JVI Accepts, published online ahead of print on 31 December 2014
J. Virol. doi:10.1128/JVI.02420-14

JVI02420-14R1 second revision

Discovery of a novel coronavirus, China *Rattus* coronavirus HKU24, from Norway rats supports murine origin of *Betacoronavirus 1* with implications on the ancestor of *Betacoronavirus* lineage A

Susanna K. P. Lau,[a,b,c,d†] Patrick C. Y. Woo,[a,b,c,d†] Kenneth S. M. Li,[d†] Alan K. L. Tsang,[d] Rachel Y. Y. Fan,[d] Hayes K. H. Luk,[d] Jian-Piao Cai,[d] Kwok-Hung Chan,[d] Bo-Jian Zheng,[a,b,c,d] Ming Wang,[e] Kwok-Yung Yuen[a,b,c,d]*

State Key Laboratory of Emerging Infectious Diseases,[a] Research Centre of Infection and Immunology,[b] Carol Yu Centre for Infection,[c] Department of Microbiology,[d] The University of Hong Kong, Hong Kong; Guangzhou Center for Disease Control and Prevention, Guangzhou;[e] China

Running title: China *Rattus* coronavirus HKU24

[†]These authors contributed the same to the manuscript.

*Corresponding author. Mailing address: State Key Laboratory of Emerging Infectious Diseases, Department of Microbiology, The University of Hong Kong, University Pathology Building, Queen Mary Hospital, Hong Kong. Phone: (852) 22554892. Fax: (852) 28551241. E-mail: kyyuen@hkucc.hku.hk

ABSTRACT

We discovered a novel *Betacoronavirus* lineage A coronavirus, China *Rattus* coronavirus HKU24 (ChRCoV HKU24), from Norway rats in China. ChRCoV HKU24 occupied a deep branch at the root of members of *Betacoronavirus 1*, being distinct from murine coronavirus and HCoV HKU1. Its unique putative cleavage sites at nsp1/2 and S, and low sequence identities to other lineage A βCoVs in conserved replicase domains, support ChRCoV HKU24 as a separate species. ChRCoV HKU24 possessed genome features that resemble both *Betacoronavirus 1* and murine coronavirus, being closer to *Betacoronavirus 1* in most predicted proteins, but closer to murine coronavirus by G+C content, a single NS4 and absent TRS for E. Its N-terminal domain (NTD) demonstrated higher sequence identity to BCoV than to MHV NTDs, with three of four critical sugar-binding residues in BCoV and two of 14 contact residues at MHV NTD/mCEACAM1a interface being conserved. Molecular clock analysis dated the tMRCA of ChRCoV HKU24, *Betacoronavirus 1* and RbCoV HKU14 to ~1400. Cross reactivities were demonstrated between other lineage A and B βCoVs and ChRCoV HKU24 nucleocapsid but not spike polypeptide. Using the spike polypeptide-based western blot, we showed that only Norway rats and two Oriental house rats from Guangzhou were infected by ChRCoV HKU24. Other rats, including Norway rats from Hong Kong, only possessed antibodies against N protein but not spike, suggesting infection by βCoVs different from ChRCoV HKU24. ChRCoV HKU24 may represent the murine origin of *Betacoronavirus 1* and rodents are likely an important reservoir for ancestors of lineage A βCoVs.

IMPORTANCE

While bats and birds are hosts for ancestors of most coronaviruses (CoVs), lineage A βCoVs have never been found in these animals and the origin of *Betacoronavirus* lineage A remains obscure. We discovered a novel lineage A βCoV, China *Rattus* coronavirus HKU24 (ChRCoV HKU24), from Norway rats in China, with a high seroprevalence. The unique genome features and phylogenetic analysis supported that ChRCoV HKU24 represents a novel CoV species, occupying a deep branch at the root of members of *Betacoronavirus 1* and distinct from murine coronavirus. Nevertheless, ChRCoV HKU24 possessed genome characteristics that resemble both *Betacoronavirus 1* and murine coronavirus. Our data suggest that ChRCoV HKU24 represents the murine origin of *Betacoronavirus 1*, with interspecies transmission from rodents to other mammals having occurred centuries ago before the emergence of HCoV OC43 in late 1800s. Rodents may be an important reservoir for ancestors of lineage A βCoVs.

INTRODUCTION

Coronaviruses (CoVs) infect a wide variety of animals including humans, causing respiratory, enteric, hepatic and neurological diseases of varying severity. Based on genotypic and serological characterization, CoVs were traditionally classified into three distinct groups (1, 2). Recently, the Coronavirus Study Group of the International Committee for Taxonomy of Viruses (ICTV) has revised the nomenclature and taxonomy to re-classify the three CoV groups into three genera, *Alphacoronavirus*, *Betacoronavirus* and *Gammacoronavirus* (3). Novel CoVs, which represented a novel genus, *Deltacoronavirus*, have also been identified (4-6). As a result of the ability to use a variety of host receptors and evolve rapidly through mutation and recombination, CoVs are capable to adapt to new hosts and ecological niches, causing wide spectra of diseases (2, 7-12).

The severe acute respiratory syndrome (SARS) epidemic and identification of SARS-CoV-like viruses from palm civet and horseshoe bats in China has boosted interests in the discovery of novel CoVs in both humans and animals (13-20). It is now known that CoVs from all four genera can be found in mammals. Historically, alphacoronaviruses (αCoVs) and betacoronaviruses (βCoVs) are found in mammals while gammacoronaviruses (γCoVs) were found in birds. However, recent findings suggested the presence of γCoVs also in mammals (5, 21, 22). Although deltacoronaviruses (δCoVs) were also mainly found in birds, potential mammalian δCoVs have been reported (4, 23). In particular, a δCoVs closely related to sparrow CoV HKU17, porcine CoV HKU15, has been identified in pigs, which suggested avian-to-mammalian transmission (4). Based on current findings, a model for CoV evolution was proposed, where bat CoVs are likely the gene source of *Alphacoronavirus* and *Betacoronavirus*,

and avian CoVs are the gene source of *Gammacoronavirus* and *Deltacoronavirus* (4). However, one notable exception to this model is *Betacoronavirus* lineage A.

The genus *Betacoronavirus* consists of four lineages, A to D. While human coronavirus OC43 (HCoV OC43) and human coronavirus HKU1 (HCoV HKU1) belong to *Betacoronavirus* lineage A (20, 24-27), SARS coronavirus (SARS-CoV) belongs to *Betacoronavirus* lineage B and the recently emerged, Middle East Respiratory syndrome coronavirus (MERS-CoV) belongs to *Betacoronavirus* lineage C. No human CoV has yet been identified from *Betacoronavirus* lineage D. On the other hand, besides *Alphacoronavirus*, diverse bat CoVs have been found in *Betacoronavirus* lineage B (e.g. SARS-related *Rhinolophus* bat CoVs), lineage C (e.g. *Tylonycteris* bat CoV HKU4 and *Pipistrellus* bat CoV HKU5) and lineage D (e.g. *Rousettus* bat CoV HKU9) (8, 14, 15, 28-37), supporting that bat CoVs are likely the ancestral origin of other mammalian CoVs in these lineages. However, no bat CoVs belonging to *Betacoronavirus* lineage A have yet been identified, despite the numerous surveillance studies on bat CoVs conducted in various countries over the years (38). Therefore, the ancestral origin of the mammalian lineage A βCoVs, such as HCoV OC43 and HCoV HKU1, remains obscure.

While HCoV OC43 is likely to have originated from zoonotic transmission, sharing a common ancestor with bovine coronavirus (BCoV) dated back to 1890 (27, 30, 39), closely related CoVs belonging to the same species, *Betacoronavirus 1*, have also been found in various mammals including pigs, horses, dogs, waterbucks, sable antelope, deer, giraffes, alpaca and dromedary camels, suggesting a common ancestor in mammals with subsequent frequent interspecies transmission (40-47). Although no zoonotic origin of HCoV HKU1 has been identified, the virus is most closely related to mouse hepatitis virus (MHV) and rat coronavirus (RCoV) which, together, are now classified as murine coronavirus (3, 20, 42). We therefore

hypothesize that rodent CoVs are the ancestral origin of *Betacoronavirus* lineage A. In this study, we tested samples from various rodent species in Hong Kong and southern China for the presence of lineage A βCoVs. A novel CoV, China *Rattus* coronavirus HKU24 (ChRCoV HKU24), was discovered from Norway rats in Guangzhou. Complete genome analysis showed that ChRCoV HKU24 represents a novel species within *Betacoronavirus* lineage A, but possessed features that resemble both *Betacoronavirus 1* and murine coronavirus. High seroprevalence was also demonstrated among Norway rats from Guangzhou using western blot analysis against ChRCoV HKU24 recombinant N protein and spike polypeptide. The present results suggest that ChRCoV HKU24 likely represents the murine origin of *Betacoronavirus 1* and provides insights on the ancestor of *Betacoronavirus* lineage A.

MATERIALS AND METHODS

Sample collection. All rodent samples were collected from January 2010 to August 2012 using procedures described previously (5, 14). Samples from southern China were collected from animal markets or restaurants. Samples from Hong Kong were collected from wild and street rodents by the Agriculture, Fisheries and Conservation Department, and Food and Environmental Hygiene Department of the Hong Kong Special Administrative Region (HKSAR) respectively. Alimentary samples were placed in viral transport medium containing Earle's balanced salt solution (Invitrogen, New York, United States), 20% glucose, 4.4% NaHCO3, 5% bovine albumin, 50000 ug/ml vancomycin, 50000 ug/ml amikacin, 10000 units/ml nystatin, before transportation to the laboratory for RNA extraction. The study was approved by the Committee on the Use of Live Animals for Teaching and Research, The University of Hong Kong.

RNA extraction. Viral RNA was extracted from the samples using QIAamp Viral RNA Mini Kit (Qiagen, Hilden, Germany). The RNA was eluted in 60 μl of Buffer AVE and was used as the template for RT-PCR.

RT-PCR of RdRp gene of CoVs using conserved primers and DNA sequencing. Initial CoV screening was performed by amplifying a 440-bp fragment of the RNA-dependent RNA polymerase (RdRp) gene of CoVs using conserved primers (5'-GGTTGGGACTATCCTAAGTGTGA-3' and 5'-CCATCATCAGATAGAATCATCATA-3') designed by multiple alignments of the nucleotide (nt) sequences of available RdRp genes of known CoVs (14, 20). Reverse transcription was performed using SuperScript III kit (Invitrogen, San Diego, CA, USA). The PCR mixture (25 μl) contained cDNA, PCR buffer (10 mM Tris-HCl pH 8.3, 50 mM KCl, 2 mM $MgCl_2$ and 0.01% gelatin), 200 μM of each dNTPs and 1.0 U *Taq*

polymerase (Applied Biosystems, Foster City, CA, USA). The mixtures were amplified in 60 cycles of 94°C for 1 min, 50°C for 1 min and 72°C for 1 min and a final extension at 72°C for 10 min in an automated thermal cycler (Applied Biosystems, Foster City, CA, USA). Standard precautions were taken to avoid PCR contamination and no false-positive was observed in negative controls.

PCR products were gel-purified using the QIAquick gel extraction kit (Qiagen, Hilden, Germany). Both strands of the PCR products were sequenced twice with an ABI Prism 3700 DNA Analyzer (Applied Biosystems, Foster City, CA, USA), using the two PCR primers. The sequences of the PCR products were compared with known sequences of the RdRp genes of CoVs in the GenBank database.

Viral culture. The three rodent samples positive for ChRCoV HKU24 by RT-PCR were subject to virus isolation in Huh-7.5 (human hepatoma), Vero E6 (African green monkey kidney), HRT-18G (human rectum epithelial), BSC-1 (African green monkey renal epithelial), RK13 (rabbit kidney), MDBK (bovine kidney), NIH/3T3 (mouse embryonic fibroblast), J774 (mouse macrophage), BHK-21 (baby hamster kidney) and RK3E (rat kidney), RMC (rat kidney mesangial), RAW264.7 (mouse macrophage) and primary SD rat lung cells as described previously (48, 49).

Real-time RT-PCR quantitation. Real-time RT-PCR was performed on rodent samples positive for ChRCoV HKU24 by RT-PCR using previously described procedures (14). Reverse transcription was performed using the SuperScript III kit with random primers (Invitrogen, San Diego, CA, USA). cDNA was amplified in Lightcycler instrument with a FastStart DNA Master SYBR Green I Mix reagent kit (Roche Diagnostics GmbH, Mannheim, Germany) using specific primers 5'-ACAGGTTCTCCCTTTATAGATGAT-3') and (5'-

159 TCTCCTGTATAGTAGCAGAAGCAT-3') targeting the RdRp gene of ChRCoV HKU24 using
160 procedures described previously (14, 50). For quantitation, a reference standard was prepared
161 using pCRII-TOPO vector (Invitrogen, San Diego, CA, USA) containing the target sequence.
162 Tenfold dilutions equivalent to 3.77 to 3.77×10^9 copies per reaction were prepared to generate
163 concomitant calibration curves. At the end of the assay, PCR products (133-bp fragment of
164 RdRp) were subjected to melting curve analysis (65–95°C, 0.1°C/s) to confirm the specificity of
165 the assay. The detection limit of this assay was 3.77 copies per reaction.

166 **Complete genome sequencing.** Three complete genomes of ChRCoV HKU24 were
167 amplified and sequenced using the RNA extracted from the original alimentary samples as
168 templates. The RNA was converted to cDNA by a combined random-priming and oligo(dT)
169 priming strategy. The cDNA was amplified by degenerate primers designed by multiple
170 alignments of the genomes of other CoVs with complete genomes available, using strategies
171 described in our previous publications (14, 20, 35, 49) and the CoV database, CoVDB (51), for
172 sequence retrieval. Additional primers were designed from the results of the first and subsequent
173 rounds of sequencing. These primer sequences are available on request. The 5' ends of the viral
174 genomes were confirmed by rapid amplification of cDNA ends using the 5'/3' RACE kit (Roche
175 Diagnostics GmbH, Mannheim, Germany). Sequences were assembled and manually edited to
176 produce final sequences of the viral genomes.

177 **Genome analysis.** The nt sequences of the genomes and the deduced amino acid (aa)
178 sequences of the open reading frames (ORFs) were compared to those of other CoVs with
179 available complete genomes using the CoVDB (51). Phylogenetic tree construction was
180 performed using maximum likelihood method using PhyML, with bootstrap values calculated
181 from 100 trees. Protein family analysis was performed using PFAM and InterProScan (52, 53).

Prediction of transmembrane domains was performed using TMHMM (54). The structure of ChRCoV HKU24 N-terminal domain (NTD) was predicted using a web-based homology-modelling server, SWISS-MODEL. BLASTp search was performed against Protein Data Bank (PDB) with the default parameters to find suitable templates for homology modelling. Based on the higher sequence identity, QMEAN Z-score, coverage and lower e-value, crystal structure of the BCoV NTD (PDB code: 4h14) was selected as template. The predicted structure was visualized using Jmol.

Estimation of divergence dates. Divergence time was calculated based on complete RdRp and HE gene sequence data using a Bayesian Markov Chain Monte Carlo (MCMC) approach as implemented in BEAST (version 1.8.0) as described previously (49, 55, 56). One parametric model (Constant Size) and one nonparametric model (Bayesian Skyline) tree priors were used for inference. Analyses were performed under SRD06 model, and using both a strict and a relaxed molecular clock. MCMC run was 2×10^8 steps long with sampling every 1,000 steps. Convergence was assessed on the basis of effective sampling size after a 10% burn-in using Tracer software, version 1.5 (55). The mean time of the most recent common ancestor (tMRCA) and the highest posterior density regions at 95% (HPDs) were calculated, and the best-fitting models were selected by a Bayes factor using marginal likelihoods implemented in Tracer (56). Bayesian skyline under a relaxed-clock model with an uncorrelated exponential distribution was adopted for making inferences, as Bayes factor analysis for the RdRp and HE genes indicated that this model fitted the data better than other models tested. The tree was summarized in a target tree by the Tree Annotator program included in the BEAST package by choosing the tree with the maximum sum of posterior probabilities (maximum clade credibility) after a 10% burn-in.

Cloning and purification of $(His)_6$-tagged recombinant ChRCoV HKU24 nucleocapsid protein and spike polypeptide. To produce fusion plasmids for protein purification, primers 5'-CTAGCTAGCATGTCTCATACGCCA-3' and 5'-CTAGCTAGCTTATATTTCTGAGCTTCCC -3', and 5'-CTAGCTAGCCAACCAATAGCAGATGTGTA-3' and 5'-CTAGCTAGCTTATCTCTTGGCTCGCCATGT-3', were used to amplify the nucleocapsid gene and a partial S1 fragment encoding amino acid residues 317 to 763 of the spike protein of ChRCoV HKU24 respectively as described previously (31, 49, 57, 58). The sequences, coding for a total of 443 aa and 447 aa residues respectively, were amplified and cloned into the NheI site of expression vector pET-28b(+) (Merck, KGaA, Darmstadt, Germany) in frame and downstream of the series of six histidine residues. The $(His)_6$-tagged recombinant nucleocapsid protein and spike polypeptide were expressed and purified using the Ni-NTA affinity chromatography (Qiagen, Hilden, Germany) according to the manufacturer's instructions.

Western blot analysis. To detect the presence of antibodies against ChRCoV HKU24 N protein and spike polypeptide in rodent sera and to test for possible cross antigenicity between ChRCoV HKU24 and other βCoVs, 600 ng of purified $(His)_6$-tagged recombinant N protein or spike polypeptide of ChRCoV HKU24 was loaded into the well of a sodium dodecyl sulfate (SDS)–10% polyacrylamide gel and subsequently electroblotted onto a nitrocellulose membrane (Bio-Rad, Hercules, CA, USA). The blot was cut into strips and the strips were incubated separately with 1:2000, 1:4000 or 1:8000 dilutions of sera collected from rodents with serum samples available, human sera from two patients with HCoV OC43 infection, sera from two rabbits with RbCoV HKU14 and human sera from two patients with SARS-CoV infection respectively. Antigen-antibody interaction was detected with 1:4000 horse radish peroxidase-

conjugated anti-rat IgG, anti-human IgG or anti-rabbit IgG (Zymed) and ECL fluorescence system (GE Healthcare Life Sciences, Little Chalfont, UK) as described previously (14, 58).

Nucleotide sequence accession numbers. The nt sequences of the three genomes of ChRCoV HKU24 have been lodged within the GenBank sequence database under accession no. KM349742-KM349744.

RESULTS

Identification of a novel CoV from Norway rats in China. Of 91 alimentary samples from rodents in China, RT-PCR for a 440-bp fragment in the RdRp gene of CoVs was positive for a potentially novel CoV in three samples from Norway rats (*Rattus norvegicus*) from a restaurant in Guangzhou (Table 1). None of the 573 alimentary samples from rodents in Hong Kong, including those from Norway rats, was positive for CoVs. Sequencing results suggested that the potentially novel virus was most closely related to MHV with ≤85% nt identities, and members of the species *Betacoronavirus 1* including HCoV OC43, BCoV, equine coronavirus (ECoV) and porcine hemagglutinating encephalomyelitis virus with ≤84% nt identities. Quantitative RT-PCR showed that the viral load in the positive samples ranged from 1.2×10^3 to 1.3×10^6 copies/g. Attempts to stably passage ChRCoV HKU24 in cell cultures were unsuccessful, with no cytopathic effect or viral replication being detected.

Genome organization and coding potential of ChRCoV HKU24. Complete genome sequence data of three strains of ChRCoV HKU24 were obtained by assembly of the sequences of RT-PCR products from the RNA directly extracted from the corresponding individual specimens. The three genomes shared >99% nt sequence similarity. Their genome size was 31234 bases, with the G + C content (40%) closer to that of murine coronavirus than to that of *Betacoronavirus 1* (Table 2). The genome organization is similar to that of other lineage A βCoVs, with the characteristic gene order 5'-replicase ORF1ab, haemagglutinin-esterase (HE), spike (S), envelope (E), membrane (M), nucleocapsid (N)-3' (Table 2 and Fig. 1). Moreover, additional ORFs coding for non-structural proteins, NS2a, NS4, NS5 and N2, are found. A putative transcription regulatory sequence (TRS) motif, 5'-CUAAAC-3', similar to that of αCoVs and the motif, 5'-UCUAAAC-3', in other lineage A βCoVs, was identified at the 3' end

257 of the leader sequence and precedes each ORF except NS4, E and N2 genes (Table 3) (26, 49,
258 59-61). However, there were base mismatches for HE and NS5, with an alternative TRS motif,
259 5'-CUGAAC-3' and 5'-GUAAAC-3' respectively.

260 The coding potential and characteristics of putative non-structural proteins (nsps) of
261 ORF1 of ChRCoV HKU24 were shown in Tables 3 and 4. The ORF1 polyprotein possessed
262 68.6-75.0% aa identities to the polyproteins of other lineage A βCoVs. It possessed a unique
263 putative cleavage site, G/L, between nsp1 and nsp2, in contrast to G/V found in other lineage A
264 βCoVs except HCoV HKU1 with G/I (Table 4 and Fig. 1). Other predicted cleavage sites were
265 mostly conserved between ChRCoV HKU24 and other lineage A βCoVs. However, the lengths
266 of nsp1, nsp2, nsp3, nsp13, nsp15 and nsp16 in ChRCoV HKU24 differed from those of
267 corresponding nsps in members of *Betacoronavirus 1* and murine coronavirus, as a result of
268 deletions or insertions.

269 All lineage A βCoVs, except HCoV HKU1, possess NS2a gene between ORF1ab and HE.
270 Unlike RbCoV HKU14 with the NS2a broken into several small ORFs (49), ChRCoV HKU24 is
271 predicted to possess a single NS2a protein as in other lineage A βCoVs. This NS2a protein
272 displayed 43.7-62.0% aa identities to those of *Betacoronavirus 1* and 45.7-47.3% aa identities to
273 those of murine coronavirus. Although the βCoV-specific NS2 protein has been shown to be
274 non-essential for in vitro viral replication (62), cyclic phosphodiesterase domains have been
275 predicted in the NS2 proteins of some CoVs and toroviruses, and a possible role in viral
276 pathogenicity has been suggested in MHV (63, 64). In contrast to MHV and RCoV, such domain
277 was not found in ChRCoV HKU24.

278 Similar to other CoV S protein, the S of ChRCoV HKU24 is predicted to be a type I
279 membrane glycoprotein, with most of the protein (residues 16-1302) exposed on the outside of

the virus and with a transmembrane domain (residues 1303-1325) at the C terminus (Fig. 2). Two heptad repeats (HR), important for membrane fusion and viral entry, were located at residues 1045-1079 (HR1) and 1253-1285 (HR2). The S protein of ChRCoV HKU24 possessed 66.7-69.6% aa identities to those of members of *Betacoronavirus 1* and 62.4-64.3% identities to those of members of murine coronavirus. The aa sequence identity between the ChRCoV HKU24 NTD and BCoV and MHV NTDs was 61 and 56%, respectively. BCoV and HCoV OC43 utilize N-acetyl-9-O acetyl neuramic acid as receptor for initiation of infection (65, 66). In contrast, MHV utilizes carcinoembryonic antigen-related cell adhesion molecule 1 (CEACAM1) as receptor and its receptor-binding domain does not bind sugars (10, 67, 68). Recent structural studies showed that, among the four critical sugar-binding residues in BoV, a Glu→Gly substitution was found in one residue in MHV, which may explain the reduction in sugar-binding affinity. In ChRCoV HKU24, a Glu→Ser substitution is found at this position (Fig. 2). Comparison of the aa sequences between the S proteins of ChRCoV HKU24 and MHV showed that ChRCoV HKU24 possessed many aa substitutions in the region corresponding to the MHV NTD (Fig. 2). In particular, 12 of the 14 important contact residues at the MHV NTD/mCEACAM1a interface were not conserved between ChRCoV HKU24 and MHV. Similar to the MHV and BCoV NTDs, the ChRCoV HKU24 NTD is also predicted to contain a core structure with β-sandwich fold as human galectins (galactose-binding lectins) using homology modelling (10). Modelling showed that the β-sandwich core structure of ChRCoV HKU24 consists of one six-stranded β-sheet and one seven-stranded β-sheet that are stacked together through hydrophobic interactions (Fig. 2). In addition, the S of ChRCoV HKU24 possessed a unique predicted cleavage site, RAKR, among lineage A βCoVs.

Other predicted domains in HE, S, NS4, NS5, E, M and N proteins of ChRCoV HKU24 are summarized in Table 3 and Fig. 1. The NS4 of ChRCoV HKU24 shared 37-42% aa identity to the NS4 proteins of murine coronavirus. In most members of *Betacoronavirus 1*, the NS4 is split into smaller proteins. The NS5 of ChRCoV HKU24 is homologous to the NS5/NS5a of members of *Betacoronavirus 1* with 47.7% to 51.4% aa identities, but to the NS5 of MHV with only 39.5% aa identity. Interestingly, NS5 is not found in the genome of RCoV. The absence of a preceding TRS upstream of the E of ChRCoV HKU24 suggests that the translation of this E protein may be cap-independent, via an internal ribosomal entry site (IRES), as demonstrated in MHV (69). Similarly, the E of RCoV and HCoV HKU1 was also not preceded by TRS. This is in contrast to members of *Betacoronavirus 1* which possess a preceding TRS upstream of their E proteins (49, 61). Downstream to N gene, the 3'-untranslated region contains a predicted bulged stem-loop structure of 69 nt (nt position 30944-31012) that is conserved in βCoVs (70). Overlapping with the bulged stem-loop structure by 5 nt, a conserved pseudoknot structure (nt position 31008–31059) that is important for CoV replication is found. Since non-structural proteins in CoVs may possess unique function for replication and virulence (71, 72), further studies are warranted to understand the potential function of the nsps and NS proteins in ChRCoV HKU24.

Phylogenetic analyses. Phylogenetic trees constructed using the aa sequences of RdRp, S and N proteins of ChRCoV HKU24 and other CoVs are shown in Fig. 3, and the corresponding pairwise aa identities shown in Table 2. For all three genes, the three ChRCoV HKU24 strains formed a distinct cluster among lineage A βCoVs, occupying a deep branch at the root of and being most closely related to members of the species *Betacoronavirus 1*. Comparison of the aa sequences of the seven conserved replicase domains, ADRP, nsp5 ($3CL^{pro}$), nsp12

(RdRp), nsp13 (Hel), nsp14 (ExoN), nsp15 (NendoU) and nsp16 (O-MT), for CoV species demarcation (3) showed that ChRCoV HKU24 possessed 69.5-81.7%, 82.2-86.8%, 88.1-92.6%, 88.9-94.8%, 80.2-88.7%, 70.1-79.5% and 83.8-89.7% aa identities to other lineage A βCoVs respectively (Table 5). Based on the present results, we propose a novel species, ChRCoV HKU24, to describe this virus under *Betacoronavirus* lineage A and distinguish it from RCoV.

HE proteins are glycoproteins that mediate reversible attachment to O-acetylated sialic acids by acting as both lectins and receptor-destroying enzymes which aid viral detachment from sugars on infected cells (68, 73). Related HEs have been found in influenza C viruses, toroviruses and lineage A βCoVs, but not other CoVs. It has been suggested that HEs of lineage A βCoVs have arisen from an influenza C-like HE fusion protein, likely as a result of relatively recent lateral gene transfer events (73). Phylogenetic analysis of the HE proteins of lineage A βCoVs, toroviruses and influenza C viruses showed that they fell into three separate clusters (Fig. 3). The HE of ChRCoV HKU24 also forms a deep branch at the root of members of the species *Betacoronavirus 1* except ECoV and is distinct from members of murine coronavirus. Previous studies have demonstrated heterogeneity of gene expression of HE proteins among different MHV strains (74). Since the HE of ChRCoV HKU24 is not preceded by a perfectly matched TRS, further studies are required if it is expressed and functional.

Estimation of divergence dates. Using the uncorrelated relaxed clock model on complete RdRp gene sequences, the date of tMRCA of ChRCoV HKU24, members of *Betacoronavirus 1* and RbCoV HKU14 was estimated to be 1402 (HPDs, 918.05 to 1749.91) (Fig. 4). The date of divergence between HCoV OC43 and BCoV was estimated to be 1897 (HPDs, 1826.15 to 1950.05), consistent with results from previous molecular clock studies (27). Using the uncorrelated relaxed clock model on complete HE gene sequences, the date of tMRCA

of ChRCoV HKU24, members of *Betacoronavirus 1* and RbCoV HKU14 was estimated to be 1337 (HPDs, 724.59 to 1776.78) (Fig. 4). The date of divergence between HCoV OC43 and BCoV was estimated to be 1871 (HPDs, 1764.55 to 1944.37). The estimated mean substitution rates of the RdRp and HE data set were 1.877×10^{-4} and 4.016×10^{-4} substitution per site per year respectively, which are comparable to previous estimation in other lineage A βCoVs (26, 27, 39).

Serological studies. Western blot analysis using recombinant ChRCoV HKU24 N protein was performed using sera from 144 rodents with serum samples available, human sera from two patients with HCoV OC43 infection, sera from two rabbits with RbCoV HKU14 and human sera from two patients with SARS-CoV infection. Among tested sera from 74 Norway rats from Guangzhou with serum samples available, 60 (81.1%) were positive for antibody against recombinant ChRCoV HKU24 N protein with prominent immunoreactive bands of about 50 kDa (Table 1 and Fig. 5). These 60 positive samples include three serum samples collected from the three Norway rats positive for ChRCoV HKU24 in their alimentary samples. In addition, 15 (48.4%) of 31 Norway rats from Hong Kong were also positive for antibody against recombinant ChRCoV HKU24 N protein, although the virus was not detected in alimentary samples from these rats. Moreover, seven (77.8%) of nine oriental house rats but only four (0.13%) of 30 black rats were positive for antibody against recombinant ChRCoV HKU24 N protein. Possible cross antigenicity between ChRCoV HKU24 and other βCoVs, including lineage A and B βCoVs, was found. Human sera from two patients with HCoV OC43 infection, sera from two rabbits with RbCoV HKU14 infection and human sera from two patients with SARS-CoV infection were also positive for antibody against recombinant ChRCoV HKU24 N protein by western blot assay (Fig. 5).

Western blot analysis using recombinant ChRCoV HKU24 spike polypeptide was performed to verify the specificity of antibodies against ChRCoV HKU24 N protein using positive rodent sera and human sera from two patients with HCoV OC43 infection, sera from two rabbits with RbCoV HKU14 and human sera from two patients with SARS-CoV infection. Among sera from the 60 Norway rats with positive antibodies against ChRCoV HKU24 N protein, 21 were positive for antibodies against ChRCoV HKU24 spike polypeptide with prominent immunoreactive bands of about 50 kDa (Table 1 and Fig. 5). However, serum samples from the three Norway rats positive for ChRCoV HKU24 in their alimentary samples were negative for anti-ChRCoV HKU24 spike polypeptide antibody. Of the seven oriental house rats with positive antibodies against ChRCoV HKU24 N protein, two were positive for antibodies against ChRCoV HKU24 spike polypeptide. However, serum samples from the four black rats and 15 Norway rats from Hong Kong with positive antibodies against ChRCoV HKU24 N protein were negative for antibodies against ChRCoV HKU24 spike polypeptide. In contrast to N protein, no cross antigenicity was detected between ChRCoV HKU24 spike polypeptide and positive sera against other βCoVs, including lineage A and B βCoVs. Human sera from two patients with HCoV OC43 infection, sera from two rabbits with RbCoV HKU14 infection and human sera from two patients with SARS-CoV infection were all negative for antibody against recombinant ChRCoV HKU24 spike polypeptide by western blot assay (Fig. 5).

DISCUSSION

We discovered a novel lineage A βCoV, ChRCoV HKU24, from Norway rats in southern China. *Betacoronavirus* lineage A comprises the traditional "group 2 CoVs" including members of murine coronavirus and *Betacoronavirus 1*, HCoV HKU1 and RbCoV HKU14. ChRCoV HKU24 possessed <90% aa identities to all other lineage A βCoVs in five of the seven conserved replicase domains for CoV species demarcation by ICTV (3), supporting that ChRCoV HKU24 belongs to a separate species. The genome of ChRCoV HKU24 also possesses features distinct from those of other lineage A βCoVs, including a unique putative nsp1/nsp2 cleavage site and a unique putative cleavage site in S protein. Phylogenetically, its position at the root of *Betacoronavirus 1*, being distinct from murine coronavirus and HCoV HKU1, suggested that ChRCoV HKU24 may represent the murine ancestor for *Betacoronavirus 1*, after branching off from the common ancestor of murine coronavirus and HCoV HKU1. Interestingly, the genome of ChRCoV HKU24 possessed features that resemble both *Betacoronavirus 1* and murine coronavirus. It is more similar to *Betacoronavirus 1* than murine coronavirus by the higher sequence identities in most predicted proteins including NS2a, NS5 and S. On the other hand, it is more similar to murine coronavirus than to *Betacoronavirus 1* in terms of its G + C content, the presence of a single NS4 and absence of TRS upstream of E gene. Therefore, it is most likely that ChRCoV has evolved from the ancestor of murine coronavirus to infect other mammals, resulting in the generation of *Betacoronavirus 1* with the acquisition of TRS for E gene. The tMRCA of ChRCoV HKU24, members of *Betacoronavirus 1* and RbCoV HKU14 was estimated to be 1402 (HPDs, 918.05 to 1749.91) and 1337 (HPDs, 724.59 to 1776.78) using complete RdRp and HE gene analysis respectively, suggesting that interspecies transmission from rodents

414 to other mammals occurred at least several centuries ago before the emergence of HCoV OC43
415 in humans at approximately1890s .

416 Western blot assays based on recombinant ChRCoV HKU24 N protein and spike
417 polypeptide showed a high seroprevalence of ChRCoV HKU24 infection among Norway rats
418 from Guangzhou. We evaluated cross reactivities of both N protein and spike polypeptide assays
419 using sera from infections by other lineage A βCoVs, HCoV OC43 in humans and RbCoV
420 HKU14 in rabbits, as well as SARS-CoV, a lineage B βCoV. Cross-reacting antibodies against N
421 proteins were observed, which is in line with previous findings on cross-reactivity between N
422 proteins of different βCoVs (49, 57). In contrast, no cross reactivities were detected against spike
423 polypeptide, supporting the specificity of CoV spike polypeptide-based assays and their ability to
424 rectify cross reactivities (57, 58). Using the present assays, 60 of 74 Norway rats from
425 Guangzhou were positive for antibodies against ChRCoV HKU24 N protein, among which 21
426 were positive for antibodies ChRCoV HKU24 spike polypeptide, supporting past infections by
427 ChRCoV HKU24 in these 21 rats. Interestingly, the three Norway rats positive for ChRCoV
428 HKU24 in their alimentary samples were positive for antibodies against ChRCoV HKU24 N
429 protein but negative for antibodies against ChRCoV HKU24 spike polypeptide. This is likely due
430 to delay in mounting neutralizing antibodies against spike protein during acute infection in these
431 three rats, while antibodies against N protein may rise earlier as a result of the high abundance
432 and antigenicity of CoV N proteins or may be a result of cross-reactions from other βCoVs. The
433 finding is also in keeping with previous findings on SARS-related *Rhinolophus* bat CoV that
434 negative correlation was observed between viral load and neutralizing antibody (14). Besides
435 Norway rats, antibodies against ChRCoV HKU24 N protein and spike polypeptide were also
436 detected in two oriental house rats from Guangzhou, although antibodies against spike

polypeptide were relatively weak. This suggests possible cross-species infection of ChRCoV HKU24 or cross reactivity from a very close lineage A βCoV. Four black rats and 15 Norway rats in Hong Kong were also positive for antibodies against ChRCoV HKU24 N protein but not spike polypeptide. This suggests possible past infection by other βCoV(s) with cross-reactivities between their N proteins and that of ChRCoV HKU24. More studies on diverse rodent species from China and other countries are required to determine the natural reservoir and host range of ChRCoV HKU24 and other murine lineage A βCoVs.

The present results extend our knowledge on the evolutionary origin of CoVs. While birds are important sources for γCoVs and δCoVs, bats host diverse αCoVs and βCoVs that may be the ancestral origins of various mammalian CoVs including human CoVs. For human αCoVs, both HCoV NL63 and HCoV 229E were likely to be originated from bat CoVs. HCoV NL63 has been shown to share common ancestry with αCoVs from North American tricolored bat, with the most recent common ancestor between these viruses occurring approximately 563 to 822 years ago (75). Moreover, immortalized lung cell lines derived from this bat species allowed replication of HCoV NL63, supporting potential zoonotic-reverse zoonotic transmission cycles between bats and humans. HCoV 229E also shared a common ancestor with diverse αCoVs from leaf-nosed bats in Ghana, with the most recent common ancestor dated to 1686-1800 (76). However, no complete genomes are available for the putative bat ancestors of HCoV NL63 and HCoV-229E. For human βCoVs, SARS-CoV and MERS-CoV are also known to share common ancestors with bat CoVs. Soon after the SARS epidemic, horseshoe bats in China were found to be the reservoir for SARS-CoV-like viruses, which were postulated to have jumped from bats to civet and later humans (8, 14, 15). A recent study also reported the isolation of a SARS-like bat CoV in Vero E6 cells, and the ability of this bat virus to use the angiotensin-converting enzyme 2

(ACE2) from humans, civets and Chinese horseshoe bats for cell entry (77). MERS-CoV belongs to *Betacoronavirus* lineage C which was only known to consist of two bat viruses, *Tylonycteris* bat CoV HKU4 and *Pipistrellus* bat CoV HKU5, before the MERS epidemic (35-37). This has led to the speculation that bats may be the zoonotic origin of MERS-CoV. However, recent evidence supported dromedary camels as the immediate source of human MERS-CoV (78-80). Nevertheless, a conspecific virus from a South African *Neoromicia capensis* bat has been found to share 85% nt identity to MERS-CoV genome, suggesting acquisition of MERS-CoV by camels from bats in Sub-Saharan Africa from where camels on the Arabian peninsula are imported (81). In contrast, there has been no evidence for bats as the origin of human lineage A βCoVs such as HCoV OC43 and HCoV HKU1. HCoV OC43, being closely related to BCoV, is believed to have emerged relatively recently from bovine-to-human transmission at around 1890 (27, 30, 39). Both viruses belonged to the promiscuous CoV species, *Betacoronavirus 1*, which consists of many closely related mammalian CoVs, implying a low threshold for cross-mammalian species transmission and a complex evolutionary history among these viruses (40-47, 49). However, the ancestral origin of members of *Betacoronavirus 1* remains elusive. As for HCoV HKU1, no recent zoonotic ancestor has yet been identified, although the virus is most closely related to members of murine coronavirus (20, 42). Although rodents constitute approximately 40% of all mammalian species, murine coronavirus has been the only CoV species known to exist in rodents. This is in contrast to the large diversity of CoVs found in bats which make up another 20% of all species of mammals (6, 33, 36). The present results suggest that rodents may be an important reservoir for lineage A βCoVs and may harbor other ancestral viruses of *Betacoronavirus 1* and HCoV HKU1 (Fig. 6). Nevertheless, many mysteries remain unresolved in the evolution of lineage A βCoVs, such as the origin of their HE proteins. For

example, both toroviruses and influenza C viruses can be found in bovine and porcine samples. Further studies are required to determine if the HE of potential rodent CoV ancestors of *Betacoronavirus* lineage A may have been acquired from cattle or pigs.

The potential pathogenicity and tissue tropism of ChRCoV HKU24 remains to be determined. While CoVs are associated with a wide spectrum of diseases in animals, some CoVs, especially those from bats, were detected in apparently healthy individuals without obvious signs of disease (8, 14, 15, 31, 33). The detection of ChRCoV HKU24 in the alimentary samples of Norway rats suggested possible enteric tropism. However, the three positive rats did not show obvious diseases. MHV, the prototype CoV most extensively studied before the SARS epidemic, can cause a variety of neurological, hepatic, gastrointestinal and respiratory diseases in mice, depending on the strain tropism and route of inoculation. The virus, originally isolated from a mouse with spontaneous encephalomyelitis, causes disseminated encephalomyelitis with extensive destruction of myelin and focal necrosis of the liver in experimentally infected mice (82-84). Strain MHV-A59 is primarily hepatotropic, while strain MHV-JHM is neurotropic. Enterotropic strains can spread quickly as a result of high level of excretion in feces and cause significant environmental contamination in animal houses. Respiratory-tropic or polytropic strains, although uncommon, are the strains that commonly contaminate cell lines. As for RCoV, it causes diseases primarily in the respiratory tract, with strain sialodacryoadenitis (SDAV) being more associated with upper respiratory tract, salivary and lacrimal gland, and eye infections, and strain RCoV-Parker causing pneumonia in experimentally infected rats (85, 86). Further investigations are required to study the tissue tropism and pathogenicity of ChRCoV HKU24 in Norway rats and other potential rodent reservoirs.

Elucidating the receptor of ChRCoV HKU24 will be important to understand the mechanism of host adaptation and interspecies transmission from rodents to other mammals. The higher sequence identity to *Betacoronavirus 1* than to murine coronavirus in the S protein and NTD of ChRCoV HKU24 is in line with other regions of the genome. Homology modelling showed that the conformation of the sugar binding loop in BCoV NTD is conserved in ChRCoV HKU24 NTD. Moreover, three of the four critical sugar-binding residues in BCoV but only two of the 14 contact residues at the MHV NTD/mCEACAM1a interface are conserved in ChRCoV HKU24. While it remains to be ascertained if ChRCoV HKU24 may utilize sugar or CEACAM1 as receptor, its predicted NTD appears to resemble that of BCoV more than that of MHV. Based on the presence of β-sandwich fold in the NTDs of MHV and BCoV, it has been proposed that CoV NTDs may have originated from a host galectin with sugar-binding functions, but evolved new structural features in MHV for binding to CEACMA1 (10, 87). If rodents are indeed the host origin for *Betacoronavirus* lineage A including *Betacoronavirus 1*, it would be interesting to study the sugar-binding activity of NTDs of different rodent βCoVs to understand their evolutionary history. Although some lineage A βCoVs, such as *Betacoronavirus 1* and MHV, can replicate in cell lines such as BSC-1 and HRT-18 cells, attempts to isolate ChRCoV HKU24 from the three positive samples were unsuccessful. Future studies to isolate the virus from more rodent samples will allow characterization of its receptor usage and pathogenicity.

ACKNOWLEDGEMENTS

We thank Dr. Wing-Man Ko, Secretary for Food and Health Bureau; Ms. Vivian Lau, Mr. Kwok-Hau Sin and Mr. M. C. Yuen of the FEHD and Mr. Alan C. K. Wong, Dr. Siu-Fai Leung, Thomas Hon-Chung Sit and Howard Kai-Hay Wong, Chung-Tong Shek and Joseph W. K. So of the AFCD for facilitation and assistance on sample collection. Views expressed in this paper are those of the authors only, and may not represent the opinion of the FEHD, AFCD or the Government of the HKSAR. We are grateful to the generous support of Mrs. Carol Yu, Professor Richard Yu, Mr. Hui Hoy and Mr. Hui Ming in the genomic sequencing platform. This work is partly supported by the Research Grant Council Grant, University Grant Council; Committee for Research and Conference Grant, Strategic Research Theme Fund, and University Development Fund, The University of Hong Kong; Health and Medical Research Fund of the Food and Health Bureau of HKSAR; and Consultancy Service for Enhancing Laboratory Surveillance of Emerging Infectious Disease for the HKSAR Department of Health.

538 **REFERENCES**

539 1. **Brian DA, Baric RS.** 2005. Coronavirus genome structure and replication. Curr. Top.
540 Microbiol. Immunol. **287:**1-30.

541 2. **Lai MM, Cavanagh D.** 1997. The molecular biology of coronaviruses. Adv. Virus. Res.
542 **48:**1-100.

543 3. **de Groot RJ, Baker SC, Baric R, Enjuanes L, Gorbalenya A, Holmes KV, Perlman**
544 **S, Poon L, Rottier PJ, Talbot PJ, Woo PC, Ziebuhr J.** 2011. *Coronaviridae*. In: Virus
545 Taxonomy, Classification and Nomenclature of Viruses, Ninth Report of the International
546 Committee on Taxonomy of Viruses, International Union of Microbiological Societies,
547 Virology Division, King AMQ, Adams MJ, Carstens EB, Lefkowitz EJ, eds. Elsevier
548 Academic Press, pp. 806-828.

549 4. **Woo PC, Lau SK, Lam CS, Lau CC, Tsang AK, Lau JH, Bai R, Teng JL, Tsang CC,**
550 **Wang M, Zheng BJ, Chan KH, Yuen KY.** 2012. Discovery of seven novel mammalian
551 and avian coronaviruses in *Deltacoronavirus* supports bat coronaviruses as the gene
552 source of *Alphacoronavirus* and *Betacoronavirus* and avian coronaviruses as the gene
553 source of *Gammacoronavirus* and *Deltacoronavirus*. J. Virol. **86:**3995-4008

554 5. **Woo PC, Lau SK, Lam CS, Lai KK, Huang Y, Lee P, Luk GS, Dyrting KC, Chan**
555 **KH, Yuen KY.** 2009. Comparative analysis of complete genome sequences of three
556 avian coronaviruses reveals a novel group 3c coronavirus. J. Virol. **83:**908-917.

557 6. **Woo PC, Lau SK, Huang Y, Yuen KY.** 2009. Coronavirus diversity, phylogeny and
558 interspecies jumping. Exp. Biol. Med. (Maywood) **234:**1117-1127.

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

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

9. **Woo PC, Lau SK, Yip CC, Huang Y, Tsoi HW, Chan KH, Yuen KY.** 2006. Comparative analysis of 22 coronavirus HKU1 genomes reveals a novel genotype and evidence of natural recombination in coronavirus HKU1. J. Virol. **80:**7136-7145.

10. **Peng G, Sun D, Rajashankar KR, Qian Z, Holmes KV, Li F.** 2011. Crystal structure of mouse coronavirus receptor-binding domain complexed with its murine receptor. Proc. Natl. Acad. Sci. U. S. A. **108:**10696-10701.

11. **Yang Y, Du L, Liu C, Wang L, Ma C, Tang J, Baric RS, Jiang S, Li F.** 2014. Receptor usage and cell entry of bat coronavirus HKU4 provide insight into bat-to-human transmission of MERS coronavirus. Proc. Natl. Acad. Sci. U. S. A. pii:201405889. [Epub ahead of print]

12. **Yeh SH, Wang HY, Tsai CY, Kao CL, Yang JY, Liu HW, Su IJ, Tsai SF, Chen DS, Chen PJ; National Taiwan University SARS Research Team.** 2004. Characterization of severe acute respiratory syndrome coronavirus genomes in Taiwan: molecular epidemiology and genome evolution. Proc. Natl. Acad. Sci. U. S. A. **101:**2542-2547.

13. **Guan Y, Zheng BJ, 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.

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

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

16. **Peiris JS, Lai ST, Poon LL, Guan Y, Yam LY, Lim W, Nicholls J, Yee WK, Yan WW, Cheung MT, Cheng VC, Chan KH, Tsang DN, Yung RW, Ng TK, Yuen KY.** 2003. Coronavirus as a possible cause of severe acute respiratory syndrome. Lancet **361:**1319-1325.

17. **Rota PA, Oberste MS, Monroe SS, Nix WA, Campagnoli R, Icenogle JP, Penaranda S, Bankamp B, Maher K, Chen MH, Tong S, Tamin A, Lowe L, Frace M, DeRisi JL, Chen Q, Wang D, Erdman DD, Peret TC, Burns C, Ksiazek TG, Rollin PE, Sanchez A, Liffick S, Holloway B, Limor J, McCaustland K, Olsen-Rasmussen M, Fouchier R, Gunther S, Osterhaus AD, Drosten C, Pallansch MA, Anderson LJ, Bellini WJ.** 2003. Characterization of a novel coronavirus associated with severe acute respiratory syndrome. Science **300:**1394-1399.

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

19. **van der Hoek L, Pyrc K, Jebbink MF, Vermeulen-Oost W, Berkhout RJ, Wolthers KC, Wertheim-van Dillen PM, Kaandorp J, Spaargaren J, Berkhout B.** 2004. Identification of a new human coronavirus. Nat. Med. **10:**368-373.

20. **Woo PC, Lau SK, Chu CM, Chan KH, Tsoi HW, Huang Y, Wong BH, Poon RW, Cai JJ, Luk WK, Poon LL, Wong SS, Guan Y, Peiris JS, Yuen KY.** 2005. Characterization and complete genome sequence of a novel coronavirus, coronavirus HKU1, from patients with pneumonia. J. Virol. **79:**884-895.

21. **Jonassen CM, Kofstad T, Larsen IL, Lovland A, Handeland K, Follestad A, Lillehaug A.** 2005. Molecular identification and characterization of novel coronaviruses infecting graylag geese (*Anser anser*), feral pigeons (*Columbia livia*) and mallards (*Anas platyrhynchos*). J. Gen. Virol. **86:**1597-1607.

22. **Mihindukulasuriya KA, Wu G, St Leger J, Nordhausen RW, Wang D.** 2008. Identification of a novel coronavirus from a beluga whale by using a panviral microarray. J. Virol. **82:**5084-5088.

23. **Dong BQ, Liu W, Fan XH, Vijaykrishna D, Tang XC, Gao F, Li LF, Li GJ, Zhang JX, Yang LQ, Poon LL, Zhang SY, Peiris JS, Smith GJ, Chen H, Guan Y.** 2007. Detection of a novel and highly divergent coronavirus from Asian leopard cats and Chinese ferret badgers in Southern China. J. Virol. **81:**6920-6926.

24. **Lau SK, Woo PC, Yip CC, Tse H, Tsoi HW, Cheng VC, Lee P, Tang BS, Cheung CH, Lee RA, So LY, Lau YL, Chan KH, Yuen KY.** 2006. Coronavirus HKU1 and other coronavirus infections in Hong Kong. J. Clin. Microbiol. **44:**2063-2071.

25. **Woo PC, Lau SK, Tsoi HW, Huang Y, Poon RW, Chu CM, Lee RA, Luk WK, Wong GK, Wong BH, Cheng VC, Tang BS, Wu AK, Yung RW, Chen H, Guan Y, Chan KH, Yuen KY.** 2005. Clinical and molecular epidemiological features of coronavirus HKU1-associated community-acquired pneumonia. J. Infect. Dis. **192:**1898-1907.

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

27. **Vijgen L, Keyaerts E, Moës 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:**1595-1604.

28. **Dominguez SR, O'Shea TJ, Oko LM, Holmes KV.** 2007. Detection of group 1 coronaviruses in bats in North America. Emerg. Infect. Dis. **13:**1295-1300.

29. **Gloza-Rausch F, Ipsen A, Seebens A, Göttsche M, Panning M, Felix Drexler J, Petersen N, Annan A, Grywna K, Müller M, Pfefferle S, Drosten C.** 2008. Detection and prevalence patterns of group I coronaviruses in bats, northern Germany. Emerg. Infect. Dis. **14:**626-631.

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

31. **Lau SK, Poon RWS, Wong BHL, Wang M, Huang Y, Xu H, Guo R, Li KSM, Gao K, Chan KH, Zheng BJ, Woo PCY, Yuen KY.** 2010. Coexistence of different genotypes in the same bat and serological characterization of *Rousettus* bat coronavirus HKU9 belonging to a novel *Betacoronavirus* subgroup. J. Virol. **84:**11385-11394.

32. **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:**2001-2009.

33. **Tang XC, Zhang JX, Zhang SY, Wang P, Fan XH, Li LF, Li G, Dong BQ, Liu W, Cheung CL, Xu KM, Song WJ, Vijaykrishna D, Poon LL, Peiris JS, Smith GJ, Chen H, Guan Y.** 2006. Prevalence and genetic diversity of coronaviruses in bats from China. J. Virol. **80:**7481-7490.

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

35. **Woo PC, Wang M, Lau SK, Xu H, Poon RW, Guo R, Wong BH, Gao K, Tsoi HW, Huang Y, Li KS, Lam CS, Chan KH, Zheng BJ, Yuen KY.** 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.

668 36. **Woo PC, Lau SK, Li KS, Poon RW, Wong BH, Tsoi HW, Yip BC, Huang Y, Chan**
669 **KH, Yuen KY.** 2006. Molecular diversity of coronaviruses in bats. Virology **351:**180-
670 187.
671 37. **Lau SK, Li KS, Tsang AK, Lam CS, Ahmed S, Chen H, Chan KH, Woo PC, Yuen**
672 **KY.** 2013. Genetic characterization of *Betacoronavirus* lineage C viruses in bats reveals
673 marked sequence divergence in the spike protein of *Pipistrellus* bat coronavirus HKU5 in
674 *Japanese pipistrelle*: implications for the origin of the novel Middle East respiratory
675 syndrome coronavirus. J. Virol. **87:**8638-8650.
676 38. **Drexler JF, Corman VM, Drosten C.** 2014. Ecology, evolution and classification of bat
677 coronaviruses in the aftermath of SARS. Antiviral Res. **101:**45-56.
678 39. **Vijgen L, Keyaerts E, Lemey P, Maes P, Van Reeth K, Nauwynck H, Pensaert M,**
679 **Van Ranst M.** 2006. Evolutionary history of the closely related group 2 coronaviruses:
680 porcine hemagglutinating encephalomyelitis virus, bovine coronavirus, and human
681 coronavirus OC43. J. Virol. **80:**7270-7274.
682 40. **Alekseev KP, Vlasova AN, Jung K, Hasoksuz M, Zhang X, Halpin R, Wang S,**
683 **Ghedin E, Spiro D, Saif LJ.** 2008. Bovine-like coronaviruses isolated from four species
684 of captive wild ruminants are homologous to bovine coronaviruses, based on complete
685 genomic sequences. J. Virol. **82:**12422-12431.
686 41. **Hasoksuz M, Alekseev K, Vlasova A, Zhang X, Spiro D, Halpin R, Wang S, Ghedin**
687 **E, Saif LJ.** 2007. Biologic, antigenic, and full-length genomic characterization of a
688 bovine-like coronavirus isolated from a giraffe. J. Virol. **81:**4981-4990.

42. **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:**560-572.

43. **Guy JS, Breslin JJ, Breuhaus B, Vivrette S, Smith LG.** 2000. Characterization of a coronavirus isolated from a diarrheic foal. J. Clin. Microbiol. **38:**4523-4526

44. **Erles K, Toomey C, Brooks HW, Brownlie J.** 2003. Detection of a group 2 coronavirus in dogs with canine infectious respiratory disease. Virology **310:**216-223.

45. **Tsunemitsu H, el-Kanawati ZR, Smith DR, Reed HH, Saif LJ.** 1995. Isolation of coronaviruses antigenically indistinguishable from bovine coronavirus from wild ruminants with diarrhea. J. Clin. Microbiol. **33:**3264-3269

46. **Mengeling WL, Boothe AD, Ritchie AE.** 1972. Characteristics of a coronavirus (strain 67N) of pigs. Am. J. Vet. Res. **33:**297-308.

47. **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.

48. **Li IW, Chan KH, To KW, Wong SS, Ho PL, Lau SK, Woo PC, Tsoi HW, Chan JF, Cheng VC, Zheng BJ, Chen H, Yuen KY.** 2009. Differential susceptibility of different cell lines to swine-origin influenza A H1N1, seasonal human influenza A H1N1, and avian influenza A H5N1 viruses. J. Clin. Virol. **46:**325-330.

49. **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:**5481-5496.

50. **Lau SK, Chan KH, Yip CCY, Ng TK, Tsang OTY, Woo PCY, Yuen KY.** 2009. Confirmation of the first Hong Kong case of human infection by novel swine-origin influenza A (H1N1) virus diagnosed using ultrarapid, real-time reverse transcriptase PCR. J. Clin. Microbiol. **47:**2344-2346.

51. **Huang Y, Lau SK, Woo PC, Yuen KY.** 2008. CoVDB: a comprehensive database for comparative analysis of coronavirus genes and genomes. Nucleic Acids Res. **36:**D504-511.

52. **Apweiler R, Attwood TK, Bairoch A, Bateman A, Birney E, Biswas M, Bucher P, Cerutti L, Corpet F, Croning MD, Durbin R, Falquet L, Fleischmann W, Gouzy J, Hermjakob H, Hulo N, Jonassen I, Kahn D, Kanapin A, Karavidopoulou Y, Lopez R, Marx B, Mulder NJ, Oinn TM, Pagni M, Servant F, Sigrist CJ, Zdobnov EM.** 2001. The InterPro database, an integrated documentation resource for protein families, domains and functional sites. Nucleic Acids Res. **29:**37-40.

53. **Bateman A, Birney E, Cerruti L, Durbin R, Etwiller L, Eddy SR, Griffiths-Jones S, Howe KL, Marshall M, Sonnhammer EL.** 2002. The Pfam protein families database. Nucleic Acids Res. **30:**276-280.

54. **Sonnhammer EL, von Heijne G, Krogh A.** 1998. A hidden Markov model for predicting transmembrane helices in protein sequences. Proc. Int. Conf. Intell. Syst. Mol. Biol. **6:**175-182.

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

56. **Suchard MA, Weiss RE, Sinsheimer JS.** 2001. Bayesian selection of continuous-time Markov chain evolutionary models. Mol. Biol. Evol. **18:**1001-1013.

734 57. **Woo PC, Lau SK, Wong BHL, Chan KH, Hui WT, Kwan GSW, Peiris JSM, Couch**
735 **RB, Yuen KY.** 2004. False positive results in a recombinant Severe Acute Respiratory
736 Syndrome-Associated Coronavirus (SARS-CoV) nucleocapsid enzyme-linked
737 immunosorbent assay due to HCoV-OC43 and HCoV-229E rectified by western blotting
738 with recombinant SARS-CoV spike polyprotein. J. Clin. Microbiol. **42:**5885-5888.
739 58. **Woo PC, Lau SK, Tsoi HW, Chan KH, Wong BH, Che XY, Tam VK, Tam SC,**
740 **Cheng VC, Hung IF, Wong SS, Zheng BJ, Guan Y, Yuen KY.** 2004. Relative rates of
741 non-pneumonic SARS coronavirus infection and SARS coronavirus pneumonia. Lancet
742 **363:**841-845.
743 59. **Jeong YS, Repass JF, Kim YN, Hwang SM, Makino S.** 1996. Coronavirus
744 transcription mediated by sequences flanking the transcription consensus sequence.
745 Virology **217:**311-322.
746 60. **Wu HY, Ozdarendeli A, Brian DA.** 2006. Bovine coronavirus 5'-proximal genomic
747 acceptor hotspot for discontinuous transcription is 65 nucleotides wide. J. Virol. **80:**2183-
748 2193.
749 61. **Zhang J, Guy JS, Snijder EJ, Denniston DA, Timoney PJ, Balasuriya UB.** 2007.
750 Genomic characterization of equine coronavirus. Virology **369:**92-104.
751 62. **Schwarz B, Routledge E, Siddell SG.** 1990. Murine coronavirus nonstructural protein
752 ns2 is not essential for virus replication in transformed cells. J. Virol. **64:**4784-4791.
753 63. **de Haan CA, Masters PS, Shen X, Weiss S, Rottier PJ.** 2002. The group-specific
754 murine coronavirus genes are not essential, but their deletion, by reverse genetics, is
755 attenuating in the natural host. Virology **296:**177-189.

64. **Snijder EJ, Bredenbeek PJ, Dobbe JC, Thiel V, Ziebuhr J, Poon LL, Guan Y, Rozanov M, Spaan WJ, Gorbalenya AE.** 2003. Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage. J. Mol. Biol. **331:**991-1004.

65. **Schultze B, Herrler G.** 1992. Bovine coronavirus uses N-acetyl-9-O-acetylneuraminic acid as a receptor determinant to initiate the infection of cultured cells. J. Gen. Virol. **73:**901-906.

66. **Vlasak R, Luytjes W, Spaan W, Palese P.** 1988. Human and bovine coronaviruses recognize sialic acid-containing receptors similar to those of influenza C viruses. Proc. Natl. Acad. Sci. U. S. A. **85:**4526-4529.

67. **Williams RK, Jiang GS, Holmes KV.** 1991. Receptor for mouse hepatitis virus is a member of the carcinoembryonic antigen family of glycoproteins. Proc. Natl. Acad. Sci. U. S. A. **88:**5533-5536.

68. **Langereis MA, van Vliet AL, Boot W, de Groot RJ**. 2010. Attachment of mouse hepatitis virus to O-acetylated sialic acid is mediated by hemagglutinin-esterase and not by the spike protein. J. Virol. **84:**8970-8974.

69. **Thiel V, Siddell SG.** 1994. Internal ribosome entry in the coding region of murine hepatitis virus mRNA 5. J. Gen. Virol. **75:**3041-3046.

70. **Goebel SJ, Hsue B, Dombrowski TF, Masters PS.** 2004. Characterization of the RNA components of a putative molecular switch in the 3' untranslated region of the murine coronavirus genome. J. Virol. **78:**669-682.

71. **Netland J, Ferraro D, Pewe L, Olivares H, Gallagher T, Perlman S.** 2007. Enhancement of murine coronavirus replication by severe acute respiratory syndrome

coronavirus protein 6 requires the N-terminal hydrophobic region but not C-terminal sorting motifs. J. Virol. **81:**11520-11525.

72. **Brockway SM, Lu XT, Peters TR, Dermody TS, Denison MR.** 2004. Intracellular localization and protein interactions of the gene 1 protein p28 during mouse hepatitis virus replication. J. Virol. **78:**11551-11562.

73. **Zeng Q, Langereis MA, van Vliet AL, Huizinga EG, de Groot RJ.** 2008. Structure of coronavirus hemagglutinin-esterase offers insight into corona and influenza virus evolution. Proc. Natl. Acad. Sci. U. S. A. **105:**9065-9069.

74. **Yokomori K, Banner LR, Lai MM.** 1991. Heterogeneity of gene expression of the hemagglutinin-esterase (HE) protein of murine coronaviruses. Virology **183:**647-657.

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

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

77. **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:**535-538.

78. **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:**140-145.

79. **Reusken CB, Haagmans BL, Müller 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, Gortázar-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:**859-866.

80. **Azhar EI, El-Kafrawy SA, Farraj SA, Hassan AM, Al-Saeed MS, Hashem AM, Madani TA**. 2014. Evidence for camel-to-human transmission of MERS coronavirus. N. Engl. J. Med. **370:**2499-2505.

81. **Corman VM, Ithete NL, Richards LR, Schoeman MC, Preiser W, Drosten C, Drexler JF.** 2014. Rooting the phylogenetic tree of MERS-Coronavirus by characterization of a conspecific virus from an African Bat. J. Virol. pii:JVI.01498-14. [Epub ahead of print]

82. **Bailey OT, Pappenheimer AM, Cheerer FS, Daniels JB.** 1949. A murine virus (JHM) causing disseminated encephalomyelitis with extensive destruction of myelin: II. Pathology. J. Exp. Med. **90:**195-212.

822 83. **Cheever FS, Daniels JB, Pappenheimer AM, Bailey OT.** 1949. A murine virus (JHM)
823 causing disseminated encephalomyelitis with extensive destruction of myelin. J. Exp.
824 Med. **90:**181-194.
825 84. **Sun N, Perlman S.** 1995. Spread of a neurotropic coronavirus to spinal cord white matter
826 via neurons and astrocytes. J. Virol. **69:**633-641.
827 85. **Miura TA, Wang J.** 2007. Rat coronaviruses infect rat alveolar type I epithelial cells and
828 induce expression of CXC chemokines. Virology **369:**288-298.
829 86. **Bhatt PN, Percy DH, Jonas AM.** 1972. Characterization of the virus of
830 sialodacryoadenitis of rats: a member of the coronavirus group. J. Infect. Dis. **126:**123-
831 130.
832 87. **Peng G, Xu L, Lin YL, Chen L, Pasquarella JR, Holmes KV, Li F.** 2012. Crystal
833 structure of bovine coronavirus spike protein lectin domain. J. Biol. Chem. **287:**41931-
834 41938.
835

836 **LEGENDS TO FIGURES**

837 **FIG 1** Comparison of genome organizations of ChRCoV HKU24, MHV, HCoV OC43 and
838 HCoV HKU1. Papain-like proteases ($PL1^{pro}$ and $PL2^{pro}$) are represented by orange boxes. The
839 residues at the cleavage site are indicated above or below the boundary of each nonstructural
840 protein. Unique cleavage site in ChRCoV HKU24 is in bold.

841 **FIG 2** Predicted model of ChRCoV HKU24 spike protein and NTD using Swiss-Model tool. (A)
842 Predicted domain structure of ChRCoV HKU24 spike protein. NTD, N-terminal domain; RBD,
843 receptor-binding domain; HR, heptad-repeat; TM, transmembrane anchor. The signal peptide
844 corresponds to residues 1–15 and is cleaved during molecular maturation. (B) Sequence
845 alignment of ChRCoV HKU24 NTD with BCoV, HCoV-OC43 and MHV NTD, performed
846 using PROMALS3D. The three strains of ChRCoV HKU24 characterized in this study are
847 bolded. Beta strands are shown as yellow arrows, and the alpha helix is shown as a coiled ribbon.
848 Loop 10-11 is boxed. The 14 contact residues at the MHV NTD/mCEACAM1a interface are
849 highlighted in blue, the four BCoV critical sugar-binding residues are highlighted in brown, and
850 BCoV non-critical sugar-binding residues are highlighted in yellow. Location of residue
851 substitution that might decrease the sugar-binding affinity of BCoV NTD is marked by inverted
852 triangle. Asterisks indicate positions that have fully conserved residues. Colons indicate
853 positions that have strongly conserved residues. Periods indicate positions that have weakly
854 conserved residues. (C) Predicted structure of the ChRCoV HKU24 NTD constructed through
855 homology modelling from BCoV NTD (4h14) and close-up of the pocket above the β-sandwich
856 core. The Global Model Quality Estimation score of 0.83 and QMEAN4 Z-score of -1.82
857 indicated reliable overall model quality.

858 **FIG 3** Phylogenetic analyses of RdRp, S, N and HE proteins of ChRCoV HKU24. The trees
859 were constructed by the maximum likelihood method using WAG+I+G substitution model and
860 bootstrap values calculated from 100 trees. Bootstrap values below 70% are not shown. Nine
861 hundred and twenty-eight, 1358, 443 and 425 aa positions in RdRp, S, N and HE, respectively,
862 were included in the analyses. The scale bar represents 0.3 substitutions per site. The three
863 strains of ChRCoV HKU24 characterized in this study are bolded.
864 **FIG 4** Estimation of tMRCA of ChRCoV HKU24 strains, BCoV/HCoV-OC43, and ChRCoV
865 HKU24/members of *Betacoronavirus 1*/RbCoV HKU14 based on the complete RdRp and HE
866 genes. The mean estimated dates (above the branch) and Bayesian posterior probabilities (below
867 the branch) are labeled and are represented by gray squares. The taxa are labeled with their
868 sampling dates.
869 **FIG 5** Western blot analysis for antibodies against purified $(His)_6$–tagged recombinant ChRCoV
870 HKU24 N protein (~50kDa) (A) and spike polypeptide (~50kDa) (B) in rodent serum samples
871 and serum samples from other animals or humans infected by different βCoVs including HCoV
872 OC43 (*Betacoronavirus* lineage A), RbCoV HKU14 (*Betacoronavirus* lineage A) and SARS-
873 CoV (*Betacoronavirus* lineage B). Lanes: 1, negative control; 2, oriental house rat serum sample
874 negative for antibody against ChRCoV HKU24 N protein and spike polypeptide; 3, Norway rat
875 serum sample negative for antibody against ChRCoV HKU24 N protein and spike polypeptide; 4,
876 oriental house rat serum sample positive for antibody against ChRCoV HKU24 N protein and
877 spike polypeptide; 5, Norway rat serum sample positive for antibody against ChRCoV HKU24 N
878 protein and spike polypeptide; 6 and 7, serum samples from rabbits infected by RbCoV HKU14;
879 8 and 9, serum samples from patients with HCoV-OC43 infection; 10 and 11, serum samples
880 from patients with SARS-CoV infection; 12, positive control (anti-His antibody).

881 **FIG 6** Evolution of CoVs from their ancestors in bat, bird and rodent hosts to virus species that
882 infect other animals. The dashed arrows indicate possible routes of transmission from bats or
883 birds to rodents before establishment of *Betacoronavirus* lineage A.

884 **Table 1.** Detection of ChRCoV HKU24 in rodents by RT-PCR and serological studies by Western blot analysis

Scientific name	Common name	No. of rodents tested	No. (%) of rodents positive for ChRCoV HKU24 in alimentary samples by RT-PCR	No. (%) of rodents positive for ChRCoV HKU24 antibody by N-Western blot analysis	No. (%) of rodents positive for ChRCoV HKU24 antibody by S1-Western blot analysis
Crocidura attenuata	Asian gray shrew	5	0/5 (0%)	NA	NA
Niviventer fulvescens	Chestnut white-bellied rat	97	0/97 (0%)	NA	NA
Rattus andamanensis	Indochinese forest rat	170	0/170 (0%)	NA	NA
Rattus norvegicus[a]	Norway rat	82	3/82 (3.6%)	60/74 (81.1%)	21/60 (35%)
Rattus norvegicus[b]	Norway rat	308	0/277 (0%)	15/31 (48.4%)	0/15 (0%)
Rattus rattus	Black rat	54	0/24 (0%)	4/30 (0.13%)	0/4 (0%)
Rattus tanezumi	Oriental house rat	9	0/9 (0%)	7/9 (77.8%)	2/7 (2.9%)

885 [a]Norway rats from Guangzhou
886 [b]Norway rats from Hong Kong

887 **Table 2.** Comparison of genomic features of ChRCoV HKU24 and other CoVs with complete
888 genome sequences available and aa identities between the predicted chymotrypsin-like protease
889 (3CL^{pro}), RNA dependent RNA polymerase (RdRp), helicase (Hel), haemagglutinin-esterase
890 (HE), spike (S), envelope (E), membrane (M) and nucleocapsid (N) proteins of ChRCoV HKU24
891 and the corresponding proteins of other CoVs

Coronaviruses[a]	Genome features		Pairwise amino acid identity (%)							
	Size	G + C	**ChRCoV HKU24-R05005I**							
	(bases)	content	3CL^{pro}	RdRp	Hel	HE	S	E	M	N
Alphacoronavirus										
TGEV	28586	0.38	45.5	58.3	59.1		26.1	22.4	36.4	27.1
MCoV	28894	0.38	46.5	59.3	57.2		25.8	24.7	32.0	27.6
CCoV	29363	0.38	44.6	58.3	58.7		26.3	23.5	36.4	27.6
FIPV	29355	0.38	45.2	58.4	58.7		25.6	22.4	33.8	26.1
PRCV	27550	0.37	45.5	58.3	58.9		26.7	23.5	35.7	27.8
HCoV-229E	27317	0.38	44.4	56.3	57.7		26.9	26.5	32.5	26.9
HCoV-NL63	27553	0.34	42.8	56.5	57.6		25.8	31.0	32.8	25.4
PEDV	28033	0.42	42.4	59.1	58.7		25.6	30.1	38.5	21.6
Rh-BatCoV HKU2	27165	0.39	43.6	57.6	55.8		24.6	30.6	35.9	26.4
Mi-BatCoV 1A	28326	0.38	42.4	58.0	58.4		25.1	31.3	32.9	26.9
Mi-BatCoV HKU8	28773	0.42	43.1	58.8	56.1		25.5	29.3	35.1	26.5
Sc-BatCoV 512	28203	0.40	41.1	58.3	58.1		25.2	26.8	38.0	24.9
Ro-BatCoV HKU10	28494	0.39	43.1	56.9	57.1		26.6	34.5	36.2	26.4
Hi-BatCoV HKU10	28492	0.38	43.1	56.7	57.0		25.8	34.5	35.7	25.6
Betacoronavirus lineage A										
Betacoronavirus 1										
HCoV-OC43	30738	0.37	85.8	91.8	93.5	70.1	67.1	78.6	88.7	74.0
BCoV	31028	0.37	86.8	92.6	93.7	69.6	68.0	78.6	89.2	74.9
PHEV	30480	0.37	86.5	92.0	93.7	68.9	67.0	77.4	89.2	74.0
ECoV	30992	0.37	86.8	92.6	94.7	66.2	69.5	76.2	85.7	73.1
SACoV	30995	0.37	86.8	92.6	93.7	69.6	68.2	80.5	89.6	74.9
CRCoV	31028	0.37	86.5	92.3	93.5	69.9	67.6	77.4	90.0	74.7
GiCoV	30979	0.37	86.8	92.6	93.7	69.6	68.4	78.6	89.6	74.9
DcCoV UAE-HKU23	31036	0.37	86.8	92.6	93.4	69.6	68.1	77.4	90.5	74.4
Murine coronavirus										
MHV	31357	0.42	82.8	90.3	90.5	39.9	63.8	63.9	82.7	67.9
RCoV	31250	0.41	82.5	90.3	90.5	59.3	63.3	62.5	80.5	67.5
HCoV-HKU1	29926	0.32	82.2	88.1	88.9	50.1	60.4	53.0	78.4	62.8
RbCoV HKU14	31084	0.38	86.8	92.5	94.7	69.9	67.9	74.2	91.3	73.9
ChRCoV HKU24-R05009I	31234	0.40	100	100	99.8	99.8	100	100	100	100
ChRCoV HKU24-R05010I	31324	0.40	100	100	100	99.8	100	100	100	100
Betacoronavirus lineage B										
SARS-CoV	29751	0.41	49.0	66.8	68.6		29.9	26.5	37.7	34.3
SARSr-Rh-BatCoV HKU3	29728	0.41	48.4	66.7	68.8		29.5	26.5	38.1	34.1
Betacoronavirus lineage C										
Ty-BatCoV HKU4	30286	0.38	52.3	68.6	68.6		33.0	25.6	42.4	36.7
Pi-BatCoV HKU5	30488	0.43	52.0	68.6	67.1		31.4	25.6	42.9	35.9
MERS-CoV	30107	0.41	53.3	68.7	67.1		31.9	29.3	43.3	37.7
Betacoronavirus lineage D										
Ro-BatCoV HKU9	29114	0.41	46.9	67.1	68.4		28.6	25.6	42.4	33.3
Gammacoronavirus										
IBV	27608	0.38	43.9	62.0	59.8		27.2	21.6	31.5	27.6
BWCoV SW1	31686	0.39	44.3	60.2	57.7		25.4	24.7	26.7	29.2
BdCoV HKU22	31759	0.39	44.3	60.6	57.9		25.2	23.1	25.1	29.2

Deltacoroanvirus									
BuCoV HKU11	26476	0.39	37.5	51.1	48.9	26.3	25.6	28.9	24.5
ThCoV HKU12	26396	0.38	38.0	51.8	48.4	26.2	23.6	30.6	22.1
MunCoV HKU13	26552	0.43	38.5	53.1	50.3	26.0	21.3	28.8	21.7
PorCoV HKU15	25421	0.43	40.4	52.2	49.0	25.6	25.3	26.9	24.2
WECoV HKU16	26027	0.40	39.1	51.9	49.3	25.6	23.3	28.2	22.2
SpCoV HKU17	26067	0.45	40.8	52.0	49.0	25.5	21.6	27.3	25.7
MRCoV HKU18	26674	0.47	38.8	51.9	49.3	26.1	22.5	28.9	23.7
NHCoV HKU19	26064	0.38	35.2	53.7	48.0	24.2	23.9	30.8	23.1
WiCoV HKU20	26211	0.39	36.9	51.6	48.8	26.8	28.6	27.8	23.2
CMCoV HKU21	26216	0.35	37.6	51.6	50.2	25.1	24.7	26.1	22.2

[a]TGEV, porcine transmissible gastroenteritis virus; MCoV, mink coronavirus; CCoV, canine coronavirus; FIPV, feline infectious peritonitis virus; PRCV, porcine respiratory coronavirus; HCoV-229E, human coronavirus 229E; HCoV-NL63, human coronavirus NL63; PEDV, porcine epidemic diarrhea virus; Rh-BatCoV HKU2, *Rhinolophus* bat coronavirus HKU2; Mi-BatCoV 1A, *Miniopterus* bat coronavirus 1A; Mi-BatCoV HKU8, *Miniopterus* bat coronavirus HKU8; Sc-BatCoV 512, *Scotophilus* bat coronavirus 512; Ro-BatCoV HKU10, *Rousettus* bat coronavirus HKU10; Hi-BatCoV HKU10, *Hipposideros* bat coronavirus HKU10; HCoV-OC43, human coronavirus OC43; BCoV, bovine coronavirus; PHEV, porcine hemagglutinating encephalomyelitis virus; ECoV, equine coronavirus; SACoV, sable antelope CoV; CRCoV, canine respiratory coronavirus; GiCoV, giraffe coronavirus; DcCoV UAE-HKU23, dromedary camel coronavirus UAE-HKU23; MHV, murine hepatitis virus; RCoV, rat coronavirus; HCoV-HKU1, human coronavirus HKU1; SARS-CoV, SARS coronavirus; SARSr-Rh-BatCoV HKU3; SARS-related *Rhinolophus* bat coronavirus HKU3; Ty-BatCoV HKU4, *Tylonycteris* bat coronavirus HKU4; Pi-BatCoV HKU5, *Pipistrellus* bat coronavirus HKU5; MERS-CoV, middle east respiratory syndrome coronavirus; Ro-BatCoV HKU9, *Rousettus* bat coronavirus HKU9; IBV, infectious bronchitis virus; BWCoV SW1, beluga whale coronavirus SW1; BdCoV HKU22, bottlenose dolphin coronavirus HKU22; BuCoV HKU11, Bulbul coronavirus HKU11; ThCoV HKU12, Thrush coronavirus HKU12; MunCoV HKU13, Munia coronavirus HKU13; PorCoV HKU15, porcine coronavirus HKU15; WECoV HKU16, white-eye coronavirus HKU16; SpCoV HKU17, sparrow coronavirus HKU17; MRCoV HKU18, magpie robin coronavirus HKU18; NHCoV HKU19, night heron coronavirus HKU19; WiCoV HKU20, wigeon coronavirus HKU20; CMCoV HKU21, common moorhen coronavirus HKU21.

909 **Table 3.** Coding potential and predicted domains in different proteins of ChRCoV HKU24

ORFs	Nucleotide positions (start-end)	No. of nucleotides	No. of amino acids	Frame	Putative function or domain[a]	Positions (aa)	Putative TRS	
							Nucleotide position in genome	TRS sequence (distance in bases to AUG)[b]
1ab	213-21637	21425	7141	+3,+2			63	**CUAAAC**(144)AUG
nsp1	213-950	738	246	+3	Unknown	1-246		
nsp2	951-2714	1764	588	+3	Unknown	247-834		
nsp3	2715-8603	5889	1963	+3	Acidic domain, Hydrophobic domain, ADRP, Putative PL^{pro} domain $PL1^{pro}$, $PL2^{pro}$	835-2797		
nsp4	8604-10091	1488	496	+3	Hydrophobic domain	2798-3293		
nsp5	10092-11000	909	303	+3	$3CL^{pro}$	3294-3596		
nsp6	11001-11861	861	287	+3	Hydrophobic domain	3597-3883		
nsp7	11862-12128	267	89	+3	Unknown	3884-3972		
nsp8	12129-12719	591	197	+3	Unknown	3973-4169		
nsp9	12720-13049	330	110	+3	Unknown	4170-4279		
nsp10	13050-13460	411	137	+3	Unknown	4280-4416		
nsp11	13461-13505	45	14	+3	Unknown (short peptide at the end of ORF1a)	4417-4430		
nsp12	13461-16243	2783	928	+2	RdRp	4417-5344		
nsp13	16244-18042	1797	599	+2	Hel	5345-5943		
nsp14	18041-19603	1563	521	+2	ExoN, N7-MTase	5944-6464		
nsp15	19604-20728	1125	375	+2	NendoU	6465-6839		
nsp16	20729-21637	909	302	+2	O-MT	6840-7141		
NS2a	21639-22469	831	276	+3			21629	**CUAAAC**(4)AUG
HE	22484-23761	1278	425	+2	Hemagglutinin domain	129-266	22466	**CUGAAC**(12)AUG
					Cleavage site	Between 1 and 18		
					Active site for neuraminate O-acetyl-esterase activity, FGDS	38-41		
S	23777-27853	4077	1358	+2	Type I membrane glycoprotein		23771	**CUAAAC**AUG
					N terminal domain	16-299		
					Cleavage site	Between 763 and 764		
					2 heptad repeats	1045–1079 (HR1), 1253-1285 (HR2)		

						Transmembrane domain Cytoplasmic tail rich in cysteine residues	1303-1325		
	NS4	27946-28356	411	136	+1	Transmembrane domain	7-29		
	NS5	28338-28652	315	104	+3			28286	**GUAAAC**(46)AUG
	E	28645-28893	249	82	+1	2 transmembrane domains	13-37 and 38-82		
	M	28908-29603	696	231	+3	3 transmembrane domains	26-45, 50-72 and 79-101	28899	**CUAAAC**(3)AUG
	N2	29596-30288	693	230	+1				
	N	29613-30944	1332	443	+3			29600	**CUAAAC**(7)AUG

[a]ADRP: adenosine diphosphate-ribose 1''-phosphatase; $PL1^{Pro}$, $PL2^{Pro}$: Papain-like protease 1 and papain-like protease 2; $3CL^{pro}$: 3C-like protease; RdRp: RNA-dependent RNA polymerase; Hel: Helicase; ExoN: 3'-to-5' exonuclease; N7-MTase, (guanine-N7)-methyltransferase; NendoU, nidoviral uridylate-specific endoribonuclease; O-MT: 2'-*O*-ribose methyltransferase.

[b]Boldface indicates putative TRS sequences.

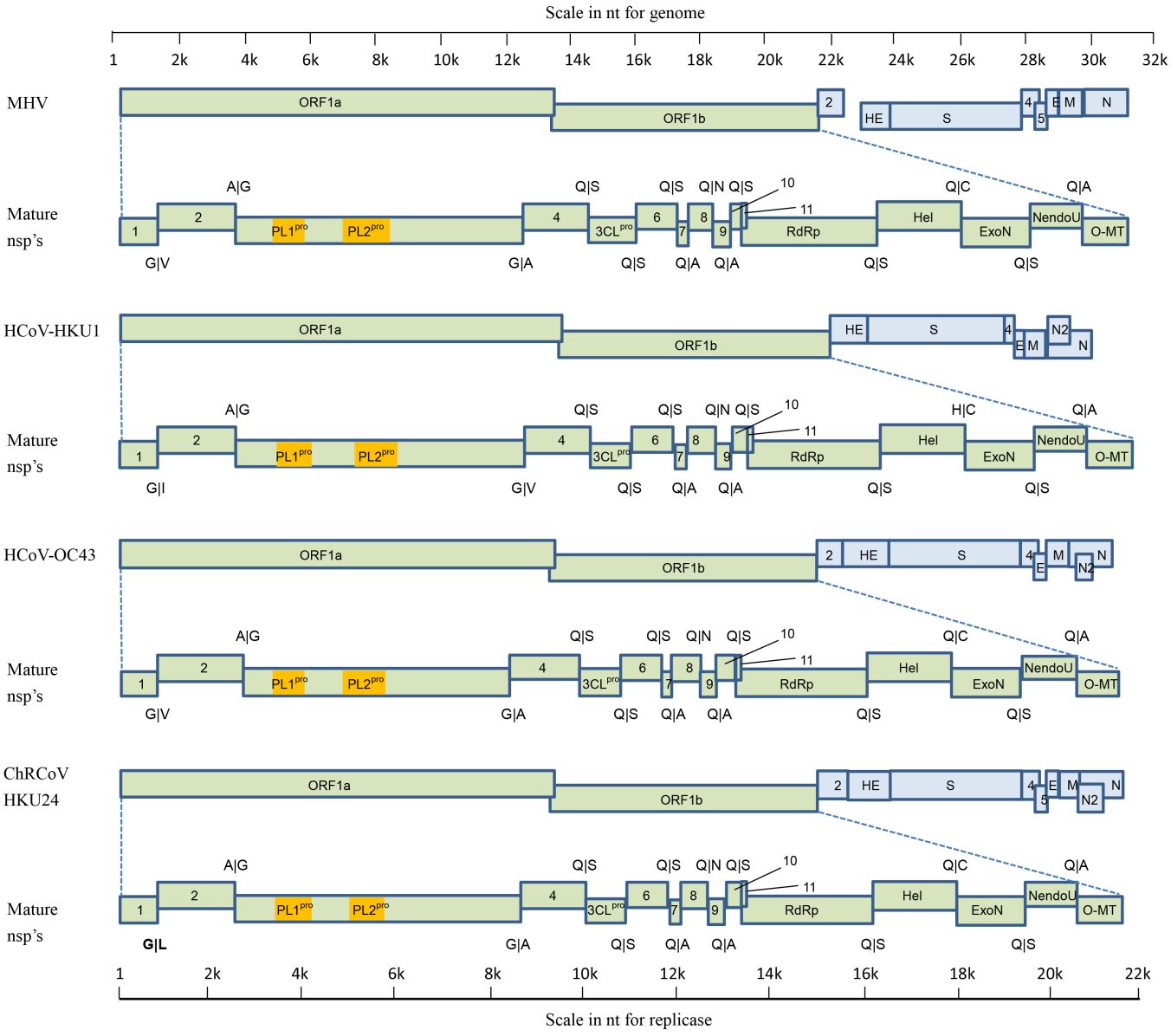
914 **Table 4.** Cleavage site used between nsps in lineage A betacoronaviruses

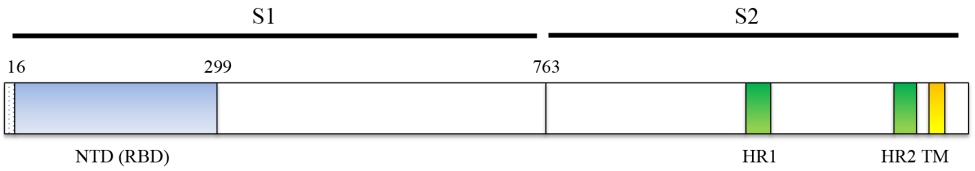
	ChRCoV HKU24[a]	*Betacoronavirus 1*	RbCoV HKU14	MHV	RCoV	HCoV-HKU1
nsp1|nsp2	**G|L**	G|V	G|V	G|V	G|V	G|I
nsp2|nsp3	A|G	A|G	A|G	A|G	A|G	A|G
nsp3|nsp4	G|A	G|A	G|A	G|A	G|A	G|V
nsp4|nsp5	Q|S	Q|S	Q|S	Q|S	Q|S	Q|S
nsp5|nsp6	Q|S	Q|S	Q|S	Q|S	Q|S	Q|S
nsp6|nsp7	Q|S	Q|S	Q|S	Q|S	Q|S	Q|S
nsp7|nsp8	Q|A	Q|A	Q|A	Q|A	H|A	Q|A
nsp8|nsp9	Q|N	Q|N	Q|N	Q|N	Q|N	Q|N
nsp9|nsp10	Q|A	Q|A	Q|A	Q|A	Q|A	Q|A
nsp10|nsp12	Q|S	Q|S	Q|S	Q|S	Q|S	Q|S
nsp12|nsp13	Q|S	Q|S	Q|S	Q|S	Q|S	Q|S
nsp13|nsp14	Q|C	Q|C	Q|C	Q|C	Q|C	H|C
nsp14|nsp15	Q|S	Q|S	Q|S	Q|S	Q|S	Q|S
nsp15|nsp16	Q|A	Q|A	Q|A	Q|A	Q|A	Q|A

915 [a]Unique cleavage site in ChRCoV HKU24 is in bold.
916
917

Table 5. Pairwise comparisons of *Coronaviridae*-wide conserved domains in replicase polyprotein 1ab between ChRCoV HKU24 and other lineage A betacoronaviruses

	Pairwise amino acid identity of ChRCoV HKU24 (%)			
Replicase polyprotein domains	*Betacoronavirus 1*	RbCoV HKU14	Murine coronavirus	HCoV-HKU1
nsp3 (ADRP)	74.8-81.7	74.8	69.5-70.2	71
nsp5 ($3CL^{pro}$)	85.8-86.8	86.8	82.5-82.8	82.2
nsp12 (RdRp)	91.8-92.6	92.5	90.3	88.1
nsp13 (Hel)	93.4-94.8	94.7-94.8	90.5-90.7	88.9-89.1
nsp14 (ExoN)	86.4-88.7	88.7	83.9-84.1	80.2
nsp15 (NendoU)	77.6-79.2	79.5	72.0-73.6	70.1
nsp16 (O-MT)	88.7-89.7	89.1	83.8-85.1	84.1






```
ChRCoV HKU24-R05005I IIGDFKCT--QSWINSAASPQPPISTEVVDVSNGVGTYYVLNRVYLNTSLLLTGYYPVSG
ChRCoV HKU24-R05009I IIGDFKCT--QSWINSAASPQPPISTEVVDVSNGVGTYYVLNRVYLNTSLLLTGYYPVSG
ChRCoV HKU24-R05010I IIGDFKCT--QSWINSAASPQPPISTEVVDVSNGVGTYYVLNRVYLNTSLLLTGYYPVSG
BCoV Mebus            VIGDLKCT--TVSINDVDTGAPSISTDIVDVTNGLGTYYVLDRVYLNTTLLLNGYYPTSG
HCoV-OC43 VR-759      VIGDLKCTSDTSYINDKDTGPPPISTDTVDVTNGLGTYYVLDRVYLNTTLFLNGYYPTSG
MHV A59               YIGDFRCIQ-LVNSNGANVSAPSISTETVEVSQGLGTYYVLDRVYLNATLLLTGYYPVDG
MHV JHM               YIGDFRCIQ-TVNYNGNNASAPSISTEAVDVSKGLGTYYVLDRVYLNATLLLTGYYPVDG
                       ***::*       *.     *.***: *:*::*:******:*****::*:*.****..*

ChRCoV HKU24-R05005I SIYRNLLLKGTQWLSTNWFLPPFLSEFNSGIFVKARNSKPVLNGITHSEFGTIVFGTSFV
ChRCoV HKU24-R05009I SIYRNLLLKGTQWLSTNWFLPPFLSEFNSGIFVKARNSKPVLNGITHSEFGTIVFGTSFV
ChRCoV HKU24-R05010I SIYRNLLLKGTQWLSTNWFLPPFLSEFNSGIFVKARNSKPVLNGITHSEFGTIVFGTSFV
BCoV Mebus            STYRNMALKGTLLLSRLWFKPPFLSDFINGIFAKVKNTKVIKKGVMYSEFPAITIGSTFV
HCoV-OC43 VR-759      STYRNMALKGSVLLSRLWFKPPFLSDFINGIFAKVKNTKVIKDRVMYSEFPAITIGSTFV
MHV A59               SKFRNLALTGTNSVSLSWFQPPYLSQFNDGIFAKVQNLKTSTPSGATAYFPTIVIGSLFG
MHV JHM               SNYRNLALTGTNTLSLTWFKPPFLSEFNDGIFAKVQNLKTNTPTGATSYFPTIVIGSLFG
                      * :**: *.*:  :*  ** **:**:* .***.*.:* *       : * :*.:*: *

ChRCoV HKU24-R05005I NTTYTIVIEPSTQIVN----GKLIGTLSASVCQYTMCEYPNTACNPVL-GNGRPSLWHAS
ChRCoV HKU24-R05009I NTTYTIVIEPSTQIVN----GKLIGTLSASVCQYTMCEYPNTACNPVL-GNGRPSLWHAS
ChRCoV HKU24-R05010I NTTYTIVIEPSTQIVN----GKLIGTLSASVCQYTMCEYPNTACNPVL-GNGRPSLWHAS
BCoV Mebus            NTSYSVVVQPHTINLD----NKLQGLLEISVCQYTMCEYPHTICHPNL-GNKRVELWHWD
HCoV-OC43 VR-759      NTSYSVVVQPRTINSTQDGYNKLQGLLEVSVCQYNMCEYPQTICHPNL-GNHRKELWHLD
MHV A59               YTSYTVVIEPYN----------GVIMASVCQYTICLLPYTDCKPNTNGNKLIGFWHTD
MHV JHM               NTSYTVVLEPYN----------NIIMASVCTYTICQLPYTPCKPNTNGNRVIGFWHTD
                       *:*::*::* .          . :  *** *.:*  * * *:*   **    :** .

ChRCoV HKU24-R05005I IGIVPCLYQRNFTYNVAADNIYFHFYQDGGTFYAYVGDKSPITTLLFQVYLGTVVTHYYV
ChRCoV HKU24-R05009I IGIVPCLYQRNFTYNVAADNIYFHFYQDGGTFYAYVGDKSPITTLLFQVYLGTVVTHYYV
ChRCoV HKU24-R05010I IGIVPCLYQRNFTYNVAADNIYFHFYQDGGTFYAYVGDKSPITTLLFQVYLGTVVTHYYV
BCoV Mebus            TGVVSCLYKRNFTYDVNADYLYFHFYQEGGTFYAYFTDTGVVTKFLFNVYLGTVLSHYYV
HCoV-OC43 VR-759      TGVVSCLYKRNFTYDVNADYLYFHFYQEGGTFYAYFTDTGVVTKFLFNVYLGMALSHYYV
MHV A59               VKPPICVLKRNFTLNVNADAFYFHFYQHGGTFYAYYADKPSATTFLFSVYIGDILTQYYV
MHV JHM               VKPPICLLKRNFTFNVNAPWLYFHFYQQGGTFYAYYADKPSATTFLFSVYIGDILTQYFV
                         *: :**** :* *  :******.*******  *.   *.:**.**:*  :::*:*

ChRCoV HKU24-R05005I LPLVCNA----RQTYEYWVTPLIKREYLLVFDGNGVITNAVDCASDHMSEIQCMT
ChRCoV HKU24-R05009I LPLVCNA----RQTYEYWVTPLIKREYLLVFDGNGVITNAVDCASDHMSEIQCMT
ChRCoV HKU24-R05010I LPLVCNA----RQTYEYWVTPLIKREYLLVFDGNGVITNAVDCASDHMSEIQCMT
BCoV Mebus            LPLTCSS----AMTLEYWVTPLTSKQYLLAFNQDGVIFNAVDCKSDFMSEIKCKT
HCoV-OC43 VR-759      MPLTCNSKVKNGFTLEYWVTPLTSRQYLLAFNQDGIIFNAVDCMSDFMSEIKCKT
MHV A59               LPFICNPTAGSTFAPRYWVTPLVKRQYLFNFNQKGVITSAVDCASSYTSEIKCKT
MHV JHM               LPFICTPTAGSTLLPLYWVTPLLKRQYLFNFNEKGVITSAVDCASSYISEIKCKT
                      :*: *..          ****** .::**: *: .*:* .**** *.. ***:* *
```