JVI Accepted Manuscript Posted Online 11 January 2017
J. Virol. doi:10.1128/JVI.01953-16

Title:

Surveillance of bat coronaviruses in Kenya identifies relatives of human coronaviruses NL63 and 229E and their recombination history

Running title:

Bat origin of human coronaviruses

Authors:

Ying Tao[1#], Mang Shi[2#], Christina Chommanard[1], Krista Queen[1], Jing Zhang[1], Wanda Markotter[3], Ivan V. Kuzmin[4]†, Edward C. Holmes[2], Suxiang Tong[1*]

Affiliations:

[1] Division of Viral Diseases, Centers for Disease Control and Prevention, Atlanta, GA 30333, USA; [2]Marie Bashir Institute for Infectious Diseases and Biosecurity, Charles Perkins Centre, School of Life and Environmental Sciences and Sydney Medical School, The University of Sydney, Sydney, Australia; [3]Centre for Viral Zoonoses, Department of Medical Virology, Faculty of Health Sciences, University of Pretoria, Pretoria, South Africa; [4]Division of High Consequence Pathogens and Pathology, Centers for Disease Control and Prevention, Atlanta, GA 30333, USA.

# Y.T. and M.S. contributed equally to this work

† Present address: Department of Pathology, University of Texas Medical Branch, Galveston, TX 77555, USA.

* Correspondence to: Dr. Suxiang Tong, 11600 Clifton Rd, mail stop G18, CDC, Atlanta, GA 30333; Tel: 4046391372; Email: sot1@cdc.gov.

The findings and conclusions in this report are those of the author(s) and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

Type of Publication: ‘Full length’ paper

Word count: Abstract (155); Importance (106); Text body (4618)

ABSTRACT

Bats harbor a large diversity of coronaviruses (CoVs), several of which are related to zoonotic pathogens that cause severe disease in humans. Our screening of bat samples collected in Kenya during 2007-2010 not only detected RNA from several novel CoVs but, more significantly, identified sequences that were closely related to human CoVs NL63 and 229E, suggesting that these two human viruses originate from bats. We also demonstrated that human CoV NL63 is a recombinant between NL63-like viruses circulating in *Triaenops* bats and 229E-like viruses circulating in *Hipposideros* bats, with the break-point located near 5’ and 3’ end of the spike (S) protein gene. In addition, two further inter-species recombination events involving the S gene were identified, suggesting that this region may represent a recombination “hotspot” in CoV genomes. Finally, using a combination of phylogenetic and distance-based approaches we showed that genetic diversity of bat CoVs is primarily structured by host species and subsequently by geographic distances.

IMPORTANCE

Understanding the driving forces of cross-species virus transmission is central to understanding the nature of disease emergence. Previous studies have demonstrated that bats are the ultimate reservoir hosts for a number of coronaviruses (CoVs) including ancestors of SARS-CoV, MERS-CoV, and HCoV-229E. However, the evolutionary pathways of bat CoVs remain elusive. We provide evidence for natural recombination between distantly-related African bat coronaviruses associated with *Triaenops afer* and *Hipposideros* sp. bats that resulted in a NL-63 like virus, an ancestor of the human pathogen HCoV-NL63. These results suggest that inter-species recombination may play an important role in CoV evolution and the emergence of novel CoVs with zoonotic potential.

INTRODUCTION

Coronaviruses (CoVs) (subfamily *Coronavirinae*, family *Coronaviridae*, order Nidovirales) are common infectious agents that infect a wide range of hosts including humans, causing respiratory, gastrointestinal, liver, and neurologic diseases, and that possess the largest genomes of any RNA viruses described to date (1). The subfamily *Coronavirinae* is currently classified into four genera: *Alphacoronavirus, Betacoronavirus, Gammacoronavirus*, and *Deltacoronavirus* (2). The alphacoronaviruses (alpha-CoV) and betacoronaviruses (beta-CoV) are exclusively found in mammals while the gammacoronaviruses (gamma-CoV) and deltacoronaviruses (delta-CoV) are mainly associated with birds. Presently, the greatest diversity of alpha- and beta-CoVs has been documented in bats, which in part reflects the more intensive surveillance of these animals since *Rhinolophus* spp. bats were implicated as the reservoir hosts for SARS-related CoVs (3, 4). This surveillance resulted in the discovery of a potential reservoir host (bat) species for another two human CoVs: Human CoV 229E (HCoV-229E), a relative of which is present in *Hipposideros* bats (5, 6), and Middle East respiratory syndrome coronavirus (MERS-CoV), for which related viruses are present in *Pipistrellus*, *Tylonycteris*, and *Neoromicia* bats (7-10), although the most likely reservoir host of human MERS-CoV identified to date is the dromedary camel (11). Most recently HCoV-229E-like CoVs were also identified in camels, although their role in human infection is unknown (12).

Africa is a major hotspot of zoonotic emerging diseases. With its rich biodiversity, Africa is inhabited by many bats of different species including those that serve as reservoirs of important zoonotic diseases such as Marburg hemorrhagic fever and rabies (13). Our initial screening demonstrated the presence of diverse CoVs in African bats, including those collected in the southern parts of Kenya during 2006 (14, 15), and in other countries including South Africa, Nigeria, and Ghana (16). Furthermore, recent studies have provided strong evidence that HCoV-229E originated from bat viruses circulating in Africa (5), underscoring the zoonotic potential of bat-borne CoVs from this continent.

One human coronavirus, HCoV-NL63, was first isolated in 2004 from the aspirate of a 8-month-old boy suffering from pneumonia in the Netherlands (17). While the clinical significance of this virus is debated, it has a worldwide distribution and is known to infect both the upper and lower respiratory tract (18). Based on a phylogeny of the RNA-dependent RNA polymerase (RdRp), HCoV-NL63 is related to another human virus HCoV-229E and had no close relatives identified in bats (16). Although Huynh et al. (19) suggested that a virus (ARCoV.2/2010/USA) isolated from the American tricolored bat (*Perimyotis subflavus*) may share common ancestry with HCoV-NL63, the genetic distance between the two viruses is large, and their close relationship has not been corroborated in other phylogenetic analyses (16, 20). Nevertheless, the successful passage of HCoV-NL63 in an immortalized bat cell line suggests its potential association with bats (19).

As is well appreciated, recombination leads to rapid changes of genetic diversity in RNA viruses (21). CoVs represent a classic example of viruses with high frequencies of homologous recombination through discontinuous RNA synthesis (22). Indeed, under experimental conditions, the recombination frequency can be as high as 25% for the entire CoV genome (23). Recombination in CoVs is also frequently reported under natural conditions, including some emerging human pathogens such as SARS-CoV (24, 25), MERS-CoV (11), HCoV-OC43 (26), and HCoV-NL63 (27), although most reports are between closely related viruses.

The Global Disease Detection Program (GDD) of the Centers for Disease Control and Prevention (CDC, Atlanta, GA) is focused on the detection of emerging infectious agents worldwide. One of the GDD projects was directed toward the detection of such potential zoonotic pathogens in African bats. Since the initial study performed during 2006 in Kenya

(14, 15), an expanded surveillance of bat CoVs has been performed in the same and other countries including Kenya, Nigeria, Democratic Republic of Georgia, Democratic Republic of Congo, Guatemala, and Peru. The project included more bat species and geographic locations, allowing a more thorough investigation of the genetic diversity and ecological dynamics of CoVs circulation in bats. In this study, we performed an ecological and evolutionary characterization of CoVs circulating in Kenya and identified distinct CoVs from *Triaenops afer* and *Hipposideros* sp. bats that are phylogenetically related to HCoV-NL63 in different parts of the genome. Based on this data, we propose a scenario for the origin and evolutionary history of HCoV-NL63 and related viruses.

MATERIALS AND METHODS

Sample collection. Between 2007 and 2010 a total of 2050 bat specimens were collected from 30 different locations in Kenya (Table S1) in collaboration with the CDC GDD regional country office in Kenya and National Museums of Kenya. The bats were captured using mist-nets, hand nets or manually. The protocol (2096FRAMULX-A3) was approved by the CDC IACUC and by Kenya Wildlife Services. Upon capture, each bat was measured, sexed and identified to species by a trained field biologist. Subsequently, fecal and oral swabs (if possible) were collected in compliance with field protocol and were then transported on dry ice from the field to -80°C storage before further processing.

CoV RNA detection. Each fecal and oral swab was suspended in 200 µL of a phosphate buffered saline. Viral total nucleic acids (TNA) were extracted using the QIAamp Mini Viral Spin kit (Qiagen, Valencia, CA, USA) according to the manufacturer’s instructions, followed by semi-nested RT-PCR (SuperScript III One-Step RT-PCR kit and Platinum Taq kit, Invitrogen, San Diego, CA, USA) using primer sets designed to target the conserved genome region of alpha-, beta-, gamma- and delta-CoVs, respectively (15). PCR products of the expected size (~ 400 nucleotides) were purified by gel extraction using the QIAquick Gel Extraction kit (Qiagen, Valencia, CA, USA) and sequenced in both directions on an ABI Prism 3130 automated sequencer (Applied Biosystems, Foster City, CA, USA). As validation, the RT-PCR procedure was repeated for each of the CoV positive specimens.

Bat mitochondrial gene sequencing. Bat species were further confirmed by sequencing the host mitochondrial cytochrome b (cytB) gene in each of the CoV-positive specimens. Both the method and the primers used have been described previously, and a final 1104 bp fragment of the cytB gene was amplified and sequenced as described previously (14, 15).

Phylogenetic analyses. This study generated a total of 240 CoV RdRP sequences (402 bp) from Kenyan bats. These sequences were first aligned in MAFFT v7.013 (28), using amino acid sequences as a guide for the nucleotide sequence alignment. Phylogenetic trees were then inferred using the maximum likelihood (ML) method available in PhyML version 3.0 (29) assuming a general time-reversible (GTR) model with a discrete gamma distributed rate variation among sites (Γ_4) and the SPR branch-swapping algorithm. To produce a more condensed data set, we clustered the highly similar sequences from the same geographic location and host species, and randomly chose one or two to represent each cluster. This condensed data set was subsequently combined with 121 reference sequences representative of the genetic diversity of alpha- and beta-CoVs on a global scale taken from GenBank. ML phylogenetic trees of these final alignments were inferred using the same procedure and substitution models as described above.

Comparisons of viral genetic, geographic, and host genetic distance matrices. To determine the relationship between viral genetic, geographic, and host genetic distances, we compiled a data set containing the Kenyan CoV samples generated in this study. The genetic distance matrices were produced from pairwise comparisons either in the form of uncorrected percentage differences or calculated from the phylogenetic trees (patristic distance) using the Patristic v1.0 program (30) The geographic distances (Euclidean distance) were calculated using the formula “distance = (acos((sin(latitude1) * sin(latitude2)) + (cos(latitude1) * cos(latitude2) * cos(longitude2 - longitude1)))) * 6371”, with spatial coordinates of the samples derived from the geographic location information.

We used Mantel correlation analyses to test the extent of the correlation between these matrices (31). Both simple Mantel’s test and partial Mantel’s test were performed, and the correlation was evaluated with 10000 permutations. To access which of the two factors – geographic or host genetic distance – best explained total variation in the virus genetic distance matrices, we performed multiple linear regression on these distance matrices (32). The statistical significance of each regression was evaluated by performing 10000 permutations. To examine whether the degree of virus genetic relatedness corresponded to the scale of geographic distance or host relatedness, we generated Mantel correlograms. In each correlogram, 10-12 distance classes were assigned following an equal-frequency criterion: each class had similar number of pairwise comparisons. All statistical analyses were performed using the Ecodist package implemented in R3.0.2 (33), and all statistical results were considered significant at the P = 0.05 level.

Full genome sequencing and sequence analyses. Five viruses representative of the full diversity of the CoVs newly described here were selected for full genome sequencing: BtKY229E-1, BtKY229E-8, BtKYNL63-9a, BtKYNL63-9b, and BtKYNL63-15. We first sequenced a number of conserved regions throughout the genome using several semi-nested or nested consensus degenerate RT-PCR amplicons. These regions were then bridged using sequence-specific RT-PCR followed by Sanger sequencing (< 2 kb) or using the PacBio platform (> 2 kb). The assembled consensus genome sequences from PacBio sequencing were later confirmed by sequence-specific RT-PCR and Sanger sequencing (GenBank accession numbers KY073744-KY073748). The 5’ and 3’ genome termini were not determined due to the limited RNA remaining, and were derived with PCR primers based on the conserved genome regions in alpha-CoVs.

For each complete genome sequence, potential ORFs were predicted based on the conserved core sequence, 5′-CUAAAC-3′, with a minimum length of 66 amino acids. Ribosomal frameshifts were identified based on the presence of the conserved slippery sequence, “UUUAAAC”. For phylogenetic analyses, the data set was first separated into six ORFs, namely; ORF1a, ORF1b, Spike (S), Envelope (E), Membrane (M), and Nucleoprotein (N) genes. The data set for each gene was translated into amino acid sequences and aligned using MAFFT v7.013. Phylogenetic trees were then inferred using PhyML as described above. Recombination events were first identified from the occurrence of incongruent topologies in these initial phylogenies, and were then confirmed and characterized using Simplot v3.5.1 (34). In the Simplot analysis, seven sequences were analyzed, including the potential recombinant, the parental viruses, as well as an outgroup. The similarity comparisons of recombinant and the other sequences were plotted using a sliding window with a size of 1000 bp and a step size of 10 bp.

RESULTS

Prevalence of CoV in Kenyan bats. We examined bats from at least 27 species (17 genera) collected over a four year period (2007-2010) from 30 locations across the southern part of Kenya (Figure 1). A total of 2,050 bats samples were screened for CoV RNA using a pan-coronavirus RT-PCR assay. We found an overall prevalence of 11.7% (240/2,050 bats) (Table S1). This overall prevalence is in line with recent reports of CoVs in bats from numerous locations including South Africa, Mexico, Philippines, Kenya, United Kingdom, Japan, Italy, and Ghana (6, 14, 15, 35-40).

Bats of the species tested (*Chaerephon pumilus*, *Coleura afra*, *Lissonycteris angolensis*, *Miniopterus africanus*, *Neoromicia tenuipinnis*, *Neoromicia* sp., *Nycteris* sp., *Pipistrellus* sp., and *Scotoecus* sp.) did not yield CoV positive samples although the sample number was limited and might not reflect the real prevalence (Table S1). Conversely, in bats of several other species the CoV prevalence was high (*Cardioderma cor*, 25%; *Eidolon helvum*, 21%; *Epomophorus labiatus*, 28.6%; *Hipposideros* sp., 27.6%; *Miniopterus minor*, 22.6%; *Otomops martiensseni, 28.6%*; *Rhinolophus hildebrandtii*, 31.3%; *Rhinolophus* sp., 28.9%; *Triaenops afer*, 26.7%). Most species (21/27) were sampled at more than one location. Of note, we detected CoVs in 21% of *E. helvum* bats tested in Kenya, whereas a previous study in Ghana failed to detect any CoVs in a similar number of bats from this species (6).

Phylogenetic diversity of Kenyan bat CoVs. The viral sequences identified in Kenyan bats showed a remarkable diversity within both alpha- and beta-CoVs (Figure 2). Based on our phylogenetic analysis, the CoVs newly identified here can be grouped into 20 phylogenetic lineages (Figure 2). Many of the sampled bat genera are associated with more than one viral lineage. Furthermore, in some cases, the divergence of the CoVs within the same host genera may also be associated with possible differences in sample types. For example, we found two lineages of CoV in *Rousettus aegyptiacus* bats, one of which was present in oral swabs (Figure 2: L7 *Rousettus*) while the other one was identified in fecal swabs (L17 *Rousettus*). The default tissue tropism for bat CoVs is believed to be intestinal and samples of choice are fecal swabs. In agreement with this, only four viruses were identified from oral swab samples (L7 *Rousettus*) as indicated in the phylogeny (Figure 2).

Our phylogenetic analyses also revealed a number of cross-species transmission events at the genus level, many of which appeared to be transient spill-overs with no evidence of onward transmission. This pattern was observed as CoV sequences recovered from bats of a particular genus located as tree tips within the phylogenetic diversity that is mainly associated with a different bat genus. From our Kenyan data set, there were seven such cross-species transmission events in total, each represented by a single sequence (dotted red in Figure 2), suggesting these are most likely viruses with limited transmission within new hosts, although this hypothesis requires confirmation on a larger set of samples.

A more comprehensive and informative phylogeny (Figure 3) was obtained after including the representative global CoV sequences from GenBank, which also included the Kenyan viruses previously reported (15). The phylogeny, which included viral sequences recovered from bats of more than 50 species (30 genera), resulted in an accurate phylogenetic assignment of the viruses described in this study (Figure 3). Importantly, the newly discovered viruses from Kenya have greatly extended our previous work (15) in terms of: (i) expanding the diversity of existing lineages, including the *Miniopterus*, *Rhinolophus*, and *Scotophilus* associated CoV clusters in the genus *Alphacoronavirus*, and the *Rousettus* and *Rhinolophus* associated CoVs clusters in the genus *Betacoronavirus*; and (ii) the discovery of new viruses from either a novel bat host (i.e. *Triaenops*) or new divergent CoV clusters in known hosts (i.e. *Rhinolophus*, *Rousettus*, *Chaerephon*, etc) (Figure 3).

The phylogeny suggests both ancient virus-host co-divergence and recent cross-species transmission of CoVs between bats and other mammalian hosts. The phylogeny clearly demonstrates that CoVs from two host groups, one dominated by bats and the other exclusively by non-chiropteran mammals, formed sister clades for both alpha- and beta-CoVs (Figure 3), suggestive of an ancient divergence between them. Conversely, several non-chiropteran CoVs are nested within the diversity of bat CoVs, suggesting that these viruses are relatively recent introductions from bats. These cross-species transmission events resulted in emergence of severe (SARS-CoV and MERS-CoV) and mild (HCoV-NL63 and HCoV-229E) human pathogens, as well as animal pathogens (Porcine epidemic diarrhea virus [PEDV] and Alpaca respiratory CoV). Interestingly, HCoV-NL63, previously thought to be related to North American tricolored bat (*P. subflavus*) (19), in our phylogeny is deeply nested within the newly identified CoVs from African *Triaenops afer* bats (Figure 3), while the *P. subflavus* virus (labeled green in Figure 3) grouped with a North American CoV sampled from a *Myotis volans* bat (Figure 3). Therefore, *Triaenops afer* bats likely represent the most recent chiropteran reservoir host of viruses ancestral to HCoV-NL63. In addition, our results identified 16 additional 229E-like viruses (L14, Figure 2), providing further evidence that *Hipposideros* bats in Africa harbor viruses that are ancestral to HCoV-229E (5, 6).

Host and spatial dynamics of bat CoVs in Kenya. We used Mantel's test to compare the virus and host genetic distance matrices, as well as virus and geographic distance matrices. Notably, the correlation values were positive and highly significant in both comparisons (Table 1), suggesting that both host and geography have shaped the structure of virus genetic diversity. This conclusion remained following partial Mantel analyses and multiple linear regression analyses in which we tested the effect between two matrices while controlling for the third (Tables 1 and 2). Importantly, however, in both simple and partial Mantel analyses, the virus genetic distance matrices had much higher correlation with host genetic distance matrices than with geographic distance matrices (Table 1), indicating that bat CoV diversity is more structured by host than by geographic distance.

Next, we used Mantel autocorrelograms to examine the effect of (i) geographic distance (Figure 4A) and (ii) host genetic distance on virus diversity (Figure 4B). Host genetic distance decreased from highly significantly positive at short taxonomic distances to highly significantly negative at long distances. Importantly, the crossing-over point was at a host genetic distance of around 0.15-0.19, which marks the boundary of intra- and inter-genera host diversity (Figure 4B). However, no obvious clinal patterns in geographic distance were observed within the Kenyan data set.

Full genome characterization and recombination analyses of NL63-like and 229E-like viruses. To further explore the evolution of the NL63-like and 229E-like viruses, we generated the complete genome sequences of five representative bat-derived CoVs: three (BtKYNL63-9a, BtKYNL63-9b, and BtKYNL63-15) were from the NL63-like group and two (BtKY229E-1 and BtKY229E-8) from the 229E-like group (L12-L14, Figure 2). For all the viruses newly described here, the genome structures follow an identical ORF arrangement: ORF1ab-S-ORF4-E-M-N-ORF8 in 229E-related viruses and ORF1ab-S-ORF3-E-M-N-ORFx in NL63-related viruses (Figure 5, Tables 3 and 4). The additional ORF8/ORF$_X$ was identified at the 3' end of the genome in all bat NL63-like and 229E-like viruses characterized in this study, although it was missing in both human viruses (HCoV-229E and HCoV-NL63). The ORF8 in bat 229E-like genomes is named in analogy with the ORF8 of Ghanaian bat and dromedary 229E-like CoVs (5, 12). The ORF8 of BtKY229E-1 shared 60% protein identity with its closest relatives while BtKY229E-8 has a shorter and highly divergent ORF8. The ORFx of NL63-like viruses shared very low identity (21-33% at the amino acid level). Similarly to the bat 229E-like CoVs recently discovered in Ghana (5), the S genes in our bat 229E-like CoVs have a considerably longer 5' S1 portion (additional 185 amino acids) compared to HCoV-229E and alpaca and dromedary 229E viruses (12).

For comparison, we also included 21 genome sequences representative of the diversity in the genus *Alphacoronavirus*. The phylogeny based on the ORF1b protein alignment confirmed that NL63-like and 229E-like groups are monophyletic (Figure 6). Given that each group is associated with a specific bat genus, it is likely that the ORF1b genes of the human viruses (i.e. HCoV-NL63 and HCoV-229E) were ultimately derived from *Triaenops*-associated CoVs and *Hipposideros*-associated CoVs, respectively. The relationship between *Hipposideros* bat CoVs and HCoV-229E was also demonstrated by Corman et al. (5) based on specimens obtained in Ghana. Compared to the viruses described in that study, the newly identified Kenyan viruses (BtKY229E-1 and BtKY229E-8) were among those more distantly related to HCoV-229E (Figure 6 and Table 3). As for the NL63-

like group, HCoV-NL63 was nested within the diversity of three lineages of *Triaenops*-associated CoVs, among which BtKYNL63-9a showed the closest relationship in all genome regions with the exception of the S gene (Figure 6 and Table 3).

Strikingly, the phylogeny of the S protein suggested an entirely different evolutionary history for HCoV-NL63 compared to the rest of the genome (Figure 6). Specifically, for all the proteins with the exception of S, HCoV-NL63 clustered with the NL63-like group. However, in the S protein, HCoV-NL63 was deeply nested within the 229E-like group, associated exclusively with viruses from *Hipposideros* bats, and particularly similar to the sequences BtKY229E-1 and BtKY229E-8 newly identified during this study (Figure 6). Interestingly, BtKY229E-1 exhibited the closest resemblance to HCoV-NL63 in the receptor binding domain (RBD, (41)), especially in the three receptor binding motifs (RBM), whereas other viruses exhibited less similarity in these regions (Figure 7A). A phylogeny based on the RBD region confirmed our observation (Figure 7B), although it remains uncertain whether these bat viruses utilize the same host cell receptor.

To further characterize this recombination event, we performed genome-scale similarity comparisons between HCoV-NL63 and related viruses (Figure 8). The analysis confirmed the chimeric nature of the HCoV-NL63 genome with only the spike protein involved in recombination via two break-points: one located near the 5' end of the S gene and the other at around 200 nucleotides upstream of the 3' end. To exclude the possibility of any artificial recombination, the break-point was further confirmed by RT-PCR and Sanger sequencing, using a single amplicon to cover each break-point. Collectively, these data show that HCoV-NL63 evolved from a recombination event between CoVs from the NL63-like and 229E-like groups.

In addition to HCoV-NL63, we identified a number of other recombination events between divergent CoVs involving the S gene. One example is the BtKYNL63-15 newly identified here. Throughout the genome, BtKYNL63-15 showed strong similarity (79% - 99% protein identities in the ORF1ab, ORF4, M, E, and N genes) with BtKYNL63-9b. In contrast, the genetic identity between S protein sequences of these viruses was only 53%. In the S protein phylogeny, BtKYNL63-15 did not cluster with NL63-like viruses but instead clustered with *Miniopterus* bat CoV HKU8 and *Chaerophon* bat CoV KY22 (Figure 6). Interestingly, HKU8 itself is a recombinant in the S gene region (Figure 6). These results suggest that the spike protein of CoVs is subject to relatively frequent recombination even between divergent viruses.

DISCUSSION

In this study we significantly extended existing knowledge on CoV diversity, their association with specific bat species, the relatedness between bat and human CoVs, and natural recombination events in the CoV spike (S) protein gene between viruses from different lineages.

Notably, we found that host species poses a greater influence on CoV diversity in bats than the geographic distance, which can be explained by the ability of bats to fly (including long-distance migrations typical for some species) and disperse their pathogens over vast territories (42). A closer inspection of the Mantel correlogram suggests the presence of less structured (homogenous, mantel statistic r>0), and highly structured (mantel statistic r<0) diversity which, strikingly, corresponds to the division between intra-genera (10% ~ 20%) and inter-genera (> 20%) host genetic distances (Figure 4B). This suggests that within-genus virus transmissions occur significantly more frequently than between-genera transmissions, which is consistent with the previous observations that phylogenetic clustering is less constrained at the host species level than at the genus level (16, 43). While it is commonly accepted that host phylogeny constrains virus cross-species transmission to some extent (44), the stronger demarcation at the genus level is of particular interest. In fact, bats of different species, genera, and families frequently roost together (in caves, tree holes, and other shelters), sometimes in dense aggregations, which provide abundant opportunity for mechanical transmission of pathogens between host species. Therefore, our data suggests that distinctions between bats at the genus level might mark a threshold where the differences in cellular and immunological environment become a major challenge for a virus to switch hosts. This, in turn, will lead to the pattern of ‘preferential host switching’ that has been observed in a number of other viruses (45).

The detection of distinctive HCoV-NL63-like and HCoV-229E-like sequences in bats sheds new light on CoV evolution. In particular, we provide strong evidence that HCoV-NL63 has a zoonotic recombinant origin. Although the majority of the HCoV-NL63 genome originates from the viruses circulating in *Triaenops afer* bats, its spike protein gene is derived from a 229E-like virus circulating in *Hipposideros* spp. bats. However, despite the strong signal for recombination, both putative parental strains show substantial genetic distances from human CoVs. This most likely reflects extensive post-recombination sequence divergence, which in turn suggests that the recombination event has occurred prior to the emergence of HCoV-NL63 in humans.

Most of the recombination events reported here involve breakpoints around the S gene. Indeed, similar breakpoints are also reported for SARS-CoV and SARS-like CoVs (24, 25), HCoV-OC43 (26), and a feline CoV (46) such that it is seemingly a recombination ‘hotspot’ in many CoVs. It has been argued that a strong secondary structure between ORF1a and S gene may promote transcriptional pulsing, facilitating recombination (47). However, there is also evidence that this recombination hotspot does not exist under non-selective conditions (48), such that it may reflect the successful spread of beneficial recombinants rather than an elevated rate of recombination per se. This hypothesis is supported by the fact that the spike protein is intimately involved in the interaction with the host immune system.

Importantly, our results also revealed that recombination has resulted in similar S proteins in the two human viruses HCoV-NL63 and HCoV-229E, such that acquisition of a 229E-like S protein may have contributed to the emergence of NL63-like viruses in humans. However, despite this similarity of S protein sequences, these two human viruses utilize different receptors (ACE2 and aminopeptidase-N for HCoV-NL63 and HCoV-229E, respectively) to enter human cells. Within the 229E-like group, the RBD of HCoV-NL63 is more closely related to BtKY229E-8 than to HCoV-229E. The RBD of BtKY229E-8 exhibits greater similarity with that of HCoV-NL63 (Figure 7), and is therefore more likely to be the prototype of RBD in HCoV-NL63.

Until recently, most reported recombination events in CoVs involved viruses associated with closely related host species, although recombination between highly divergent CoVs has been demonstrated experimentally (49-51). The apparent lack of interspecies recombination under natural conditions is most likely due to the insufficient collection of complete genome sequences that are truly representative of coronavirus diversity. Indeed, a number of viruses, such as HKU2, display phylogenetic incongruence across different parts of the genome (52), although the lack of one of the putative parental strains has prevented clear identification of a recombinant history.

Finally, our study provides insights into the evolution history of CoVs. Although it is unclear whether bats are direct ancestors of all alpha- or beta-CoVs due to the presence of non-bat CoV clades at the basal phylogenetic positions of both genera (Figure 3), bat-borne CoVs constitute a substantial part of the diversities of alpha- or beta-CoVs. In addition, six lineages of non-bat CoVs are nested within the bat-borne clades. These likely represent independent and successful adaptations via shifts from the progenitor reservoir species (bats) to other mammals. Four well-characterized human CoVs lie within these clades. However, it is worth noting that bats may not have directly transmitted the viruses to humans. Indeed, HCoV-229E is more closely related to viruses circulating in camels than those in bats, suggesting that camels may be intermediate hosts between bats and humans (12). Similarly, other human CoVs such as SARS-CoV and MERS-CoV all use terrestrial mammals rather than bats as intermediate hosts, which have an increased chance of contact with humans. This underlines a typical zoonotic link of bat-associated CoV to humans via terrestrial mammals.

ACKNOWLEDGEMENTS

We thank Michael Niezgoda, Richard Franka, Amy T. Gilbert, Charles E. Rupprecht (CDC, Atlanta, GA), and Bernard Agwanda (National Museums of Kenya, Nairobi) for field sampling of bats, as well as Evelyne Mulama, Edwin Danga, Robert F. Breiman, and Joel M. Montgomery (CDC-Kenya, Nairobi) for logistical support during field expeditions.

REFERENCES

1. **Weiss SR, Leibowitz JL.** 2011. Coronavirus pathogenesis. Adv Virus Res **81:**85-164.

2. **Adams MJ, Lefkowitz EJ, King AM, Bamford DH, Breitbart M, Davison AJ, Ghabrial SA, Gorbalenya AE, Knowles NJ, Krell P, Lavigne R, Prangishvili D, Sanfacon H, Siddell SG, Simmonds P, Carstens EB.** 2015. Ratification vote on taxonomic proposals to the International Committee on Taxonomy of Viruses (2015). Arch Virol **160:**1837-1850.

3. **Lau SK, Woo PC, Li KS, Huang Y, Tsoi HW, Wong BH, Wong SS, Leung SY, Chan KH, Yuen KY.** 2005. Severe acute respiratory syndrome coronavirus-like virus in Chinese horseshoe bats. Proc Natl Acad Sci U S A **102:**14040-14045.

4. **Li W, Shi Z, Yu M, Ren W, Smith C, Epstein JH, Wang H, Crameri G, Hu Z, Zhang H, Zhang J, McEachern J, Field H, Daszak P, Eaton BT, Zhang S, Wang LF.** 2005. Bats are natural reservoirs of SARS-like coronaviruses. Science **310:**676-679.

5. **Corman VM, Baldwin HJ, Tateno AF, Zerbinati RM, Annan A, Owusu M, Nkrumah EE, Maganga GD, Oppong S, Adu-Sarkodie Y, Vallo P, da Silva Filho LV, Leroy EM, Thiel V, van der Hoek L, Poon LL, Tschapka M, Drosten C, Drexler JF.** 2015. Evidence for an Ancestral Association of Human Coronavirus 229E with Bats. J Virol **89:**11858-11870.

6. **Pfefferle S, Oppong S, Drexler JF, Gloza-Rausch F, Ipsen A, Seebens A, Muller MA, Annan A, Vallo P, Adu-Sarkodie Y, Kruppa TF, Drosten C.** 2009. Distant relatives of severe acute respiratory syndrome coronavirus and close relatives of human coronavirus 229E in bats, Ghana. Emerg Infect Dis **15:**1377-1384.

7. **Annan A, Baldwin HJ, Corman VM, Klose SM, Owusu M, Nkrumah EE, Badu EK, Anti P, Agbenyega O, Meyer B, Oppong S, Sarkodie YA, Kalko EK, Lina PH, Godlevska EV, Reusken C, Seebens A, Gloza-Rausch F, Vallo P, Tschapka M, Drosten C, Drexler JF.** 2013. Human betacoronavirus 2c EMC/2012-related viruses in bats, Ghana and Europe. Emerg Infect Dis **19:**456-459.

8. **Lau SK, Li KS, Tsang AK, Lam CS, Ahmed S, Chen H, Chan KH, Woo PC, Yuen KY.** 2013. Genetic characterization of Betacoronavirus lineage C viruses in bats reveals marked sequence divergence in the spike protein of pipistrellus bat coronavirus HKU5 in Japanese pipistrelle: implications for the origin of the novel Middle East respiratory syndrome coronavirus. J Virol **87:**8638-8650.

9. **Corman V, Ithete N, Richards L, Schoeman M, Preiser W, Drosten C, Drexler J.** 2014. Rooting the Phylogenetic Tree of Middle East Respiratory Syndrome Coronavirus by Characterization of a Conspecific Virus from an African Bat. J Virol **88:**11297–11303.

10. **Ithete N, Stoffberg S, Corman V, Cottontail V, Richards L, Schoeman M, Drosten C, Drexler J, Preiser W.** 2013. Close Relative of Human Middle East Respiratory Syndrome Coronavirus in Bat, South Africa. Emerg Infect Dis **19:**1697–1699. .

11. **Sabir JS, Lam TT, Ahmed MM, Li L, Shen Y, Abo-Aba SE, Qureshi MI, Abu-Zeid M, Zhang Y, Khiyami MA, Alharbi NS, Hajrah NH, Sabir MJ, Mutwakil MH, Kabli SA, Alsulaimany FA, Obaid AY, Zhou B, Smith DK, Holmes EC, Zhu H, Guan Y.** 2016. Co-circulation of three camel coronavirus species and recombination of MERS-CoVs in Saudi Arabia. Science **351:**81-84.

12. **Corman V, Eckerle I, Memish Z, Liljander A, Dijkman R, Jonsdottir H, Juma NK, Kamau E, Younan M, Al Masri M, Assiri A, Gluecks I, Musa B, Meyer B, Müller M, Hilali M, Bornstein S, Wernery U, Thiel V, Jores J, Drexler J, Drosten C.** 2016. Link of a ubiquitous human coronavirus to dromedary camels. Proc Natl Acad Sci U S A.

13. **O'Shea TJ, Cryan PM, Cunningham AA, Fooks AR, Hayman DT, Luis AD, Peel AJ, Plowright RK, Wood JL.** 2014. Bat flight and zoonotic viruses. Emerg Infect Dis **20:**741-745.

14. **Tao Y, Tang K, Shi M, Conrardy C, Li KS, Lau SK, Anderson LJ, Tong S.** 2012. Genomic characterization of seven distinct bat coronaviruses in Kenya. Virus Res **167:**67-73.

15. **Tong S, Conrardy C, Ruone S, Kuzmin IV, Guo X, Tao Y, Niezgoda M, Haynes L, Agwanda B, Breiman RF, Anderson LJ, Rupprecht CE.** 2009. Detection of novel SARS-like and other coronaviruses in bats from Kenya. Emerg Infect Dis **15:**482-485.

16. **Drexler JF, Corman VM, Drosten C.** 2014. Ecology, evolution and classification of bat coronaviruses in the aftermath of SARS. Antiviral Res **101:**45-56.

17. **Fouchier RA, Hartwig NG, Bestebroer TM, Niemeyer B, de Jong JC, Simon JH, Osterhaus AD.** 2004. A previously undescribed coronavirus associated with respiratory disease in humans. Proc Natl Acad Sci U S A **101:**6212-6216.

18. **Fielding BC.** 2011. Human coronavirus NL63: a clinically important virus? Future Microbiol **6:**153-159.

19. **Huynh J, Li S, Yount B, Smith A, Sturges L, Olsen JC, Nagel J, Johnson JB, Agnihothram S, Gates JE, Frieman MB, Baric RS, Donaldson EF.** 2012. Evidence supporting a zoonotic origin of human coronavirus strain NL63. J Virol **86:**12816-12825.

20. **Corman VM, Rasche A, Diallo TD, Cottontail VM, Stocker A, Souza BF, Correa JI, Carneiro AJ, Franke CR, Nagy M, Metz M, Knornschild M, Kalko EK, Ghanem SJ, Morales KD, Salsamendi E, Spinola M, Herrler G, Voigt CC, Tschapka M, Drosten C, Drexler JF.** 2013. Highly diversified coronaviruses in neotropical bats. J Gen Virol **94:**1984-1994.

21. **Holmes EC.** 2009. The evolution and emergence of RNA viruses, *on* Oxford University Press,.

22. **Lai MM.** 1992. RNA recombination in animal and plant viruses. Microbiol Rev **56:**61-79.

23. **Baric RS, Fu K, Schaad MC, Stohlman SA.** 1990. Establishing a genetic recombination map for murine coronavirus strain A59 complementation groups. Virology **177:**646-656.

24. **Hon CC, Lam TY, Shi ZL, Drummond AJ, Yip CW, Zeng F, Lam PY, Leung FC.** 2008. Evidence of the recombinant origin of a bat severe acute respiratory syndrome (SARS)-like coronavirus and its implications on the direct ancestor of SARS coronavirus. J Virol **82:**1819-1826.

25. **Lau SK, Li KS, Huang Y, Shek CT, Tse H, Wang M, Choi GK, Xu H, Lam CS, Guo R, Chan KH, Zheng BJ, Woo PC, Yuen KY.** 2010. Ecoepidemiology and complete genome comparison of different strains of severe acute respiratory syndrome-related Rhinolophus bat coronavirus in China reveal bats as a reservoir for acute, self-limiting infection that allows recombination events. J Virol **84:**2808-2819.

26. **Lau SK, Lee P, Tsang AK, Yip CC, Tse H, Lee RA, So LY, Lau YL, Chan KH, Woo PC, Yuen KY.** 2011. Molecular epidemiology of human coronavirus OC43 reveals evolution of different genotypes over time and recent emergence of a novel genotype due to natural recombination. J Virol **85:**11325-11337.

27. **Pyrc K, Dijkman R, Deng L, Jebbink MF, Ross HA, Berkhout B, van der Hoek L.** 2006. Mosaic structure of human coronavirus NL63, one thousand years of evolution. J Mol Biol **364:**964-973.

28. **Katoh K, Standley DM.** 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol **30:**772-780.

29. **Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O.** 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol **59:**307-321.

30. **Fourment M, Gibbs M.** 2006. PATRISTIC: a program for calculating patristic distances and graphically comparing the components of genetic change. BMC Evol Biol **6:**1-5.

31. **Mantel N.** 1967. The detection of disease clustering and a generalized regression approach. Cancer Res **27:**209-220.

32. **Lichstein JW.** 2007. Multiple regression on distance matrices: a multivariate spatial analysis tool. Plant Ecology **188:**117-131.

33. **Goslee SC, Urban DL.** 2007. The ecodist package for dissimilarity-based analysis of ecological data. Journal of Statistical Software **22:**1-19.

34. **Lole KS, Bollinger RC, Paranjape RS, Gadkari D, Kulkarni SS, Novak NG, Ingersoll R, Sheppard HW, Ray SC.** 1999. Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. J Virol **73:**152-160.

35. **Anthony SJ, Ojeda-Flores R, Rico-Chavez O, Navarrete-Macias I, Zambrana-Torrelio CM, Rostal MK, Epstein JH, Tipps T, Liang E, Sanchez-Leon M, Sotomayor-Bonilla J, Aguirre AA, Avila-Flores R, Medellin RA, Goldstein T,**

Suzan G, Daszak P, Lipkin WI. 2013. Coronaviruses in bats from Mexico. J Gen Virol **94:**1028-1038.

36. **August TA, Mathews F, Nunn MA.** 2012. Alphacoronavirus detected in bats in the United Kingdom. Vector Borne Zoonotic Dis **12:**530-533.

37. **Balboni A, Palladini A, Bogliani G, Battilani M.** 2011. Detection of a virus related to betacoronaviruses in Italian greater horseshoe bats. Epidemiol Infect **139:**216-219.

38. **Geldenhuys M, Weyer J, Nel LH, Markotter W.** 2013. Coronaviruses in South African bats. Vector Borne Zoonotic Dis **13:**516-519.

39. **Shirato K, Maeda K, Tsuda S, Suzuki K, Watanabe S, Shimoda H, Ueda N, Iha K, Taniguchi S, Kyuwa S, Endoh D, Matsuyama S, Kurane I, Saijo M, Morikawa S, Yoshikawa Y, Akashi H, Mizutani T.** 2012. Detection of bat coronaviruses from Miniopterus fuliginosus in Japan. Virus Genes **44:**40-44.

40. **Tsuda S, Watanabe S, Masangkay JS, Mizutani T, Alviola P, Ueda N, Iha K, Taniguchi S, Fujii H, Kato K, Horimoto T, Kyuwa S, Yoshikawa Y, Akashi H.** 2012. Genomic and serological detection of bat coronavirus from bats in the Philippines. Arch Virol **157:**2349-2355.

41. **Wu K, Li W, Peng G, Li F.** 2009. Crystal structure of NL63 respiratory coronavirus receptor-binding domain complexed with its human receptor. Proc Natl Acad Sci U S A **106:**19970-19974.

42. **Peel AJ, Sargan DR, Baker KS, Hayman DT, Barr JA, Crameri G, Suu-Ire R, Broder CC, Lembo T, Wang LF, Fooks AR, Rossiter SJ, Wood JL, AA. C.** 2013. Continent-wide panmixia of an African fruit bat facilitates transmission of potentially zoonotic viruses. Nat Commun **4:**2770.

43. **Drexler JF, Gloza-Rausch F, Glende J, Corman VM, Muth D, Goettsche M, Seebens A, Niedrig M, Pfefferle S, Yordanov S, Zhelyazkov L, Hermanns U, Vallo P, Lukashev A, Muller MA, Deng H, Herrler G, Drosten C.** 2010. Genomic characterization of severe acute respiratory syndrome-related coronavirus in European bats and classification of coronaviruses based on partial RNA-dependent RNA polymerase gene sequences. J Virol **84:**11336-11349.

44. **Streicker DG, Turmelle AS, Vonhof MJ, Kuzmin IV, McCracken GF, Rupprecht CE.** 2010. Host phylogeny constrains cross-species emergence and establishment of rabies virus in bats. Science **329:**676-679.

45. **Charleston MA, Robertson DL.** 2002. Preferential host switching by primate lentiviruses can account for phylogenetic similarity with the primate phylogeny. Syst Biol **51:**528-535.

46. **Herrewegh AA, Smeenk I, Horzinek MC, Rottier PJ, de Groot RJ.** 1998. Feline coronavirus type II strains 79-1683 and 79-1146 originate from a double recombination between feline coronavirus type I and canine coronavirus. J Virol **72:**4508-4514.

47. **Mills DR, Dobkin C, Kramer FR.** 1978. Template-determined, variable rate of RNA chain elongation. Cell **15:**541-550.

48. **Banner LR, Lai MM.** 1991. Random nature of coronavirus RNA recombination in the absence of selection pressure. Virology **185:**441-445.

49. **Baric RS, Yount B, Hensley L, Peel SA, Chen W.** 1997. Episodic evolution mediates interspecies transfer of a murine coronavirus. J Virol **71:**1946-1955.

50. **Becker MM, Graham RL, Donaldson EF, Rockx B, Sims AC, Sheahan T, Pickles RJ, Corti D, Johnston RE, Baric RS, Denison MR.** 2008. Synthetic recombinant bat SARS-like coronavirus is infectious in cultured cells and in mice. Proc Natl Acad Sci U S A **105:**19944-19949.

51. **Masters PS, Rottier PJ.** 2005. Coronavirus reverse genetics by targeted RNA recombination. Curr Top Microbiol Immunol **287:**133-159.

52. **Lau SK, Woo PC, Li KS, Huang Y, Wang M, Lam CS, Xu H, Guo R, Chan KH, Zheng BJ, Yuen KY.** 2007. Complete genome sequence of bat coronavirus HKU2 from Chinese horseshoe bats revealed a much smaller spike gene with a different evolutionary lineage from the rest of the genome. Virology **367:**428-439.

Figure Legends

Figure 1. Map of Kenya showing the geographic locations of 30 bat collection sites.

Figure 2. Phylogeny of RdRp of all CoVs discovered in this study. The host (bat genus), number of sequences, and operational classification (lineage) are shown on the right of the tree. Branches that represent the minority host genera within the lineage defined by a single dominant host genus are marked with red and labeled with solid circle. The tree is mid-point rooted for clarity only and support values are only shown for internal branches.

Figure 3. Phylogenies of RdRp of alphacoronaviruses and betacoronaviruses. The trees are inferred using representative CoV sequences from this study as well as those obtained from the GenBank. The sequences are labeled with accession number/strain name, host (species) and geographic origin (three letter country code). Different colors are used to distinguish the following groups: Kenyan bat CoVs discovered during this study (orange), CoVs identified from non-bat mammals (blue), the *Perimyotis subflavus* virus previously reported to be related to HCoV-NL63 (green), and the remaining bat viruses (black). The lineage information for Kenyan CoVs is shown to the right of the phylogeny and matches that in Figure 2.

Figure 4. Mantel correlograms showing the Kenyan bat CoV RdRp sequences stratified by (A) geographic distances and (B) host genetic distances. A Mantel correlation index (r) was calculated for each of the distance classes. Under the null hypothesis of no relationship between the distance matrices, r values would be close to zero. Positive r values suggest lower genetic distances between case pairs, whereas negative r values suggest higher genetic distances between case pairs. Solid dots: r significantly different from zero; hollow dots: r not significantly different from zero. The graph (B) also shows kernel density plots for intra-genus host distances density (grey solid line) and inter-genus host distances density (grey dotted line). The corresponding y-axis for the plot is shown on the right of the figure (B). The grey box in between the two plots marked the transition area between the intra-genus and inter-genus host genetic distances

Figure 5. Genome organization of 2 bat 229E-like and 3 bat NL63-like viruses sampled from Kenyan bats. A unified length scale is used for all the genomes. Within each genome, the ORFs (arrow boxes) and ribosomal frame shift sites (vertical lines) are indicated at their corresponding positions.

Figure 6. Phylogenetic analyses of major open reading frames of NL63-like and 229E-like CoVs in the context of alphacoronaviruses revealing evidence of recombination. Viruses sequenced in this study are shown in orange. Three potential recombinant genomes of HCoV-NL63, BtKYNL63-15, and HKU8 are indicated with red circles, blue triangles, and brown squares.

Figure 7. The relationships between HCoV-NL63 and related viruses at the receptor binding domain. (A) Alignment of NL63-like and 229E-like viruses and related viruses at the receptor binding domain. The positions of three receptor binding motifs (RBMs) are marked with double arrowed line. Residues in the NL63-CoV RBMs that directly contact the ACE2 receptor are marked with red downward arrows. (B) Phylogenetic relationships of NL63-like and 229E-like viruses at receptor binding domain of HCoV-NL63. The tree is based on an amino acid alignment and mid-point rooted.

Figure 8. Recombination analyses of HCoV-NL63 using Simplot. Genome-scale similarity comparisons of HCoV-NL63 (query) against BtKYNL63-9a (major parental group, blue), BtKYNL63-9b (green), BtKY229E-8 (minor parental group, red), HCoV-229E (orange), BtCoV/FO1A-F2/Hip_aba/GHA/2010 (pink), and Alaca respiratory CoV (brown). A full genome structure, with reference to HCoV-NL63, is shown above the similarity plot, marking the positions and boundaries of the major open reading frames. At the beginning of the S gene, the flat-line followed by a sudden drop of similarity is due to a gap (deletion within HCoV-229E S gene) in the alignment.

Tables

Table 1. Results of Mantel tests and partial Mantel tests comparing two factors (host genetic distance and geographic distance) that predict the structure of virus genetic diversity

Model	***r* value for Kenyan bats (*P* value)**
Host[a]	0.5265 ($P < 0.0001$)[c]
Host | Geography[b]	0.5055 ($P < 0.0001$)[c]
Geography[a]	0.2122 ($P < 0.0001$)[c]
Geography | Host[b]	0.1285 ($P = 0.0005$)[c]

[a]Mantel test; [b]partial Mantel test; [c]significant at 0.001.

Table 2. Multiple regression of virus genetic distance against host genetic distance and geographic distance in Kenyan bat CoVs (2007-2010)

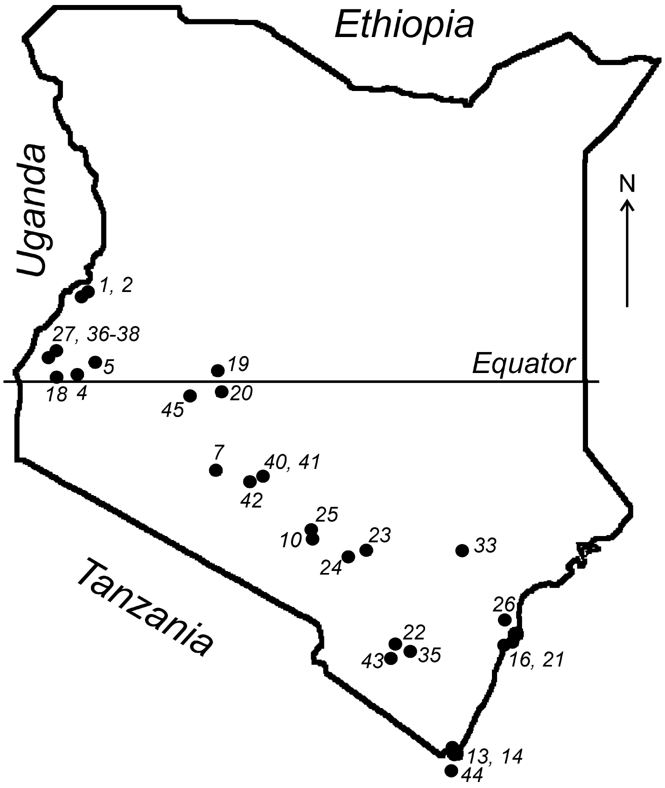
Variable	**Correlation coefficient**	***P*-value**
Host	7.58E-01	1.00E-04
Geography	1.19E-06	1.00E-02

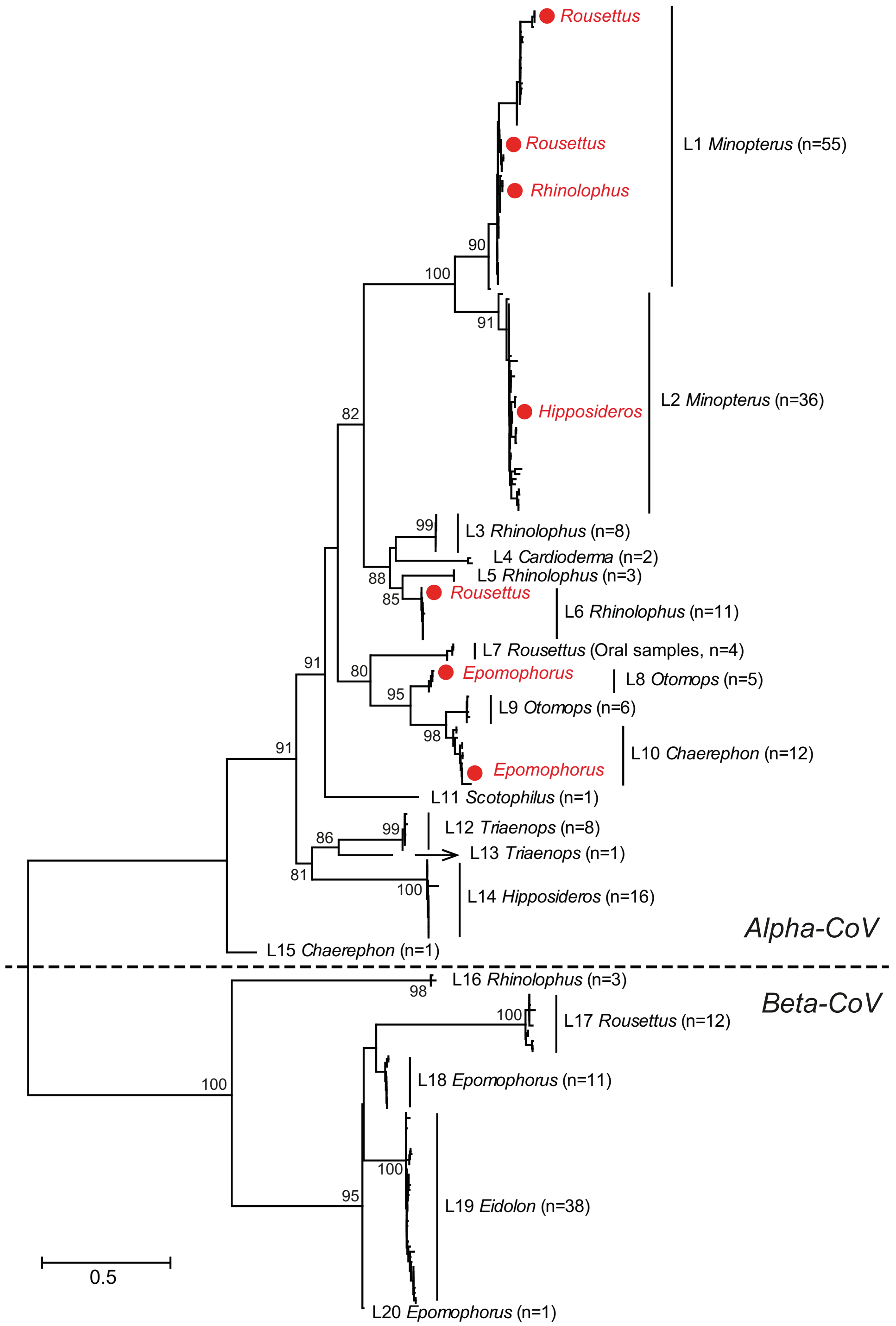
Table 3. Sequence comparisons of the Kenyan bat CoVs with HCoV-229E or HCoV-NL63

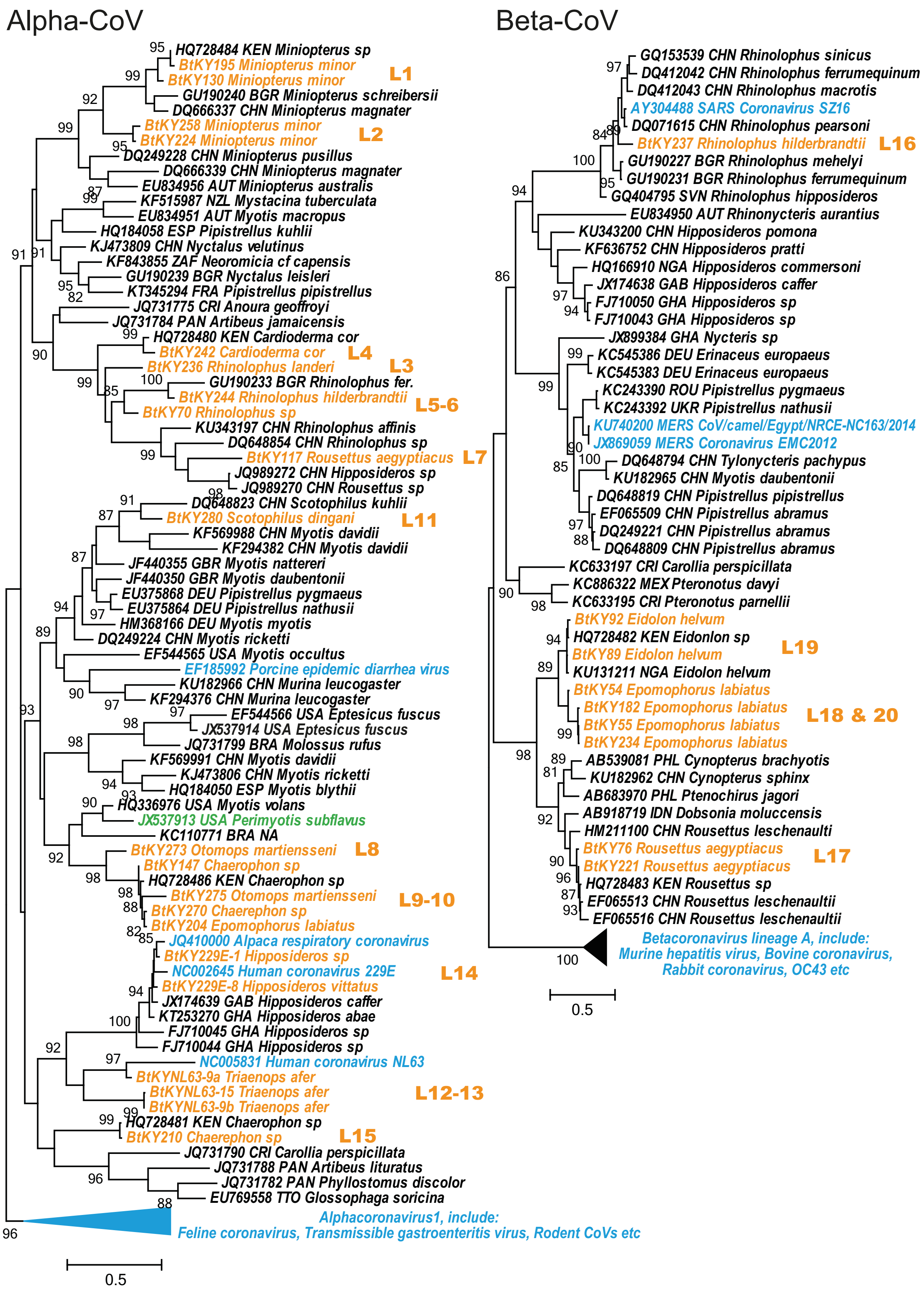
	Genome identity	**Concatenated domains**	**ADRP**	**nsp5**	**nsp12**	**nsp13**	**nsp14**	**nsp15**	**nsp16**	**1ab**	**S**	**ORF3/4**	**E**	**M**	**N**
	Nucleotide %	Amino acid % identity to HCoV-229E													
BtKY229E-1	88	98	92	98	97	99	97	96	94	95	75	92	93	90	78
BtKY229E-8	88	97	89	98	98	98	97	97	94	96	74	94	97	90	68
	Nucleotide %	Amino acid % identity to HCoV-NL63													
BtKYNL63-9a	78	91	75	89	93	94	89	88	94	86	53	67	80	82	69
BtKYNL63-9b	68	83	51	76	88	91	82	81	84	72	52	55	64	61	51
BtKYNL63-15	68	84	51	76	88	91	82	81	87	72	49	55	62	58	52

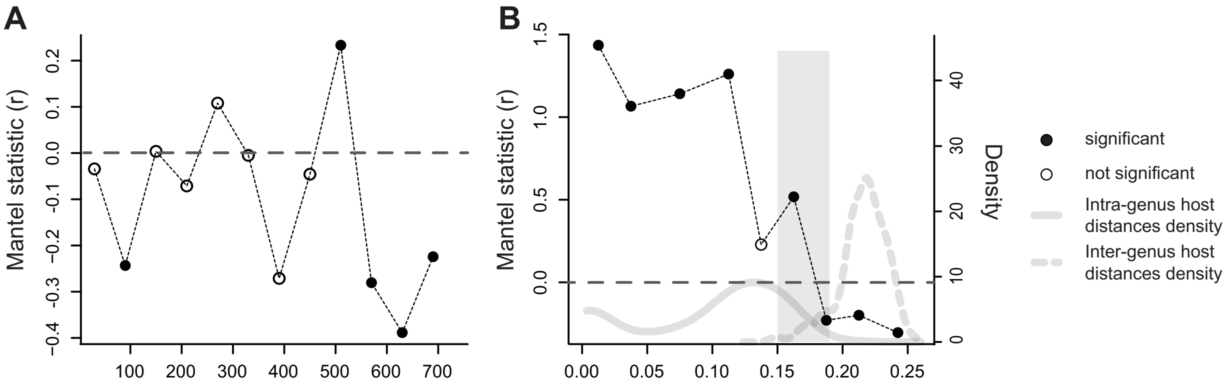
Table 4. Genomic features of the open reading frames (ORF) in the Kenyan bat CoVs and their putative transcription regulatory sequences (TRS).

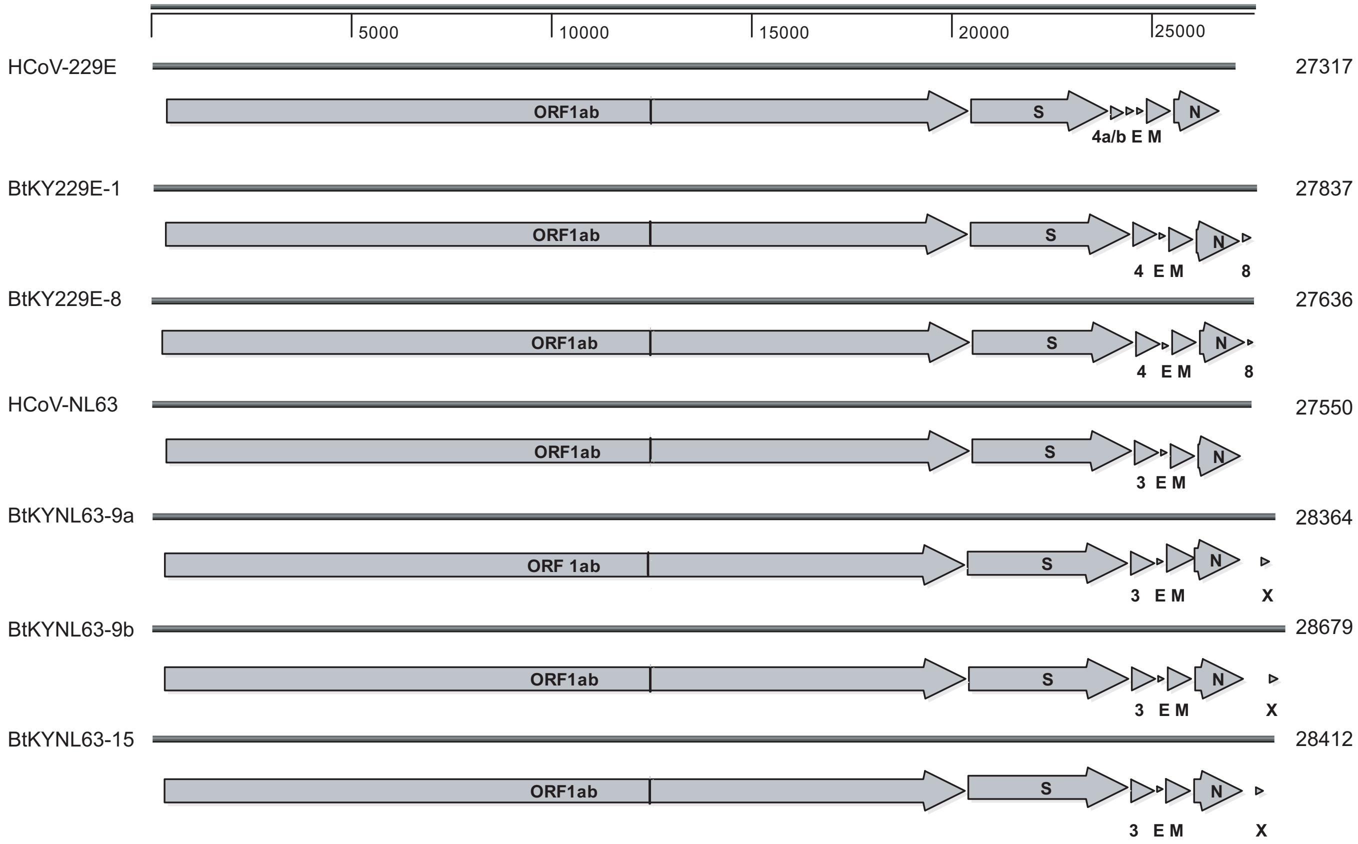
Virus		229E-like virus		NL63-like virus		
		BtKY229E-1	BtKY229E-8	BtKYNL63-9a	BtKYNL63-9b	BtKYNL63-15
Sequences (nt)		27837	27666	28363	28677	28479
GC%		38	39	39	43	43
ORF1ab (nt)	ORF size (nt)	20286	20304	20277	20349	20355
	Putative TRS	TCTCAACTAAAC(N219)AUG	TCTCAACTAAAC(N219)AUG	TCAACTAAAC(N214)AUG	CTCAACTAAAC(N215)AUG	TCTCAACTAAAC(N215)AUG
S	ORF size (nt)	4095	4095	4119	4122	4134
	Putative TRS	TCTCAACTAAATAAAAUG	UCTCAACUAAA(4)AUG	TCAACTAAAC(N1)AUG	CTCAACTAAAUG	TCAACTAAAC(N1)AUG
ORF3/4	ORF size (nt)	681	684	684	684	684
	Putative TRS	TCAACTAAAC(N37)AUG	TCAACTAAAC(N37)AUG	TCAACTAAAC(N37)AUG	TCAACTAAAC(N37)AUG	CAACUAAAC(N37)AUG
E	ORF size (nt)	234	234	234	234	234
	Putative TRS	TCTCAACTAAAC(N151)AUG	TCTTCAATGTAAC(N281)AUG	TTATAAC(N79)AUG	TCTGCTAAC(N151)AUG	TCTGATAAC(N151)AUG
M	ORF size (nt)	681	681	693	681	684
	Putative TRS	CTAAACTAAAC(N4)AUG	CTAAACTAAAC(N4)AUG	CTAAAC(N6)AUG	TCTAAACTAAAC(N4)AUG	UCUAAACUAAA(N4)AUG
N	ORF size (nt)	1161	1200	1225	1302	1302
	Putative TRS	TTAATCTAAAC(N11)AUG	ATCTAAAC(N11)AUG	TCTAAACTAAAC(N3)AUG	CTAAACCAAAC(N4)AUG	UCUAAACUAAAC(N4)AUG
ORF8/ ORFx	ORF size (nt)	288	198	287	291	270
	Putative TRS	UCAACUAAAAC(1)AUG	UCAACUAAAAC(4)AUG	CAAAACCUAAC(N12)AUG	TCAACTAAAC(N567)AUG	CAACUAAAC(N234)AUG

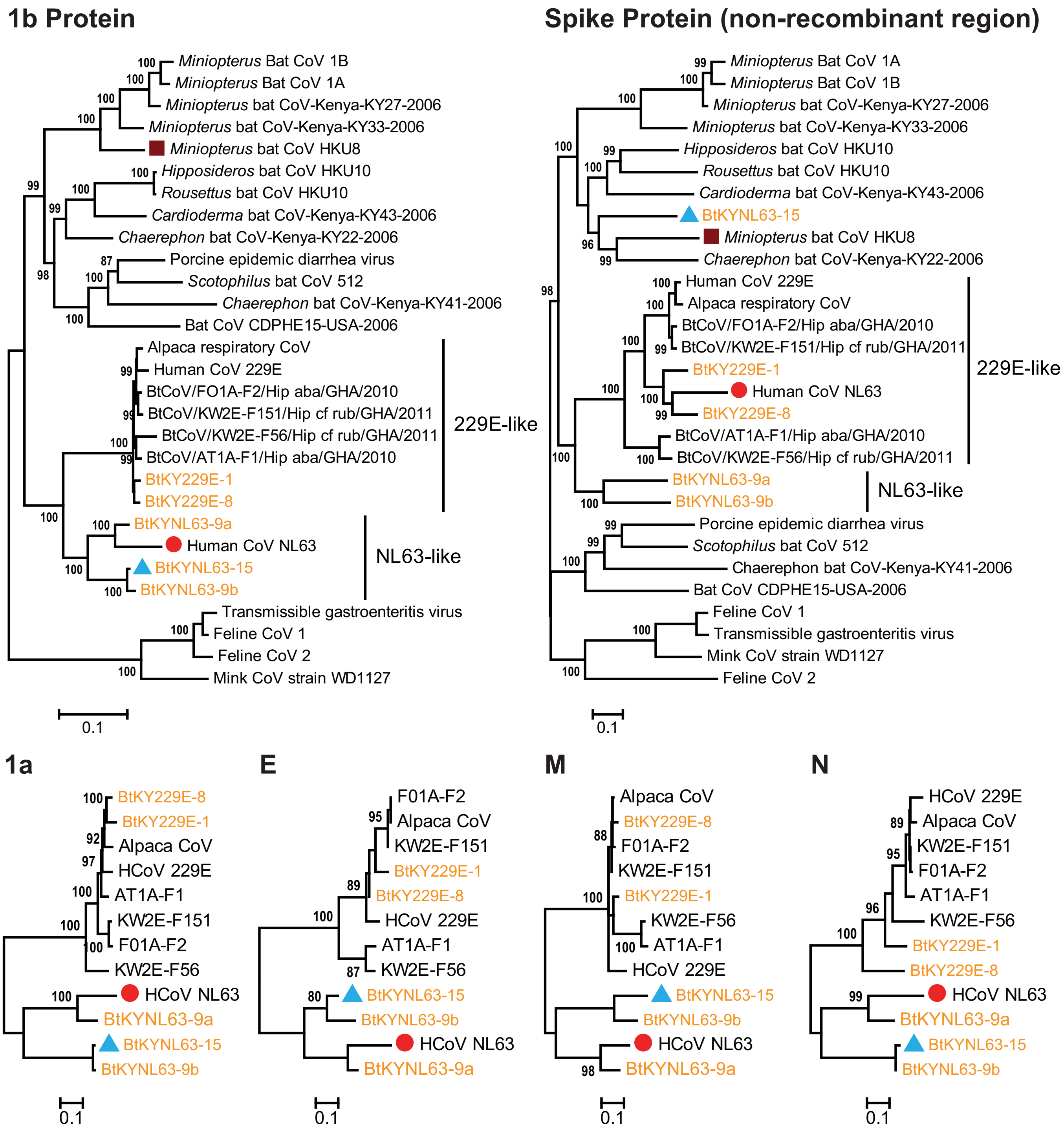

Ethiopia
Uganda
N
1, 2
27, 36-38
5
19
Equator
18
4
20
45
7
40, 41
42
25
10
23
24
33
Tanzania
26
22
35
43
16, 21
13, 14
44

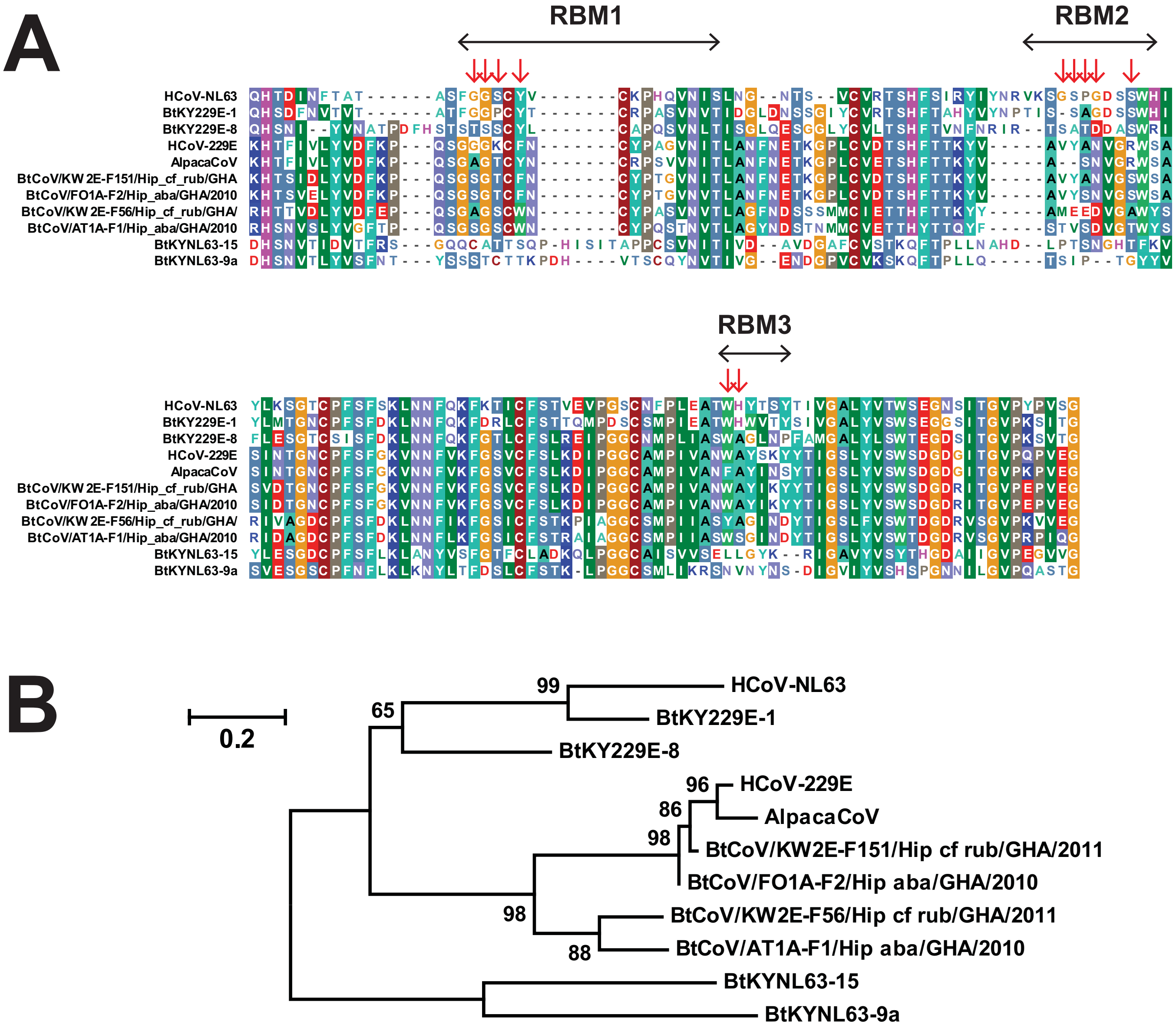

Rousettus
Rousettus
Rhinolophus
L1 Minopterus (n=55)
90
100
91
Hipposideros
L2 Minopterus (n=36)
82
99
L3 Rhinolophus (n=8)
L4 Cardioderma (n=2)
88
L5 Rhinolophus (n=3)
85
Rousettus
L6 Rhinolophus (n=11)
91
L7 Rousettus (Oral samples, n=4)
80
Epomophorus
L8 Otomops (n=5)
95
L9 Otomops (n=6)
98
L10 Chaerephon (n=12)
91
Epomophorus
L11 Scotophilus (n=1)
99
L12 Triaenops (n=8)
86
L13 Triaenops (n=1)
81
100
L14 Hipposideros (n=16)
Alpha-CoV
L15 Chaerephon (n=1)
L16 Rhinolophus (n=3)
98
Beta-CoV
100
L17 Rousettus (n=12)
100
L18 Epomophorus (n=11)
100
95
L19 Eidolon (n=38)
0.5
L20 Epomophorus (n=1)

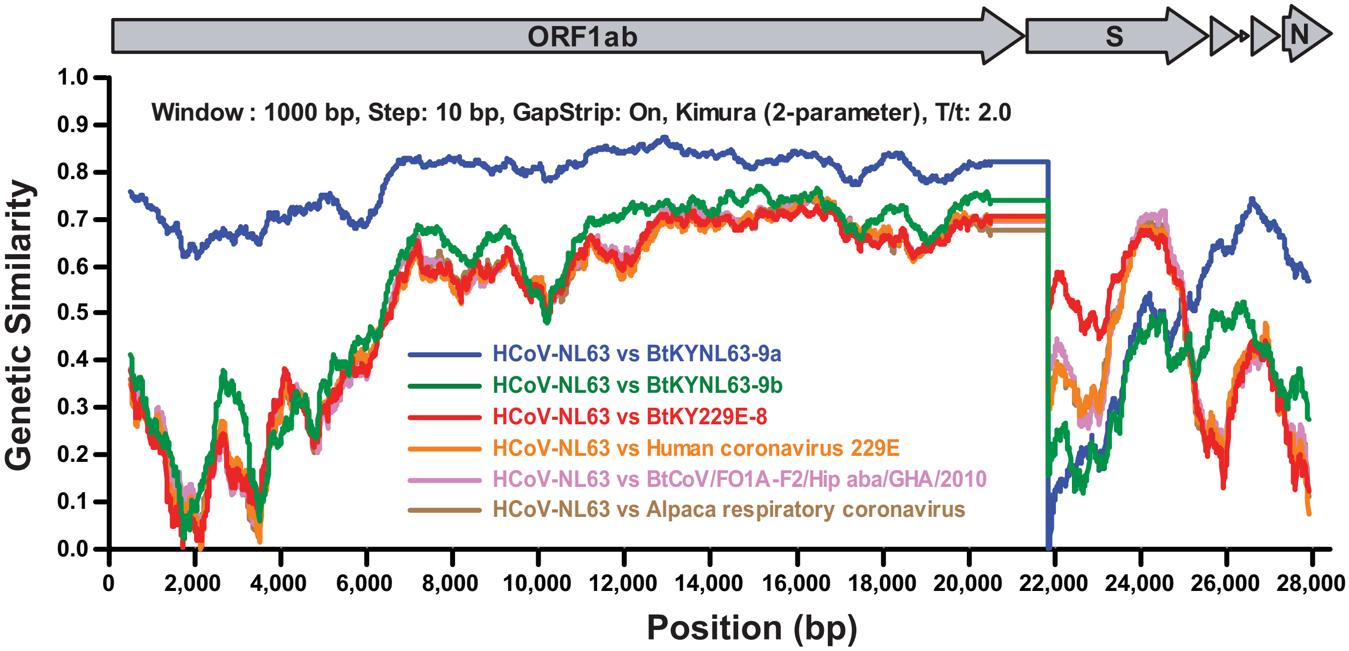

Alpha-CoV
95
HQ728484 KEN Miniopterus sp
BtKY195 Miniopterus minor
99
BtKY130 Miniopterus minor
L1
92
GU190240 BGR Miniopterus schreibersii
DQ666337 CHN Miniopterus magnater
99
BtKY258 Miniopterus minor
BtKY224 Miniopterus minor
L2
95
DQ249228 CHN Miniopterus pusillus
DQ666339 CHN Miniopterus magnater
EU834956 AUT Miniopterus australis
87
99
KF515987 NZL Mystacina tuberculata
EU834951 AUT Myotis macropus
HQ184058 ESP Pipistrellus kuhlii
91
91
KJ473809 CHN Nyctalus velutinus
KF843855 ZAF Neoromicia cf capensis
GU190239 BGR Nyctalus leisleri
95
82
KT345294 FRA Pipistrellus pipistrellus
JQ731775 CRI Anoura geoffroyi
JQ731784 PAN Artibeus jamaicensis
99
HQ728480 KEN Cardioderma cor
BtKY242 Cardioderma cor
L4
90
99
BtKY236 Rhinolophus landeri
L3
100
GU190233 BGR Rhinolophus fer.
85
BtKY244 Rhinolophus hilderbrandtii
BtKY70 Rhinolophus sp
L5-6
KU343197 CHN Rhinolophus affinis
DQ648854 CHN Rhinolophus sp
99
BtKY117 Rousettus aegyptiacus
L7
JQ989272 CHN Hipposideros sp
JQ989270 CHN Rousettus sp
91
98
DQ648823 CHN Scotophilus kuhlii
87
BtKY280 Scotophilus dingani
L11
KF569988 CHN Myotis davidii
KF294382 CHN Myotis davidii
87
JF440355 GBR Myotis nattereri
JF440350 GBR Myotis daubentonii
94
EU375868 DEU Pipistrellus pygmaeus
97
EU375864 DEU Pipistrellus nathusii
HM368166 DEU Myotis myotis
89
DQ249224 CHN Myotis ricketti
EF544565 USA Myotis occultus
EF185992 Porcine epidemic diarrhea virus
90
KU182966 CHN Murina leucogaster
97
KF294376 CHN Murina leucogaster
93
97
EF544566 USA Eptesicus fuscus
98
JX537914 USA Eptesicus fuscus
98
JQ731799 BRA Molossus rufus
KF569991 CHN Myotis davidii
KJ473806 CHN Myotis ricketti
94
93
HQ184050 ESP Myotis blythii
90
HQ336976 USA Myotis volans
JX537913 USA Perimyotis subflavus
KC110771 BRA NA
92
BtKY273 Otomops martiensseni
L8
BtKY147 Chaerophon sp
98
HQ728486 KEN Chaerophon sp
98
BtKY275 Otomops martiensseni
L9-10
88
BtKY270 Chaerephon sp
BtKY204 Epomophorus labiatus
82
85
JQ410000 Alpaca respiratory coronavirus
BtKY229E-1 Hipposideros sp
NC002645 Human coronavirus 229E
L14
94
BtKY229E-8 Hipposideros vittatus
JX174639 GAB Hipposideros caffer
KT253270 GHA Hipposideros abae
100
FJ710045 GHA Hipposideros sp
92
FJ710044 GHA Hipposideros sp
97
NC005831 Human coronavirus NL63
BtKYNL63-9a Triaenops afer
BtKYNL63-15 Triaenops afer
L12-13
BtKYNL63-9b Triaenops afer
99
99
HQ728481 KEN Chaerophon sp
BtKY210 Chaerephon sp
L15
JQ731790 CRI Carollia perspicillata
JQ731788 PAN Artibeus lituratus
96
JQ731782 PAN Phyllostomus discolor
EU769558 TTO Glossophaga soricina
88
96
Alphacoronavirus1, include:
Feline coronavirus, Transmissible gastroenteritis virus, Rodent CoVs etc
0.5
Beta-CoV
GQ153539 CHN Rhinolophus sinicus
97
DQ412042 CHN Rhinolophus ferrumequinum
DQ412043 CHN Rhinolophus macrotis
AY304488 SARS Coronavirus SZ16
84
89
DQ071615 CHN Rhinolophus pearsoni
BtKY237 Rhinolophus hilderbrandtii
L16
100
GU190227 BGR Rhinolophus mehelyi
GU190231 BGR Rhinolophus ferrumequinum
95
GQ404795 SVN Rhinolophus hipposideros
94
EU834950 AUT Rhinonycteris aurantius
KU343200 CHN Hipposideros pomona
KF636752 CHN Hipposideros pratti
86
HQ166910 NGA Hipposideros commersoni
JX174638 GAB Hipposideros caffer
97
FJ710050 GHA Hipposideros sp
94
FJ710043 GHA Hipposideros sp
JX899384 GHA Nycteris sp
99
KC545386 DEU Erinaceus europaeus
KC545383 DEU Erinaceus europaeus
99
KC243390 ROU Pipistrellus pygmaeus
KC243392 UKR Pipistrellus nathusii
KU740200 MERS CoV/camel/Egypt/NRCE-NC163/2014
JX869059 MERS Coronavirus EMC2012
90
100
DQ648794 CHN Tylonycteris pachypus
85
KU182965 CHN Myotis daubentonii
DQ648819 CHN Pipistrellus pipistrellus
EF065509 CHN Pipistrellus abramus
97
DQ249221 CHN Pipistrellus abramus
88
DQ648809 CHN Pipistrellus abramus
KC633197 CRI Carollia perspicillata
90
KC886322 MEX Pteronotus davyi
98
KC633195 CRI Pteronotus parnellii
BtKY92 Eidolon helvum
94
HQ728482 KEN Eidonlon sp
BtKY89 Eidolon helvum
L19
89
KU131211 NGA Eidolon helvum
BtKY54 Epomophorus labiatus
BtKY182 Epomophorus labiatus
L18 & 20
BtKY55 Epomophorus labiatus
99
BtKY234 Epomophorus labiatus
98
89
AB539081 PHL Cynopterus brachyotis
81
KU182962 CHN Cynopterus sphinx
AB683970 PHL Ptenochirus jagori
92
AB918719 IDN Dobsonia moluccensis
HM211100 CHN Rousettus leschenaulti
BtKY76 Rousettus aegyptiacus
L17
90
BtKY221 Rousettus aegyptiacus
96
HQ728483 KEN Rousettus sp
87
EF065513 CHN Rousettus leschenaultii
93
EF065516 CHN Rousettus leschenaultii
100
Betacoronavirus lineage A, include:
Murine hepatitis virus, Bovine coronavirus,
Rabbit coronavirus, OC43 etc
0.5


A
Mantel statistic (r)
0.2
0.1
0.0
−0.1
−0.2
−0.3
−0.4
100
200
300
400
500
600
700
B
Mantel statistic (r)
1.5
1.0
0.5
0.0
0.00
0.05
0.10
0.15
0.20
0.25
Density
40
30
20
10
0
● significant
○ not significant
Intra-genus host distances density
Inter-genus host distances density


5000
10000
15000
20000
25000
HCoV-229E
27317
ORF1ab
S
N
4a/b E M
BtKY229E-1
27837
ORF1ab
S
N
4 E M
8
BtKY229E-8
27636
ORF1ab
S
N
4 E M
8
HCoV-NL63
27550
ORF1ab
S
N
3 E M
BtKYNL63-9a
28364
ORF 1ab
S
N
3 E M
X
BtKYNL63-9b
28679
ORF1ab
S
N
3 E M
X
BtKYNL63-15
28412
ORF1ab
S
N
3 E M
X


1b Protein
100 Miniopterus Bat CoV 1B
100 Miniopterus Bat CoV 1A
100 Miniopterus bat CoV-Kenya-KY27-2006
100 Miniopterus bat CoV-Kenya-KY33-2006
Miniopterus bat CoV HKU8
99
100 Hipposideros bat CoV HKU10
100 Rousettus bat CoV HKU10
99 Cardioderma bat CoV-Kenya-KY43-2006
Chaerephon bat CoV-Kenya-KY22-2006
87 Porcine epidemic diarrhea virus
98
100 Scotophilus bat CoV 512
Chaerephon bat CoV-Kenya-KY41-2006
100 Bat CoV CDPHE15-USA-2006
Alpaca respiratory CoV
99 Human CoV 229E
BtCoV/FO1A-F2/Hip aba/GHA/2010
99 BtCoV/KW2E-F151/Hip cf rub/GHA/2011
100 BtCoV/KW2E-F56/Hip cf rub/GHA/2011
99 BtCoV/AT1A-F1/Hip aba/GHA/2010
BtKY229E-1
BtKY229E-8
229E-like
100
100 BtKYNL63-9a
Human CoV NL63
100 BtKYNL63-15
100 BtKYNL63-9b
NL63-like
Transmissible gastroenteritis virus
100 Feline CoV 1
Feline CoV 2
100 Mink CoV strain WD1127
0.1
Spike Protein (non-recombinant region)
99 Miniopterus Bat CoV 1A
100 Miniopterus Bat CoV 1B
100 Miniopterus bat CoV-Kenya-KY27-2006
Miniopterus bat CoV-Kenya-KY33-2006
100
99 Hipposideros bat CoV HKU10
100 Rousettus bat CoV HKU10
Cardioderma bat CoV-Kenya-KY43-2006
BtKYNL63-15
96 Miniopterus bat CoV HKU8
99 Chaerephon bat CoV-Kenya-KY22-2006
98
100 Human CoV 229E
100 Alpaca respiratory CoV
BtCoV/FO1A-F2/Hip aba/GHA/2010
100
99 BtCoV/KW2E-F151/Hip cf rub/GHA/2011
BtKY229E-1
100
Human CoV NL63
100
99 BtKY229E-8
BtCoV/AT1A-F1/Hip aba/GHA/2010
100 BtCoV/KW2E-F56/Hip cf rub/GHA/2011
229E-like
100
BtKYNL63-9a
100 BtKYNL63-9b
NL63-like
99 Porcine epidemic diarrhea virus
99 Scotophilus bat CoV 512
100 Chaerephon bat CoV-Kenya-KY41-2006
Bat CoV CDPHE15-USA-2006
Feline CoV 1
100 Transmissible gastroenteritis virus
Mink CoV strain WD1127
100 Feline CoV 2
0.1
1a
100 BtKY229E-8
BtKY229E-1
92 Alpaca CoV
97 HCoV 229E
100 AT1A-F1
100 KW2E-F151
100 F01A-F2
KW2E-F56
100 HCoV NL63
BtKYNL63-9a
BtKYNL63-15
BtKYNL63-9b
0.1
E
F01A-F2
95 Alpaca CoV
KW2E-F151
BtKY229E-1
89 BtKY229E-8
100 HCoV 229E
AT1A-F1
87 KW2E-F56
80 BtKYNL63-15
BtKYNL63-9b
HCoV NL63
BtKYNL63-9a
0.1
M
Alpaca CoV
BtKY229E-8
88 F01A-F2
KW2E-F151
100 BtKY229E-1
KW2E-F56
100 AT1A-F1
HCoV 229E
BtKYNL63-15
BtKYNL63-9b
HCoV NL63
98 BtKYNL63-9a
0.1
N
HCoV 229E
89 Alpaca CoV
95 KW2E-F151
F01A-F2
AT1A-F1
96 KW2E-F56
100 BtKY229E-1
BtKY229E-8
99 HCoV NL63
BtKYNL63-9a
BtKYNL63-15
BtKYNL63-9b
0.1


A
RBM1
RBM2
HCoV-NL63 QHTDINFTAT------ASFGGSCYV-------CKPHQVNISLNG--NTS--VCVRTSHFSIRYIYNRVKSGSPGDSSWHI
BtKY229E-1 QHSDFNVTVT------ATFGGPCYT-------CRPASVNVTIDGLDNSSGIYCVRTSHFTAHYVYNPTIS-SAGDSSWHI
BtKY229E-8 QHSNI--YVNATPDFHSTSTSSCYL-------CAPQSVNLTISGLQESGGLYCVLTSHFTVNFNRIR--TSATDDASWRI
HCoV-229E KHTFIVLYVDFKP---QSGGGKCFN-------CYPAGVNITLANFNETKGPLCVDTSHFTTKYV-----AVYANVGRWSA
AlpacaCoV KHTFIVLYVDFKP---QSGAGTCYN-------CRPSVVNITLANFNETKGPLCVETSHFTTKYV-----A--SNVGRWSA
BtCoV/KW 2E-F151/Hip_cf_rub/GHA KHTSIDLYVDFKP---QSGSGTCFN-------CYPTGVNITLANFNETKGPLCVDTSHFTTKYV-----AVYANVGSWSA
BtCoV/FO1A-F2/Hip_aba/GHA/2010 KHTSVELYVDFKP---QSGSGTCFN-------CYPTGVNVTLANFNETKGPLCVDTSHFTTKYV-----AVYSNVGSWSA
BtCoV/KW 2E-F56/Hip_cf_rub/GHA/ RHTTVDLYVDFEP---QSGAGSCWN-------CYPASVNVTLAGFNDSSSMMCIETTHFTTKYY-----AMEEDVGAWYS
BtCoV/AT1A-F1/Hip_aba/GHA/2010 RHSNVSLYVGFTP---QSGSGSCWN-------CYPQSTNVTLAGYNDSTNMMCVETTHYTTQYF-----STVSDVGTWYS
BtKYNL63-15 DHSNVTIDVTFRS---GQQCATTSQP-HISITAPPCSVNITIVD--AVDGAFCVSTKQFTPLLNAHD--LPTSNGHTFKV
BtKYNL63-9a DHSNVTLYVSFNT---YSSSTCTTKPDH----VTSCQYNVTIVG--ENDGPVCVKSKQFTPLLQ-----TSIP--TGYYV
RBM3
HCoV-NL63 YLKSGTCPFSFSKLNNFQKFKTICFSTVEVPGSCNFPLEATWHYTSYTIVGALYVTWSEGNSITGVPYPVSG
BtKY229E-1 YLMTGNCPFSFDKLNNFQKFDRLCFSTTQMPDSCSMPIEATWHWVTYSIVGALYVSWSEGGSITGVPKSITG
BtKY229E-8 FLESGTCSISFDKLNNFQKFGTLCFSLREIPGGCNMPLIASWAGLNPFAMGALYLSWTEGDSITGVPKSVTG
HCoV-229E SINTGNCPFSFGKVNNFVKFGSVCFSLKDIPGGCAMPIVANWAYSKYYTIGSLYVSWSDGDGITGVPQPVEG
AlpacaCoV SINTGNCPFSFGKVNNFVKFGSVCFSLKDIPGGCAMPIVANFAYINSYTIGSLYVSWSDGDGITGVPKPVEG
BtCoV/KW 2E-F151/Hip_cf_rub/GHA SVDTGNCPFSFGKVNNFVKFGSVCFSLKDIPGGCAMPIVANWAYIKYYTIGSLYVSWSDGDRITGVPEPVEG
BtCoV/FO1A-F2/Hip_aba/GHA/2010 SIDTGNCPFSFGKVNNFVKFGSVCFSLKDIPGGCAMPIVANWAYIKYYTIGSLYVSWSDGDRITGVPEPVEG
BtCoV/KW 2E-F56/Hip_cf_rub/GHA/ RIVAGDCPFSFDKLNNFIKFGSICFSTKPIAGGCSMPIIASYAGINDYTIGSLFVSWTDGDRVSGVPKVVEG
BtCoV/AT1A-F1/Hip_aba/GHA/2010 RIDAGDCPFSFDKLNNFLKFGSICFSTRAIAGGCSMPIIASWSGINDYTIGSLYVSWTDGDRVSGVPRPIQG
BtKYNL63-15 YLESGDCPFSFLKLANYVSFGTFCLADKQLPGGCAISVVSELLGYK--RIGAVYVSYTHGDAIIGVPEGVVG
BtKYNL63-9a SVESGSCPFNFLKLKNYLTFDSLCFSTK-LPGGCSMLIKRSNVNYNS-DIGVIYVSHSPGNNILGVPQASTG
B
0.2
99
65
HCoV-NL63
BtKY229E-1
BtKY229E-8
96
86
98
HCoV-229E
AlpacaCoV
BtCoV/KW2E-F151/Hip cf rub/GHA/2011
BtCoV/FO1A-F2/Hip aba/GHA/2010
98
88
BtCoV/KW2E-F56/Hip cf rub/GHA/2011
BtCoV/AT1A-F1/Hip aba/GHA/2010
BtKYNL63-15
BtKYNL63-9a


ORF1ab
S
N
Window : 1000 bp, Step: 10 bp, GapStrip: On, Kimura (2-parameter), T/t: 2.0
Genetic Similarity
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
HCoV-NL63 vs BtKYNL63-9a
HCoV-NL63 vs BtKYNL63-9b
HCoV-NL63 vs BtKY229E-8
HCoV-NL63 vs Human coronavirus 229E
HCoV-NL63 vs BtCoV/FO1A-F2/Hip aba/GHA/2010
HCoV-NL63 vs Alpaca respiratory coronavirus
0
2,000
4,000
6,000
8,000
10,000
12,000
14,000
16,000
18,000
20,000
22,000
24,000
26,000
28,000
Position (bp)