronavirus clades b and d, with an emphasis on the sampling bias that exists among bat species and other mammals, then present examples of CoVs infecting different hosts on the one hand and viruses apparently confined to host genera on the other. We also demonstrate a geographic bias within available studies on bat CoVs, and identify a critical lack of information from biodiversity hotspots in Africa, Asia and Latin America. We then present evidence for a zoonotic origin of four of the six known human CoVs (HCoV), three of which likely involved bats, namely SARS-CoV, MERS-CoV and HCoV-229E; compare the available data on CoV pathogenesis in bats to that in other mammalian hosts; and discuss hypotheses on the putative insect origins of CoV ancestors. Finally, we suggest caution with conclusions on the zoonotic potential of bat viruses, based only on genomic sequence data, and emphasize the need to preserve these ecologically highly relevant animals. This paper forms part of a symposium in Antiviral Research on “from SARS to MERS: 10 years of research on highly pathogenic human coronaviruses”.

⁎ Corresponding author. Address: Institute of Virology, University of Bonn Medical Centre, 53127 Bonn, Germany. Tel.: +49 228 287 11697; fax: +49 228 287 19144.
E-mail address: drexler@virology-bonn.de (J.F. Drexler).

© 2013 Elsevier B.V. All rights reserved.

Contents

1. Introduction ............................................................................................................... 46
1.1. Coronavirus genomic organization ................................................................. 46
1.2. Bat coronavirus detection and challenges for their taxonomic classification .................................................. 46
1.3. Increase in coronavirus genomic sequences .................................................. 48
1.4. Coronavirus phylogeny .................................................................................... 48
1.5. Bats as coronavirus hosts worldwide ............................................................. 49
1.6. Association of coronavirus clades with mammalian hosts ............................... 49
1.6.1. Promiscuous versus host-specific coronaviruses ........................................ 49
1.6.2. Evidence for phylogenetic co-segregation of coronaviruses and bat hosts ........................................................................... 49
1.7. Coronavirus host switches from bats ............................................................... 49
1.7.1. Host switches suggested by phylogenetically-related viruses in humans and bats .............................................................. 49
1.7.2. Mechanisms of host switches ................................................................... 51
1.8. Coronavirus pathology in bats compared to other mammalian hosts ................ 52
1. Introduction

In 2002/2003, an outbreak of severe respiratory disease occurred, infecting 8096 people worldwide and killing 774 (9.5%) of them. A novel human coronavirus (CoV) named severe acute respiratory syndrome (SARS)-CoV was identified as the causative agent (Drosten et al., 2003). This virus was much more pathogenic than previously known human coronaviruses (HCoV) mainly causing mild respiratory disease (van der Hoek, 2007). The clinical aspects of the SARS epidemic and research on the SARS-CoV have been reviewed in two other articles in the present series (Cheng et al., 2013; Hilgenfeld and Peiris, 2013).

Animal CoV have been known since the late 1930s, including viruses that are highly pathogenic for livestock, pet and laboratory animals, such as transmissible gastroenteritis virus of swine (TGEV), bovine CoV (BCoV), feline infectious peritonitis virus (FIPV), mouse hepatitis virus (MHV) and infectious bronchitis virus (IBV) (Saif, 2004). The earliest descriptions of two HCoVs termed HCoV-229E and -OC43 were already made in the 1960s (Hamre and Procknow, 1966; McIntosh et al., 1967). In the SARS aftermath, two additional HCoV termed -NL63 and -HKU1 were detected (van der Hoek et al., 2004; Woo et al., 2005). In 2012, a highly pathogenic sixth HCoV, termed Middle East respiratory syndrome (MERS)-CoV emerged in the Arabian peninsula (de Groot et al., 2013; Zaki et al., 2012; Hilgenfeld and Peiris, 2013).

During the SARS epidemic, first hints pointed to a zoonotic origin of the SARS-CoV, with civets as the suspected source of human infection (Guan et al., 2003; Song et al., 2005; Xu et al., 2004). Genetically diversified CoVs related to SARS-CoV were then found in Chinese rhinolophid bats, indicating these animals may constitute the animal reservoir of this novel HCoV (Lau et al., 2005; Li et al., 2005). These findings and the concomitant description of Ebola viruses in African flying foxes (Leroy et al., 2005) triggered research into bats as hosts of emerging pathogens. Among the factors promoting bats as animal reservoirs for mammalian viruses are their longevity, densely packed colonies, close social interaction and their ability to fly (Calisher et al., 2006; Luis et al., 2013). The numerous descriptions of novel viruses in bats and other animals have drastically changed our perception of the relevance of animal reservoirs for the understanding of emerging zoonoses (Drosten, 2013; Karesh et al., 2012; Morse et al., 2012). In analogy to SARS-CoV, the novel MERS-CoV may share a putative origin in bats (Annan et al., 2013). This review focuses on bats and the CoVs they host.

1.1. Coronavirus genomic organization

CoVs belong to the order Nidovirales, family Coronaviridae and subfamily Coronavirusae, comprising four genera termed Alpha-, Beta-, Gamma- and Deltacoronavirus (Adams and Carstens, 2012; Perlman and Netland, 2009). Betacoronaviruses are further separated into clades a–d, while the separation between alphacoronavirus clades a and b has been discontinued (de Groot et al., 2012). CoV genomes are composed of one continuous RNA strand of positive polarity, ranging between approximately 27 and 32,000 nucleotides, constituting the largest continuous RNA genomes among mammalian viruses (Woo et al., 2009). Starting from the 5′-end of these genomes, approximate two-thirds encodes a large open reading frame (ORF) 1, producing up to 16 nonstructural proteins (nsp). A conserved sequence UUUAAAC, located around genome positions 12–14,000 in alpha- and betacoronaviruses, provides the characteristic ribosomal slippage leading to transcription of ORF1ab (Brian and Baric, 2005; Masters, 2006). The functions of individual nsp are only partially understood.

Known ORF1ab gene products include, among others, a papain-like protease (PLpro) and main protease (Mpro), a helicase, two methytransferases, a RNA-dependent RNA polymerase (RdRp) and several innate immunity antagonists (Perlman and Netland, 2009). Downstream of the ORF1ab, all CoVs contain genes coding for the structural proteins spike, envelope, membrane and nucleocapsid and several accessory genes in variable number and genomic location (Woo et al., 2009). A unique CoV feature is the discontinuous transcription of genes downstream of the ORF1ab from subgenomic mRNAs, involving interaction between characteristic transcription-regulatory sequences (TRS) found in the most 5′-genomic region (leader TRS) and upstream of individual genes (body TRS) (Brian and Baric, 2005; Pasternak et al., 2006; Perlman and Netland, 2009; Woo et al., 2009).

1.2. Bat coronavirus detection and challenges for their taxonomic classification

Almost all CoV field studies investigating bats or other animals are based on PCR assays targeting parts of the ORF1ab, typically the RdRp. In these studies, PCR amplicon sizes range from as little as 121 to around 404 base-pairs (Anthony et al., 2013a; da Silva Filho et al., 2012; de Souza Luna et al., 2007; Moes et al., 2005; Tong et al., 2009). It should be noted that these small amplicon sizes drastically hinder CoV phylogenetic reconstructions. Therefore, caution should be taken when drawing conclusions on the relationships within the subfamily Coronavirusae, based on such small sequence fragments.

There are only few characterizations of complete CoV genomes from bat feces, as summarized in Table 1. The lack of virus isolates obtained directly from bats challenges the complete genomic characterization of these large and highly variable RNA viruses. The information originating from the four available metagenomic studies reporting CoV sequences offers even less phylogenetic power, because the resulting genomic fragments are very variable in genomic location and length (Donaldson et al., 2010; Ge et al., 2012; Li et al., 2010; Wu et al., 2012b).

To provide order in CoV taxonomy, the current proposal of the International Committee for the Taxonomy of Viruses (ICTV) is based on pairwise amino acid distances in seven concatenated partial or complete nps domains, encompassing about 50% of the CoV genome. In technical analogy to an approach that has been validated for the RNA virus family Picornaviridae (Lauber and Gorbaleny, 2012b), amino acid identity below 90% was found to be discriminatory for the designation of novel CoV species (de Groot et al., 2012). A different criterion, relying on RdRp–grouping units (RGU), also used pairwise amino acid distances, but was restricted to a translated 816 nucleotide RdRp (nsp12) fragment. This was chosen as an amenable approach to account for the partial genomic sequences generated by most PCR-based field studies. It was found that defined alphacoronavirus RGU differed by at least 4.8% and betacoronavirus RGU by at least 6.3% in this RdRp fragment. This allowed a surrogate criterion for species definition in the absence of complete genomic sequences (Drexler et al., 2010;
Tao et al., 2012). However, the increase in partial CoV sequences has caused a loss of discriminatory power of these RGU thresholds. For example, the novel HCoV species MERS-CoV (de Groot et al., 2013; van Boheemen et al., 2012) and the genetically closely related Pipistrellus bat CoVs clearly belong to one RGU (Annan et al., 2013; Reusken et al., 2010). However, these novel bat viruses also differ by only 5.1% from the established CoV species HKU5, implying that they share sufficient RdRp sequence identity to be classifiable within both an RGU defined either by MERS-CoV or by HKU5. Re-analysis of all available RdRp sequence data for the present paper indicated that the 4.8% threshold previously determined for alphacoronaviruses could be maintained. In contrast, a revised lower threshold of at least 5.1% amino acid sequence distance in this RdRp fragment was necessary to accommodate all betacoronavirus species included in the current ICTV proposal (de Groot et al., 2012) and the novel partial CoV sequences.

### Table 1

<table>
<thead>
<tr>
<th>Continent</th>
<th>Bat family yielding CoV sequences</th>
<th>Country</th>
<th>Reference</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>America</td>
<td>Vespertilionidae</td>
<td>Brazil</td>
<td>Brandao et al. (2008)</td>
<td>No full genome information available</td>
</tr>
<tr>
<td></td>
<td>Molossidae</td>
<td>Brazil</td>
<td>Lima et al. (2013)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Phyllostomidae, Molossidae</td>
<td>Brazil</td>
<td>Corman et al. (2013b)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Vespertilionidae</td>
<td>Brazil</td>
<td>Goes et al. (2013)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Vespertilionidae, Mormoroidae</td>
<td>Canada</td>
<td>Misra et al. (2009)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Phyllostomidae, Emballonuridae</td>
<td>Costa Rica</td>
<td>Corman et al. (2013b)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Vespertilionidae, Mormoroidae, Molossidae</td>
<td>Mexico</td>
<td>Goes et al. (2013)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Rhinolophidae</td>
<td>Panama</td>
<td>Corman et al. (2013b)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Vespertilionidae</td>
<td>Trinidad</td>
<td>Carrington et al. (2008)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Vespertilionidae, Molossidae</td>
<td>USA</td>
<td>Donaldson et al. (2010)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Vespertilionidae</td>
<td>USA</td>
<td>Osborne et al. (2011)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Vespertilionidae</td>
<td>USA</td>
<td>Dominguez et al. (2007)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Vespertilionidae</td>
<td>USA</td>
<td>Li et al. (2010)</td>
<td></td>
</tr>
<tr>
<td>Europe</td>
<td>Rhinolophidae, Vespertilionida</td>
<td>Bulgaria</td>
<td>Drexler et al. (2010)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Vespertilionidae</td>
<td>UK</td>
<td>August et al. (2012)</td>
<td>SARS-related CoV; Full genome</td>
</tr>
<tr>
<td></td>
<td>Vespertilionidae</td>
<td>Germany</td>
<td>Gloza-Rausch et al. (2008)</td>
<td>No full genome information available</td>
</tr>
<tr>
<td></td>
<td>Vespertilionidae</td>
<td>Germany</td>
<td>Drexler et al. (2010)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Rhinolophidae</td>
<td>Germany</td>
<td>Drexler et al. (2011)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Rhinolophidae</td>
<td>Italy</td>
<td>Balbani et al. (2011)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Vespertilionidae</td>
<td>Italy</td>
<td>Balbani et al. (2012)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Vespertilionidae</td>
<td>Netherlands</td>
<td>Reusken et al. (2010)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Vespertilionidae</td>
<td>Netherlands</td>
<td>Annan et al. (2013)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Vespertilionidae</td>
<td>Romania</td>
<td>Annan et al. (2013)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Vespertilionidae</td>
<td>Romania</td>
<td>Annan et al. (2013)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Vespertilionidae</td>
<td>Romania</td>
<td>Rihtari et al. (2010)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Vespertilionidae</td>
<td>Spain</td>
<td>Falcon et al. (2011)</td>
<td>No full genome information available</td>
</tr>
<tr>
<td></td>
<td>Vespertilionidae</td>
<td>Ukraine</td>
<td>Annan et al. (2013)</td>
<td></td>
</tr>
<tr>
<td>Africa</td>
<td>Hipposideridae</td>
<td>Ghana</td>
<td>Pfefferle et al. (2009)</td>
<td>No full genome information available</td>
</tr>
<tr>
<td></td>
<td>Nycteridae</td>
<td>Ghana</td>
<td>Annan et al. (2013)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pteropodidae, Hipposideridae, Vespertilionidae, Molossidae</td>
<td>Kenya</td>
<td>Tong et al. (2009)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pteropodidae, Megadermatidae, Vespertilionidae, Molossidae</td>
<td>Kenya</td>
<td>Tao et al. (2012)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Hipposideridae</td>
<td>Nigeria</td>
<td>Quan et al. (2010)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Vespertilionidae, Molossidae</td>
<td>South Africa</td>
<td>Geldenhuys et al. (2013)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Vespertilionidae</td>
<td>South Africa</td>
<td>Ihette et al. (2013)</td>
<td></td>
</tr>
<tr>
<td>Asia</td>
<td>Pteropodidae</td>
<td>Bangladesh</td>
<td>Anthony et al. (2013b)</td>
<td>No full genome information available</td>
</tr>
<tr>
<td></td>
<td>Vespertilionidae</td>
<td>China</td>
<td>Poon et al. (2005)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Vespertilionidae</td>
<td>China</td>
<td>Woo et al. (2006)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Vespertilionidae</td>
<td>China</td>
<td>Chu et al. (2006)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pteropodidae</td>
<td>China</td>
<td>Lau et al. (2010b)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Rhinolophidae</td>
<td>China</td>
<td>Yuen et al. (2012)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Hipposideridae</td>
<td>China</td>
<td>Ge et al. (2012)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Rhinolophidae, Vespertilionida</td>
<td>China</td>
<td>Wu et al. (2012b)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Vespertilionidae</td>
<td>Japan</td>
<td>Shirato et al. (2012)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Vespertilionidae</td>
<td>Saudi Arabia</td>
<td>Memish et al. (2013)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pteropodidae, Vespertilionida</td>
<td>Philippines</td>
<td>Watanabe et al. (2010)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Pteropodidae, Hipposiderida</td>
<td>Philippines</td>
<td>Tsuda et al. (2012)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Hipposideridae</td>
<td>Thailand</td>
<td>Goulah et al. (2011)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Molossidae</td>
<td>Thailand</td>
<td>Wacharapluesadee et al. (2013)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Vespertilionidae</td>
<td>China</td>
<td>Chu et al. (2008)</td>
<td>1A/1B, HKU8; full genome</td>
</tr>
<tr>
<td></td>
<td>Pteropodidae, Hipposiderida</td>
<td>China</td>
<td>Lau et al. (2012)</td>
<td>HKU10; full genomes</td>
</tr>
<tr>
<td></td>
<td>Rhinolophidae</td>
<td>China</td>
<td>Lau et al. (2007)</td>
<td>HKU2; full genome</td>
</tr>
<tr>
<td></td>
<td>Pteropodidae, Rhinolophidae, Vespertilionidae</td>
<td>China</td>
<td>Woo et al. (2007)</td>
<td>HKU4, HKU5, HKU9; full genome</td>
</tr>
<tr>
<td></td>
<td>Rhinolophidae, Vespertilionida</td>
<td>China</td>
<td>Tang et al. (2006)</td>
<td>SARS-related CoV, 512, HKU4; full genome</td>
</tr>
<tr>
<td></td>
<td>Rhinolophidae</td>
<td>China</td>
<td>Lau et al. (2005)</td>
<td>SARS-related CoV; full genome</td>
</tr>
<tr>
<td></td>
<td>Rhinolophidae</td>
<td>China</td>
<td>Li et al. (2005)</td>
<td>SARS-related CoV; full genome</td>
</tr>
<tr>
<td></td>
<td>Rhinolophidae</td>
<td>China</td>
<td>Ren et al. (2006)</td>
<td>SARS-related CoV; full genome</td>
</tr>
<tr>
<td></td>
<td>Rhinolophidae</td>
<td>China</td>
<td>Yuan et al. (2010)</td>
<td>SARS-related CoV; full genome</td>
</tr>
<tr>
<td></td>
<td>Rhinolophidae</td>
<td>China</td>
<td>Lau et al. (2010a)</td>
<td>SARS-related CoV; full genome</td>
</tr>
<tr>
<td></td>
<td>Rhinolophidae, Molossidae</td>
<td>China</td>
<td>Yang et al. (2013)</td>
<td>SARS-related CoV; full genomes</td>
</tr>
</tbody>
</table>
1.3. Increase in coronavirus genomic sequences

Fig. 1A shows the increase in bat CoV sequence deposition into GenBank that closely followed the 2002–3 SARS pandemic. An even bigger increase of overall CoV sequence entries can be observed after 2009. This is likely due to both the advent of next-generation sequencing techniques and changed GenBank policies requiring host information from submitters. Fig. 1B shows that the proportion of bat CoV sequences among the 5489 overall CoV entries containing host information is small, compared to those from hosts such as ungulates, birds and carnivores. Possible explanations for this difference are likely the veterinary relevance of livestock and pet CoVs and the usage of prototype viruses as laboratory models leading to sequence entries, e.g., for BCoV, TGEV, FIPV and IBV. Still, the number of bat CoV entries \( n = 390 \) almost equalled that of HCoVs \( n = 460 \). However, the lack of host information for HCoVs probably introduces a bias in this comparison.

An analysis of published CoV research reports is shown in Fig. 1c. The impact of the SARS-epidemic in 2002–3 on the number of CoV publications was tremendous, leading to a near four-fold increase to around 700 published studies per year after 2002. Similarly, the identification of SARS-related CoVs in bats led to around 15 studies on bat CoVs per year after 2005. Of note, this figure also shows that scientific publications began to decrease a few years after SARS-CoV disappeared. The emergence of the MERS-CoV is likely to counteract this phenomenon in the years to come.

1.4. Coronavirus phylogeny

The rapid increase in bat CoV studies enabled hypotheses of bats as reservoir hosts for alpha and betacoronaviruses only a few years after the SARS epidemic (Vijaykrishna et al., 2007; Woo et al., 2009). Fig. 2 shows the clear separation between the mammalian genera Alpha- and Betacoronavirus and the bird-associated genera Gamma- and Deltacoronavirus (Weiss and Navas-Martin, 2005; Woo et al., 2012) in a Bayesian phylogeny based on an 816 RdRp sequence fragment of the complete subfamily Coronavirus. Fig. 3 provides details of this phylogenetic analysis for the genera Alpha- and Betacoronavirus. The large number of deep branches leading to bat viruses (shown in red) emphasizes the association of these two CoV genera with bat hosts. This is particularly true for the Alphacoronavirus clade formerly designated 1b, which includes HCoV-229E, -NL63 and the Betacoronavirus clades b–d.

The relevance of bat CoVs for these genera is also reflected in the current taxonomic proposal of the ICTV (de Groot et al., 2012). Of the 15 recognized species in these two genera, six were only found in bats. These viruses are shown with bat pictograms.
in Fig. 3 and include Miniopterus bat coronavirus 1, Miniopterus bat coronavirus HKU8, Rhinolophus bat coronavirus HKU2, Scotophilus bat coronavirus 512, Pipistrellus bat coronavirus HKU5, Rousettus bat coronavirus HKU9, and Tylonycteris bat coronavirus HKU4. Many partial bat CoV RdRp sequences were not included in Fig. 3, because only small sequence fragments reducing the phylogenetic resolution were available (Corman et al., 2013b; Drexler et al., 2010). Still, the sequences from these studies (detailed in Table 1) do not alter the overall picture of bat CoV-associated clades within the genera Alpha- and Betacoronavirus.

For most of these bat CoVs, lack of complete genomic sequences prevents their taxonomic designation as species. Still, many of the partial sequences included in Fig. 3 branch deeply in the phylogenetic tree and likely represent not only new species, but even new genetic clades. This is exemplified by the unclassified African Hipposideros betacoronaviruses (Pfefferle et al., 2009; Quan et al., 2010; Tong et al., 2009), which putatively represent a yet to be defined Betacoronavirus clade e and unclassified neotropical Carollia and Pteronotus viruses (Corman et al., 2013b) putatively corresponding to additional Betacoronavirus clades.

Fig. 4 shows that 11 of the 18 extant bat families already contain CoV descriptions, including the two major bat lineages Yinpterochiroptera and Yangochiroptera (Teeling et al., 2005). In most bat families, both alpha- and betacoronaviruses are known, and these detections have originated from both frugivorous and insectivorous bat hosts. Lack of detection in the remaining bat families is likely due to non-exhaustive sampling of the almost 1200 extant bat species (Schiper et al., 2008; Simmons, 2005; Teeling et al., 2005). This void may be filled in future studies.

1.5. Bats as coronavirus hosts worldwide

Fig. 5 shows the geographic origin of all 53 studies characterizing novel bat CoVs and Table 1 provides details for these studies. The figure highlights that studies from all continents are now available, but there is a drastic lack of studies from resource-limited or politically unstable settings. Specifically, several biodiversity hotspots linked to the emergence of zoonotic viruses (Jones et al., 2008) are not covered at all, including the Congo basin, large parts of South-East Asia and the Neotropical ecozone. Future sampling of bats from these other poorly studied areas will likely complete bat species coverage and further increase the known CoV genomic diversity.

1.6. Association of coronavirus clades with mammalian hosts

1.6.1. Promiscuous versus host-specific coronaviruses

Only a small fraction of the currently known mammalian CoVs originates from primate, ungulate, lagomorph, carnivore and rodent hosts. As shown in Fig. 3, bats outnumber any other mammalian host in terms of virus diversity. Throughout the CoV phylogeny, examples can be found of both "promiscuous" and very host-restricted viruses. The paramount example of a promiscuous CoV is probably Betacoronavirus 1 (the species including BCoV, HCoV-OC43 and related viruses), which has been detected in cows, horses, dogs, humans, waterbucks, deer, antelopes, camels and giraffes worldwide (Alkesev et al., 2008; Guy et al., 2000; Hasokszu et al., 2007; Jin et al., 2007; Lim et al., 2013; Majhdi et al., 1997; Zhang et al., 1994). Similarly, FIPV, Canine coronavirus (CCoV) and TGEV are now included in a single species termed Alphacoronavirus 1, and MHV and Rat coronavirus together are now termed Murine coronavirus (de Groot et al., 2012). Another example of an apparently promiscuous CoV is the unclassified bat virus HKU10, which has been detected in the bat families Hipposideridae and Pteropodidae (Lau et al., 2012).

Most other CoVs have been confined to single host genera, exemplified by the detection of SARS-related CoVs and several alphacoronaviruses in Rhinolophus, Myotis, Miniopterus, Nyctalus and Carollia bat hosts, including detections of closely related viruses in individual bats separated by thousands of miles (Corman et al., 2013b; Drexler et al., 2010; Tang et al., 2006). Similarly, Hipposideros betacoronaviruses from Thailand, Kenya, Nigeria and Ghana are closely related (Gouilh et al., 2011; Pfefferle et al., 2009; Quan et al., 2010; Tong et al., 2009) and the betacoronavirus HKU9 has been detected in different species of flying foxes in Africa and Asia (Anthony et al., 2013b; Lau et al., 2010b; Tao et al., 2012; Watanabe et al., 2010; Woo et al., 2007). Of note, the detection of both host-specific and -nonspecific mammalian CoVs parallels what can be observed in the avian Coronavirinae genera. For example, infectious bronchitis virus (IBV, genus Gammacoronavirus) has been detected in a wide range of birds, while the recently described deltacoronaviruses appear to be more host-specific (Chu et al., 2011).

1.6.2. Evidence for phylogenetic co-segregation of coronaviruses and bat hosts

Co-segregation of CoVs and their bat hosts is most visible for HKU9, the Hipposideros betacoronaviruses and the SARS-related CoV, compared to Pteropodidae, Hipposideros and Rhinolophus hosts (Figs. 3 and 4). A striking counter-example is the large number of bat alphacoronaviruses that cluster together with the prototype viruses HCoV-229E, -NL63 and PEDV (Fig. 3). Viruses from numerous bat hosts, together with ungulate and human viruses, are contained in this part of the Alphacoronavirus tree, and in contrast to the betacoronaviruses, the designation of clearly separated subclades is challenging. However, more work needs to be done to formally analyze the degree of phylogenetic co-segregation in the Coronavirinae subfamily.

1.7. Coronavirus host switches from bats

1.7.1. Host switches suggested by phylogenetically-related viruses in humans and bats

The most well-studied CoV bat host switches have probably occurred from bats to humans. The foremost example is the paradigmatic host switch of SARS-CoV from rhinolophid bats into humans.
or potentially civets (Balboni et al., 2011; Drexler et al., 2010; Lau et al., 2010a, 2005; Li et al., 2005; Ren et al., 2006; Rihtaric et al., 2010; Yang et al., 2013). These human, civet and bat viruses are now officially summarized by the ICTV in one species, termed SARS-related coronavirus (de Groot et al., 2012).

The genomic relatedness of human and bat SARS-related coronaviruses is greatest in the ORF1ab, while a bat ancestor containing the structural proteins of human SARS-CoV has so far not been detected. Bat SARS-related coronaviruses fail to interact with the human SARS-CoV receptor molecule ACE2, possibly associated with small deletions in their receptor-binding domain (RBD), compared to human SARS-CoV (Li, 2013; Ren et al., 2008). In line with these differences, a bat SARS-related coronavirus synthesized by reverse genetics was only infectious in cell culture and mice when the spike gene was exchanged by the human SARS-CoV homologue (Becker et al., 2008). Because the RBD of European rhinolophid bat SARS-related coronaviruses was more related to that of the human SARS-CoV than the RBD from Chinese bat viruses (Drexler et al., 2010), recombination may have played a role in the emergence of the human pathogenic virus.

However, not all rhinolophid bat species have been tested for SARS-related coronaviruses. For example, only 12 of the at least 19 rhinolophid bat species that occur in China have been tested and SARS-related coronavirus sequence information is only available from 5 of these species (Lau et al., 2005; Li et al., 2005; Poon et al., 2005; Tang et al., 2006; Woo et al., 2006, 2007; Yang et al., 2013; Yuan et al., 2010). Therefore, further studies of Rhinolophus species in Africa, Europe and Asia may provide more insight into the ancestral bat viruses that were the source of the emergence of human SARS-CoV.

It should also be mentioned that the Hipposideros betacoronaviruses detected in Africa and Asia are clearly distinct from SARS-related coronaviruses. These viruses can be distinguished by both sequence distance-based taxonomic approaches described above. Additionally, the phylogenetic position and genomic properties of the unclassified Hipposideros betacoronaviruses differ from SARS-related coronaviruses. These genomic properties include chiefly their different viral 3'-genome ends and accessory ORFs downstream from the membrane gene in the Hipposideros CoVs (Pfefferle et al., 2009; Quan et al., 2010).

![Fig. 3. Phylogenetic relationships between coronaviruses and bat hosts. Details of the phylogeny shown in Fig. 2 for the genera Alpha- and Betacoronavirus. ICTV species are given to the right of clade designations and bat symbols, when applicable. Virus designations include strain names, GenBank accession numbers and host information as the first three letters of the latin genus and species names. Bat viruses are shown in red. Boxes indicate Alpha- and Betacoronavirus genera, according to the coloring in Fig. 2.](image_url)
a virus termed PML/2011 from a South-African *Neoromicia* bat (Annan et al., 2013; Ithete et al., 2013); CoVs from a Spanish *Hypsugo savii* and *Eptesicus isabellinus* (Falcon et al., 2011); sequences from Thai bat guano (Wacharapluesadee et al., 2013); a CoV from a Mexican *Nyctinomops* bat (Anthony et al., 2013a); and the fully sequenced betacoronavirus clade c prototype bat CoVs HKU4 and HKU5 from China (Woo et al., 2007). Clade c betacoronaviruses from Ghanaian *Nycteris* bats were more distantly related to the MERS-CoV (Annan et al., 2013). A short 203-nt RdRp sequence fragment 100% identical to the MERS-CoV prototype strain EMC/2012 was described in a single Saudi-Arabian *Taphozous perforatus* bat, belonging to the family Emballonuridae (Memish et al., 2013). Although the high sequence identity should have facilitated characterization of other CoV genomic regions, no further virus sequence could be obtained from this specimen.

Detection of closely related CoVs in different bat families is rare, but may occur (Lau et al., 2012; Yang et al., 2013). However, even if closely related MERS-CoV bat ancestors existed in both the distantly related bat families Vespertilionidae and Emballonuridae (Fig. 4), the lack of further CoV sequence information and the single detection of a short genome fragment require further confirmation of this finding. Interestingly, in the same study, a 202-nt sequence from a *Rhinopoma hardwickii* bat was detected which was 100% identical to BCoV and other Betacoronavirus 1 reference strains (Memish et al., 2013). Again, whether clade c betacoronaviruses also exist in bats requires confirmation by detection in more individual bats and more CoV sequence information.

Regarding potential intermediate hosts of MERS-CoV, camels have been shown to have antibodies against this virus at high rates and titers (Perera et al., 2013; Reusken et al., 2013). Conclusions on the passage of putative bat ancestors of MERS-CoV to humans via camels will only be possible upon genomic characterization of the viruses eliciting this strong antibody response.

Finally, no direct bat ancestor of HCoV-NL63 has ever been found, although the phylogenetic clade containing this HCoV is enclosed by bat CoVs. While the recent success in cultivating HCoV-NL63 on immortalized bat cells may hint at some link between bats and this virus (Huynh et al., 2012), the bat viruses described in that and all previous studies are genetically rather distant from HCoV-NL63 (Corman et al., 2013b).

### 1.7.2. Mechanisms of host switches

The exact mechanisms by which CoVs adapt to new hosts are unclear. Clearly, virus entry and innate immune responses are among the paramount obstacles to be overcome during viral host switching. The high degree of conservation of the MERS-CoV receptor Dipeptidyl peptidase 4 in different hosts is thus a worrying scenario (Muller et al., 2012; Raj et al., 2013). Generally, it is unclear whether direct zoonotic transmission from bats to humans has occurred for any HCoV. Alternative scenarios involve intermediate hosts, such as carnivores or ungulates (Annan et al., 2013; Cotten et al., 2013; Enserink, 2013; Graham and Baric, 2010; Lau et al., 2013; Song et al., 2005; Wu et al., 2012a).

One way to rapidly acquire novel genes that potentially facilitate host switching is recombination between different viruses. A canonical example of recombination in CoVs is FIPV type 2. The prototype strains of this feline virus are recombinants between FIPV type 1 and CoCV in different parts of the ORF1b and Spike genes (Herrewegh et al., 1998). Recombination has also been hypothesized to be involved in the emergence of the SARS-CoV (Graham and Baric, 2010; Hon et al., 2008; Lau et al., 2010a; Yuan et al., 2010) and HCoV-OC43 genotypes (Lau et al., 2011). However, the observed recombination events in SARS-CoV and HCoV-OC43 are restricted to genetically closely related viruses. The exact recombination partners giving rise to SARS-CoV have never been detected.
Another phenomenon putatively associated with CoV adaptation to distinct hosts and cells is exemplified by the variable deletions spanning over 600 nucleotides in the globular S1 domain of the Spike gene that is associated with a change from enteric (TGEV) to respiratory tract (Porcine respiratory coronavirus, PRCV) tropism. Whether the 290-nucleotide deletion downstream from the Spike gene in HCoV-OC43, in comparison to its likely ancestral virus BCoV (Vijgen et al., 2006, 2005), represents an adaptation to a human host remains unknown.

1.8. Coronavirus pathology in bats compared to other mammalian hosts

There is little information on the clinical consequences of CoV infections for bat hosts. In humans, HCoVs-NL63, -OC43, -HKU1 and -229E circulate constantly. SARS-CoV disappeared a few years after its epidemic transmission, while the novel MERS-CoV may only be starting to spread (Ensberink, 2013). All HCoV cause primarily respiratory symptoms in humans (Hamre and Procknow, 1966; McIntosh et al., 1967; Peiris et al., 2003; van der Hoek, 2007; van der Hoek et al., 2006; Woo et al., 2005), although the MERS-CoV has also been associated with severe renal complications (Zaki et al., 2012). HCoV shedding in feces is not uncommon (Liu et al., 2004), and there are sporadic reports on CoVs in humans with gastrointestinal, e.g., a case report on BCoV in a pediatric patient leading to the tentative designation Human enteric coronavirus (Zhang et al., 1994). However, CoVs are apparently not generally related to gastrointestinal in humans (Esper et al., 2010).

In comparison, the clinical picture in other animals is considerably more variable, ranging from mild respiratory and gastrointestinal symptoms to systemic disease with hepatitis, multi-organ failure and death (summarized in Saif, 2004). The high CoV concentrations in bat feces (Annan et al., 2013; Drexler et al., 2011), together with the recovery of coronaviral RNA from the small and large intestine of frugivorous bats (Watanabe et al., 2010) point to replication in the enteric tract. This is compatible both with the use of bat fecal material or intestinal specimens to detect bat CoVs in almost all published studies and with the high CoV RNA concentrations in the lower intestine of naturally infected hedgehogs (Corman et al., 2013a). Although there are no apparent clinical signs of gastrointestinal (Watanabe et al., 2010) or any other disease in bats, it should be noted that bats appear to raise antibodies against their coronaviruses at high rates (Lau et al., 2010b, 2005; Muller et al., 2007; Tsuda et al., 2012). Whether this correlates with the severity of infection remains unknown.

From an ecologic point of view, CoVs seem to rely on massive amplification on a population level during bat reproductive seasonal cycles, potentially associated with fecal-oral transmission (Drexler et al., 2011; Gloza-Rausch et al., 2008; Osborne et al., 2011). Similar phenomena have been observed for other bat-associated viruses, such as the filo-, henip-, astro- and lyssaviruses (Amman et al., 2012; Drexler et al., 2011; George et al., 2011; Wacharapluesadee et al., 2010). Still, respiratory or vertical CoV transmission cannot be ruled out at this point, mostly due to the necessity of shielding these protected animals from human interference. Hypothetically, one could question whether insectivorous bats that rely on a functional larynx for echolocation of prey could even tolerate respiratory infection. This highly speculative scenario could point to an evolution of bat (corona-) viruses towards infection outside the respiratory tract. Alternatively, it could represent a sample bias due to the rapid death and decay of bats suffering from any respiratory disease that impairs predation, a likely scenario given the very active metabolism of bats, which for example requires insectivorous bats to consume several grams of insects per night (Encarnação and Dietz, 2006).

2. Discussion

Ten years after the SARS epidemic, there is an overwhelming body of evidence highlighting the relevance of bats for the evolution of mammalian CoVs (Vijaykrishna et al., 2007; Woo et al., 2009). From a taxonomic perspective, there is a clear need for the establishment of reliable criteria for these viruses, such as those established for the family Picornaviridae (Lauber and Gorbalenya, 2012b). The challenges observed in attempts to transfer these distance-based approaches to less well characterized virus families such as the Filoviridae (Lauber and Gorbalenya, 2012a) highlight that the current set of CoV taxonomic criteria will likely have to be optimized and thoroughly validated.

Both the existing ICTV and the RUG-based approaches use a similar methodological background and offer more reliable criteria than phylogeny alone. However, both suffer from several shortcomings. Loss of discriminatory sharpness upon the increment in known CoV genetic diversity will likely limit these approaches, although the ICTV criteria may prove more robust, due to the larger genomic fragments incorporated. From a technical point of view, it should be noted that the extension of the RdRp fragment size used for RUG definitions can be challenging when field specimens contain low RNA concentrations (Corman et al., 2013b; Drexler et al., 2010). A similar difficulty can be observed for the ICTV proposal, which requires the characterization of approximately 50% of a typical CoV genome. The more than 20 unclassified CoVs species in the current ICTV proposal demonstrate that this task is not easily fulfilled for viruses which have not been grown in culture (de Groot et al., 2012). Furthermore, both approaches do not incorporate any sequence distance-based data from genomic regions encoding the structural proteins. Besides the technical constraints on characterizing these genes that were stated above, this is likely also due to the poor sequence alignments some of these genes entail. The high variability, e.g., of the Spike gene, complicates the establishment of criteria applicable to the full CoV genome. Besides the technical constraints on characterizing these genes, this is likely also due to the poor sequence alignments some of these genes entail.

Second, intermediate hosts adapting bat viruses to humans or other mammals have neither been conclusively found for SARS-CoV nor for any other HCoV. Despite the high seroprevalence of Chinese animal handlers against SARS-CoV, it is not entirely clear whether civets indeed adapted SARS-CoV for infection in humans (Graham and Baric, 2010; Guan et al., 2003). Similarly, the high seroprevalence of camels against MERS-CoV suggests that these animals are a potential source of the initial introduction of this virus into the human population (Reusken et al., 2013). However, the current lack of CoV data from NHPs prevents more definite assertions on this topic.

Third, many mammalian orders have not been or are only poorly studied. This is exemplified by the single CoV detected in lagomorphs (rabbit coronavirus HKU14), which likely belongs to the Betacoronavirus 1 species, and by Marine Coronavirus (the former MHV), the only known rodent CoV species. Whether CoVs exist in other mammalian hosts, and to what extent these viruses...
contribute to our understanding of CoV evolution, for now remain obscure.

Finally, the recent description of distantly related nidoviruses in insects has enabled hypotheses on a putative insect origin of *Coronaviridae* ancestors (Nga et al., 2011; Zirkel et al., 2011). Interestingly, the number of CoV descriptions from insectivorous bats is large, and a clade c betacoronavirus was recently described from insectivorous hedgehogs (Corman et al., 2013a). Insectivorous mammals, such as shrews, moles and hedgehogs from the order *Eulipotyphla* could thus be specifically investigated to bolster hypotheses on the putative ancient insect origins of the CoVs harbored by insectivorous mammalian hosts.

It has been hypothesized previously that large genome size, frequent recombination and high mutation rate make coronaviruses particularly prone to zoonotic host switching (Woo et al., 2009). These hallmarks require further consideration. First, the large CoV genomes indeed allow for the accommodation of several genes, putatively enhancing virus pathogenicity. For example, the evolutionary origin of the highly variable SARS-CoV accessory proteins is largely unknown, and deletion of these genes strikingly decreases virulence (McBride and Fielding, 2012; Narayanani et al., 2008). Similarly, among CoV genomes, only members of the betacoronavirus clade contain a hemagglutinin-esterase related to that of influenza C viruses (Zeng et al., 2008), compatible with putative inter-viral gene exchange. Still, little is known about how and when CoVs can incorporate novel genes into their genomes. The technical difficulties observed with the usage of CoV-based vectors for heterologous gene expression indicate that, at least in vitro, the introduction of foreign genes into a CoV genome is not an easy task (de Haan et al., 2005; Sola et al., 2003). Additionally, there are many examples of smaller RNA viruses containing numerous accessory genes that enhance virulence, such as HIV-1 at only a third of the typical CoV genome size (Malim and Emerman, 2008).

Second, it should be kept in mind that CoVs possess proofreading enzymes which limit their mutation rates drastically, compared to other RNA viruses (Eckerle et al., 2010; Hanada et al., 2004; Minskaia et al., 2006). CoVs have thus likely evolved a fine balance between necessary mutation rates and RNA proofreading (Denison et al., 2011). Assumptions of high CoV mutation rates are therefore probably only adequate in specific scenarios, such as the recent host switches of SARS- and MERS-CoV.

Third, recombination may certainly occur during CoV replication, and the prototype example of the recombinant origins of FIPV type 2 was discussed above (Herrewegh et al., 1998; Pasternak et al., 2006). It should be noted, however, that other RNA viruses, such as the negative-stranded paramyxoviruses, which probably recombine much less than CoVs, still give rise to a huge number of likely zoonotic host switches from bats and birds into other mammals, at least rivalling the CoVs (Drexler et al., 2012).

The cumulative evidence suggests that at least four HCoVs share a likely zoonotic origin: HCoV-OC43 from bovines and HCoV-229E, SARS- and MERS-CoV from bats. In comparison, three alphacoronaviruses have so far successfully established circulation in the human population (Palese, 2004). In contrast to CoVs, influenza A viruses cause regular epidemics or even pandemics. Still, the large number of zoonotic HCoVs emphasizes their potential to infect animals other than their original mammalian hosts, and demonstrates the value of continuous screening of humans and wildlife for potentially emerging pathogens. However, it should be kept in mind that sequence detection alone does not warrant premature conclusions as to the zoonotic potential of such viruses (Deeks et al., 2013; Drosten, 2013). Authentic risk assessment strategies would, for example, rely on infection experiments employing isolated or synthetically reconstructed viruses. In the specific case of the CoVs, identification of genomic markers associated with the viral potential to switch hosts or with increased pathogenicity may enable alternative studies. These markers could include viral interferon antagonists and the RBD, both of which could be analyzed in vitro, even in the absence of virus isolates or completely characterized genomes.

Finally, any attempt to eradicate bats or other wildlife in reaction to virus descriptions should be strongly discouraged, as this counteracts the ecological relevance of these animals, e.g., for pollination and pest reduction (Boyles et al., 2011; Kalka et al., 2008).

Acknowledgements

We are grateful to Monika Eschbach-Bludau, Tobias Bleicker and Sebastian Brünink for technical assistance.

References

...


