Title: Diversity of dromedary camel coronavirus HKU23 in African camels revealed multiple recombination events among closely related Betacoronaviruses of the subgenus Embecovirus.

Authors: Ray T. Y. So, [a] Daniel K. W. Chu,[a] Eve Miguel,[b,c,] Ranawaka A. P. M. Perera,[a] Jamiu O. Oladipo,[a,d] Ouafaa Fassi-Fihri,[e] Gelagay Aylet,[f] Ronald L. W. Ko,[a] Ziqi Zhou,[a] Mo-Sheung Cheng,[a] Sulyman A. Kuranga,[d] François L. Roger,[b,g] Veronique Chevalier, [b,h] Richard J. Webby,[i] Patrick C. Y. Woo,[a] Leo L. M. Poon,[a] Malik Peiris. [a#]

[a]School of Public Health, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Pokfulam, Hong Kong Special Administrative Region, Republic of China;

[b] Animal, Santé, Territoires, Risques et Ecosystèmes, Centre de Coopération Internationale en Recherche Agronomique pour le Développement, Institut National de la Recherche Agronomique, Université de Montpellier, F-34398 Montpellier, France;

[c]Maladies Infectieuses et Vecteurs: Ecologie, Génétique, Evolution et Contrôle, L'Institut de Recherche pour le Développement, CNRS, Universitè de Montpellier, F-34398 Montpellier, France;

[d] Department of Surgery, Faculty of Clinical Sciences, University of Ilorin, Ilorin, Nigeria;

[e] Institut Agronomique et Vétérinaire, Hassan II Université, B.P. 6202 Rabat-Instituts, Rabat, Morocco;

[f] Bacterial, Parasitic and Zoonotic Diseases Research Directorate, Ethiopian Public Health Institute, Addis Ababa, Ethiopia;

[g] Kasetsart University, 10900 Bangkok, Thailand;

[h] Institut Pasteur du Cambodge, Phnom Penh, Cambodia

[i] St. Jude Children's Research Hospital, Memphis, TN 38105;

[j] Department of Microbiology, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Pokfulam, Hong Kong Special Administrative Region, Republic of China;

Running Head: Recombination of coronavirus HKU23 in African camels.

Keywords: Coronaviruses, Betacoronaviruses, recombination, phylogeny, Dromedary camels

#Address correspondence to: Malik Peiris, School of Public Health, The University of Hong Kong, E mail: malik@hku.hk

Word counts: 132 words for abstract, 72 words for importance and 5021 words for main text.

Abstract

Genetic recombination has frequently been observed in coronaviruses. Here, we sequenced multiple complete genomes of dromedary camel coronavirus HKU23 (DcCoV-HKU23) from Nigeria, Morocco and Ethiopia and identified several genomic positions indicative of cross species virus recombination events among other Betacoronaviruses of the subgenus Embecovirus (clade A β-CoVs). Recombinant fragments of a rabbit coronavirus (RbCoV-HKU14) were identified at the hemagglutinin esterase gene position. Homolog fragments of a rodent CoV were also observed at the 8.9 kDa open reading frame 4a at the 3' end of the spike gene. The patterns of recombination varied geographically across the African region, highlighting a mosaic structure of DcCoV-HKU23 genomes circulating in dromedaries. Our results highlighted active recombination of coronaviruses circulating in dromedaries and is also relevant to the emergence and evolution of other Betacoronaviruses including MERS-coronavirus (MERS-CoV).

Importance

Genetic recombination is often demonstrated in coronaviruses and can result in host range expansion or alteration in tissue tropism. Here, we showed interspecies recombination events of an endemic dromedary camel coronavirus HKU23 with other clade A Betacoronaviruses. Our results supported the possibility that the zoonotic pathogen, MERS-CoV, which also co-circulates in the same camel species, may have undergone similar recombination events facilitating its emergence or may do so in its future evolution.

Introduction

Emerging infectious disease outbreaks usually arise by inter-species jumps of viruses between animal species, sometimes including humans. Coronaviruses have repeatedly made species jumps between animal species (e.g. SADS coronavirus from bats to swine) (1) and from animals to humans (2). Two human coronaviruses (HCoVs) 229E and OC43 now endemic in the human population emerged from camels and bovines respectively, within the past few hundred years (2, 3). SARS coronavirus emerged from bats via intermediate mammalian hosts in live game animal markets in Guangdong to spread to over 25 countries across 5 continents sickening almost 8000 people and leading to almost 800 deaths(4-6). The

ability of coronaviruses to make inter-species jumps is facilitated by complex virus-host interactions. High frequency of virus genetic recombination is one strategy for the virus to adapt to new host.

Virus genetic recombination is frequently observed in coronaviruses and other positive sense RNA viruses. Murine hepatitis virus (MHV), a clade A β-CoV, is a well-studied example of homologous recombination, with up to 25% of its progeny in infected cells being demonstrated to be recombinants (7). High frequency of recombination is believed to be contributed by the large genome size, the intrinsic template-switching property of the viral RNA-dependent RNA polymerase (RdRp) during replication and the abundance of subgenomic RNA strands for template switching (8, 9). The role of RdRp in RNA recombination has been shown in poliovirus where a single amino acid residue mutation in the RdRp of poliovirus can result in a decrease in RNA recombination frequency (10). The exoribonucleases (ExoN) activity in replicase nonstructural protein (nsp) 14 of CoVs that constitutes the proofreading activity of genome replication has been suggested to be a potential regulator of RNA recombination (11). The presence of group-specific genes in CoVs is assumed to be a result of heterologous recombination, which involves exchange of non-homologus viral or cellular RNAs. The hemagglutinin esterase (HE) gene that is only expressed in clade A β-CoVs is believed to be acquired from influenza C virus through such heterologous recombination (12).

In 2012, a novel respiratory pathogen, Middle East respiratory syndrome coronavirus (MERS-CoV), was isolated from a patient with severe respiratory illness in Jeddah, Saudi Arabia (13). It was a zoonotic virus found in dromedary camels which occasionally transmits to human (14-16). MERS-CoV is enzootic in dromedary camels in the Middle East as well as Africa with the greatest virus diversity found in Africa (17). Circulation of multiple lineages of clade B MERS-CoV in dromedary camels eventually resulted in a recombinant lineage 5 virus that caused major outbreaks during 2015 both within Saudi Arabia and in South Korea, following introduction of the virus by a returning traveler (18). Recent studies have shown that two other coronaviruses, an alphacoronavirus dromedary camel coronavirus 229E (DcCoV-229E), and a β-CoV dromedary camel coronavirus HKU23 (DcCoV-HKU23) co-circulate in dromedaries in Saudi Arabia (18, 19). The co-circulation of at least three coronaviruses within camels provides an opportunity for the emergence of novel emerging infections via recombination. It is therefore important to investigate for evidence of recombination between coronaviruses co-circulating in dromedary camels because this may

have contributed to the emergence MERS-CoV and to future emergence of viruses of zoonotic and epidemic potential. Here we report the genetic diversity of DcCoV-HKU23 in the African camel population and identify several recombination events that have taken place with clade A β-CoVs, including bovine coronavirus (BCoV) and more distant species such as rabbit coronavirus (RbCoVHKU14) and rodent coronavirus (RodentCoV). We carried out our studies in West (Nigeria), East (Ethiopia) and North (Morocco) Africa because over 70% of the global population of dromedaries are found in Africa and this is likely where the greatest diversity of these dromedary coronaviruses is likely to be manifest and where MERS-CoV emerged.

Methods & materials

Sample collection.

Nasal swabs and sera were collected from dromedary camels sampled in Nigeria, Morocco and Ethiopia in previous studies of MERS-CoV in 2015 and 2016 (17, 20). Camel nasal swabs were collected from a camel abattoir in Kano, Nigeria (n=2529) in 2015 and 2016 (17, 21), from dromedary herds and abattoirs in Morocco (n=1569) in 2015 and 2016 and Ethiopia (n=621) in 2015 (20) (Table 1). The camels from Morocco and Ethiopia were mostly raised for meat, milk production or transport. Camel sera were concurrently collected from the abattoir in Nigeria (n=150) and from abattoirs and farms in both Ethiopia (n=100) and Morocco (n=100). Sampled camels were aged from 1 month-old to 20 years-old (median age of 3 years).

DcCoV-HKU23 detection and genome sequencing.

Total nucleic acid was extracted from swab samples using EasyMAG (bioMerieux, France) system. RNA was reverse transcribed into cDNA with random hexamers using PrimeScript™ RT reagent Kit (Perfect Real Time) (Takara, Japan), according to manufacturer's protocol. cDNAs were screened for DcCoV-HKU23 using a broad-range pancoronavirus nested PCR assay designed to detect known and unknown CoVs targeting the consensus region of the RNA-dependent RNA-polymerase (RdRP) gene (22). RT-PCR positive amplicons were purified using the ExoSap-IT® reagent (USB, USA) and Sanger sequenced to identify the CoV identity. Samples with DcCoV-HKU23-like virus sequences were identified and

subjected to viral load quantification using a reverse transcription quantitative PCR (RT-qPCR) assay. Oligonucleotide sequences were designed to target both the N gene of DcCoV-HKU23 and Bovine CoV (Forward primer: 5'-GTCAATACCCCGGCTGAC-3', Probe: 5'-(FAM)TCGGGACCCAAGTAGCGATGAGGC(BHQ)-3' and Reverse primer: 5'-AACCCTGAGGGAGTACCG-3'). RT-qPCR reaction was performed using the TaqMan® Fast Virus 1-Step Master Mix (Thermo Fisher Scientific, USA), with the cycling protocol: 5 min at 50°C for reverse transcription, followed by 20 seconds at 95°C and 40 cycles of 3 seconds at 95°C and 30 s at 60°C. Samples with low cycle threshold (ct) values were selected for full genome sequencing. Reverse transcription with HKU23 specific primers targeting different regions of the genome were used to generate cDNA, which were subsequently amplified by PCR with primers designed to generate overlapping amplicons that can cover the whole genome. The primer sequences are available upon request. PCR amplicons from each sample were pooled for next generation sequencing and processed with Nextera XT library preparation kit following the protocol provided by the manufacturer. Sequencing was performed using the Illumina MiSeq instrument with read length of paired ends of approx. 300bp. Raw sequence reads generated were mapped to a reference DcCoV-HKU23 genome (KF906250.1) using BWA (23). Sequence of the target virus was generated by taking the majority consensus of the mapped reads with sequencing coverage at each position of higher than 100 times.

Genomic and Phylogenetic analysis.

Opening reading frames (ORFs) of the virus genome encoding for proteins were predicted using ORFfinder (NIH, USA). Full genome of DcCoV-HKU23 with previous sequences from Saudi Arabia, Bovine CoV and human CoV OC43 were aligned using MAFFT. Gaps and poorly aligned regions in the alignment were manually edited. Pairwise genetic distances were calculated using MEGA 7 (24). Phylogenetic analysis of DcCoV-HKU23 was performed by maximum likelihood method using IQ-Tree, version 1.6.8 (25).

Recombination analysis was performed using Simplot version 3.5.1 (26). Bootscan analysis for a recombination event was performed on an alignment of the genome sequences as described above, with a 50% consensus sequence of 4 DcCoV-HKU23 in Nigeria with the same genotype C2/$C_{Outlier}$/C3 (NV1010, NV1092, NV1097 and NV1385) as the query

sequence. A sliding window of 600 nucleotides and a step of 100 nucleotide was used as the scanning setting.

Microneutralization assays.

Heat-inactivated (56^{0}C for 30 minutes) camel sera were first diluted 1:10, then serially two-fold diluted and mixed with equal volumes of virus at a dose of 200 50% tissue culture infective doses ($TCID_{50}$) of DcCoV-HKU23 isolate 368F (27). After 1h of incubation at 37^{0}C, 35 µL of the virus–serum mixture was added in quadruplicate to HRT-18G cell monolayers in 96-well microtiter plates. After 1 h of adsorption, the virus–serum mixture was removed and replaced with 150 µL of virus growth medium to each well. The plates were incubated for 5 days at 37 ^{0}C in 5% CO2 in a humidified incubator. Cytopathic effect was observed at day 5 post-inoculation. The highest serum dilution protecting ≥50% of the replicate wells was denoted the neutralizing antibody titer. A virus back titration of the input virus was included in each batch of tests.

Result

Screening of DcCoV-HKU23 in African camels by rt-qPCR and microneutralization assay.

Nasal swab samples of dromedary camels in Nigeria (n=2529), Morocco (n=1569) and Ethiopia (n=621) were tested for coronaviruses using the pan-CoV RT-PCR and identified by sequencing PCR amplicons (Table 1). (22). The overall prevalence of HKU23 viruses at each location ranged from 0.4% of 1569 samples tested in Morocco to 2.2% of 2529 from Nigeria (Table 1). A DcCoV-HKU23-specific quantitative real-time RT-PCR assay was subsequently performed to identify samples with a high viral RNA copy number for whole genome sequencing.

In Morocco, DcCoV-HKU23 RNA positivity in young camels aged ≤2 years (n=584) was 1.0% and not significantly different from adults (n=577) with 0.17% positive (Fisher's exact test, P=0.124). In Nigeria, DcCoV-HKU23 RNA positivity in young camels was 4.1% (n=194) compared to 2.0% in adults (n=2335) (Fisher's exact test, P=0.0674). In Ethiopia, young camels (n=136) had 1 positive swab while adults (n=314) had 5 positive swabs, which

was also without significant difference (Fisher's exact test, P=0.673). Swab specimens were collected during the months October – April with virus detection in most months (Table 2).

To study the seroprevalence of Dc-CoV-HKU23 in African camels, dromedary sera were also collected from a subset of camels during the same sampling occasions and were tested by micro-neutralization assay. A high seroprevalence was detected in dromedary camels in all three countries, with a seroprevalence of 92% of 150 sera in Nigeria, 91% of 100 sera in Ethiopia and 79% of 100 sera in Morocco respectively (Table 3). A lower seropositive rate was observed in younger (≤2 years) compared to older Moroccan camels from abattoirs (48% vs. 92%; Fisher's exact test, P=0.0036) and farms (76% vs. 100%; Fisher's exact test, P=0.0223). There was no marked difference in seroprevalence of young and old camels in the Nigerian abattoir or in abattoirs or farms in Ethiopia.

Cross-neutralizing antibody response of DcCoV-HKU23 and BCoV were evaluated by testing camel sera with high, medium, low and no neutralizing DcCoV-HKU23 titers by neutralization tests with BCoV-Mebus strain. There was significant correlation between titers of DcCoV-HKU23 and BCoV, suggesting likely serological cross-reactivity between the two viruses (Figure 1).

Evolutionary divergence and genetic diversity of DcCoV-HKU23.

Full genomes of DcCoV-HKU23 were obtained from four swab-samples in Nigeria (NV1010, NV1092, NV1097 and NV1385) and one sample each from Morocco (CAC2586) and Ethiopia (CAC1019). The African virus genomes were found closely related with pairwise base substitutions per site below 0.0270 (Table 4). These full genomes were compared with those previously reported from Saudi Arabia (18). DcCoV-HKU23 in the African region differed from those from Saudi Arabia by a range of 0.0223 – 0.0270 pairwise base substitutions per site, comparable to the divergence observed among the regions in Africa. Compared to other closely related species within clade A β-CoVs, they were distanced to BCoV by a range of 0.0249 – 0.0300 pairwise base substitutions per site and to HCoV-OC43 by a range of 0.0445 – 0.0468 pairwise base substitutions per site. The closest species related to HCoV-OC43 remains BCoV, rather than DcCoV-HKU23.

The genetic diversity of DcCoV-HKU23 across Africa and the Middle East was studied based on the ORF1ab gene by a distance plot using SSE version 1.3 (28) (Figure 2). As

recombination has previously been shown to increase progressively from the 5' to the 3' end of the genome (29), ORF1ab was selected to study the genetic diversity with minimal confounding by recombination. Along the position of ORF1ab gene, a mean pairwise distance of about 0.01 was observed within the 6 DcCoV-HKU23 sequences from Africa and the 4 reference sequences available from Saudi Arabia. The observed diversity of DcCoV-HKU23 was comparable to BCoV, suggesting both viruses were introduced into their animal hosts at similar points in time. Another circulating CoV in the same camel populations, MERS-CoV, was included in the analysis and showed a diversity of about 0.004 by the same analysis (Figure 2), relatively lower compared to DcCoV-HKU23.

Phylogenetic analysis of DcCoV-HKU23 with BCoV and HCoV-OC43 sequences.

To infer the phylogenetic relationship of the newly identified African DcCoV-HKU23 with previously reported DcCoV-HKU23 viruses from Saudi Arabia and bovine coronaviruses which are closely related, phylogenetic trees based on the complete coding sequences of RdRp (2783 nt), Spike (4101 nt) and nucleocapsid (1347 nt) gene were constructed. In addition to the 6 full genomes, four more virus sequences with complete RdRp, S, and N genes of DcCoV-HKU23 from Ethiopia (CAC1320, CAC1452) and Morocco (CAC2505, CAC2753), were obtained and included in the analysis. Using the genotyping nomenclature previously described for HCoV-OC43 and BCoV (30, 31) as a reference-point, our sequences of DcCoV-HKU23 in this study were mapped into the 3 main sub-clusters of BCoV, namely C1, C2 and C3, in which C1 contains BCoVs from the Americas, C2 contains BCoVs from Europe and C3 contains the prototype BCoV (Figure 3). In the phylogenetic analysis of the RdRp gene, all the African and Saudi DcCoV-HKU23 clustered within clade C2, which includes BCoVs from Europe. African DcCoV-HKU23 sequences did not form a monophyletic clade with Saudi Arabia strain, instead these sequences were scattered within clade C2, suggesting a multiple ancestral origin of DcCoV-HKU23 across different geographic areas.

The analysis of the spike gene showed that the 8 African and Saudi DcCoV-HKU23 were clustered together and grouped into a clade distinct from BCoV, which we designated as clade $C_{Outlier}$ and has a basal phylogenetic relationship to the BCoV clades. The phylogeny of the viruses was geographically structured so that most of the sequences were grouped within the $C_{Outlier}$ clade into subclades of Saudi Arabia, Ethiopia and Western / Northern Africa

(Nigeria and Morocco) viruses. The phylogenetic tree of the spike gene resembles the region dependent diversity of MERS-CoV as observed in these camels, in which viruses from Africa and Middle East were grouped into two separate clades (17). However, two sequences from Morocco (CAC2505 and CAC2753) were distinct from other sequences and fell into the clade C2 of BCoVs, sharing a common ancestor with a cluster of BCoVs from France.

The phylogeny of the N gene of DcCoV-HKU23 was more diverse with virus sequences being distributed in BCoV clade C3 which included 4 Nigerian sequences (NV1010, NV1092, NV1097 and NV1385), one Moroccan sequence (CAC2586) and two sequences from Ethiopia (CAC1320). These sequences clustered together monophyletically and were related to the human enteric coronavirus strain 4408. The other 3 sequences, Ethiopia CAC1019 and Morocco (CAC2753, CAC2505) were grouped within BCoV clade C2 together with the strains from Saudi Arabia.

Combining the clade classification of these three gene regions, there were 3 circulating genotypes of DcCoV-HKU23, viz C2/C2/C2; C2/$C_{Outlier}$/C2 and C2/$C_{Outlier}$/C3 (Table 5) suggesting multiple genetic recombination occurred in the past. This contrasts with BCoV which does not appear to exhibit such genetic instability (31). However, there is a lack of BCoV sequence data from Africa. These recombinant genotypes C2/$C_{Outlier}$/C2 and C2/$C_{Outlier}$/C3 were observed in DcCoV-HKU23 across the African region without a distinct geographic pattern. Genotype C2/$C_{Outlier}$/C2 was observed in one sample from Saudi Arabia and one from Ethiopia, while C2/$C_{Outlier}$/C3 were observed in 4 samples from Nigeria, one from Ethiopia and one from Morocco. The BCoV genotype C2/C2/C2 was observed in two Moroccan strains (CAC2753, CAC2505), possibly suggesting a direct spill-over of BCoV genotype C2 to the camel population.

Distinct genetic region upstream of NS5a among DcCoV-HKU23.

The genomic organization of African DcCoV-HKU23 is almost identical to Saudi Arabia DcCoV-HKU23 and BCoV, except for a 400nt region between S gene and NS5a that was found to be highly divergent among DcCoV-HKU23 and other clade A β-CoVs (Figure 4). In BCoV, this region contains 2 ORFs (4a and 4b) that encode a size of 4.9 kDa and 4.8 kDa non-structural proteins respectively. The absence of ORF4a and 4b in HCoV-OC43 suggested they are not essential for viral replication (32). Pairwise comparison of this region among all DcCoV-HKU23 with BCoV-DB2, RbCoV HKU14 and HCoV-OC43 revealed nonsense

mutations in both ORF4a and 4b in DcCoV-HKU23 from Saudi Arabia and Africa, resulting in premature stop codons and a truncated protein. A 200nt deletion was observed in NV1385 after the premature stop codon of ORF4a. Similar, though not identical patterns of deletion were also found in RbCoV-HKU14 and HCoV-OC43 that resulted in more truncated protein sequences. Although these protein sequences are varied, the nucleotide sequences of DcCoV-HKU23/362F, DcCoV-HKU23/CAC1019, RbCoV-HKU14 and HCoV-OC43 in fact share a high pairwise similarity with the BCoV sequences that contain both full length ORF4a and 4b. Interestingly, DcCoV/NV1097 and DcCoV/CAC2586 contains a distinct ORF4a that encodes a 8.9 kDa non-structural protein among other DcCoVs in this region. Homology of this protein was BLAST searched and mapped to another 8.6 kDa non-structural protein encoded by a Rodent CoV RtMm-CoV-1/IM2014 (Accession No.: KY370052.1) with about 60% amino acid similarity (Figure 4b). The BCoV ORF4a and 4b have previously been suggested to be counterparts of the 11 kDa non-structural protein in mouse hepatitis virus (MHV) (33). The discovery of this rodent-like ORF4a encoded in DcCoV-HKU23 illustrated a possible homologous recombination with rodent coronaviruses that highlighted its distinct evolutionary history as compared to BCoV.

Recombination analysis with other clade A β-CoVs.

To study the possibility of cross species recombination, Bootscan analysis of full genomes of DcCoV-HKU23 in Africa with other clade A β-CoVs was performed using Simplot, version 3.5.1. A multiple sequence alignment of the 6 African stain DcCoV-HKU23 with RbCoV-HKU14, PHEV, BCoV-DB2, EqCoV-NC99, RodentCoV-IM2014, DcCoV-HKU23 from Saudi Arabia and HCoV-OC43 was made. Using the Nigeria strain DcCoV-HKU23 as the query, recombination signals were observed with BCoV-DB2 at the NS2a gene (position 21901-22600), with RbCoV-HKU14 at the hemagglutinin esterase (HE) gene (position 22601-23600) and with the RodentCoV-IM2014 at the region of ORF4a, 4b and NS5a gene (position 27901-28800) (Figure 5a). Phylogenetic analysis of the BCoV signal region showed the BCoV-DB2 clustered with the group of DcCoV-HKU23 from Africa (Figure 5b). The signal at NS2a gene extended the region showing the mixing of BCoV with DcCoV-HKU23, in addition to the RdRp and N gene. The tree of the RbCoV signal region showed the RbCoV-HKU14 changed its phylogenetic position and linked to the cluster of DcCoV-HKU23 from Nigeria and Ethiopia. This clustering suggests a recombination event between a common ancestor of DcCoV-HKU23 sequences from Nigeria or Ethiopia with the RbCoV-HKU14 or a RbCoV-like virus. The RodentCoV signal at the region of ORF4a and 4b

supported the homologous recombinant of West African DcCoV-HKU23 with a RodentCoV in the genetic organization analysis. A tree plotted from position 27901-28800 showed DcCoV-HKU23 were split into two separate evolutionary pathways (as illustrated as BCoV-like and RodentCoV-like in Figure 4a), in which West Africa DcCoV-HKU23 were clustered outgroup with RodentCoV-IM2014, while East Africa DcCoV-HKU23 were clustered with BCoV.

Discussion

Our data provides an enhanced understanding of the diversity and circulation of an endemic DcCoV-HKU23 in African dromedaries in East (Ethiopia), West (Nigeria) and North (Morocco) Africa in comparison with viruses in Saudi Arabia and for the evolutionary relationships between HKU23 and BCoV, an important pathogen of cattle (34). In this study, HKU23 viral RNA was detected in 2.2% of dromedary nasal swabs tested in Nigeria, 0.5% in Morocco and 1.4% in Ethiopia respectively. The rates were comparatively higher than a previous study with year-around sampling of camel nasal swabs done in Saudi Arabia where a virus detection rate of 0.2% was reported (18). We detected DcCoV-HKU23 in most sampling months from October to April, which were the only periods of the year we investigated suggesting there was no clear seasonality of virus activity. While co-circulation of three CoVs, MERS-CoV, DcCoV-229E and DcCoV-HKU23 were reported in Saudi Arabia camels, similar virus circulation was also observed in Nigeria, with a positive rate of 2.2% for MERS-CoV (21) and 1.0% for DcCoV-229E (data not published). Serological prevalence of DcCoV-HKU23 antibodies in camels were 92% in Nigeria, 91% in Ethiopia and 79% in Morocco respectively, suggesting a widespread circulation of this or an antigenically related virus over a broad geographical area across Africa, in a manner comparable with MERS-CoV (20, 35). Younger camels had lower seropositive rates than adults in Morocco, but no age related difference was observed in Nigeria and Ethiopia. Differences in DcCoV-HKU23 seroprevalence may be attributed to the variations in the husbandry practices, co-habitant animal hosts or climatic factors between these countries. Similar seroprevalence were observed in camels in abattoirs and farms, suggesting virus circulation has already been established at the farm or herd levels and did not solely reflect amplification in the camel marketing chain. There is extensive cross neutralization of both DcCoV-HKU23 and BCoV, with a trend to higher titers to DcCoV-HKU23. The high amino

351 acid sequence identity of the Spike protein (92% - 97%) of DcCoV-HKU23 and BCoV very
352 likely contributed the cross neutralization between the two viruses.

353

354 The full genome sequences of DcCoV-HKU23 across the Africa and Saudi Arabia allowed us
355 to study the genetic diversity of this virus in camel populations. The distance plot using the
356 ORF1ab gene can modestly evaluate the diversity within DcCoV-HKU23 sequences due to
357 random mutations, with lesser effect contributed by recombination. The observed diversity of
358 DcCoV-HKU23 was comparable to BCoV, suggesting both viruses have established in their
359 hosts for a similar period of time. It is of interest to observe DcCoV-HKU23 to have a much
360 higher diversity compared with MERS-CoV. At present there are three CoVs (Camel CoV
361 229E, DcCoV-HKU23 and MERS-CoV) co-circulating in dromedary camels, the narrower
362 genetic diversity of MERS-CoV possibly indicates a more recent introduction into camels or
363 a purifying selection event in its more recent evolutionary history.

364 Phylogenetic analysis of the full genomes of DcCoV-HKU23 in Africa with other published
365 BCoV sequences available at Genbank identified incongruent topologies in the phylogenetic
366 trees of RdRp, S and N, indicating events of recombination. With reference to the BCoV
367 genotyping method described previously, DcCoV-HKU23 obtained in this study were
368 classified into 3 genotypes: C2/C2/C2, C2/$C_{Outlier}$/C2 and C2/$C_{Outlier}$/C3. The $C_{Outlier}$ clade has
369 no BCoV sequences and is a uniquely DcCoV-HKU23 clade. Two DcCoVs-HKU23 from
370 Morocco (CAC2505, CAC2753) showed a genotype C2/C2/C2, which is a non-recombinant
371 variant of BCoV clade2, suggesting a BCoV evolutionary origin and the possibility of a
372 BCoV spill-over into the Moroccan camel population. This is the first time that a non-
373 recombinant BCoV variant has been detected in camels. Previously, BCoV has been detected
374 in a wide range of ungulate hosts, including bovine, waterbuck, sambar deer, white-tail deer,
375 alpaca, giraffe, stable antelope, buffalo and yak (36-39). The expansion of BCoV to camel
376 hosts further illustrates its ability to cross species and infect other similar ungulate hosts.
377 Recently, surveillance of BCoV has been expanded into regions of East Asia, including
378 China (40), Korea (41)and Vietnam (42), and Caribbean in Cuba (43). The Spike gene of
379 BCoV from these regions were all phylogenetically clustered into C1. Figure 6 summarizes
380 the currently known geographic distribution of different genotypes of BCoV/HKU23-like
381 viruses in camels, bovines and other species.

382

383 The phylogeny of the S gene of DcCoV exhibited a region dependent diversity that is also
384 noted in BCoV and MERS-CoV. The C1 and C2 genotypes of BCoV correspond to the
385 American-Asia and European cluster respectively. In this study, the C2 genotype of the two
386 Morocco strains that clustered to BCoVs detected in France suggesting a divergence from the
387 common ancestor of C2 genotype BCoVs. Other DcCoVs identified in this study as well as
388 those previously identified in Saudi Arabia are recombinants, because their S gene showed an
389 outgroup topology to BCoVs and HCoV-OC43, indicating DcCoV-HKU23 acquired its S
390 gene through recombination with an ancestor yet to be identified. However, we cannot infer
391 whether the recombination occurred prior or after the introduction of the virus to camels with
392 the present data. The phylogeny of the N gene of DcCoV-HKU23 also varied among the 8
393 recombinant sequences with $C_{Outlier}$ genotype in the S gene. 7 DcCoV-HKU23 sequences
394 (NV1010, NV1092, NV1097, NV1385, CAC1320, CAC1452 and CAC2586) were grouped
395 to C3 with the N gene tree, indicating a close relatedness to prototype BCoV in the N gene.
396 DcCoV-HKU23/CAC1019 from Ethiopia and DcCoV-HKU23 from Saudi Arabia were
397 grouped to the C2 European clade. These C2 genotype sequences suggest a possible
398 recombination event between the N gene of a non-recombinant BCoV strain with a
399 recombinant DcCoV-HKU23 strain as the backbone. Overall, DcCoV-HKU23 exhibited a
400 broader diversity that contrasted to the genetic stability as observed in BCoV. However, a
401 limitation in the analysis is the lack of BCoV sequence data in Africa. With more BCoV
402 genetic datasets from Africa, the recombination events of DcCoV-HKU23 and BCoV may be
403 more clearly resolved.

404

405 Cross species recombination of DcCoV-HKU23 was also observed with other clade A β-CoV
406 species that involved a rabbit-CoV HKU14-like virus. Rabbit-CoV14 is a virus initially
407 discovered through surveillance in China (44), but similar viruses may be present in a much
408 wider geographic region. Bootscan analysis showed DcCoV-HKU23 from Nigeria and
409 Ethiopia showed a recombination signal with RbCoV-HKU14-like virus at the position
410 22601-23600, encoding the NS2a and HE gene. The HE gene in clade A β-CoVs has been
411 suggested to be acquired by a heterologous recombination with influenza C virus. The
412 recombination here was located in a similar region suggesting a possible recombination
413 hotspot. The phylogeny of the signal region showed the RbCoV sequence was linked to the
414 DcCoV-HKU23 sequences from Ethiopia and Nigeria at a basal position, suggesting the
415 recombination event may have occurred with the ancestral sequence from these two regions.

The HE protein encoded in clade A β-CoVs plays a role in the receptor binding through its receptor binding or receptor destroying activity to glycan components (45). An example of the functional property of HE has been demonstrated in the adaptation of HCoV-OC43 and HKU1 to infect humans, when the HE lectin domain was progressively lost through accumulated mutations (46). The diversity of HE gene between parental and recombinant DcCoVs deserves further research to characterize the glycoconjugates targeted by the receptor destroying activity and study how such virion-glycan interactions may contribute to its host tropism.

We further identified multiple mutations in the downstream ORF4a and 4b encoded between S and NS5a gene that provided insight on the evolutionary origin of clade A β-CoVs. The deletion patterns of ORF4a and 4b observed in DcCoV-HKU23, RbCoV-HKU14 and HCoV-OC43 revealed a stepwise deletion among these sequences. While these patterns may suggest a BCoV-origin of these sequences, it is also possible that an ancestral virus infected multiple hosts and bovines preferentially retained those ORFs. Nonsense mutations and deletions of these ORFs in DcCoV-HKU23, RbCoV-HKU14 and HCoV-OC43 supported the contention that this region may not contain essential genetic sequences and the loss of such genetic information will not impair virus fitness in dromedaries. In fact, ORF4a and 4b in BCoV has previously been suggested as vestiges of an 11-kDa protein encoded by mouse hepatitis virus (MHV) resulting from a nonsense mutation in the middle of the ORF4 (33). The region between the S and E gene may suggest a mouse or murine CoV origin of this region. As additional evidence, we also observed another 8.9 kDa protein encoded by the ORF4a in DcCoV-HKU23 identified in this study which mapped to a similar protein in rodent CoV with 60% amino acid similarity. Bootscan analysis showed DcCoV from Morocco (CAC2586) and Nigeria (NV1010, 1092, 1097 & 1385) were phylogenetically outgrouped with rodent CoV at the position from 27901-28800, suggesting a possible homologous recombination. These sequences altogether illustrated a multiple origin of clade A β-CoVs from rodent like species. Recent surveillance of rodent species have identified many more novel CoV species, including ChRCoV-HKU24, LAMV, LRLV and rodent CoV (47-49). These sequences are phylogenetically positioned at the deep branch rooting members of clade A β-CoVs. The discovery of sequence remnants of rodent CoV in DcCoV-HKU23 further support the involvement of rodent like species in the evolutionary history of clade A β-CoVs (47).

448 The occurrence of the recombination events identified in this study requires the co-infection
449 of two or more different CoVs in the same host and same cell and this requires the parental
450 viruses to be co-circulating in the same geographic region. The wide host range and
451 geographic range of BCoV-like viruses could potentially facilitate recombination with other
452 coronaviruses. Recombination events that led to the emergence of SARS-like CoV likely
453 occurred in its natural reservoirs in bats, and the virus then spilled over to intermediate hosts
454 such as civets, raccoon dogs and humans. Since rodents harbor the highest diversity of clade
455 A β-CoVs (49), one may speculate that it may be possible for recombination events to occur
456 in rodents with the recombinant virus subsequently spilling over to other mammalian hosts if
457 it has competitive advantages over pre-existing strains. Thus, one may speculate that the
458 recombination of DcCoV-HKU23 with RabbitCoV-HKU14 and RodentCoV-IM2014 could
459 have occurred in a rodent species.

460

461 Limitations of this study include the lack of availability of rectal swabs to evaluate the tissue
462 tropism associated with DcCoV-HKU23 infection. Although the virus seems not to cause
463 significant disease as it was detected in apparently healthy camels in abattoirs, it is unclear
464 that whether infection is limited to the upper respiratory mucosa similar to MERS-CoV or
465 whether it spreads more systemically. Specimens from sites of the body other than the upper
466 respiratory tract may provide information on the tropism of the virus. The lack of year round
467 sampling precludes conclusions on the seasonality of virus activity. On the other hand, the
468 strength of the study is sampling across East, West and North Africa which allows an
469 understanding of the virus diversity across a large geographic region. The lack of BCoV
470 sequences from Africa precludes a more definitive analysis of the origins of the different
471 genotypes of HKU23.

472 In conclusion, the study showed a mosaic structure of DcCoV-HKU23 that is likely to be
473 contributed by several recombination events among clade A β-CoVs. Among the three
474 identified DcCoVs that circulate in dromedary camels, MERS-CoV has so far demonstrated
475 intraspecies recombination, while DcCoV-HKU23, in addition, further demonstrated inter-
476 species recombination. Our study highlighted the importance of studying recombination of
477 CoVs to understand its evolutionary history and cross species transmission of coronaviruses
478 in dromedaries.

479

Acknowledgements

The study was supported by research grants from NIH Contract HHSN272201400006C (to M.P. and R.J.W.), a Commissioned Grant from the Health and Medical Research Fund, Food and Health Bureau, Government of the Hong Kong Special Administrative Region (to M.P.), the Theme-based Research Scheme (Project No. T11/707/15), University Grant Committee, Hong Kong (to PCW).

References

1. Zhou P, Fan H, Lan T, Yang XL, Shi WF, Zhang W, Zhu Y, Zhang YW, Xie QM, Mani S, Zheng XS, Li B, Li JM, Guo H, Pei GQ, An XP, Chen JW, Zhou L, Mai KJ, Wu ZX, Li D, Anderson DE, Zhang LB, Li SY, Mi ZQ, He TT, Cong F, Guo PJ, Huang R, Luo Y, Liu XL, Chen J, Huang Y, Sun Q, Zhang XL, Wang YY, Xing SZ, Chen YS, Sun Y, Li J, Daszak P, Wang LF, Shi ZL, Tong YG, Ma JY. 2018. Fatal swine acute diarrhoea syndrome caused by an HKU2-related coronavirus of bat origin. Nature 556:255-258.
2. Vijgen L, Keyaerts E, Moes E, Thoelen I, Wollants E, Lemey P, Vandamme AM, Van Ranst M. 2005. Complete genomic sequence of human coronavirus OC43: molecular clock analysis suggests a relatively recent zoonotic coronavirus transmission event. J Virol 79:1595-604.
3. Corman VM, Eckerle I, Memish ZA, Liljander AM, Dijkman R, Jonsdottir H, Juma Ngeiywa KJ, Kamau E, Younan M, Al Masri M, Assiri A, Gluecks I, Musa BE, Meyer B, Muller MA, Hilali M, Bornstein S, Wernery U, Thiel V, Jores J, Drexler JF, Drosten C. 2016. Link of a ubiquitous human coronavirus to dromedary camels. Proc Natl Acad Sci U S A 113:9864-9.
4. Peiris JS, Yuen KY, Osterhaus AD, Stohr K. 2003. The severe acute respiratory syndrome. N Engl J Med 349:2431-41.
5. 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 102:14040-5.
6. Ge XY, Li JL, Yang XL, Chmura AA, Zhu G, Epstein JH, Mazet JK, Hu B, Zhang W, Peng C, Zhang YJ, Luo CM, Tan B, Wang N, Zhu Y, Crameri G, Zhang SY, Wang LF, Daszak P, Shi ZL. 2013. Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. Nature 503:535-8.
7. Makino S, Keck JG, Stohlman SA, Lai MM. 1986. High-frequency RNA recombination of murine coronaviruses. J Virol 57:729-37.
8. Cheng CP, Nagy PD. 2003. Mechanism of RNA recombination in carmo- and tombusviruses: evidence for template switching by the RNA-dependent RNA polymerase in vitro. J Virol 77:12033-47.
9. Fu K, Baric RS. 1992. Evidence for variable rates of recombination in the MHV genome. Virology 189:88-102.
10. Kempf BJ, Peersen OB, Barton DJ. 2016. Poliovirus Polymerase Leu420 Facilitates RNA Recombination and Ribavirin Resistance. J Virol 90:8410-21.
11. Denison MR, Graham RL, Donaldson EF, Eckerle LD, Baric RS. 2011. Coronaviruses: an RNA proofreading machine regulates replication fidelity and diversity. RNA Biol 8:270-9.

12. Luytjes W, Bredenbeek PJ, Noten AF, Horzinek MC, Spaan WJ. 1988. Sequence of mouse hepatitis virus A59 mRNA 2: indications for RNA recombination between coronaviruses and influenza C virus. Virology 166:415-22.
13. Zaki AM, van Boheemen S, Bestebroer TM, Osterhaus AD, Fouchier RA. 2012. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. N Engl J Med 367:1814-20.
14. Perera RA, Wang P, Gomaa MR, El-Shesheny R, Kandeil A, Bagato O, Siu LY, Shehata MM, Kayed AS, Moatasim Y, Li M, Poon LL, Guan Y, Webby RJ, Ali MA, Peiris JS, Kayali G. 2013. Seroepidemiology for MERS coronavirus using microneutralisation and pseudoparticle virus neutralisation assays reveal a high prevalence of antibody in dromedary camels in Egypt, June 2013. Euro Surveill 18:pii=20574.
15. Reusken CB, Haagmans BL, Muller MA, Gutierrez C, Godeke GJ, Meyer B, Muth D, Raj VS, Smits-De Vries L, Corman VM, Drexler JF, Smits SL, El Tahir YE, De Sousa R, van Beek J, Nowotny N, van Maanen K, Hidalgo-Hermoso E, Bosch BJ, Rottier P, Osterhaus A, Gortazar-Schmidt C, Drosten C, Koopmans MP. 2013. Middle East respiratory syndrome coronavirus neutralising serum antibodies in dromedary camels: a comparative serological study. Lancet Infect Dis 13:859-66.
16. Haagmans BL, Al Dhahiry SH, Reusken CB, Raj VS, Galiano M, Myers R, Godeke GJ, Jonges M, Farag E, Diab A, Ghobashy H, Alhajri F, Al-Thani M, Al-Marri SA, Al Romaihi HE, Al Khal A, Bermingham A, Osterhaus AD, AlHajri MM, Koopmans MP. 2014. Middle East respiratory syndrome coronavirus in dromedary camels: an outbreak investigation. Lancet Infect Dis 14:140-5.
17. Chu DKW, Hui KPY, Perera R, Miguel E, Niemeyer D, Zhao J, Channappanavar R, Dudas G, Oladipo JO, Traore A, Fassi-Fihri O, Ali A, Demissie GF, Muth D, Chan MCW, Nicholls JM, Meyerholz DK, Kuranga SA, Mamo G, Zhou Z, So RTY, Hemida MG, Webby RJ, Roger F, Rambaut A, Poon LLM, Perlman S, Drosten C, Chevalier V, Peiris M. 2018. MERS coronaviruses from camels in Africa exhibit region-dependent genetic diversity. Proc Natl Acad Sci U S A doi:10.1073/pnas.1718769115.
18. 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 351:81-4.
19. Woo PC, Lau SK, Wernery U, Wong EY, Tsang AK, Johnson B, Yip CC, Lau CC, Sivakumar S, Cai JP, Fan RY, Chan KH, Mareena R, Yuen KY. 2014. Novel betacoronavirus in dromedaries of the Middle East, 2013. Emerg Infect Dis 20:560-72.
20. Miguel E, Chevalier V, Ayelet G, Ben Bencheikh MN, Boussini H, Chu DK, El Berbri I, Fassi-Fihri O, Faye B, Fekadu G, Grosbois V, Ng BC, Perera RA, So TY, Traore A, Roger F, Peiris M. 2017. Risk factors for MERS coronavirus infection in dromedary camels in Burkina Faso, Ethiopia, and Morocco, 2015. Euro Surveill 22.
21. So RT, Perera RA, Oladipo JO, Chu DK, Kuranga SA, Chan KH, Lau EH, Cheng SM, Poon LL, Webby RJ, Peiris M. 2018. Lack of serological evidence of Middle East respiratory syndrome coronavirus infection in virus exposed camel abattoir workers in Nigeria, 2016. Euro Surveill 23.
22. Poon LL, Chu DK, Chan KH, Wong OK, Ellis TM, Leung YH, Lau SK, Woo PC, Suen KY, Yuen KY, Guan Y, Peiris JS. 2005. Identification of a novel coronavirus in bats. J Virol 79:2001-9.
23. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754-60.
24. Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol Biol Evol 33:1870-4.
25. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol 32:268-74.

26. 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 73:152-60.
27. Woo PC, Lau SK, Fan RY, Lau CC, Wong EY, Joseph S, Tsang AK, Wernery R, Yip CC, Tsang CC, Wernery U, Yuen KY. 2016. Isolation and Characterization of Dromedary Camel Coronavirus UAE-HKU23 from Dromedaries of the Middle East: Minimal Serological Cross-Reactivity between MERS Coronavirus and Dromedary Camel Coronavirus UAE-HKU23. Int J Mol Sci 17.
28. Simmonds P. 2012. SSE: a nucleotide and amino acid sequence analysis platform. BMC Res Notes 5:50.
29. Fu K, Baric RS. 1994. Map locations of mouse hepatitis virus temperature-sensitive mutants: confirmation of variable rates of recombination. J Virol 68:7458-66.
30. 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 85:11325-37.
31. Kin N, Miszczak F, Diancourt L, Caro V, Moutou F, Vabret A, Ar Gouilh M. 2016. Comparative molecular epidemiology of two closely related coronaviruses, bovine coronavirus (BCoV) and human coronavirus OC43 (HCoV-OC43), reveals a different evolutionary pattern. Infect Genet Evol 40:186-191.
32. Mounir S, Talbot PJ. 1993. Human coronavirus OC43 RNA 4 lacks two open reading frames located downstream of the S gene of bovine coronavirus. Virology 192:355-60.
33. Abraham S, Kienzle TE, Lapps WE, Brian DA. 1990. Sequence and expression analysis of potential nonstructural proteins of 4.9, 4.8, 12.7, and 9.5 kDa encoded between the spike and membrane protein genes of the bovine coronavirus. Virology 177:488-95.
34. Saif LJ. 2010. Bovine respiratory coronavirus. Vet Clin North Am Food Anim Pract 26:349-64.
35. Reusken CB, Messadi L, Feyisa A, Ularamu H, Godeke GJ, Danmarwa A, Dawo F, Jemli M, Melaku S, Shamaki D, Woma Y, Wungak Y, Gebremedhin EZ, Zutt I, Bosch BJ, Haagmans BL, Koopmans MP. 2014. Geographic distribution of MERS coronavirus among dromedary camels, Africa. Emerg Infect Dis 20:1370-4.
36. Hasoksuz M, Alekseev K, Vlasova A, Zhang X, Spiro D, Halpin R, Wang S, Ghedin E, Saif LJ. 2007. Biologic, antigenic, and full-length genomic characterization of a bovine-like coronavirus isolated from a giraffe. J Virol 81:4981-90.
37. Jin L, Cebra CK, Baker RJ, Mattson DE, Cohen SA, Alvarado DE, Rohrmann GF. 2007. Analysis of the genome sequence of an alpaca coronavirus. Virology 365:198-203.
38. Decaro N, Martella V, Elia G, Campolo M, Mari V, Desario C, Lucente MS, Lorusso A, Greco G, Corrente M, Tempesta M, Buonavoglia C. 2008. Biological and genetic analysis of a bovine-like coronavirus isolated from water buffalo (Bubalus bubalis) calves. Virology 370:213-22.
39. Alekseev KP, Vlasova AN, Jung K, Hasoksuz M, Zhang X, Halpin R, Wang S, Ghedin E, Spiro D, Saif LJ. 2008. Bovine-like coronaviruses isolated from four species of captive wild ruminants are homologous to bovine coronaviruses, based on complete genomic sequences. J Virol 82:12422-31.
40. Keha A, Xue L, Yan S, Yue H, Tang C. 2019. Prevalence of a novel bovine coronavirus strain with a recombinant hemagglutinin/esterase gene in dairy calves in China. Transbound Emerg Dis doi:10.1111/tbed.13228.
41. Jeong JH, Kim GY, Yoon SS, Park SJ, Kim YJ, Sung CM, Shin SS, Lee BJ, Kang MI, Park NY, Koh HB, Cho KO. 2005. Molecular analysis of S gene of spike glycoprotein of winter dysentery bovine coronavirus circulated in Korea during 2002-2003. Virus Res 108:207-12.
42. Shin J, Tark D, Le VP, Choe S, Cha RM, Park GN, Cho IS, Nga BTT, Lan NT, An DJ. 2019. Genetic characterization of bovine coronavirus in Vietnam. Virus Genes 55:415-420.

43. Martinez N, Brandao PE, de Souza SP, Barrera M, Santana N, de Arce HD, Perez LJ. 2012. Molecular and phylogenetic analysis of bovine coronavirus based on the spike glycoprotein gene. Infect Genet Evol 12:1870-8.

44. Lau SK, Woo PC, Yip CC, Fan RY, Huang Y, Wang M, Guo R, Lam CS, Tsang AK, Lai KK, Chan KH, Che XY, Zheng BJ, Yuen KY. 2012. Isolation and characterization of a novel Betacoronavirus subgroup A coronavirus, rabbit coronavirus HKU14, from domestic rabbits. J Virol 86:5481-96.

45. Schultze B, Wahn K, Klenk HD, Herrler G. 1991. Isolated HE-protein from hemagglutinating encephalomyelitis virus and bovine coronavirus has receptor-destroying and receptor-binding activity. Virology 180:221-8.

46. Bakkers MJ, Lang Y, Feitsma LJ, Hulswit RJ, de Poot SA, van Vliet AL, Margine I, de Groot-Mijnes JD, van Kuppeveld FJ, Langereis MA, Huizinga EG, de Groot RJ. 2017. Betacoronavirus Adaptation to Humans Involved Progressive Loss of Hemagglutinin-Esterase Lectin Activity. Cell Host Microbe 21:356-366.

47. Lau SK, Woo PC, Li KS, Tsang AK, Fan RY, Luk HK, Cai JP, Chan KH, Zheng BJ, Wang M, Yuen KY. 2015. Discovery of a novel coronavirus, China Rattus coronavirus HKU24, from Norway rats supports the murine origin of Betacoronavirus 1 and has implications for the ancestor of Betacoronavirus lineage A. J Virol 89:3076-92.

48. Wang W, Lin XD, Guo WP, Zhou RH, Wang MR, Wang CQ, Ge S, Mei SH, Li MH, Shi M, Holmes EC, Zhang YZ. 2015. Discovery, diversity and evolution of novel coronaviruses sampled from rodents in China. Virology 474:19-27.

49. Wu Z, Lu L, Du J, Yang L, Ren X, Liu B, Jiang J, Yang J, Dong J, Sun L, Zhu Y, Li Y, Zheng D, Zhang C, Su H, Zheng Y, Zhou H, Zhu G, Li H, Chmura A, Yang F, Daszak P, Wang J, Liu Q, Jin Q. 2018. Comparative analysis of rodent and small mammal viromes to better understand the wildlife origin of emerging infectious diseases. Microbiome 6:178.

650 Table 1. Screening of DcCoV-HKU23 in dromedary camels from Nigeria, Ethiopia and
651 Morocco.

Country	Sampling year	Sample positive / tested * (% positive)	Ct range	Field collections reported in Ref.
Nigeria	2015, 2016	Total: 55/2529 (2.2%) Young camels: 8/194 (4.1%) Adult camels: 47/2335 (2.0%)	18.8 – 36.9	So et al,2018
Morocco	2015, 2016	Total: 7/1569* (0.45%) Young camels: 6/584 (1.0%) Adult camels: 1/577 (0.17%)	26.5 – 33.4	Miguel et al, 2017
Ethiopia	2015	Total: 9/621* (1.4%) Young camels: 1/136 (0.74%) Adult camels: 5/314 (1.6%)	23.2 – 30.2	Miguel et al, 2017

652 * Note: Age information were not available for all sampled camels.

653

654

655 Table 2. The monthly RNA rate of Dc-CoV-HKU23 in camels from Africa and Saudi Arabia.

Month	RNA rate (%)			
	2014-2015		2015-2016	
	Morocco	Ethiopia	Nigeria	Morocco
May - September	ND	ND	ND	ND
October	ND	ND	1/526 (0.2%)	ND
November	ND	ND	15/739 (2.0%)	ND
December	ND	ND	1/35 (2.9%)	ND
January	ND	2/120 (1.7%)	12/531 (2.3%)	0/349 (0%)
February	0/195 (0%)	7/501 (1.4%)	26/698 (3.7%)	ND
March	0/186 (0%)	ND	ND	5/385 (1.3%)
April	ND	ND	ND	2/453 (0.4%)

656 ND, No data;

Table 3. The seropositive rate of Dc-CoV-HKU23 in camel sera from Nigeria, Ethiopia and Morocco by micro-neutralization assay.

Region	Nigeria		Ethiopia				Morocco			
Location type	Abattoir		Abattoir		Farm		Abattoir		Farm	
Camel age group	Young (age <=2)	Adult (age >2)	Young	Adult	Young	Adult	Young	Adult	Young	Adult
Seropositive rate	12/16 (75%)	123/134 (92%)	22/24 (92%)	22/26 (85%)	23/25 (92%)	24/25 (96%)	12/25 (48%)	23/25 (92%)	19/25 (76%)	25/25 (100%)
Median titre	1:40	1:20	1:40	1:40	1:40	1:40	<1:10	1:20	1:20	1:160

Table 4. Estimates of evolutionary divergence between the complete genome sequences of DcCoV-HKU23 identified in Africa (Nigeria, Ethiopia and Morocco) and Saudi Arabia, Bovine CoV and Human CoV-OC43. Analyses were conducted using the Tamura-Nei model using MEGA7.

Host	Country	Strain	Genome size (bp)	%GC	Pairwise evolutionary divergence in the number of base substitutions per site											
					NV1010	NV1092	NV1097	NV1385	CAC1019	CAC2586	368F	362F	265F	Ry123	DB2	OC43
Camel	Nigeria	DcCoV-HKU23-NV1010	30780	36.90%	-											
		DcCoV-HKU23-NV1092	30799	36.90%	0.0001	-										
		DcCoV-HKU23-NV1097	31075	37.00%	0.0167	0.0168	-									
		DcCoV-HKU23-NV1385	30798	36.90%	0.0004	0.0004	0.0168	-								
	Ethiopia	DcCoV-HKU23-CAC1019	31021	36.90%	0.0206	0.0207	0.0246	0.0206	-							
	Morocco	DcCoV-HKU23-CAC2586	31062	37.00%	0.0191	0.0191	0.0129	0.0191	0.0270	-						
	Saudi Arabia	DcCoV-HKU23-368F *KF906251.1*	31052	37.00%	0.0263	0.0265	0.0231	0.0263	0.0223	0.0225	-					
		DcCoV HKU23-362F *KF906250.1*	31052	37.00%	0.0263	0.0265	0.0231	0.0263	0.0223	0.0225	0.0000	-				
		DcCoV-HKU23-265F *KF906249.1*	31052	37.00%	0.0264	0.0266	0.0235	0.0264	0.0227	0.0229	0.0017	0.0017	-			
		DcCoV-HKU23-Ry123 *KT368891.1*	31041	37.00%	0.0269	0.0270	0.0237	0.0268	0.0229	0.0231	0.0017	0.0017	0.0026	-		
Bovine		BCoV-DB2 *DQ811784.2*	31007	37.10%	0.0299	0.0300	0.0288	0.0299	0.0249	0.0279	0.0197	0.0197	0.0200	0.0204	-	
Human		HCoV-OC43 *AY391777.1*	30738	36.80%	0.0466	0.0467	0.0468	0.0468	0.0445	0.0463	0.0420	0.0420	0.0421	0.0425	0.0333	-

Table 5. Summary of the genotypes of DcCoV-HKU23 identified in this study.

			RdRp	S	N	Genotype
DcCoV-HKU23	Nigeria	NV1010	C2	C outlier	C3	C2/C outlier/C3
		NV1092	C2	C outlier	C3	C2/C outlier/C3
		NV1097	C2	C outlier	C3	C2/C outlier/C3
		NV1385	C2	C outlier	C3	C2/C outlier/C3
	Ethiopia	CAC1019	C2	C outlier	C2	C2/C outlier/C2
		CAC1320	C2	C outlier	C3	C2/C outlier/C3
		CAC1452	C2	C outlier	C3	C2/C outlier/C3
	Morocco	CAC2505	C2	C2	C2	C2/C2/C2
		CAC2586	C2	C outlier	C3	C2/C outlier/C3
		CAC2753	C2	C2	C2	C2/C2/C2
	Saudi Arabia	362F	C2	C outlier	C2	C2/C outlier/C2
BCoV	Europe	BCoV/FRA	C2	C2	C2	C2/C2/C2
	Americas	BCoV ENT	C1	C1	C1	C1/C1/C1

Journal of Virology JVI Journal of Virology

Figure 1. Scatter plot showing camel sera (n=13) with different neutralizing titres against DcCoV HKU23 were tested for cross neutralization against BCoV.

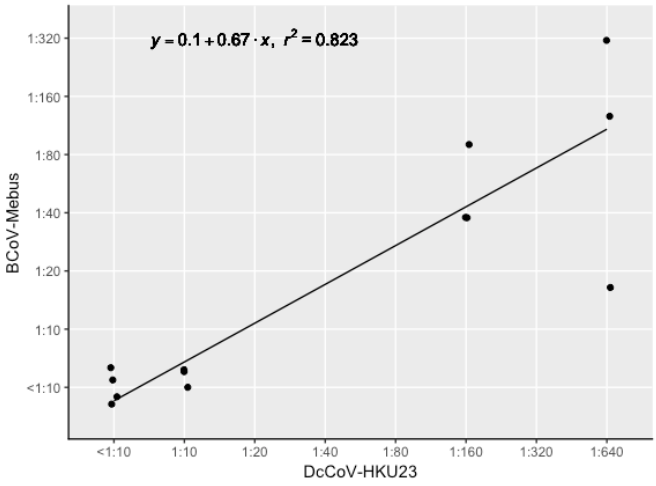


Figure 2. Comparison of the genetic diversity of the ORF1ab gene of DcCoV-HKU23, BCoV and MERS-CoV. Sequence distance plot were generated using SSE version 1.3 using a sliding window of 250 and a step size of 25 nucleotides. Sequences were obtained from Genbank database and closely related sequences with a pairwise distance < 0.001 were excluded in the analysis. A total of 10 DcCoV-HKU23 sequences, 28 BCoV sequences and 88 MERS-CoV sequences were included in the analysis.

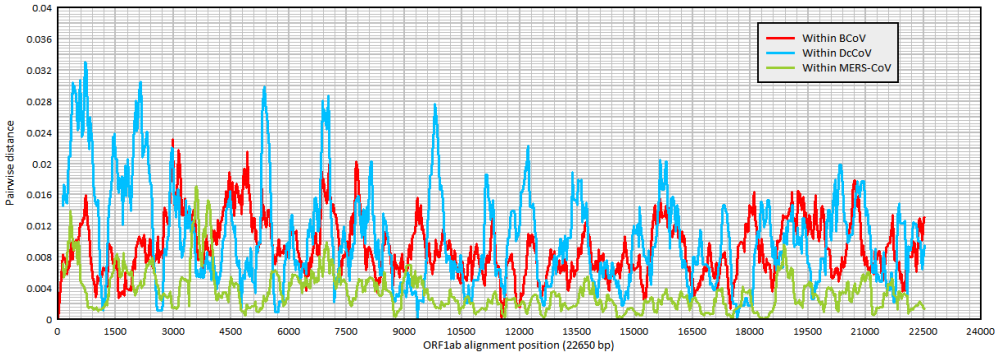
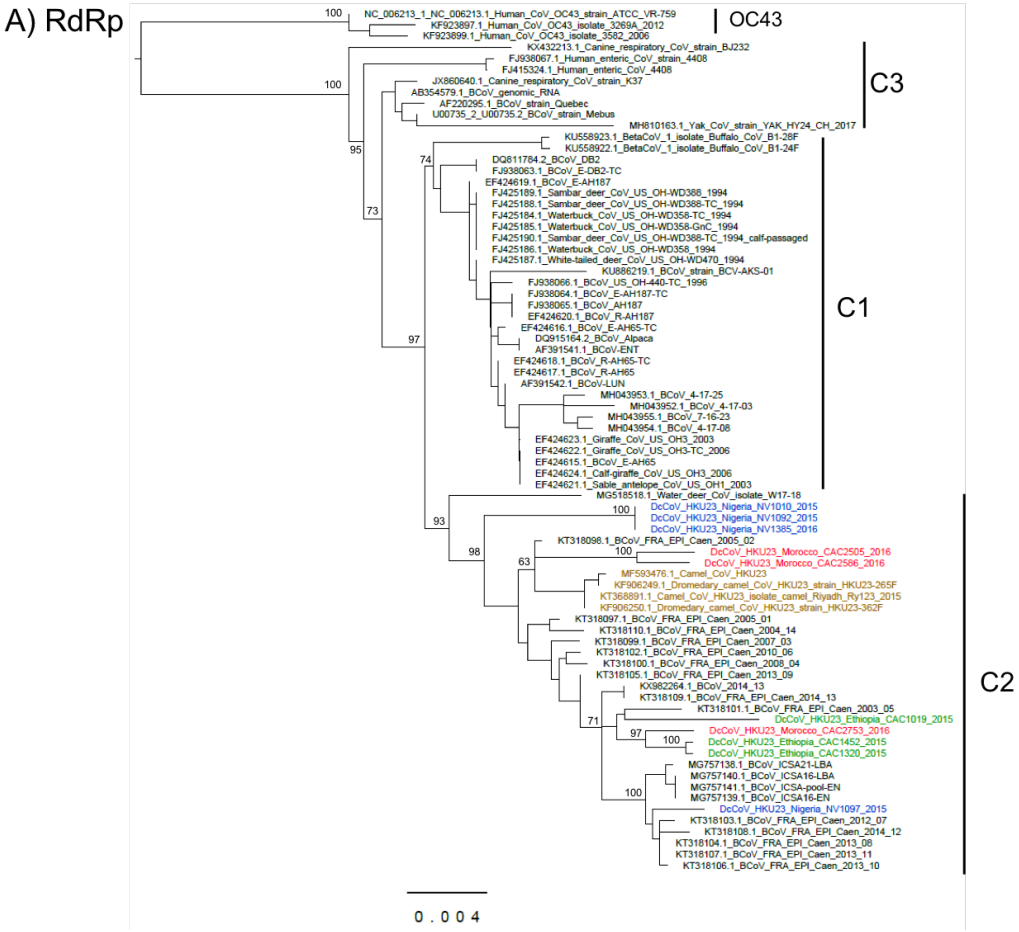
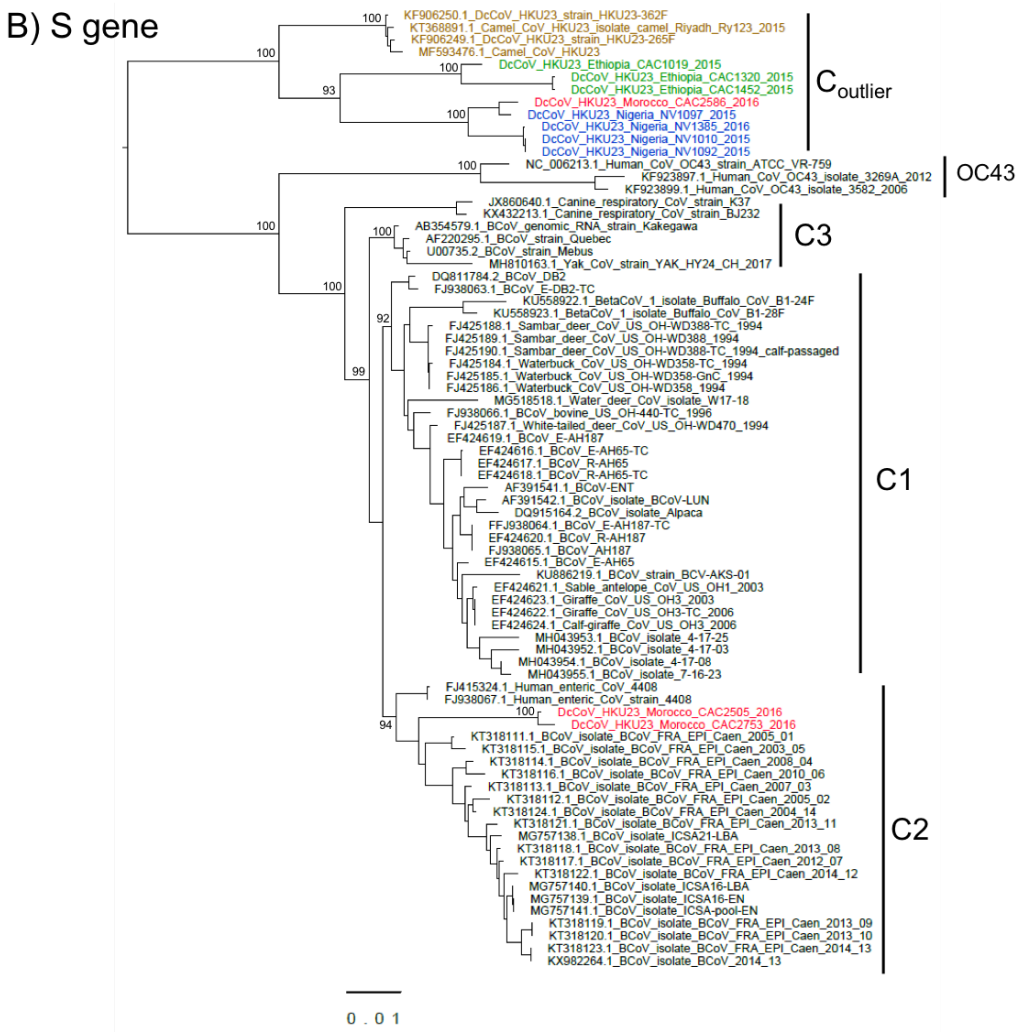


Figure 3. Phylogenetic analysis of a) RdRp, b) spike and c) nucleocapsid gene of DcCoV-HKU23 from Nigeria (colored blue), Morocco (colored red) and Ethiopia (colored green). Reference DcCoV-HKU23 sequences from Saudi Arabia were colored brown. Alignment of each gene is manually trimmed to obtain an alignment of 2769nt for RdRp, 4137nt for spike and 1347nt for nucleocapsid gene respectively. Tree was constructed by maximum likelihood method using IQ-Tree with the best-fit model automatically selected by ModelFinder. Nodes indicated bootstrap values calculated using ultrafast bootstrap with 1000 replicates. Trees were mid-point rooted.





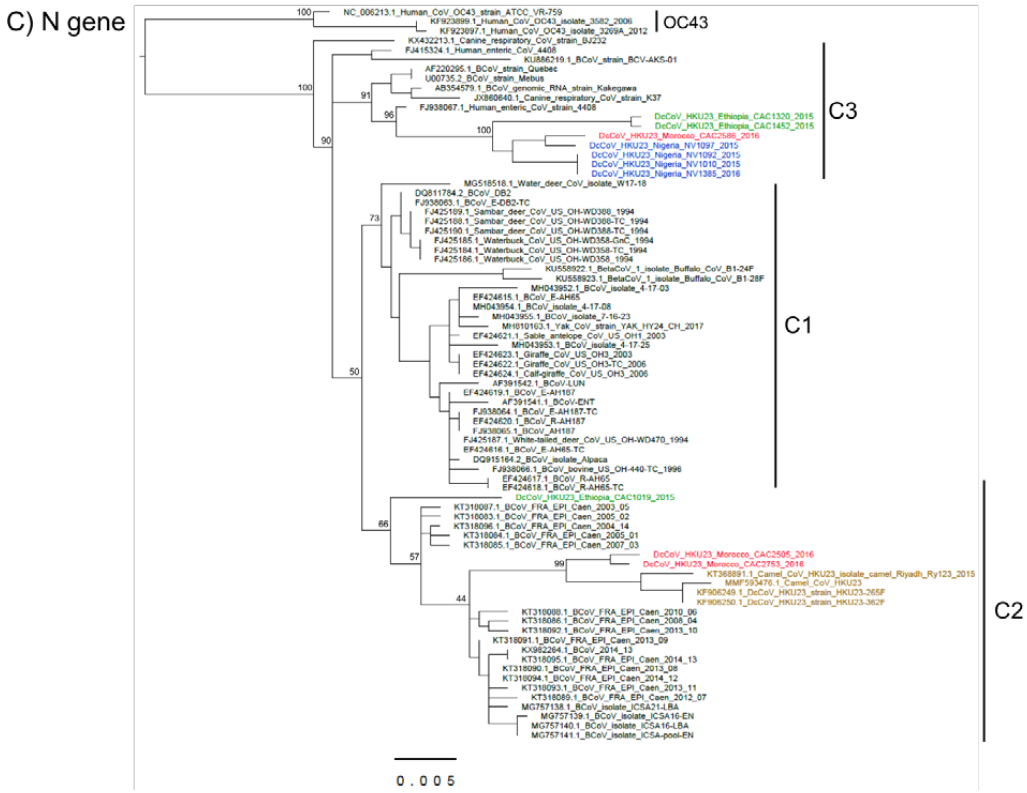
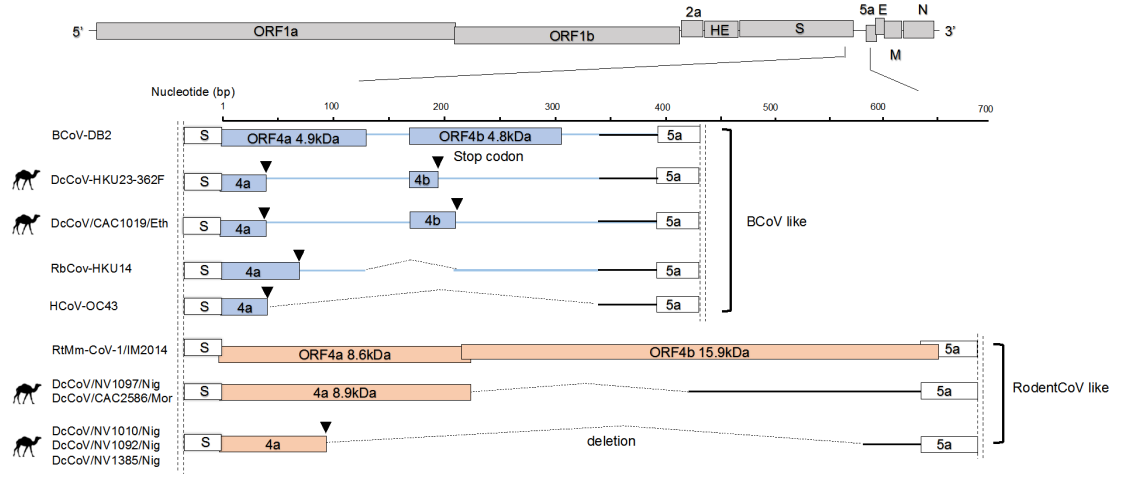


Figure 4. A) Genomic organization of DcCoV-HKU23 and other clade A β-CoVs of the region between spike gene and NS5a. Distinct ORFs patterns were found in DcCoV-HKU23 in this region. Stop codons in the ORFs were labelled by black triangle. Horizontal dotted lines indicated region of deletion. B) The amino acid sequence alignment of the 8.9kDa ORF4a of rodent coronaviruses (RtAs-CoV/IM2014, accession no. KY370044; RtMm-CoV-1/IM2014, accession no. KY370052; RtMruf-CoV-2/JL2014, accession no. KY370046) and DcCoV-HKU23.

A



B

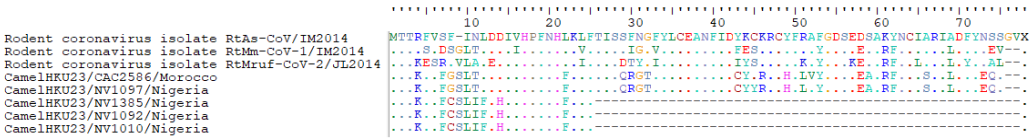


Figure 5. A) Recombination analysis of the genomes of DcCoV-HKU23, BCoV-DB2, PHEV, HCoV-OC43, RbCoV-HKU14, EquineCoV-NC99 and RodentCoV-RtMm-CoV-1/IM2014. Bootscan analysis was performed by Simplot, version 3.5.1, using a 50% consensus sequence of the DcCoV-HKU23 in Nigeria with the genotype C2/$C_{outlier}$/C3 as the query. B) Phylogenetic trees from representative regions were constructed by maximum likelihood method using IQ-Tree, version 1.6.8. Trees were midpoint rooted. Accession number of the CoVs used in this analysis: DcCoV-HKU23/362F (KF906250.1), BCoV-DB2 (DQ811784.2), PHEV (KY994645), HCoV-OC43 (AY391777.1), RbCoV-HKU14 (JN874559), EquineCoV-NC99 (EF446615) and RodentCoV-IM2014 (KY370052).

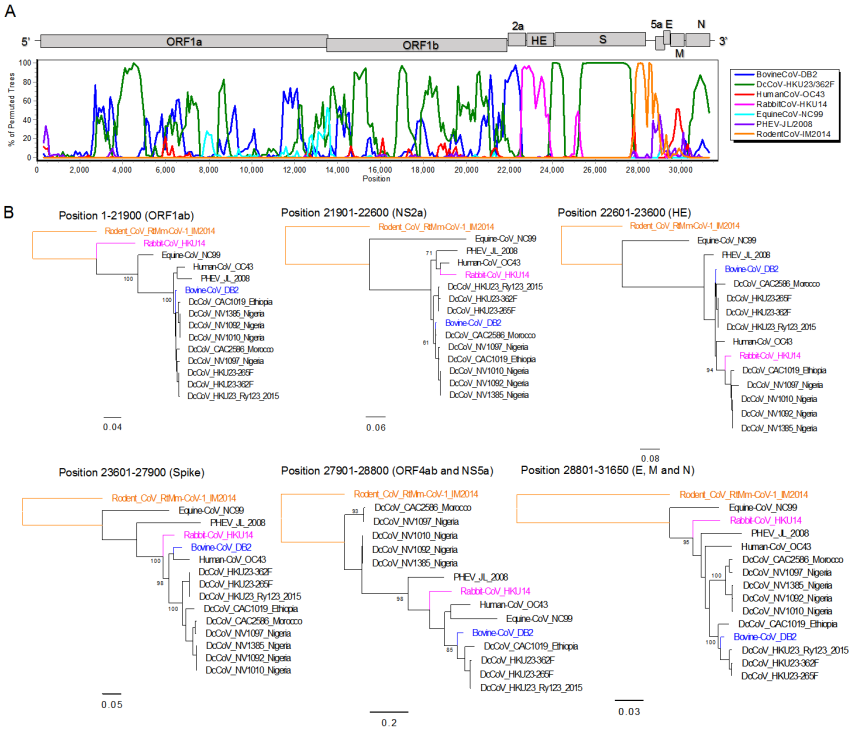


Figure 6. The geographic distribution of different genotypes of BCoV/HKU23-like viruses in camels, bovines and other species. The map was drawn using R software.