RESEARCH ARTICLE

Identification of a Severe Acute Respiratory Syndrome Coronavirus-Like Virus in a Leaf-Nosed Bat in Nigeria

Phenix-Lan Quan,[a] Cadhla Firth,[a] Craig Street,[a] Jose A. Henriquez,[a] Alexandra Petrosov,[a] Alla Tashmukhamedova,[a] Stephen K. Hutchison,[b] Michael Egholm,[b] Modupe O. V. Osinubi,[c] Michael Niezgoda,[c] Albert B. Ogunkoya,[d] Thomas Briese,[a] Charles E. Rupprecht,[c] and W. Ian Lipkin[a]

Center for Infection and Immunity, Mailman School of Public Health, Columbia University, New York, New York, USA[a]; 454 Life Sciences, Branford, Connecticut, USA[b]; Division of Viral and Rickettsial Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia, USA[c]; and Department of Veterinary Surgery and Medicine, Ahmadu Bello University, Zaria, Nigeria[d]

ABSTRACT Bats are reservoirs for emerging zoonotic viruses that can have a profound impact on human and animal health, including lyssaviruses, filoviruses, paramyxoviruses, and severe acute respiratory syndrome coronaviruses (SARS-CoVs). In the course of a project focused on pathogen discovery in contexts where human-bat contact might facilitate more efficient interspecies transmission of viruses, we surveyed gastrointestinal tissue obtained from bats collected in caves in Nigeria that are frequented by humans. Coronavirus consensus PCR and unbiased high-throughput pyrosequencing revealed the presence of coronavirus sequences related to those of SARS-CoV in a Commerson's leaf-nosed bat (*Hipposideros commersoni*). Additional genomic sequencing indicated that this virus, unlike subgroup 2b CoVs, which includes SARS-CoV, is unique, comprising three overlapping open reading frames between the M and N genes and two conserved stem-loop II motifs. Phylogenetic analyses in conjunction with these features suggest that this virus represents a new subgroup within group 2 CoVs.

IMPORTANCE Bats (order Chiroptera, suborders Megachiroptera and Microchiroptera) are reservoirs for a wide range of viruses that cause diseases in humans and livestock, including the severe acute respiratory syndrome coronavirus (SARS-CoV), responsible for the global SARS outbreak in 2003. The diversity of viruses harbored by bats is only just beginning to be understood because of expanded wildlife surveillance and the development and application of new tools for pathogen discovery. This paper describes a new coronavirus, one with a distinctive genomic organization that may provide insights into coronavirus evolution and biology.

Address correspondence to Phenix-Lan Quan, pq2106@columbia.edu.

Coronaviruses (order *Nidovirales*, family *Coronaviridae*, subfamily *Coronavirinae*) infect a wide range of vertebrates and cause respiratory, enteric, or less frequently, neurological diseases (1, 2). Coronaviruses were originally divided into three groups based on their antigenic cross-reactivities and nucleotide sequences (3). They have been recently reclassified by the International Committee on Taxonomy of Viruses into 3 genera, designated *Alphacoronavirus* (former group 1), *Betacoronavirus* (former group 2), and *Gammacoronavirus* (former group 3) (4). Whereas the alphacoronaviruses and betacoronaviruses are associated with diseases of mammals, including humans, the gammacoronaviruses are implicated chiefly in diseases of birds. Interest in coronaviruses was largely focused on their impact on domestic porcine and avian husbandry and their utility in animal models of virus-induced demyelination (5) until the emergence of severe acute respiratory syndrome (SARS) in 2003 (6). Thereafter, with recognition of the causative agent SARS coronavirus (SARS-CoV) (7–10) and of the presence of SARS-CoV-like viruses in Chinese horseshoe bats (*Rhinolophus* spp.) (11), efforts to explore the genetic diversity of coronaviruses and their host range intensified (12).

Bats are suggested to be important reservoir hosts of many zoonotic viruses with significant impact on human and animal health, including lyssaviruses, henipaviruses, filoviruses, and coronaviruses (13–17). Viruses of bats may be transmitted to humans directly through bites or via exposure to saliva, fecal aerosols, or infected tissues as well as indirectly through contact with infected intermediate hosts, such as swine (18). In the course of a project focused on pathogen discovery in situations where human-bat contact might facilitate more efficient interspecies transmission of emerging viruses, we surveyed bats in Nigeria. Through consensus PCR (cPCR) and unbiased high-throughput pyrosequencing (UHTS) of bat tissue samples, we identified a coronavirus that is most closely related to the genus *Betacoronavirus* (subgroup 2b), which includes SARS-CoV and SARS-CoV-like viruses. However, the genomic organization of this coronavirus, obtained from a Commerson's leaf-nosed bat (*Hipposideros commersoni*), is unique in that it is comprised of three overlapping open reading frames (ORFs) between the M and N genes and two conserved stem-loop II motifs (s2m). Based on these observations and phylogenetic analyses, we propose that this new member of
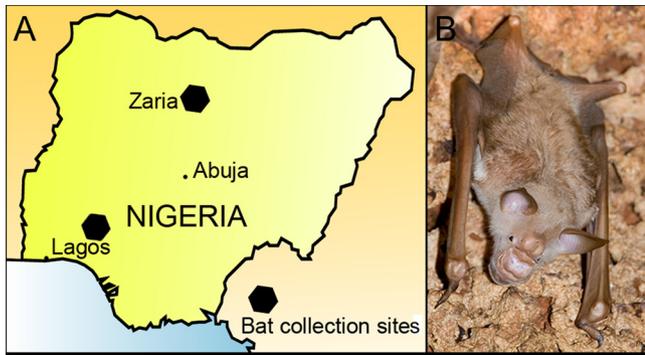
FIG 1 (A) Map of Nigeria showing the locations of bat collection sites. (B) Photograph of a male Commerson's leaf-nosed bat (*Hipposideros commersoni*), courtesy of Ivan V. Kuzmin, reproduced with permission.

the family *Coronaviridae*, tentatively named Zaria bat coronavirus (ZBCoV) after the city near to where the bat was captured, represents a new subgroup of group 2 CoVs.

RESULTS

Identification of a coronavirus in intestinal tissue of a Commerson's leaf-nosed bat (*Hipposideros commersoni*). Total RNA extracts from gastrointestinal tract (GIT) specimens obtained from 33 bats of 6 different species (*Eidolon helvum*, *Hipposideros com-*

mersoni, *Pipistrellus* sp., *Rousettus aegyptiacus*, *Scotophilus nigrita*, and *Scotophilus leucogaster*) captured at 2 different sites from a roost inside a cave in Nigeria (Fig. 1A) were screened for the presence of coronaviruses by consensus PCRs of a 400-nucleotide (nt) fragment of the RNA-dependent RNA polymerase (RdRp) gene. One specimen obtained from a Commerson's leaf-nosed bat (Fig. 1B) yielded products that shared no more than 70% nt identity to any known coronavirus. RNA from ZBCoV was submitted for UHTS, resulting in a library comprising 74,133 sequence reads. Alignment of unique singleton and assembled contiguous sequences to the GenBank database (http://www.ncbi.nlm.nih.gov/) using the Basic Local Alignment Search Tool (Blastn and Blastx) (19) indicated coverage of approximately 6,500 nt of sequence distributed along coronavirus genome scaffolds and homology to regions of replicase, spike (S), and nucleocapsid (N) sequences.

Genome organization and coding potential of ZBCoV. The additional genomic sequence of ZBCoV was determined by filling in gaps between UTHS reads, applying consensus PCRs, and 3′ and 5′ rapid amplification of cDNA ends (RACE). Overlapping primer sets based on the draft genome were synthesized to facilitate sequence validation by conventional dideoxy sequencing. Due to exhaustion of the sample, we were unable to completely sequence the open reading frame 1ab (ORF 1ab) region (Fig. 2A).

ZBCoV has a genome organization similar to that of other coronaviruses, with the following characteristic gene order:
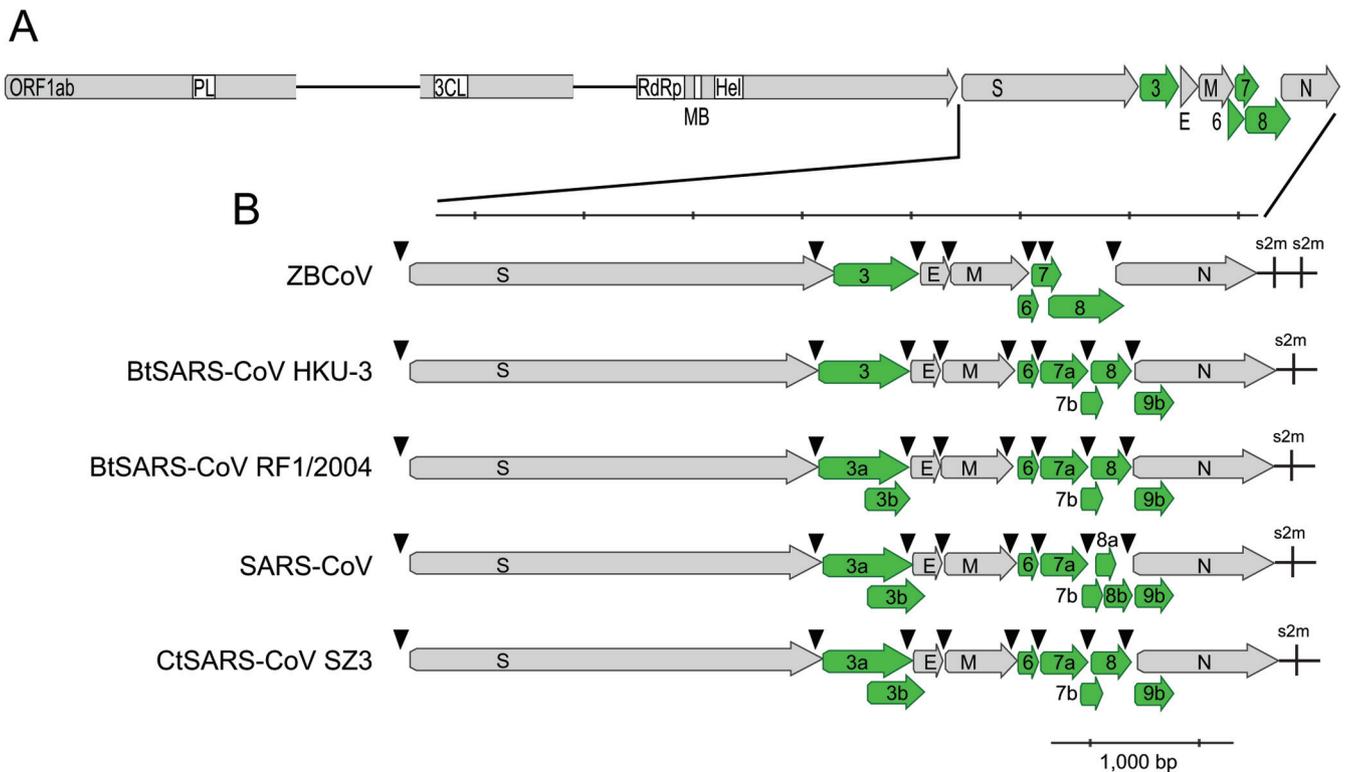

FIG 2 Genome organization of ZBCoV in comparison to that of representative coronaviruses from subgroup 2b. (A) Overall genome organization of ZBCoV. The ORF 1ab, spike (S), envelope (E), membrane (M), and nucleocapsid (N) genes are shown in gray arrows, whereas putative accessory genes ORF 3, ORF 6, ORF 7, and ORF 8 are indicated as 3, 6, 7, and 8 and illustrated by green arrows. The following conserved functional domains in ORF 1ab are represented in boxes: papain-like protease (PL), 3C-like protease (3CL), RNA-dependent RNA polymerase (RdRp), metal ion-binding domain (MB), and helicase (Hel). The two regions in ORF 1ab where sequences are incomplete are indicated by black lines. (B) Expanded diagram of the 3′ region of the ZBCoV genome in comparison to representative CoVs from subgroup 2b. TRS motifs and s2m are represented by black arrowheads and vertical lines, respectively.

TABLE 1 ORFs and putative TRS motifs

ORF	Length in:		TRS
	nt	aa	
1ab	NC[a]	NC	ACGAAC$_{221}$AUG
Spike	3,897	1,299	ACGAACAUG
3	750	250	ACGAAC$_{28}$AUG
Envelope	237	79	ACGAAC$_{23}$AUG
Membrane	729	243	ACGAAC$_{30}$AUG
6	147	49	NA[b]
7	237	79	ACGAAC$_3$AUG
8	654	218	ACGAAC$_{10}$AUG
Nucleocapsid	1,260	420	ACGAAC$_{12}$AUG

[a] NC, not complete.
[b] NA, not applicable.

5′-replicase ORF 1ab-spike (S)-envelope (E)-membrane (M)-nucleocapsid (N)-3′. Both the 5′ and 3′ ends contain short untranslated regions of 297 nt and 363 nt, respectively. The conserved putative transcription regulatory sequence (TRS) motif 5′-ACGAAC-3′ identified in subgroup 2b, 2c, and 2d viruses (2) is present in ZBCoV at the 3′ end of the leader sequence and upstream of potential initiating methionine residues of each ORF except ORF 6 (Table 1).

All domains within replicase polyproteins of coronaviruses that are implicated in viral replication are found in ZBCoV, including the papain-like protease (PL^{pro}), 3C-like protease (3CL^{pro}), RNA-dependent RNA polymerase (RdRp), and helicase (Hel) domains (Fig. 2A). ORFs consistent with the S, E, M, and N proteins present in all other coronaviruses are also present in ZBCoV (Table 1; Fig. 2). Pairwise identity (I) and similarity (S) comparisons of a deduced amino acid sequence of ZBCoV to that of representative coronaviruses in other groups showed that the predicted proteins of ZBCoV are more similar to those of subgroup 2b CoVs than to those of other subgroups, with Hel and RdRp having the highest homologies (Hel: I, 80%; S, 90%; RdRp: I, 74%; S, 85%) and the S protein having the lowest (I, 36 to 38%; S, 50 to 53%) (http://cait.cumc.columbia.edu:88/dept/greeneidlab/IdentificationofaSARS-Coronavirus-likevirusinaleaf-nosedbatinNigeria.html).

The putative spike (S) protein of ZBCoV, comprising 1,299 amino acids (aa) in length, is slightly larger than those of other subgroup 2b CoVs (see Table S8 in the supplemental material). ZBCoV showed the highest amino acid conservation to human and civet SARS-CoV (I, 38%; S, 53%) (http://cait.cumc.columbia.edu:88/dept/greeneidlab/IdentificationofaSARS-Coronavirus-likevirusinaleaf-nosedbatinNigeria.html). Pfam (20) analysis identified a spike receptor binding domain (PF09408) that corresponds to the immunogenic receptor binding domain that binds to angiotensin-converting enzyme 2 (ACE2) and the coronavirus S1 (PF01600) and S2 (PF01601) spike glycoprotein domains. Transmembrane region prediction (TMHMM 2.0) (21) revealed a long ectodomain (aa 1 to 1240), a transmembrane domain near the C-terminal end (aa 1241 to 1263), and a short cytoplasmic tail (aa 1264 to 1298). A predicted signal peptide (SignalP 3.0) ($P = 1$) (22) was identified with a cleavage site ($P = 0.768$) between residues A$_{16}$ and A$_{17}$. NetNGlyc 1.0 identified 25 putative N-linked glycosylation sites. The S protein of ZBCoV displays major sequence differences compared to that of subgroup 2b CoVs, especially in the S1 domain involved in receptor binding. The critical residues suggested to be important for the cleavage of the SARS-CoV S protein are present in the S protein of ZBCoV (23–25) (see

Fig. S1A in the supplemental material). Motifs at the carboxyl terminus of the S protein that are conserved among coronaviruses are also found in the ZBCoV S protein, including the conserved motif Y(X)KWPW(Y/W)(V/I)WL present as Y$_{1237}$EKWPWYIWL and the cysteine-rich cytoplasmic tail (10) (see Fig. S1B in the supplemental material).

In addition to the five genes present in all genomes, coronaviruses also have several group-specific genes between the S gene and the 3′ end of the genome that encode accessory proteins (Fig. 2) (26, 27).

An ORF (ORF 3) encoding a putative 250-aa protein was observed between the S and E proteins of ZBCoV (Table 1). ORF 3 corresponds to the genomic position of ORF 3a in subgroup 2b CoVs. Similar to subgroup 2b CoVs, ORF 3 is the largest accessory gene of ZBCoV and is 75 nt shorter than ORF 3a of subgroup 2b CoVs (see Table S8 in the supplemental material). ORF 3 shows 21 to 23% aa identity and 31 to 35% aa similarity to the ORF 3a protein of subgroup 2b CoVs (see Table S9 in the supplemental material). Pfam analysis showed a relationship with PF11289, a viral family protein of an unknown function; TMHMM analysis predicts the presence of 4 transmembrane regions, spanning residues P$_{43}$ to L$_{65}$, A$_{72}$ to E$_{94}$, V$_{99}$ to L$_{121}$, and Y$_{196}$ to V$_{218}$. NetOGlyc 3.1 predicted two potential O glycosylation sites in ZBCoV. ORF 3 contains only a portion of the cysteine-rich domain identified in the ORF 3a protein of SARS-CoV; however, the cysteine potentially involved in ORF 3a protein polymerization (28) is present in ORF 3. No signal peptide, YXXΦ, or diacidic motifs were identified in ORF 3 of ZBCoV (29).

ZBCoV has a set of ORFs located between the M and N genes that are not shared by any of the known coronaviruses. These ORFs, ORF 6, ORF 7, and ORF 8, encode predicted proteins of 49, 79, and 218 aa, respectively (Table 1). A TRS was identified upstream of ORF 7 and ORF 8 but not ORF 6. ORF 6 overlaps with the M gene at the 3′ end by 101 nt, ORF 7 overlaps with ORF 6 by 31 nt, and ORF 8 overlaps with ORF 7 and the N gene by 83 and 35 nt, respectively. Blastx and Pfam analyses of ORF 6, ORF 7, and ORF 8 revealed no significant similarities or functional domains. Pfam analysis of ORF 7 indicated nonsignificant associations to the PRA1 (prenylated Rab acceptor 1) proteins (PF03208) (E value = 0.02) and the 7 transmembrane G-protein-coupled-receptor protein families (PF10323) (E value = 0.025). TMHMM analysis of ORF 7 suggested the presence of a transmembrane region between residues L$_{10}$ and I$_{32}$. No signal peptide was predicted.

TMHMM and SignalP analyses of ORF 6 indicated no transmembrane region or signal peptide. TMHMM analysis of ORF 8 predicted 2 transmembrane regions, and a third transmembrane region located downstream was predicted by TMpred (30). SignalP revealed a signal peptide ($P = 0.988$) with a putative cleaved signal sequence ($P = 0.804$) between residues G$_{29}$ and A$_{30}$.

At only 788 nt, the region in ZBCoV between the M and N genes is significantly shorter than those observed for subgroup 2b CoVs (see Table S8 in the supplemental material). Alignment of the region between the M and N genes of ZBCoV with those of subgroup 2b CoVs indicated large deletions in ZBCoV (see Fig. S2 in the supplemental material).

Another distinctive genomic feature of ZBCoV is the presence downstream from the N gene of two conserved motifs corresponding to the conserved stem-loop II motif (s2m) (31). A unique s2m is observed in coronaviruses from subgroups 2b, 3a, and 3c and in astroviruses and in the picornavirus equine rhinitis
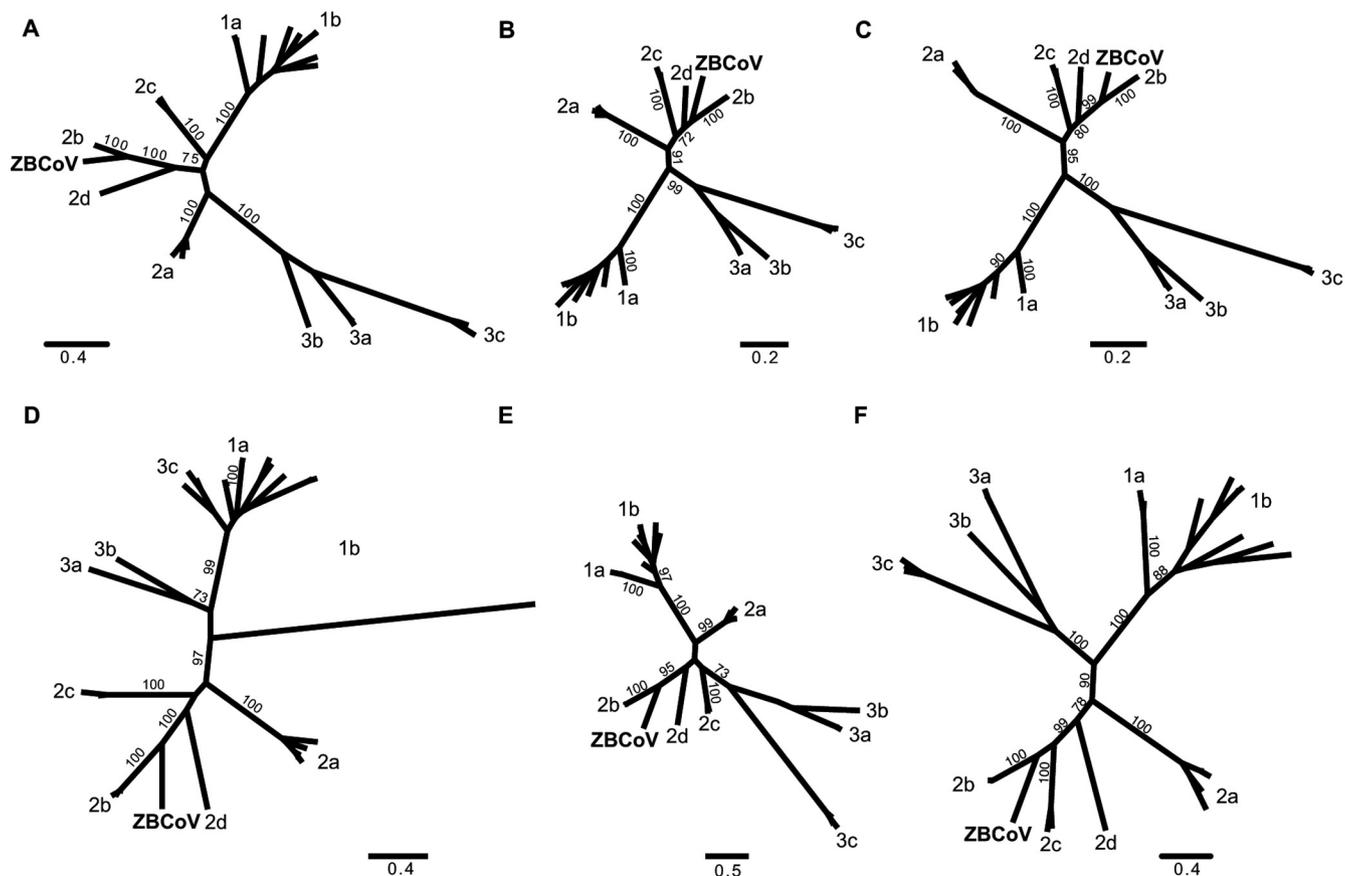
FIG 3 Phylogenetic analysis of the 3CL^pro, RdRp, Hel, S, M, and N proteins of ZBCoV. Unrooted maximum likelihood phylogenies of the 3CL^pro (A), RNA-dependent RNA polymerase (B), helicase (C), spike (D), membrane (E), and nucleocapsid (F) proteins. All phylogenies were constructed using the complete amino acid alignments of each protein, with the exception of RdRp (partial region available) and spike (only an 884-aa region could be reliably aligned). The scale bar indicates the number of substitutions per amino acid site. The numbers at each branch node represent the maximum likelihood bootstrap support; only major nodes where values exceed 70% are shown. The CoV subgroups are indicated as 1a and b, 2a to d, and 3a to c, and the following sequences obtained from GenBank were included, with the GenBank accession numbers given in parentheses: PRCV, porcine respiratory coronavirus (DQ811787); FIPV, feline infectious peritonitis virus (AY994055); HCoV-229E, human coronavirus 229E (NC_002645); HCoV-NL63, human coronavirus NL63 (NC_005831); BtCoV-512/2005, bat coronavirus 512/2005 (NC_009657); BtCoV-HKU2, bat coronavirus HKU2 (NC_009988); BtCoV-1B, bat coronavirus 1 B (NC_010436); BtCoV-1A, bat coronavirus 1A (NC_010437); BtCoV-HKU8, bat coronavirus HKU8 (NC_010438); BCoV, bovine coronavirus (NC_003045); HCoV-OC43, human coronavirus OC43 (NC_005147); HCoV-HKU1, human coronavirus HKU1 (NC_006577); MHV, mouse hepatitis virus (NC_006577); PHEV, porcine hem-agglutinating encephalomyelitis virus (NC_007732); ECoV, equine coronavirus (NC_010327); BtSARS-CoV HKU3, bat SARS coronavirus HKU3 (NC_009694); CtSARS-CoV SZ3, civet SARS coronavirus SZ3 (AY304486); SARS-CoV, SARS coronavirus (NC_004718); BtSARS-CoV Rp3, bat coronavirus Rp3 (NC_009693); BtSARS-CoV Rf1/2004, bat coronavirus Rf1/2004 (NC_009695); BtSARS-CoV RM1, bat coronavirus RM1 (NC_009696); BtCoV-HKU4, bat coronavirus HKU4 (NC_009019); BtCoV HKU5, bat coronavirus HKU5 (NC_009020); BtCoV HKU9, bat coronavirus HKU9 (NC_009021); IBV, infectious bronchitis virus (NC_001451); TCoV, turkey coronavirus (NC_010800); SW1, beluga whale coronavirus (NC_010646); BuCoV HKU11, Bulbul coronavirus HKU11 (NC_011548); ThCoV HKU12, thrush coronavirus HKU12 (NC_011549); and MuCoV HKU13, Munia coronavirus HKU13 (NC_011550).

B virus (ERBV) (31–33) (see Fig. S3A in the supplemental material). Alignment of the 3′ end of ZBCoV with subgroup 2b CoVs showed deletions in the genome of subgroup 2b CoVs where the second s2m of ZBCoV is identified (see Fig. S3B). The s2m of ZBCoV are almost identical in sequence and are separated by 19 nt (see Fig. S3B). mfold prediction (34) of RNA secondary structure indicated that both s2m fold into RNA stem-loop motifs (see Fig. S3C).

Phylogenetic analyses. Phylogenetic trees constructed from 3CL^pro, RdRp, Hel, S, M and N amino acid sequences of ZBCoV and representative coronaviruses show that ZBCoV is most closely related to but distinct from the subgroup 2b CoVs, which include SARS-CoV and SARS-CoV-like viruses (Fig. 3). This

finding is in accord with results obtained from pairwise amino acid comparisons of ZBCoV and other coronaviruses (http://cait .cumc.columbia.edu:88/dept/greeneidlab/IdentificationofaSARS -Coronavirus-likevirusinaleaf-nosedbatinNigeria.html). To fur-ther define the phylogenetic position of ZBCoV, an additional phylogeny was constructed using a conserved 659-nt sequence of RdRp, and the time to the most recent common ancestor (TMRCA) between ZBCoV and related coronaviruses was estimated. Based on the best-fit model (SRD06 with informative rate prior), the results of this analysis indicated that ZBCoV is most closely related to GhanaBt-CoV, a recently identified coronavirus found in bats in Ghana (35) (Fig. 4). Furthermore, ZBCoV and GhanaBt-CoV together form a well-supported clade distinct from that of the subgroup 2b CoVs. The
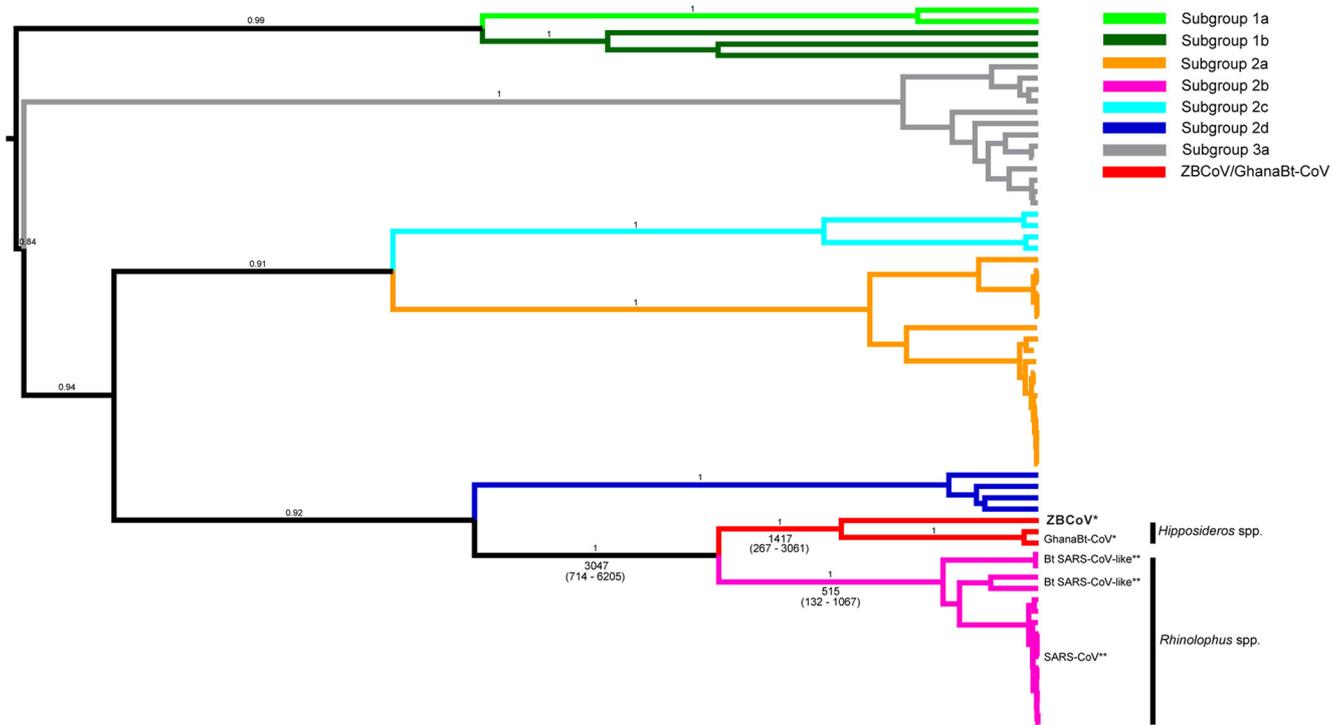
FIG 4 Estimation of the time of divergence between ZBCoV and representative coronaviruses. Bayesian MCMC phylogeny of a 659-nt region of the RNA-dependent RNA polymerase gene of ZBCoV and representative members of group 1, 2, and 3 coronaviruses. The host bat species and their geographic origins (*, Africa; **, Asia) are indicated for ZBCoV, GhanaBtCoV, and subgroup 2b CoVs. The times given at branch tips represent the dates of viral sampling, and the tree is rooted through the use of a relaxed molecular clock. Bayesian posterior probability values greater than 0.8 are shown above the branches leading to each major node. The mean TMRCAs for the taxa in subgroup 2b CoVs and ZBCoV are given below each branch, with the 95% highest probability densities indicated in parentheses. The following sequences from GenBank were included, with the GenBank accession numbers given in parentheses: for subgroup 1a CoVs, feline coronavirus (FJ938055) and canine coronavirus (GQ477367); for subgroup 1b CoVs, bat coronavirus HKU2 (DQ249213), bat coronavirus BtCoV/512/2005 (DQ648858), and human coronavirus NL63 (DQ445911); for subgroup 2a CoVs, murine hepatitis virus (AB551247), human coronavirus HKU1 (AY597011, DQ422731, DQ422728, DQ422732, DQ422737, and DQ422733), bovine respiratory coronavirus (AF220295, AF391541, EF424615, EF424620, FJ938066, and U00735), equine coronavirus (EF446615), human enteric coronavirus 4408 (FJ415324), human coronavirus OC43 (AY391777 and AY903460), and waterbuck coronavirus (FJ425184); for subgroup 2b CoVs, bat SARS coronavirus Rf1 (DQ412042 and DQ648856), SARS coronavirus (AY313906, AY545914, AY559085, AY559097, AY595412, DQ071615, FJ882929, FJ882931, FJ882941, FJ882944, FJ882959, and FJ88686), bat SARS coronavirus HKU3 (DQ084199), and bat SARS coronavirus RM1 (DQ412043); for subgroup 2c CoVs, bat coronavirus HKU5 (DQ249217 and DQ249218), bat coronavirus HKU4 (DQ074652), and bat coronavirus BtCov/133/2005 (DQ648794); for subgroup 2d CoVs, bat coronavirus HKU9-1 (EF065513), bat coronavirus HKU9-2 (EF065514), bat coronavirus HKU9-3 (EF065515), and bat coronavirus HKU9-4 (EF065516); and for subgroup 3a CoVs, avian infectious bronchitis virus (AY514485, AY641576, AY646283, DQ001339, DQ646405, EU714029, FJ888351, FN430414, FN430415, HM245923, and HM245924) and turkey coronavirus (GQ427174, GQ427175, and GQ427176).

TMRCA between ZBCoV and GhanaBt-CoV was estimated at 1,417 years before present (ybp) (95% highest population density [HPD] = 267 to 3,061 ybp). The TMRCA between the ZBCoV/GhanaBt-CoV clade and subgroup 2b CoVs was estimated at 3,047 ybp (95% HPD = 714 to 6,205 ybp), whereas the TMRCA between SARS-CoVs and SARS-CoV-like viruses was only 515 ybp (95% HPD = 132 to 1,067 ybp). Estimates of the TMRCAs between subgroup 2b CoVs and the rest of the coronavirus groups are not provided due to the potential for nucleotide site saturation at deeper phylogenetic levels to artificially create too recent TMRCA estimates.

Whereas the mean pairwise nucleotide similarity of the partial RdRp gene region was 85% (standard deviation [SD] = 9.75) within coronavirus subgroups (excluding ZBCoV/GhanaBt-CoV), the mean pairwise similarity between coronavirus subgroups was 66% (SD = 5.14) (see Fig. S4 in the supplemental material). Based on the results of the Mann-Whitney U test, these distributions are statistically different (P < 0.0001). Additionally, whereas the mean pairwise similarity within the clade ZBCoV/

GhanaBt-CoV was 85% (SD = 9.01), the pairwise similarity between the clade ZBCoV/GhanaBt-CoV and subgroup 2b CoVs was only 73% (SD = 0.84). Based on the results of the Mann-Whitney U test, these distributions are statistically different (P = 0.0092). Together, these findings indicate that the clade containing ZBCoV and GhanaBt-CoV should be considered a separate subgroup within group 2 CoVs, distinct from subgroup 2b CoVs (see Fig. S4 in the supplemental material).

DISCUSSION

Differences in phylogenetic relationships and genomic organization and the low amino acid similarities of ORF 3 and the S protein of ZBCoV compared to the ORF 3a and S proteins of subgroup 2b CoVs suggest that ZBCoV represents a new subgroup of coronaviruses within the group 2 CoVs. Although ZBCoV has features found in subgroup 2b CoVs, including the TRS, a unique PL^Pro, ORFs between the M and N genes, and the presence of the s2m, ZBCoV forms a unique branch distinct from subgroup 2b CoVs in

all phylogenetic trees analyzed. Furthermore, it differs from subgroup 2b CoVs in that ZBCoV contains three (versus four to five) ORFs between the M and N genes and has two (versus one) s2m.

Whereas the S proteins of subgroup 2b CoVs share 78 to 98% aa sequence identity, the S protein of ZBCoV has only 36 to 38% identity in the deduced amino acid sequence with those of subgroup 2b CoVs. Despite limited primary sequence conservation of the spike protein among ZBCoV and subgroup 2b CoVs, particularly in the S1 domain, Pfam analyses indicated the presence of a receptor domain that binds to the receptor ACE2, the cellular receptor for SARS-CoV (36). However, the residues in SARS-CoV that interact with the human ACE2 molecule are not conserved in ZBCoV, suggesting that human ACE2 is not a *bona fide* receptor for ZBCoV (37).

ORF 3, located between the S and E proteins of ZBCoV, is slightly shorter than the 3a proteins of subgroup 2b CoVs and has at most only 22% aa identity to the 3a proteins of subgroup 2b CoVs. In contrast, the 3a proteins of subgroup 2b CoVs share 81 to 98% aa identity. ORF 3 is predicted to contain four transmembrane domains with extracellular N and C termini. In contrast, ORF 3a of SARS-CoV is predicted to contain three transmembrane domains with extracellular N termini and intracellular C termini (28, 29). Whereas four O glycosylation sites are predicted in the ORF 3a protein of SARS-CoV (38), only two putative O glycosylation sites were identified in the ORF 3 of ZBCoV. The 3a protein of SARS-CoV has a cysteine-rich region important for polymerization and ion channel activity (28), as well as YXXΦ and diacidic motifs suggested to be involved in the intracellular trafficking (29). These domains were recently suggested to be important for the proapoptotic function of ORF 3a of SARS-CoV (39). However, ORF 3 of ZBCoV contains only a portion of the cysteine-rich domain and has no YXXΦ diacidic motifs. In contrast to human and civet SARS-CoV and bat RF1/2004, there is no ORF 3b in ZBCoV. The 3b protein may function as an interferon antagonist (40).

ZBCoV contains a unique set of ORFs located between the M and N genes. In subgroup 2b CoVs, ORF 6, ORF 7, and ORF 8 between the M and N genes do not overlap. In contrast, the three ORFs between the M and N genes overlap in ZBCoV. Alignment with subgroup 2b CoVs indicated deletions in ZBCoV, and as a result, one continuous ORF, ORF 8, is present in ZBCoV in place of ORFs 7a, 7b, 8, 8a, and 8b of subgroup 2b CoVs.

Similar to SARS-CoV, the putative products of ORF 6, ORF 7, and ORF 8 of ZBCoV show no sequence homology to other viral proteins. No TRS upstream of ORF 6 is found, suggesting that if ORF 6 encodes a *bona fide* protein, that protein is likely expressed by the subgenomic RNA M. There is precedent in SARS-CoV for functional bicistronic RNAs in the expression of ORF 3b, ORF 7b, ORF 8b, and ORF 9b (26, 41). Coronaviruses possess accessory genes, the size and location of which are group specific (2). By analogy to SARS-CoV, ORF 6, ORF 7, and ORF 8 of ZBCoV may encode accessory proteins important for virus-host interactions that may contribute to virulence and pathogenesis (26). Recent studies suggest that the SARS-CoV accessory proteins 6 and 7b are incorporated into virus particles and that 3a, 7a, and 9b are structural components of the virion (26, 41, 42). The SARS-CoV accessory proteins are suggested to have biological functions that include virus release, interferon antagonism, apoptosis induction, and inhibition of cellular protein synthesis (26, 41).

Another unique feature of ZBCoV is the presence of two highly conserved RNA sequences (s2m) downstream of the N gene. A single s2m is identified at the 3′ end of the genomes of members of several RNA virus families, including the *Coronaviridae* and *Astroviridae*, as well as the picornavirus ERBV (31–33). Recent data suggest that the SARS-CoV s2m RNA is a functional molecular mimic of the 530 stem-loop region in small-subunit ribosomal RNA, which could facilitate viral hijacking of the host's protein synthesis machinery (43). The presence of a second s2m in ZBCoV may further increase the efficiency of this process. Interestingly, secondary structures downstream of the N gene, including bulged stem-loop and pseudoknot structures, are also identified in the genomes of subgroup 2a and 2c CoVs (44, 45).

Lagos bat virus (family *Rhabdoviridae*, genus *Lyssavirus*) was initially identified in Nigeria in the 1950s. The discovery of ZBCoV in a bat of the genus *Hipposideros* (family Hipposideridae), is the first identification of a coronavirus in wildlife from Nigeria. Recently, bat coronaviruses closely related to ZBCoV were isolated from roundleaf bats (*Hipposideros caffer* and *Hipposideros ruber*) in Ghana, a country that is close to Nigeria (35). Phylogenetic analysis indicates that ZBCoV and GhanaBt-CoV form a unique clade that is distinct from those in subgroup 2b CoVs. However, as the only sequence available for GhanaBt-CoV is a fragment of the RdRp gene, a comparison of the genome organization between ZBCoV and GhanaBt-CoV is not possible. Our findings and recent published data, wherein a SARS-CoV-like virus was found to lack ORF 8, suggest that there is considerable diversity in the genome organization of SARS-CoV-like viruses (46).

SARS-CoV-like viruses have been isolated from various rhinolophid bats (family Rhinolophidae, genus *Rhinolophus*), common insectivorous bats found in Africa and Eurasia. However, despite extensive studies, no SARS-CoV-like viruses have been reported in *Hipposideros* sp. bats in China (32). The *Rhinolophus* species suggested as reservoirs of SARS-CoV-like viruses are not present in Africa. A sequence fragment of a SARS-CoV-like virus was identified in Kenya in bats of the *Chaerephon* genus (family Molossidae) (47), and antibodies reactive with SARS-CoV antigen have also been detected in the sera of seven different genera of insectivorous and fruit bats sampled in central and southern Africa (48). In concert, these findings suggest that there may be no strict species-specific host restriction of SARS-CoV-like viruses in African bats.

Our phylogenetic analysis indicates that the clade containing ZBCoV and GhanaBt-CoV occupies an ancestral position to the group 2b CoVs, which include SARS-CoV and SARS-CoV-like viruses. Similar to previous estimates, the TMRCA of these two clades was estimated at ~3,047 ybp (although with large 95% HPDs). Although SARS-CoV-like viruses have been identified exclusively in bats in China, a recent sequence fragment (~120 bp) recovered from a Kenyan bat was found to occupy a position just outside subgroup 2b and may represent the ancestral African lineage of all subgroup 2b CoVs (47). Together with the position of the African clade of ZBCoV/GhanaBt-CoV relative to subgroup 2b CoVs, this finding suggests that a migration event from Africa to China within the last 100 to 1,000 years may have resulted in the subgroup 2b lineage of CoVs. Indeed, the geographic distribution and the phylogenetic relationships of bat coronaviruses seen both here (Fig. 4) and in previous work (35) suggest the presence of multiple independent migration events between Africa and Asia

throughout the history of bat coronaviruses. Additional sequence data for the bat coronaviruses identified in Kenya along with increased sampling for coronaviruses in Africa as well as central and eastern Asia will likely be necessary to unveil the timing and origin of this diverse group of coronaviruses.

Bats are important reservoir hosts of zoonotic viruses with significant impact on human health, including rabies, Nipah virus, Hendra virus, Zaire Ebola virus, Marburg virus, and SARS-CoV. The wide genetic diversity that exists among zoonotic viruses in bats may allow an increased emergent potential of interspecies variants that may cause outbreaks of disease in humans and domestic animals. The giant leaf-nosed bat, *Hipposideros commersoni*, is widespread in sub-Saharan Africa, from Gambia to Ethiopia, Mozambique, and Madagascar, but little is known concerning its ecology, population biology, or vector competence. Clearly, in order to enhance our knowledge of the diversity and cooccurrence of potential reservoir hosts, it is essential to better understand emerging pathogen dynamics and public health relevance as a means to prevent and control future disease outbreaks.

MATERIALS AND METHODS

Bat sample collection. During June 2008, bats were collected with mist netting in caves and around human dwellings or manually from roost locations near Idanre and Zaria, Nigeria. All bats appeared clinically normal. Captured bats were anesthetized by intramuscular inoculation with ketamine hydrochloride (0.05 to 0.1 mg/g of body weight) and euthanized under sedation by intracardiac exsanguination and cervical dislocation. The species of each captured bat was recorded, as well as the sex, forearm and body lengths (in cm), and weight. All samples were initially stored, transported on ice packs, and stored thereafter at $-20°C$, until shipment on dry ice and final storage at $-80°C$. No lyssavirus-specific antigens were identified in bat brains by use of direct fluorescent antibody testing.

Coronavirus consensus PCRs. Coronavirus screening was performed by nested PCR, amplifying a 400-nt fragment of the RdRp genes of coronaviruses using consensus primer sequences 5′-CGTTGGIACW AAYBTVCCWYTICARBTRGG-3′ and 5′-GGTCATKATAGCRTCA VMASWWGCNACNACATG-3′ for the first PCR and consensus primer sequences 5′-GGCWCCWCCHGGNGARCAATT-3′ and 5′-GGWAWCCCCAYTGYTGWAYRTC-3′ for the second PCR. Primers were designed by multiple alignments of the nucleotide sequences of available RdRp genes of known coronaviruses. Reverse transcription was performed using the SuperScript III kit (Invitrogen, San Diego, CA). PCR primers were applied at 0.2-μM concentrations with 1 μl cDNA and Hot-Star polymerase (Qiagen, Valencia, CA). Cycle conditions used were as follows: 1 cycle at 95°C for 15 min; 15 cycles at 95°C for 30 s, 65°C for 30 s ($-1°C$/cycle), and 72°C for 45 s; 35 cycles at 94°C for 30 s, 50°C for 30 s, and 72°C for 45 s; and 1 cycle at 72°C for 5 min.

UHTS. Total RNA obtained from the gastrointestinal tract specimen positive for coronavirus was extracted for UHTS. Purified RNA (0.5 μg) was DNase I digested (DNA-free; Ambion, Austin, TX) and reverse transcribed using a Superscript II kit (Invitrogen) with random octamer primers linked to an arbitrary, defined 17-mer primer sequence (MWG, Huntsville, AL). cDNA was RNase H treated prior to random amplification by PCR, applying a 9:1 dilution mixture of a primer corresponding to the defined 17-mer sequence and the octamer-linked 17-mer sequence primer, respectively. Products of >70 bp were purified (MinElute; Qiagen) and ligated to linkers for sequencing on a GS FLX sequencer (454 Life Sciences, Branford, CT).

Genome sequencing. PCR primers for amplification across sequence gaps were designed (available upon request) based on the UTHS data, and the draft genome was sequenced by overlapping PCR products. Products were purified (QIAquick PCR purification kit; Qiagen) and directly dideoxy sequenced in both directions with ABI Prism BigDye Terminator

1.1 cycle sequencing kits (PerkinElmer Applied Biosystems, Foster City, CA). Additional methods applied to obtain the genome sequence included additional consensus PCR and 3′ and 5′ RACE (Invitrogen).

Phylogenetic and sequence analyses. Alignments were constructed using MUSCLE 3.7 (49) and adjusted manually using Se-Al (50). Maximum likelihood (ML) phylogenetic trees containing representative taxa from each coronavirus genus ($n = 31$) (Fig. 3, legend) were constructed using the subtree pruning and regrafting (SPR) method of branch swapping in PhyML (51). Phylogenies were constructed using amino acid alignments for the complete proteins of 3CL, Hel, M, and N and partial protein alignments for the available RdRp protein sequence and for the S protein after regions with low alignment confidence were removed. In all cases, the Whelan and Goldman model of amino acid replacement was used (52), with a gamma distribution of rate heterogeneity. The value of the shape parameter for gamma (α) was estimated from the data and approximated by six rate categories. The reliability of each branch in all phylogenies was estimated using a bootstrap resampling procedure, with 100 ML replications.

To estimate the time to the most recent common ancestor (TMRCA) for the taxa contained within subgroup 2b CoVs and including ZBCoV, an additional 659-nt alignment of the RdRp gene was constructed and chosen for homology to the gene region sequenced for the coronaviruses most closely related to ZBCoV (GhanaBt-CoV). All sequences for which time-of-sampling information was available were included ($n = 64$). TMRCAs were estimated using the Bayesian Markov chain Monte Carlo (MCMC) method with the BEAST package, version 1.5.2 (53), and both the general time-reversible (GTR) model plus Γ distribution and the SRD06 model of nucleotide substitution. A relaxed uncorrelated lognormal molecular clock was used, calibrated by the time-stamped sequences, both with and without informative rates prior on the molecular clock of $2.0 \times 10^{-4} \pm$ 0.0009 nt substitutions/site/year (35). This analysis was run until all parameters converged, with 10% of the MCMC chains discarded as burn-in. Statistical confidence in the TMRCA estimates is given by the 95% highest probability density (HPD) interval around the marginal posterior parameter mean.

The classification of ZBCoV and GhanaBt-CoV as a putative new subgroup within group 2 CoVs was determined by first calculating the percent pairwise nucleotide similarity of the same 659-nt region of RdRp genes between and within the existing subgroups of coronaviruses and then extending this comparison to include the clade ZBCoV/GhanaBt-CoV. To verify this approach, a nonparametric Mann-Whitney U test was used to assess if the pairwise nucleotide similarity within the currently accepted subgroups is different from that between subgroups. This test was then used to determine if the percent pairwise similarity within the clade ZBCoV/GhanaBt-CoV is statistically different from that of the most closely related subgroup 2b CoVs.

Protein family analysis was performed using Pfam (http://pfam.sanger .ac.uk/). Predictions of signal peptide cleavage sites, glycosylation sites, and transmembrane domains were performed using respective prediction servers available at the Center for Biological Sequence Analysis (http: //www.cbs.dtu.dk/services/ and http://www.ch.embnet.org/software /TMPRED_form.html). The percent amino acid sequence identity and similarity were calculated using the Needleman algorithm with an EBLOSUM62 substitution matrix (gap open/extension penalties of 10/0.1 for nucleotide and amino acid alignments; EMBOSS [54]), using a Perl script to iterate the process for all versus all comparisons. Prediction of RNA secondary structures was performed with the mfold program (http: //mfold.bioinfo.rpi.edu/).

Nucleotide sequence accession number. The GenBank accession number for the ZBCoV sequence is HQ166910.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at http://mbio.asm.org/lookup/suppl/doi:10.1128/mBio.00208-10/-/DCSupplemental.

Table S8, PDF file, 0.036 MB.
Table S9, PDF file, 0.027 MB.
Figure S1, PDF file, 0.034 MB.
Figure S2, PDF file, 0.020 MB.
Figure S3, PDF file, 0.140 MB.
Figure S4, PDF file, 0.297 MB.

REFERENCES

1. **Lai, M., S. Perlman, and L. Anderson.** 2007. *Coronaviridae*, p. 1306–1335. *In* D. M. Knipe, et al. (ed.), Fields virology, 5th ed. Lippincott Williams & Wilkins, Philadelphia, PA.
2. **Woo, P. C., S. K. Lau, Y. Huang, and K. Y. Yuen.** 2009. Coronavirus diversity, phylogeny and interspecies jumping. Exp. Biol. Med. (Maywood) **234:**1117–1127.
3. **Lai, M. M., and D. Cavanagh.** 1997. The molecular biology of coronaviruses. Adv. Virus Res. **48:**1–100.
4. **Carstens, E. B.** 2009. Ratification vote on taxonomic proposals to the International Committee on Taxonomy of Viruses. Arch. Virol. **155:**133–146.
5. **Lane, T. E., and M. J. Buchmeier.** 1997. Murine coronavirus infection: a paradigm for virus-induced demyelinating disease. Trends Microbiol. **5:**9–14.
6. **Baric, R. S.** 2008. SARS-CoV: lessons for global health. Virus Res. **133:**1–3.
7. **Ksiazek, T. G., D. Erdman, C. S. Goldsmith, S. R. Zaki, T. Peret, S. Emery, S. Tong, C. Urbani, J. A. Comer, W. Lim, P. E. Rollin, S. F. Dowell, A. E. Ling, C. D. Humphrey, W. J. Shieh, J. Guarner, C. D. Paddock, P. Rota, B. Fields, J. DeRisi, J. Y. Yang, N. Cox, J. M. Hughes, J. W. LeDuc, W. J. Bellini, and L. J. Anderson.** 2003. A novel coronavirus associated with severe acute respiratory syndrome. N. Engl. J. Med. **348:**1953–1966.
8. **Marra, M. A., S. J. Jones, C. R. Astell, R. A. Holt, A. Brooks-Wilson, Y. S. Butterfield, J. Khattra, J. K. Asano, S. A. Barber, S. Y. Chan, A. Cloutier, S. M. Coughlin, D. Freeman, N. Girn, O. L. Griffith, S. R. Leach, M. Mayo, H. McDonald, S. B. Montgomery, P. K. Pandoh, A. S. Petrescu, A. G. Robertson, J. E. Schein, A. Siddiqui, D. E. Smailus, J. M. Stott, G. S. Yang, F. Plummer, A. Andonov, H. Artsob, N. Bastien, K. Bernard, T. F. Booth, D. Bowness, M. Czub, M. Drebot, L. Fernando, R. Flick, M. Garbutt, M. Gray, A. Grolla, S. Jones, H. Feldmann, A. Meyers, A. Kabani, Y. Li, S. Normand, U. Stroher, G. A. Tipples, S. Tyler, R. Vogrig, D. Ward, B. Watson, R. C. Brunham, M. Krajden, M. Petric, D. M. Skowronski, C. Upton, and R. L. Roper.** 2003. The genome sequence of the SARS-associated coronavirus. Science **300:**1399–1404.
9. **Peiris, J. S., S. T. Lai, L. L. Poon, Y. Guan, L. Y. Yam, W. Lim, J. Nicholls, W. K. Yee, W. W. Yan, M. T. Cheung, V. C. Cheng, K. H. Chan, D. N. Tsang, R. W. Yung, T. K. Ng, and K. Y. Yuen.** 2003. Coronavirus as a possible cause of severe acute respiratory syndrome. Lancet **361:**1319–1325.
10. **Rota, P. A., M. S. Oberste, S. S. Monroe, W. A. Nix, R. Campagnoli, J. P. Icenogle, S. Penaranda, B. Bankamp, K. Maher, M. H. Chen, S. Tong, A. Tamin, L. Lowe, M. Frace, J. L. DeRisi, Q. Chen, D. Wang, D. D. Erdman, T. C. Peret, C. Burns, T. G. Ksiazek, P. E. Rollin, A. Sanchez, S. Liffick, B. Holloway, J. Limor, K. McCaustland, M. Olsen-Rasmussen, R. Fouchier, S. Gunther, A. D. Osterhaus, C. Drosten, M. A. Pallansch, L. J. Anderson, and W. J. Bellini.** 2003. Characterization of a novel coronavirus associated with severe acute respiratory syndrome. Science **300:**1394–1399.
11. **Lau, S. K., P. C. Woo, K. S. Li, Y. Huang, H. W. Tsoi, B. H. Wong, S. S. Wong, S. Y. Leung, K. H. Chan, and K. Y. Yuen.** 2005. Severe acute respiratory syndrome coronavirus-like virus in Chinese horseshoe bats. Proc. Natl. Acad. Sci. U. S. A. **102:**14040–14045.
12. **Guan, Y., B. J. Zheng, Y. Q. He, X. L. Liu, Z. X. Zhuang, C. L. Cheung, S. W. Luo, P. H. Li, L. J. Zhang, Y. J. Guan, K. M. Butt, K. L. Wong, K. W. Chan, W. Lim, K. F. Shortridge, K. Y. Yuen, J. S. Peiris, and L. L. Poon.** 2003. Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China. Science **302:**276–278.
13. **Calisher, C. H., J. E. Childs, H. E. Field, K. V. Holmes, and T. Schountz.** 2006. Bats: important reservoir hosts of emerging viruses. Clin. Microbiol. Rev. **19:**531–545.
14. **Chua, K. B., W. J. Bellini, P. A. Rota, B. H. Harcourt, A. Tamin, S. K. Lam, T. G. Ksiazek, P. E. Rollin, S. R. Zaki, W. Shieh, C. S. Goldsmith, D. J. Gubler, J. T. Roehrig, B. Eaton, A. R. Gould, J. Olson, H. Field, P. Daniels, A. E. Ling, C. J. Peters, L. J. Anderson, and B. W. Mahy.** 2000. Nipah virus: a recently emergent deadly paramyxovirus. Science **288:**1432–1435.
15. **Leroy, E. M., B. Kumulungui, X. Pourrut, P. Rouquet, A. Hassanin, P. Yaba, A. Delicat, J. T. Paweska, J. P. Gonzalez, and R. Swanepoel.** 2005. Fruit bats as reservoirs of Ebola virus. Nature **438:**575–576.
16. **Li, W., Z. Shi, M. Yu, W. Ren, C. Smith, J. H. Epstein, H. Wang, G. Crameri, Z. Hu, H. Zhang, J. Zhang, J. McEachern, H. Field, P. Daszak, B. T. Eaton, S. Zhang, and L. F. Wang.** 2005. Bats are natural reservoirs of SARS-like coronaviruses. Science **310:**676–679.
17. **Murray, K., P. Selleck, P. Hooper, A. Hyatt, A. Gould, L. Gleeson, H. Westbury, L. Hiley, L. Selvey, B. Rodwell, and P. Ketterer.** 1995. A morbillivirus that caused fatal disease in horses and humans. Science **268:**94–97.
18. **Wong, S., S. Lau, P. Woo, and K. Y. Yuen.** 2007. Bats as a continuing source of emerging infections in humans. Rev. Med. Virol. **17:**67–91.
19. **Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25:**3389–3402.
20. **Finn, R. D., J. Mistry, J. Tate, P. Coggill, A. Heger, J. E. Pollington, O. L. Gavin, P. Gunasekaran, G. Ceric, K. Forslund, L. Holm, E. L. Sonnhammer, S. R. Eddy, and A. Bateman.** 2010. The Pfam protein families database. Nucleic Acids Res. **38:**D211–D222.
21. **Sonnhammer, E. L., G. von Heijne, and A. Krogh.** 1998. A hidden Markov model for predicting transmembrane helices in protein sequences. Proc. Int. Conf. Intell. Syst. Mol. Biol. **6:**175–182.
22. **Nielsen, H., J. Engelbrecht, S. Brunak, and G. von Heijne.** 1997. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Protein Eng. **10:**1–6.
23. **Belouzard, S., V. C. Chu, and G. R. Whittaker.** 2009. Activation of the SARS coronavirus spike protein via sequential proteolytic cleavage at two distinct sites. Proc. Natl. Acad. Sci. U. S. A. **106:**5871–5876.
24. **Madu, I. G., S. Belouzard, and G. R. Whittaker.** 2009. SARS-coronavirus spike S2 domain flanked by cysteine residues C822 and C833 is important for activation of membrane fusion. Virology **393:**265–271.
25. **Madu, I. G., S. L. Roth, S. Belouzard, and G. R. Whittaker.** 2009. Characterization of a highly conserved domain within the severe acute respiratory syndrome coronavirus spike protein S2 domain with characteristics of a viral fusion peptide. J. Virol. **83:**7411–7421.
26. **Narayanan, K., C. Huang, and S. Makino.** 2008. SARS coronavirus accessory proteins. Virus Res. **133:**113–121.
27. **Schaecher, S. R., and A. Pekosz.** 2010. SARS coronavirus accessory gene expression and function, p. 153–166. *In* S. K. Lal (ed.), Molecular biology of the SARS-coronavirus. Springer Verlag, Berlin, Germany.
28. **Lu, W., B. J. Zheng, K. Xu, W. Schwarz, L. Du, C. K. Wong, J. Chen, S. Duan, V. Deubel, and B. Sun.** 2006. Severe acute respiratory syndrome-associated coronavirus 3a protein forms an ion channel and modulates virus release. Proc. Natl. Acad. Sci. U. S. A. **103:**12540–12545.
29. **Tan, Y. J., E. Teng, S. Shen, T. H. Tan, P. Y. Goh, B. C. Fielding, E. E. Ooi, H. C. Tan, S. G. Lim, and W. Hong.** 2004. A novel severe acute respiratory syndrome coronavirus protein, U274, is transported to the cell surface and undergoes endocytosis. J. Virol. **78:**6723–6734.

30. **Hoffman, K., and W. Stoffel.** 1993. A database of membrane spanning protein segments. Biol. Chem. Hoppe Seyler **347:**166–170.

31. **Jonassen, C. M., T. O. Jonassen, and B. Grinde.** 1998. A common RNA motif in the 3′ end of the genomes of astroviruses, avian infectious bronchitis virus and an equine rhinovirus. J. Gen. Virol. **79:**715–718.

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

33. **Woo, P. C., S. K. Lau, C. S. Lam, K. K. Lai, Y. Huang, P. Lee, G. S. Luk, K. C. Dyrting, K. H. Chan, and K. Y. Yuen.** 2009. Comparative analysis of complete genome sequences of three avian coronaviruses reveals a novel group 3c coronavirus. J. Virol. **83:**908–917.

34. **Zuker, M.** 2003. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. **31:**3406–3415.

35. **Pfefferle, S., S. Oppong, J. F. Drexler, F. Gloza-Rausch, A. Ipsen, A. Seebens, M. A. Muller, A. Annan, P. Vallo, Y. Adu-Sarkodie, T. F. Kruppa, and C. Drosten.** 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.

36. **Li, W., M. J. Moore, N. Vasilieva, J. Sui, S. K. Wong, M. A. Berne, M. Somasundaran, J. L. Sullivan, K. Luzuriaga, T. C. Greenough, H. Choe, and M. Farzan.** 2003. Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. Nature **426:**450–454.

37. **Li, F., W. Li, M. Farzan, and S. C. Harrison.** 2005. Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. Science **309:**1864–1868.

38. **Oostra, M., C. A. de Haan, R. J. de Groot, and P. J. Rottier.** 2006. Glycosylation of the severe acute respiratory syndrome coronavirus triple-spanning membrane proteins 3a and M. J. Virol. **80:**2326–2336.

39. **Chan, C. M., H. Tsoi, W. M. Chan, S. Zhai, C. O. Wong, X. Yao, W. Y. Chan, S. K. Tsui, and H. Y. Chan.** 2009. The ion channel activity of the SARS-coronavirus 3a protein is linked to its proapoptotic function. Int. J. Biochem. Cell Biol. **41:**2232–2239.

40. **Kopecky-Bromberg, S. A., L. Martinez-Sobrido, M. Frieman, R. A. Baric, and P. Palese.** 2007. Severe acute respiratory syndrome coronavirus open reading frame (ORF) 3b, ORF 6, and nucleocapsid proteins function as interferon antagonists. J. Virol. **81:**548–557.

41. **Xu, K., B. J. Zheng, R. Zeng, W. Lu, Y. P. Lin, L. Xue, L. Li, L. L. Yang, C. Xu, J. Dai, F. Wang, Q. Li, Q. X. Dong, R. F. Yang, J. R. Wu, and B. Sun.** 2009. Severe acute respiratory syndrome coronavirus accessory protein 9b is a virion-associated protein. Virology **388:**279–285.

42. **Huang, C., N. Ito, C. T. Tseng, and S. Makino.** 2006. Severe acute respiratory syndrome coronavirus 7a accessory protein is a viral structural protein. J. Virol. **80:**7287–7294.

43. **Robertson, M. P., H. Igel, R. Baertsch, D. Haussler, M. Ares, Jr., and W. G. Scott.** 2005. The structure of a rigorously conserved RNA element within the SARS virus genome. PLoS Biol. **3:**e5.

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

45. **Hsue, B., T. Hartshorne, and P. S. Masters.** 2000. Characterization of an essential RNA secondary structure in the 3′ untranslated region of the murine coronavirus genome. J. Virol. **74:**6911–6921.

46. **Drexler, J. F., F. Gloza-Rausch, J. Glende, V. M. Corman, D. Muth, M. Goettsche, A. Seebens, M. Niedrig, S. Pfefferle, S. Yordanov, L. Zhelyazkov, U. Hermanns, P. Vallo, A. Lukashev, M. A. Muller, H. Deng, G. Herrler, and C. Drosten.** 4 August 2010. Genomic characterization of SARS-related coronavirus in European bats and classification of coronaviruses based on partial RNA-dependent RNA polymerase gene sequences. J. Virol. doi:10.1128/JVI.00650-10.

47. **Tong, S., C. Conrardy, S. Ruone, I. V. Kuzmin, X. Guo, Y. Tao, M. Niezgoda, L. Haynes, B. Agwanda, R. F. Breiman, L. J. Anderson, and C. E. Rupprecht.** 2009. Detection of novel SARS-like and other coronaviruses in bats from Kenya. Emerg. Infect. Dis. **15:**482–485.

48. **Muller, M. A., J. T. Paweska, P. A. Leman, C. Drosten, K. Grywna, A. Kemp, L. Braack, K. Sonnenberg, M. Niedrig, and R. Swanepoel.** 2007. Coronavirus antibodies in African bat species. Emerg. Infect. Dis. **13:**1367–1370.

49. **Edgar, R. C.** 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. **32:**1792–1797.

50. **Rambaut, A., N. C. Grassly, S. Nee, and P. H. Harvey.** 1996. Bi-De: an application for simulating phylogenetic processes. Comput. Appl. Biosci. **12:**469–471.

51. **Guindon, S., and O. Gascuel.** 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. **52:**696–704.

52. **Whelan, S., and N. Goldman.** 2001. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. Mol. Biol. Evol. **18:**691–699.

53. **Drummond, A. J., and A. Rambaut.** 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. **7:**214.

54. **Rice, P., I. Longden, and A. Bleasby.** 2000. EMBOSS: the European molecular biology open software suite. Trends Genet. **16:**276–277.