Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Virus Research 167 (2012) 67–73

Contents lists available at SciVerse ScienceDirect

Virus Research

journal homepage: www.elsevier.com/locate/virusres

Genomic characterization of seven distinct bat coronaviruses in Kenya☆

Ying Tao[a], Kevin Tang[b], Mang Shi[a], Christina Conrardy[a], Kenneth S.M. Li[c], Susanna K.P. Lau[c], Larry J. Anderson[a], Suxiang Tong[a,*]

[a] Division of Viral Diseases, Centers for Disease Control and Prevention, Atlanta, GA 30333, United States
[b] Division of Scientific Resources, Biotechnology Core Facility, Centers for Disease Control and Prevention, Atlanta, GA 30333, United States
[c] Department of Microbiology, University of Hong Kong, Hong Kong

ARTICLE INFO

Article history:
Received 9 December 2011
Received in revised form 13 April 2012
Accepted 18 April 2012
Available online 26 April 2012

Keywords:
Coronavirus
Novel
Genome sequence
Bat
Kenya

ABSTRACT

To better understand the genetic diversity and genomic features of 41 coronaviruses (CoVs) identified from Kenya bats in 2006, seven CoVs as representatives of seven different phylogenetic groups identified from partial polymerase gene sequences, were subjected to extensive genomic sequencing. As a result, 15–16 kb nucleotide sequences encoding complete RNA dependent RNA polymerase, spike, envelope, membrane, and nucleocapsid proteins plus other open reading frames (ORFs) were generated. Sequences analysis confirmed that the CoVs from Kenya bats are divergent members of *Alphacoronavirus* and *Betacoronavirus* genera. Furthermore, the CoVs BtKY22, BtKY41, and BtKY43 in *Alphacoronavirus* genus and BtKY24 in *Betacoronavirus* genus are likely representatives of 4 novel CoV species. BtKY27 and BtKY33 are members of the established bat CoV species in *Alphacoronavirus* genus and BtKY06 is a member of the established bat CoV species in *Betacoronavirus* genus. The genome organization of these seven CoVs is similar to other known CoVs from the same groups except for differences in the number of putative ORFs following the N gene. The present results confirm a significant diversity of CoVs circulating in Kenya bats. These Kenya bat CoVs are phylogenetically distant from any previously described human and animal CoVs. However, because of the examples of host switching among CoVs after relatively minor sequence changes in S1 domain of spike protein, a further surveillance in animal reservoirs and understanding the interface between host susceptibility is critical for predicting and preventing the potential threat of bat CoVs to public health.

Published by Elsevier B.V.

1. Introduction

Coronaviruses (CoVs) are large, enveloped viruses containing linear, positive-sense, single-stranded RNA genomes. Their genomes range approximately from 27- to 32-kb in length and contain 7–14 open reading frames (ORFs) (Woo et al., 2009a). Six major ORFs encoding polymerase complex (ORF1a and ORF1b), spike glycoprotein (S), envelope protein (E), membrane glycoprotein (M), and nucleocapsid protein (N) are present in all CoVs (Poon et al., 2005). In addition, up to seven putative accessory ORFs and one ORF encoding hemagglutinin-esterase glycoprotein (HE) are interspersed between the six major ORFs. The numbers and sizes of these accessory ORFs differ markedly among CoVs (Woo et al., 2009a).

CoVs have been identified from a broad range of birds and mammals including humans in which they can cause respiratory, enteric, hepatic and neurologic diseases of varying severity (Weiss and Navas-Martin, 2005). CoVs in the subfamily *Coronavirinae* are classified into three genera, *Alphacoronavirus*, *Betacoronavirus*, and *Gammacoronavirus* (former serogroups 1–3) (de Groot et al., 2011). Alpha- and beta-coronaviruses have been exclusively isolated from mammals and majority of gamma-coronaviruses from birds. CoVs of a distinctive lineage were recently detected from birds and pigs (Chu et al., 2011; Woo et al., 2009b, 2012) and have been proposed to belong to a new genus, provisionally named *Deltacoronavirus* (de Groot et al., 2011). The finding that the outbreak of severe acute respiratory syndrome (SARS) in early 2003 was caused by a novel CoV (SARS-CoV) has boosted interest in the search for novel CoVs in humans and animals. At least 30 previously unrecognized distinctive CoVs from human and various animal reservoirs were reported during recent years, including SARS-related CoVs and CoVs from all genera in the subfamily *Coronavirinae* which have significantly expanded our understanding of CoV diversity and complexity (Woo et al., 2009a). Based on available data, bats appear to harbor a great diversity of CoVs. The frequency and diversity of CoV detection in bats, now in multiple continents, suggest that bats are likely a source for CoV introduction into other species globally and possibly play an important role in the ecology and evolution of CoVs.

☆ The findings and conclusions in this report are those of the authors and do not necessarily represent the views of the Centers for Disease Control and Prevention.

* Corresponding author at: CDC, 1600 Clifton Rd., MS G18, Atlanta, GA 30333, United States. Tel.: +1 404 639 1372; fax: +1 404 639 4005.
E-mail address: sot1@cdc.gov (S. Tong).

0168-1702/$ – see front matter. Published by Elsevier B.V.
http://dx.doi.org/10.1016/j.virusres.2012.04.007

Recently we reported the identification of 41 divergent CoVs in bats from Kenya, based on limited ORF1b sequences (Tong et al., 2009). These newly discovered bat CoVs were grouped into 8 different phylogenetic clusters. Of these, five clusters belonged to previously identified *Alphacoronavirus* genus, and three clusters belonged to previously identified *Betacoronavirus* genus, including a SARS-related CoV lineage. In the present study, we expand our sequence data for seven CoVs, representing 7 of the 8 distinctive clusters we identified in Kenya bats during 2006 summer (Tong et al., 2009). The sample representing the eighth cluster of a SARS-related CoV was a weak positive and had limited specimen amount, therefore further sequencing studies were not included in this analysis. The purpose of our study was to further characterize the genomes and refine the phylogenetic relationships of these seven CoVs with other CoVs, based on the ORFs 1b, S, E, M, and N.

2. Materials and methods

2.1. Bat sampling and RNA extraction

Kenya was chosen as a major comparative Old World study location in Africa as part of the CDC Global Disease Detection program. Detailed information on bat capture and sampling is available in the previous publication (Tong et al., 2009). The protocols for animal capture and use were approved by the CDC Animal Institutional Care and Use Committee and the Ethics and Animal Care and Use Committee of the Kenya Wildlife Service (Nairobi, Kenya). In brief, representative samples at each site were collected from bats of available species, including adult and juvenile of both sexes. After euthanasia, a complete necropsy was performed in compliance with the approved field protocols. Samples included blood, various organs (liver, lung, and kidney), rectal and oral swabs.

In this study, seven CoV-positive rectal swabs were selected as representatives of the seven different phylogenetic groups (Tong et al., 2009) for extensive genome sequencing. These are *Rousettus* bat coronavirus/Kenya/KY06/2006 (BtKY06), *Chaerephon* bat coronavirus/Kenya/KY22/2006 (BtKY22), *Eidolon* bat coronavirus/Kenya/KY24/2006 (BtKY24), *Miniopterus* bat coronavirus/Kenya/KY27/2006 (BtKY27), *Miniopterus* bat coronavirus/Kenya/KY33/2006 (BtKY33), *Chaerephon* bat coronavirus/Kenya/KY41/2006 (BtKY41), and *Cardioderma* bat coronavirus/Kenya/KY43/2006 (BtKY43). BtKY43 was not described previously, but represents a group of 4 Kenya bat CoVs (BtKY03, BtKY12, BtKY13, and BtKY29) (Tong et al., 2009). Total nucleic acids (TNA) were extracted by using the QIAamp MinElute Virus Spin Kit (Qiagen, Santa Clarita, CA) according to the manufacturer's instructions from 200 μl of phosphate buffered saline suspension of the rectal swab and homogenized organ tissues (liver, lung, and/or kidney) of each bat except for bats BtKY33 and BtKY43 whose organ tissues were not available. The TNA was eluted in 80 μl DEPC-treated water and then stored at −80 °C.

2.2. Reverse transcription-PCR (RT-PCR)

Each CoV-positive result on the rectal swab included in this study was repeated from different TNA aliquots. The presence of CoV RNA in organ tissues of these bats was determined using the pan CoV RT-PCR assays as described previously (Tong et al., 2009) and the sequence specific and/or group specific CoV RT-PCR assays (Table S1). The RT-PCR were performed as described previously (Tong et al., 2009). Standard precautions were taken to avoid cross-contamination of samples before and after RNA extraction and amplification. Purified DNA amplicons were sequenced with the RT-PCR primers on an ABI Prism 3130 automated capillary sequencer using a BigDye Terminator v3.1 Cycle Sequencing kit (Applied Biosystems, Carlsbad, CA).

2.3. Partial genome sequencing

High throughput 454 pyrosequencing on CoV RNA-positive bat samples was initially attempted, but failed to acquire any CoV-associated reads due to lower sensitivity. Therefore the RT-PCR-amplicon sequencing by Sanger chain-termination method was chosen in this study. Each of the seven contiguous sequences was obtained by using 4–6 pairs of semi-nested or nested consensus degenerate group specific primers and 4–7 pairs of semi-nested or nested sequence-specific bridging primers which generated a series of 8–13 overlapping fragments covering 15–16kb genomic sequences at the 3′ end (Table S1). The other half genome sequence containing the ORF1a, was not recovered in this analysis due to the limited amount of rectal swab samples. Consensus degenerate primers of each group were designed from conserved sequences of known members of the corresponding sequence group or its close group based on CODEHOP strategy (Rose et al., 1998). The 3′ end of genome sequence was determined using the 3′ RACE kit (Roche, Indianapolis, IN) according to the manufacturer's instructions. Semi-nested or nested primers were used to improve the PCR sensitivity. When nested primers were not available, the PCR product was re-amplified using the same RT-PCR primers. The RT-PCR reactions were performed with SuperScript III one-step RT-PCR High Fidelity kit (Invitrogen, San Diego, CA) according to the manufacturer's instructions, and the second round RCR reactions were performed with AccuPrime Taq DNA polymerase High Fidelity kit (Invitrogen, San Diego, CA). The RT-PCR products were visualized on 1% agarose gels containing 0.5 μg/mL of ethidium bromide, and purified by QIAquick PCR purification kit (QIAGEN, Santa Clarita, CA). The RT-PCR amplicons for each sample were first sequenced with the consensus degenerate RT-PCR primers in both directions, and then the remaining internal gaps and 3′ end genome were sequenced with sequence-specific bridging primers in both directions as described previously. The genomic sequences (ORF1b, S, ORF3, E, M, and N) of BtKY22, BtKY33, BtKY27, BtKY41, BtKY43, BtKY06, and BtKY24 were deposited in NCBI GenBank (HQ728480–HQ728486).

2.4. Sequence analysis

Sequences were assembled in Sequencher (Genecodes, Ann Arbor, MI). Each putative ORF was predicted using the NCBI ORF finder (http://www.ncbi.nlm.nih.gov/gorf/gorf.html). N-glycosylation sites were predicted using NetNGlyc 1.0 Server (http://www.cbs.dtu.dk/services/NetNGlyc/). BLAST analyses were performed against NCBI non-redundant protein database (Altschul et al., 1990) and against the Conserved Domain Database for protein classification (CDD) (Marchler-Bauer et al., 2005) to characterize the putative ORFs.

Alignments of the seven Kenya bat CoV gene sequences with a representative set of 43 other CoV sequences, available in the public domain, were performed using the MUSCLE v3.6 (Edgar, 2004). We constructed maximum likelihood trees for each gene alignment (ORF1b, S, E, M, and N) in MEGA software package v5.0 (Tamura et al., 2011) with 1000 bootstrap replications. We used General-Time-Reversible nucleotide (nt) substitution model with 4 categories of gamma distributed rate heterogeneity and a proportion of invariant sites (GTR + γ4 + I). To identify potential recombination events of the seven Kenya bat CoVs, three methods implemented in recombination detection program RDP version 2 (Martin et al., 2005) were used, including MaxChi (Smith, 1992), Chimaera (Posada et al., 2002), and Geneconv (Padidam et al., 1999).

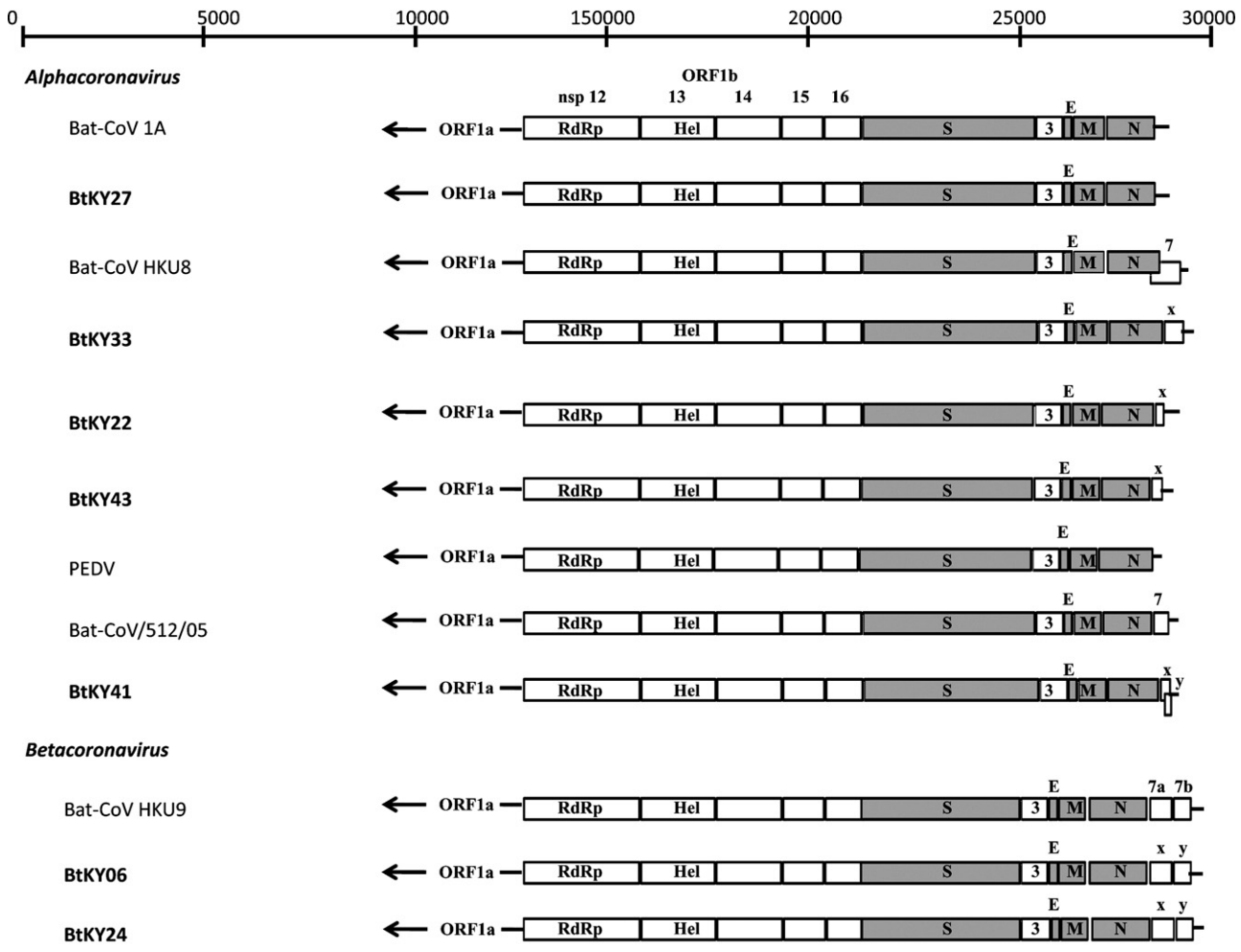

Fig. 1. Schematic representation of the genome organization of Kenya bat CoVs and representative alpha- and beta-coronaviruses. Shaded boxes represent open reading frames (ORFs) encoding structural proteins and unshaded boxes represent those encoding nonstructural proteins.

Table 1
Genomic features of open reading frames from seven bat coronaviruses and their putative transcription regulatory sequences (TRS).

Genus	Alphacoronavirus					Betacoronavirus	
Virus	BtKY27	BtKY33	BtKY22	BtKY41	BtKY43	BtKY24	BtKY06
Sequences[a] (nt)	15314	15908	15480	15578	15474	16186	16201
ORF1a (nt)	NA[b]	NA[b]	NA[b]	NA[b]	NA[b]	NA[b]	NA[b]
ORF1b (nt)	8022	8025	8025	8025	8022	8040	8067
S							
ORF size (nt)	4128	4152	4071	4161	4095	3795	3837
Putative TRS	CUAAAU	CUAAAU	CUAAAU	CGAAAU	CUAAAU	ACGAAC	ACGAAC
ORF3							
ORF size (nt)	660	672	672	687	660	717	663
Putative TRS	CGUUAC	CGUUAC	CGUUAC	CUAGAC	CUAAAC	ACGAAC	ACGAAC
E							
ORF size (nt)	225	225	225	231	243	228	249
Putative TRS	CUAUAC	CUUUAC	CUCUAC	CUAGAC	CUUUAC	UCGAAC	UCGAAC
M							
ORF size (nt)	768	780	684	690	684	666	669
Putative TRS	CUAAAC	CUAAAC	CUAAAC	CUAAAC	CUAAAC	ACGAAC	ACGAAC
N							
ORF size (nt)	1185	1296	1263	1227	1182	1404	1407
Putative TRS	CUAAAC	CUAAAC	CUAAAC	CUAAAU	CUAAAC	ACGAAC	ACGAAC
ORFx							
ORF size (nt)		486	231	264	288	567	558
Putative TRS		CAAAAU	CUAAAC	CUAAAU	CUAAAC	ACGAAC	ACGAAC
ORFy							
ORF size (nt)				195		432	450
Putative TRS				CUAAAC		ACGAAC	ACGAAC
3′ UTR (nt, excluding poly A)	269	222	251	222	221	231	217

[a] Partial genome sequence starts from the first nt position in the RdRp to the end of genome.
[b] NA, not available.

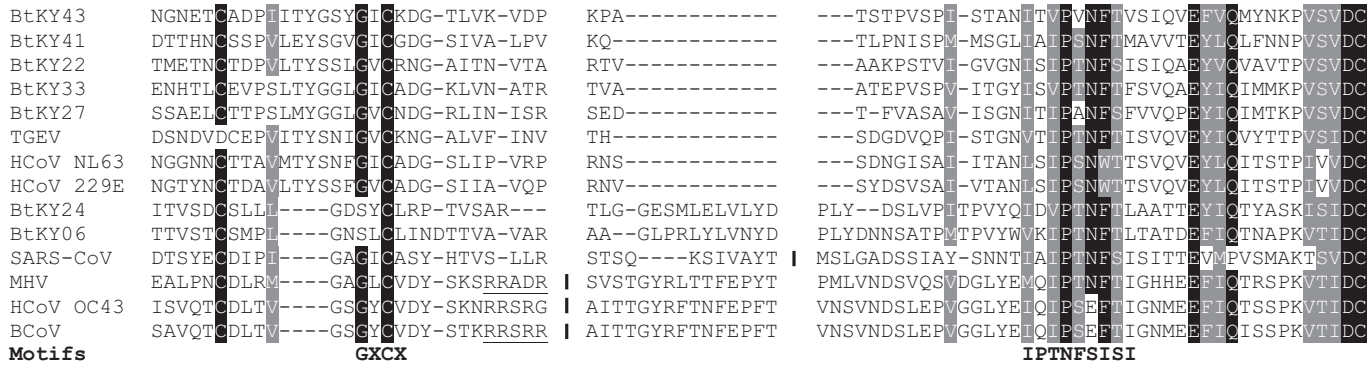

Fig. 2. Multiple amino acid sequence alignments showing the putative S1–S2 junctional region of CoV spike protein. The identical amino acids are highlighted in black and the similar amino acids are highlighted in gray. The regions containing S1 GxCx motif, conserved S2 nonamer IPTNFSISI, the furin cleavage site (in MHV, HCoV OC43, and BCoV; underlined), and cathepsin L cleavage site (in SARS-CoV) are indicated.

Events detected by all three methods with default parameters were considered as potential recombination events.

3. Results and discussion

3.1. Detection of CoV RNA in bat tissues

The aliquots of bat rectal samples for BtKY27, BtKY33, BtKY22, BtKY41, BtKY43, BtKY24, and BtKY06 were confirmed positive by the pan CoV RT-PCR assay, while among tissues (liver, lung, and/or kidney) that were available from bats BtKY27, BtKY22, BtKY41, BtKY24, and BtKY06, only the liver from bat BtKY22 (*Chaerephon* sp.) and the kidney from bat BtKY24 (*Eidolon helvum*) tested positive by RT-PCR. These data support an infection process rather than transit of ingested infected material through the digestive tract as the source of viral RNA in rectal swabs, particularly because these bat species do not feed on vertebrates. Negative results for other tissues may be explained by specific pathobiology and a limited tropism to the available tissues.

3.2. Partial genome sequence and organization

Each acquired CoV genome sequence covers the complete ORF1b, S protein, ORF3, E protein, M protein, N protein, other putative ORFs after N and the 3′ end untranslated region with a poly A tail. The genome organization and size for each of the ORFs are shown in Fig. 1 and Table 1, respectively. They are similar to other known CoV genome organization in the order of 5′-ORF1b, S, ORF3, E, M, and N-3′, but have a variable number of putative ORFs downstream of the N gene. The sizes of these seven genomic sequences from ORF1b to the 3′ end are between ~15k and ~16k and their G + C contents are between 37.6% and 42.6%. BtKY27 has no evidence of a putative ORF downstream of the N gene, but possesses a short untranslated region and poly-A tail similar to Bat-CoV 1A (Chu et al., 2008). BtKY22, BtKY33 and BtKY43 have one small putative ORF (76–161 amino acids (aa)) downstream of the N with no significant homology to previously described CoV ORFs. BtKY06 and BtKY24 have two small putative ORFs downstream of the N with sequence similarity to NS7a and NS7b in Bat-CoV HKU9, respectively (Woo et al., 2007). BtKY41 has two small putative ORFs downstream of the N, which are overlapped and have no significant sequence homology to the previously described ORFs.

Like most alphacoronaviruses, the BtKY27, BtKY33, BtKY22, BtKY41, and BtKY43 viruses share a core sequence 5′-CUAAAC-3′ or similar putative transcription regulatory sequence (TRS) upstream of ORFs S, M, N, and ORFx and ORFy (Table 1) (Chu et al., 2008; Woo et al., 2005). ORF3 and E have putative core TRSs that sometimes varied from that for the other ORFs. The BtKY06 and BtKY24 have a core sequence TRS 5′-ACGAAC-3′ in the upstream of each ORF except E which has a core sequence TRS 5′-UCGAAC-3′ (Table 1).

Spike proteins are the type I glycosylated membrane proteins, with a putative signal peptide at the N terminal. There are 31, 27,

Table 2
Pairwise sequence comparison of Kenya bat CoVs with their nearest known CoV species.

Genus	Kenya bat CoV	% identity to nearest known CoV[a]									
		3′ genome[b]	Nsp12[c]	Nsp13[c]	Nsp14[c]	Nsp15[c]	Nsp16[c]	S[c]	E[c]	M[c]	N[c]
Alphacoronavirus	BtKY27	85	97	96	94	95 Bat-CoV 1A	95 Bat-CoV 1A	87	91	93	91
	BtKY33	75	93	91	90	87 Bat-CoV HKU8	93 Bat-CoV HKU8	62	65	75	69
	BtKY22	71	86	88	80	75 Bat-CoV/512/05	87 Bat-CoV/512/05	56	70	79	58
	BtKY41	69	80	86	79	75 Bat-CoV HKU8	86 Bat-CoV HKU8	55	63	70	57
	BtKY43	69	84	88	77	71 Bat-CoV HKU9	80 Bat-CoV HKU9	53	49	73	52
Betacoronavirus	BtKY06	90	>99	99	99	99 Bat-CoV HKU9	86 Bat-CoV HKU9	83	97	95	94
	BtKY24	70	87	88	82	69	79	52	57	66	66

[a] The nearest known CoV species were chosen based on the blast search.
[b] 3′ 15–16k genome nucleotide identity.
[c] Amino acid identity.

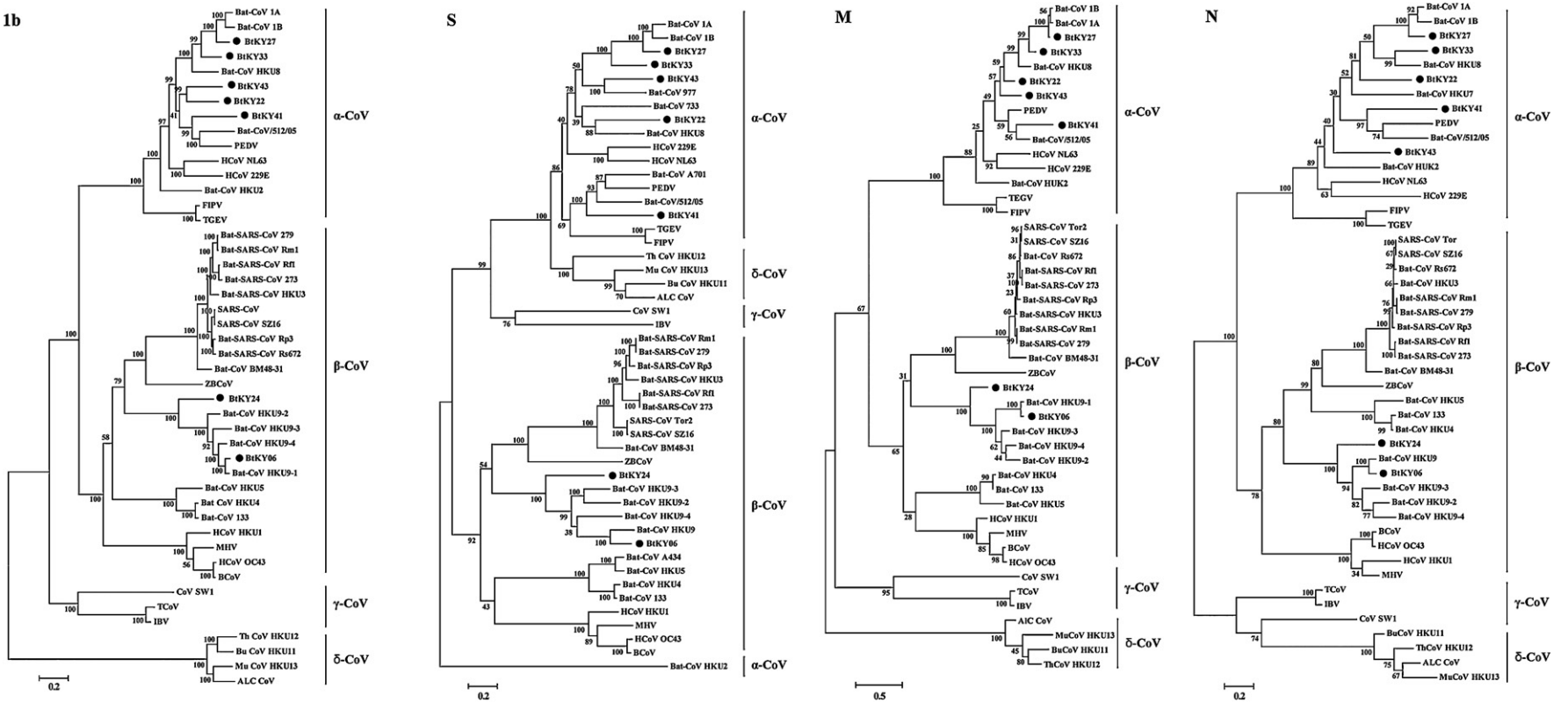

Fig. 3. Phylogenetic analysis of ORF1b, S, M and N of bat CoVs from Kenya. The unrooted trees are constructed by Maximum likelihood method with 1000 bootstrap replications after ambiguous regions from alignments of ORF1b, S, M, and N are removed. The seven Kenya CoVs are highlighted with solid circles. The genus taxonomy information is shown to the right side of the phylogeny. The maximum likelihood bootstrap is indicated next to the nodes. The scale bar indicates the estimated number of nucleotide substitutions per site.

28, 25, 31, 20, and 19 potential N-glycosylation sites in BtKY22, BtKY27, BtKY33, BtKY41, BtKY43, BtKY24, and BtKY06, respectively. As shown in Fig. 2, spike proteins of the seven bat CoVs lack furin protease recognition site, such as RRADR-S in Murine Hepatitis Virus (MHV), RRSRG-A in human CoV OC43 (HCoV OC43), RRSRR-A in bovine CoV (BCoV) (Follis et al., 2006), and cathepsin L cleavage site (VAYT-M) as in SARS-CoV (Bosch et al., 2008). In spite of lacking conserved cleavage sites, they all consist of two domains, S1 and S2, showing the conserved GxCx motif in S1 around the cleavage site and the conserved nonamer motif IPTNFSISI or similar motif in S2. These motifs have been observed in other known CoVs (Follis et al., 2006). The S1 is responsible for virus binding to the receptor on the target cells and may contain receptor binding domains (RBDs) that directly bind to host cellular receptors. For example, the RBDs of HCoV 229E, TGEV, and HCoV NL63 in *Alphacoronavirus* are mapped at the C terminus of their S1 domain (Bonavia et al., 2003; Godet et al., 1994; Lin et al., 2008). The RBDs of MHV and SARS-CoV in *Betacoronavirus* are mapped at N terminus and central region of S1 domain, respectively (Li et al., 2005; Lin et al., 2008). Alignment of aa sequences of S1 regions from BtKY22, BtKY27, BtKY33, BtKY41, and BtKY43 of *Alphacoronavirus* with the corresponding known RBD S1 regions of HCoV 229E, TGEV, and HCoV NL 63 showed 33–41% identity in S1 RBD domains to HCoV 229E and 24–29% identity to TGEV and HCoV NL63 (Fig. S1A–C). BtKY24 and BtKY06 from *Betacoronavirus* are quite different in the corresponding RBD S1 regions from SARS-CoV and MHV (17–19% identity) (Fig. S1D–E). The dissimilarity of S1 regions of these bat CoVs to other CoVs may suggest their different host specificity.

3.3. Phylogeny

We constructed phylogenetic trees using maximum likelihood method based on nt sequences of ORF1b, S, E, M and N genes with representative viruses whose corresponding sequences of their genomes were available (Fig. 3). The phylogeny of E gene is not shown due to the short length and limited value for inferring species phylogenies. Similar topologies were observed in the phylogenetic trees based on each of 5 ORFs (Fig. 3). The analysis revealed that among the seven bat CoVs, five belonged to *Alphacoronavirus* while the other two belonged to *Betacoronavirus* (Fig. 3). Phylogenetic clusterings within *Alphacoronavirus* varied slightly when different genes were analyzed. For example, BtKY22 and BtKY43 grouped into one monophyletic clade in ORF1b tree while they were grouped differently in the S and N gene trees with generally insignificant bootstrap values (Fig. 3). Although recombination was suspected, we found no evidence of recombination in the seven analyzed viruses using MaxChi (Smith, 1992), Chimaera (Posada et al., 2002), and Geneconv (Padidam et al., 1999). Since the analyses were based on representatives from each CoV species, the results suggest a lack of inter-species recombination in these viruses. One explanation is that the recombination frequency decreases significantly when the sequence divergence is high (Kleiboeker et al., 2005; van Vugt et al., 2001). Alternatively, the lack of inter-specie recombination is due to rare co-infections as the viruses adapted to different bats species. Therefore, the phylogenetic incongruence observed in the gene trees is probably due to low phylogenetic signals, which may be improved by sampling more CoVs that are related to BtKY22 and BtKY43.

The pairwise nt comparisons among these seven bat CoV gene sequences revealed 67–76% overall nt identity. Among the five alphacoronaviruses, three (BtKY22, BtKY41 and BtKY43) were distantly related to other known alphacoronaviruses with only 69–71% overall nt identity and with <90% aa identity in all five conserved domains (nsps 12–16) of ORF1b (Table 2). Since we were not able to obtain all the genome portions necessary for definite species classification (de Groot et al., 2011), we adopted the separation criteria based on the RdRp group units (RGU) (Drexler et al., 2010). The aa distances in the 816 bp fragment of the RdRp gene from the Kenya bat CoVs described in this study were compared to the aa sequences from their close reference viruses (Table S2).

BtKY22, BtKY41, and BtKY43 had >4.8% aa distance in the RdRp fragment (Table S2). This suggests that they are most likely three distinctive alphacoronvirus species. BtKY27 and BtKY33 identified in *Miniopterus* bats were closely related to Bat-CoV 1A, which was identified from bent-winged *Miniopterus* bat in Hong Kong (Chu et al., 2006) with 85% and 75% overall nt identity and with >90% aa identity in 5/5 and 4/5 conserved domains (nsps 12–16) in ORF1b, respectively (Table 2). BtKY27 and BtKY33 had <4.8% aa distance in the 816 bp RdRp to their close reference viruses indicating that they are members of the established bat CoV species in *Alphacoronavirus*.

As for the two members of *Betacoronavirus* genus identified, one (BtKY06 identified in *Rousettus aegyptiacus* bat) was likely a member of Bat-CoV HKU9 species identified from *Rousettus leschenaulti* bat in China (Woo et al., 2007), sharing 90% overall nt identity and 99% aa identity in 4/5 conserved domains (nsps12-16) in ORF1b (Table 2). The other (BtKY24) was distantly related to other known betacoronaviruses with $\leq$70% overall nt identity and <90% aa identity in all 5 conserved domains (nsps 12–16) from ORF1b (Table 2). Additionally, based on the RGU criteria, BtKY24 had >6.3% aa distance in the 816 bp RdRp fragment compared to its closest reference virus indicating that it is most likely a distinctive betacoronavirus.

In conclusion, sequence data for the structural and nonstructural ORFs in the 3′-end of the genome of seven Kenya bat CoVs confirmed the high diversity and their phylogenetical placement into *Alphacoronavirus* and *Betacoronavirus* genera. The four clusters of Kenya bat CoVs represented by BtKY22, BtKY41, BtKY43, and BtKY24 respectively, most likely belonged to novel CoV species, the two clusters represented by BtKY27 and BtKY33 were likely members of Bat-CoV 1A, and the cluster represented by BtKY06 was likely a member of Bat-CoV HKU9 species. As noted with other novel CoVs, the genome organization is similar but differences were found in the number of putative ORFs downstream from the ORF N. The present results are in line with previous findings of extensive diversity of CoVs detected in bats and confirm that bat CoVs mainly belong to the *Alphacoronavirus* and *Betacoronavirus* genera (Lau et al., 2005, 2007; Tang et al., 2006; Woo et al., 2007, 2009b). Consistent with other reports, none of the bat CoVs characterized in the present study was sufficiently similar to the human SARS-CoV and other human CoVs to be suggested their direct progenitors. The examples of host switching among CoVs after relatively minor sequence changes in S1 domain of spike protein (Haijema et al., 2003; Kuo et al., 2000; Qu et al., 2005) suggest the potential risks for introduction into humans as occurred with SARS-CoV. Therefore characterization of novel CoVs and understanding species diversity in animals should help understand and respond to emerging zoonotic infections.

Acknowledgments

We thank Ivan Kuzmin, Michael Niezgoda, and Charles E. Rupprecht from Division of High Consequence Pathogens and Pathology, CDC, Atlanta, GA; Robert F. Breiman from Global Disease Detection Division, CDC-Kenya, Nairobi, Kenya; and Bernard Agwanda from National Museum, Kenya Wildlife Service, Nairobi, Kenya for excellent technical and logistical assistance and field study. The study was supported in part by the Global Disease Detection program of CDC (Atlanta, GA).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.virusres.2012.04.007.

References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. Journal of Molecular Biology 215 (3), 403–410.

Bonavia, A., Zelus, B.D., Wentworth, D.E., Talbot, P.J., Holmes, K.V., 2003. Identification of a receptor-binding domain of the spike glycoprotein of human coronavirus HCoV-229E. Journal of Virology 77 (4), 2530–2538.

Bosch, B.J., Bartelink, W., Rottier, P.J., 2008. Cathepsin L functionally cleaves the severe acute respiratory syndrome coronavirus class I fusion protein upstream of rather than adjacent to the fusion peptide. Journal of Virology 82 (17), 8887–8890.

Chu, D.K., Leung, C.Y., Gilbert, M., Joyner, P.H., Ng, E.M., Tse, T.M., Guan, Y., Peiris, J.S., Poon, L.L., 2011. Avian coronavirus in wild aquatic birds. Journal of Virology 85 (23), 12815–12820.

Chu, D.K., Peiris, J.S., Chen, H., Guan, Y., Poon, L.L., 2008. Genomic characterizations of bat coronaviruses (1A, 1B and HKU8) and evidence for co-infections in *Miniopterus* bats. The Journal of General Virology 89 (Pt 5), 1282–1287.

Chu, D.K., Poon, L.L., Chan, K.H., Chen, H., Guan, Y., Yuen, K.Y., Peiris, J.S., 2006. Coronaviruses in bent-winged bats (*Miniopterus* spp.). The Journal of General Virology 87 (Pt 9), 2461–2466.

de Groot, R., Baker, S., Baric, R., Enjuanes, L., Gorbalenya, A., Holmes, K., Perlman, S., Poon, L., Rottier, P., Talbot, P., Woo, P., Ziebuhr, J., 2011. Family Coronaviridae. In: King, A.M.Q., Adams, M.J., Carstens, E.B., Lefkowitz, E.J. (Eds.), Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses. Elsevier, Oxford, pp. 806–828.

Drexler, J.F., Gloza-Rausch, F., Glende, J., Corman, V.M., Muth, D., Goettsche, M., Seebens, A., Niedrig, M., Pfefferle, S., Yordanov, S., Zhelyazkov, L., Hermanns, U., Vallo, P., Lukashev, A., Muller, M.A., 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. Journal of Virology 84 (21), 11336–11349.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32 (5), 1792–1797.

Follis, K.E., York, J., Nunberg, J.H., 2006. Furin cleavage of the SARS coronavirus spike glycoprotein enhances cell-cell fusion but does not affect virion entry. Virology 350 (2), 358–369.

Godet, M., Grosclaude, J., Delmas, B., Laude, H., 1994. Major receptor-binding and neutralization determinants are located within the same domain of the transmissible gastroenteritis virus (coronavirus) spike protein. Journal of Virology 68 (12), 8008–8016.

Haijema, B.J., Volders, H., Rottier, P.J., 2003. Switching species tropism: an effective way to manipulate the feline coronavirus genome. Journal of Virology 77 (8), 4528–4538.

Kleiboeker, S.B., Schommer, S.K., Lee, S.M., Watkins, S., Chittick, W., Polson, D., 2005. Simultaneous detection of North American and European porcine reproductive and respiratory syndrome virus using real-time quantitative reverse transcriptase-PCR. Journal of Veterinary Diagnostic Investigation 17 (2), 165–170.

Kuo, L., Godeke, G.J., Raamsman, M.J., Masters, P.S., Rottier, P.J., 2000. Retargeting of coronavirus by substitution of the spike glycoprotein ectodomain: crossing the host cell species barrier. Journal of Virology 74 (3), 1393–1406.

Lau, S.K., Woo, P.C., Li, K.S., Huang, Y., Tsoi, H.W., Wong, B.H., Wong, S.S., Leung, S.Y., Chan, K.H., Yuen, K.Y., 2005. Severe acute respiratory syndrome coronavirus-like virus in Chinese horseshoe bats. Proceedings of the National Academy of Sciences of the United States of America 102 (39), 14040–14045.

Lau, S.K., Woo, P.C., Li, K.S., Huang, Y., Wang, M., Lam, C.S., Xu, H., Guo, R., Chan, K.H., Zheng, B.J., Yuen, K.Y., 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 (2), 428–439.

Li, F., Li, W., Farzan, M., Harrison, S.C., 2005. Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. Science 309 (5742), 1864–1868.

Lin, H.X., Feng, Y., Wong, G., Wang, L., Li, B., Zhao, X., Li, Y., Smaill, F., Zhang, C., 2008. Identification of residues in the receptor-binding domain (RBD) of the spike protein of human coronavirus NL63 that are critical for the RBD-ACE2 receptor interaction. The Journal of General Virology 89 (Pt 4), 1015–1024.

Marchler-Bauer, A., Anderson, J.B., Cherukuri, P.F., DeWeese-Scott, C., Geer, L.Y., Gwadz, M., He, S., Hurwitz, D.I., Jackson, J.D., Ke, Z., Lanczycki, C.J., Liebert, C.A., Liu, C., Lu, F., Marchler, G.H., Mullokandov, M., Shoemaker, B.A., Simonyan, V., Song, J.S., Thiessen, P.A., Yamashita, R.A., Yin, J.J., Zhang, D., Bryant, S.H., 2005. CDD: a Conserved Domain Database for protein classification. Nucleic Acids Research 33 (Database issue), D192–D196.

Martin, D.P., Williamson, C., Posada, D., 2005. RDP2: recombination detection and analysis from sequence alignments. Bioinformatics 21 (2), 260–262.

Padidam, M., Sawyer, S., Fauquet, C.M., 1999. Possible emergence of new geminiviruses by frequent recombination. Virology 265, 218–225.

Poon, L.L., Chu, D.K., Chan, K.H., Wong, O.K., Ellis, T.M., Leung, Y.H., Lau, S.K., Woo, P.C., Suen, K.Y., Yuen, K.Y., Guan, Y., Peiris, J.S., 2005. Identification of a novel coronavirus in bats. Journal of Virology 79 (4), 2001–2009.

Posada, D., Crandall, K.A., Holmes, E.C., 2002. Recombination in evolutionary genomics. Annual Review of Genetics 36, 75–97.

Qu, X.X., Hao, P., Song, X.J., Jiang, S.M., Liu, Y.X., Wang, P.G., Rao, X., Song, H.D., Wang, S.Y., Zuo, Y., Zheng, A.H., Luo, M., Wang, H.L., Deng, F., Wang, H.Z., Hu, Z.H., Ding, M.X., Zhao, G.P., Deng, H.K., 2005. Identification of two critical amino acid residues of the severe acute respiratory syndrome coronavirus spike protein for its variation in zoonotic tropism transition via a double substitution strategy. The Journal of Biological Chemistry 280 (33), 29588–29595.

Rose, T.M., Schultz, E.R., Henikoff, J.G., Pietrokovski, S., McCallum, C.M., Henikoff, S., 1998. Consensus-degenerate hybrid oligonucleotide primers for amplification of distantly related sequences. Nucleic Acids Research 26 (7), 1628–1635.

Smith, J.M., 1992. Analyzing the mosaic structure of genes. Journal of Molecular Evolution 34 (2), 126–129.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: Molecular Evolutionary Genetics Analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Molecular Biology and Evolution 28 (10), 2731–2739.

Tang, X.C., Zhang, J.X., Zhang, S.Y., Wang, P., Fan, X.H., Li, L.F., Li, G., Dong, B.Q., Liu, W., Cheung, C.L., Xu, K.M., Song, W.J., Vijaykrishna, D., Poon, L.L., Peiris, J.S., Smith, G.J., Chen, H., Guan, Y., 2006. Prevalence and genetic diversity of coronaviruses in bats from China. Journal of Virology 80 (15), 7481–7490.

Tong, S., Conrardy, C., Ruone, S., Kuzmin, I., Guo, X., Tao, Y., et al., 2009. Detection of novel SARS-like and other coronaviruses in bats from Kenya. Emerging Infectious Diseases 15 (3), 482–485.

van Vugt, J.J., Storgaard, T., Oleksiewicz, M.B., Botner, A., 2001. High frequency RNA recombination in porcine reproductive and respiratory syndrome virus occurs preferentially between parental sequences with high similarity. The Journal of General Virology 82 (Pt 11), 2615–2620.

Weiss, S.R., Navas-Martin, S., 2005. Coronavirus pathogenesis and the emerging pathogen severe acute respiratory syndrome coronavirus. Microbiology and Molecular Biology Reviews 69 (4), 635–664.

Woo, P.C., Lau, S.K., Chu, C.M., Chan, K.H., Tsoi, H.W., Huang, Y., Wong, B.H., Poon, R.W., Cai, J.J., Luk, W.K., Poon, L.L., Wong, S.S., Guan, Y., Peiris, J.S., Yuen, K.Y., 2005. Characterization and complete genome sequence of a novel coronavirus, coronavirus HKU1, from patients with pneumonia. Journal of Virology 79 (2), 884–895.

Woo, P.C., Lau, S.K., Huang, Y., Yuen, K.Y., 2009a. Coronavirus diversity, phylogeny and interspecies jumping. Experimental Biology and Medicine (Maywood) 234 (10), 1117–1127.

Woo, P.C., Lau, S.K., Lam, C.S., Lai, K.K., Huang, Y., Lee, P., Luk, G.S., Dyrting, K.C., Chan, K.H., Yuen, K.Y., 2009b. Comparative analysis of complete genome sequences of three avian coronaviruses reveals a novel group 3c coronavirus. Journal of Virology 83 (2), 908–917.

Woo, P.C., Lau, S.K., Lam, C.S., Lau, C.C., Tsang, A.K., Lau, J.H., Bai, R., Teng, J.L., Tsang, C.C., Wang, M., Zheng, B.J., Chan, K.H., Yuen, K.Y., 2012. Discovery of seven novel Mammalian and avian coronaviruses in the genus deltacoronavirus supports bat coronaviruses as the gene source of alphacoronavirus and betacoronavirus and avian coronaviruses as the gene source of gammacoronavirus and deltacoronavirus. Journal of Virology 86 (7), 3995–4008.

Woo, P.C., Wang, M., Lau, S.K., Xu, H., Poon, R.W., Guo, R., Wong, B.H., Gao, K., Tsoi, H.W., Huang, Y., Li, K.S., Lam, C.S., Chan, K.H., Zheng, B.J., Yuen, K.Y., 2007. Comparative analysis of twelve genomes of three novel group 2c and group 2d coronaviruses reveals unique group and subgroup features. Journal of Virology 81 (4), 1574–1585.