Genomic characterizations of bat coronaviruses (1A, 1B and HKU8) and evidence for co-infections in Miniopterus bats

D. K. W. Chu,1 J. S. M. Peiris,1,2 H. Chen,1 Y. Guan1 and L. L. M. Poon1

1State Key Laboratory of Emerging Infectious Diseases, Department of Microbiology, The University of Hong Kong, Hong Kong SAR
2HKU-Pasteur Research Centre, Hong Kong SAR

We previously reported the detection of bat coronaviruses (bat CoVs 1A, 1B, HKU7, HKU8 and bat-severe acute respiratory syndrome coronavirus) in Miniopterus spp. that cohabit a cave in Hong Kong. Here, we report the full genomic sequences of bat CoVs 1A, 1B and HKU8. Bat CoVs 1A and 1B, which are commonly found in the Miniopterus, are phylogenetically closely related. Using species-specific RT-PCR assays, bat CoVs 1A and 1B were confirmed to have distinct host specificities to Miniopterus magnater and Miniopterus pusillus, respectively. Interestingly, co-infections of bat CoVs 1B and HKU8 in Miniopterus pusillus are detected in seven of 38 virus-positive specimens collected from 2004 to 2006. These findings highlight that co-infections of some coronaviruses might be common events in nature. The biological basis for the host restriction of bat coronaviruses, however, is yet to be determined.

Coronavirus (CoV) is a genus of viruses in the family Coronaviridae under the order Nidovirales. Based on antigenic and genetic analyses, CoVs are divided into three major groups. The viruses have positive-sense, single-stranded RNA genomes with sizes ranging from 27000 to 32000 nt (Gorbalenya et al., 2006). Their genomes are polycistronic and contain at least five major open reading frames (ORFs) with the characteristic gene order [5′-replicase (rep), spike (S), envelope (E), membrane (M) and nucleocapsid (N)-3′]. The severe acute respiratory syndrome (SARS) outbreak in 2003 highlighted the potential health threat of coronaviruses to humans (Peiris et al., 2003). This prompted an intensive search for the precursor of SARS CoV in wildlife and the discovery of several novel CoVs in bats, including the viruses that are closely related to SARS CoV (Dong et al., 2007; Lau et al., 2005; Poon et al., 2005; Tang et al., 2006; Woo et al., 2006). Recently, the biological importance of bats for the ecology of coronaviruses was further reiterated by the discovery of novel bat CoVs (BtCoVs) in other continents (Domínguez et al., 2007; Muller et al., 2007).

We previously reported the discovery of the first BtCoV (Poon et al., 2005). In addition, we also performed a longitudinal study on Miniopterus magnater and Miniopterus pusillus bats that cohabit an abandoned cave in Hong Kong (Chu et al., 2006). BtCoV HKU7 and bat-SARS-CoV were detected only once in our specimens, while BtCoVs 1A, 1B and HKU8 were detected repeatedly in these bats. A clear host species restriction was observed with BtCoV 1A being found in M. magnater and BtCoV 1B in M. pusillus even though these two CoVs are genetically very similar and the two bat species co-inhabit the same cave. Thus, these virus surveillance studies might allow unique opportunities to understand better the dynamics and prevalence of coronaviruses within a single geographical location. To understand better the genetic relationships of the BtCoVs that are repeatedly found in the local Miniopterus spp., representative faecal samples positive for BtCoVs 1A, 1B and HKU8 in our previous investigations were selected for this study. Genomic sequences of these three viruses were determined and analysed as described in Supplementary Material (available in JGV Online).

The genomes of BtCoVs 1A, 1B and HKU8, excluding the 3′ poly(A) tails, have sizes of 28309, 28460 and 28757 nt, respectively (GenBank accession nos EU420137–EU420139). The G+C content of 1A, 1B and HKU8 is 38.2, 38.6 and 41.8%, respectively. The genome organizations of these CoVs are typical of group 1 CoVs (Table 1) and the amino acid sequence identities of the major viral proteins between these viruses are shown in Supplementary Table S1 (available in JGV Online). The 3′ UTRs and N genes of these viruses were reported previously and will not be discussed further (Chu et al., 2006). Interestingly, HKU8 was found to contain an extra ORF at its 3′ end. This
predicted ORF (designated ORF 7) overlaps with the \( N \) gene and encodes a putative protein with 248 aa residues (Table 1).

The size of the 5′ UTR of 1A, 1B and HKU8 is 271, 272 and 269 nt, respectively. Three stem–loop structures that were proposed to be conserved in group 1 CoVs (Liu et al., 2007) and a core sequence of leader transcription regulatory sequence (5′-\( \text{CUAAAC} \)-3′) were identified in these 5′ UTR sequences (Supplementary Fig. S1 available in JGV Online).

These ORFs 1 occupy approximately 70% of the genomes. ORF1 consists of ORF1a and ORF1b that encodes viral pp1a and pp1ab, respectively. Putative features responsible for the \(-1\) ribosomal frame shifting, e.g. the ‘slippage sequence’ (5′-\( \text{UUAAAC} \)-3′), were identified in the genomes (data not shown). Based on \textit{in silico} analysis, the pp1a and pp1ab proteins are predicted to be cleaved into a total of 16 non-structural proteins by virus proteases (Table 1) (Chen et al., 2003). The predicted 3CL protease cleavage sites in these polyproteins (Supplementary Table S2 available in JGV Online) agree with previously defined substrate specificity for other 3CL proteases (Ziebuhr et al., 2000). All pp1ab of BtCoV 1A, 1B and HKU8 contain functional units that are typical for CoVs. For example, all these viruses contain two papain-like (PL) protease domains in nsp3, a 3C-like (3CL) protease in nsp5, CoV RNA-dependent RNA polymerase (RdRp) in nsp12, RNA helicase (Hel) in nsp13 and an mRNA cap-1 methyltransferase (MT) functional domain in nsp16. It was noted that the putative nsp1 of HKU8 has a size of 18 aa, which is significantly smaller than sizes of nsp1 predicted in other group 1 CoVs.

Similar to other CoVs, the N-terminal regions of these \( S \) genes are genetically diverse. The protein sequence identities between the putative S1 domains (the first 740 residues approximately) of HKU8 to 1A, HKU8 to 1B, and 1A to 1B are 38.5, 38.7 and 86.6%, respectively. The corresponding sequence identities for the putative S2 domains are 56.2 (HKU8 to 1A), 55.9 (HKU8 to 1B) and 91.3% (1A to 1B), respectively. Twenty nine, 31 and 30 potential N-glycosylation sites were predicted in 1A, 1B and HKU8 S proteins, respectively. Two putative heptad repeat (HR) regions, which are conserved in all CoVs (Bosch et al., 2003), were identified in these S proteins (both 1A and 1B: HR1, aa 973--1122; HR2, aa 1252--1311. HKU8: HR1, aa 976--1121; HR2, aa 1261--1317).

<table>
<thead>
<tr>
<th></th>
<th>BtCoV 1A</th>
<th></th>
<th>BtCoV 1B</th>
<th></th>
<th>BtCoV HKU8</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Position</td>
<td>Length (nt)</td>
<td>Position</td>
<td>Length (nt)</td>
<td>Position</td>
<td>Length (nt)</td>
</tr>
<tr>
<td>ORF1a</td>
<td>272–13078</td>
<td>12804</td>
<td>273–13247</td>
<td>12972</td>
<td>270–12965</td>
<td>12693</td>
</tr>
<tr>
<td>ORF1b</td>
<td>13048–21069</td>
<td>8019</td>
<td>13217–21238</td>
<td>8019</td>
<td>12935–20959</td>
<td>8022</td>
</tr>
<tr>
<td>nsp1 (pp1a/pp1ab)</td>
<td>272–859</td>
<td>588</td>
<td>273–860</td>
<td>588</td>
<td>270–323</td>
<td>54</td>
</tr>
<tr>
<td>nsp2 (pp1a/pp1ab)</td>
<td>860–2950</td>
<td>2091</td>
<td>861–2948</td>
<td>2088</td>
<td>324–1811</td>
<td>1488</td>
</tr>
<tr>
<td>nsp3 (pp1a/pp1ab)</td>
<td>2951–8290</td>
<td>5340</td>
<td>2949–8459</td>
<td>5511</td>
<td>1812–8174</td>
<td>6363</td>
</tr>
<tr>
<td>nsp4 (pp1a/pp1ab)</td>
<td>8291–9712</td>
<td>1422</td>
<td>8460–9881</td>
<td>1422</td>
<td>8175–9599</td>
<td>1425</td>
</tr>
<tr>
<td>nsp5 (pp1a/pp1ab)</td>
<td>9713–10618</td>
<td>906</td>
<td>9882–0787</td>
<td>906</td>
<td>9600–10505</td>
<td>906</td>
</tr>
<tr>
<td>nsp6 (pp1a/pp1ab)</td>
<td>10619–11455</td>
<td>837</td>
<td>10788–11624</td>
<td>837</td>
<td>10506–11342</td>
<td>837</td>
</tr>
<tr>
<td>nsp7 (pp1a/pp1ab)</td>
<td>11456–11704</td>
<td>249</td>
<td>11625–11873</td>
<td>249</td>
<td>11343–11591</td>
<td>249</td>
</tr>
<tr>
<td>nsp8 (pp1a/pp1ab)</td>
<td>11705–12286</td>
<td>582</td>
<td>11874–12455</td>
<td>582</td>
<td>11592–12173</td>
<td>582</td>
</tr>
<tr>
<td>nsp9 (pp1a/pp1ab)</td>
<td>12287–12616</td>
<td>330</td>
<td>12456–12785</td>
<td>330</td>
<td>12174–12503</td>
<td>330</td>
</tr>
<tr>
<td>nsp10 (pp1a/pp1ab)</td>
<td>12617–13024</td>
<td>408</td>
<td>12786–13193</td>
<td>408</td>
<td>12504–12911</td>
<td>408</td>
</tr>
<tr>
<td>nsp11 (pp1a)</td>
<td>13025–13078</td>
<td>54</td>
<td>13194–13247</td>
<td>54</td>
<td>12912–12965</td>
<td>54</td>
</tr>
<tr>
<td>nsp12 (pp1ab)</td>
<td>13025–15804</td>
<td>2781</td>
<td>13194–15973</td>
<td>2781</td>
<td>12912–15691</td>
<td>2781</td>
</tr>
<tr>
<td>nsp13 (pp1ab)</td>
<td>15805–17595</td>
<td>1791</td>
<td>15974–17764</td>
<td>1791</td>
<td>15692–17482</td>
<td>1791</td>
</tr>
<tr>
<td>nsp14 (pp1ab)</td>
<td>17596–19149</td>
<td>1554</td>
<td>17765–19318</td>
<td>1554</td>
<td>17483–19039</td>
<td>1557</td>
</tr>
<tr>
<td>nsp15 (pp1ab)</td>
<td>19150–20166</td>
<td>1017</td>
<td>19319–20335</td>
<td>1017</td>
<td>19040–20056</td>
<td>1017</td>
</tr>
<tr>
<td>nsp16 (pp1ab)</td>
<td>20167–21066</td>
<td>900</td>
<td>20336–21235</td>
<td>900</td>
<td>20057–20956</td>
<td>900</td>
</tr>
<tr>
<td>S</td>
<td>21071–25198</td>
<td>4125</td>
<td>21240–25397</td>
<td>4155</td>
<td>20956–25083</td>
<td>4125</td>
</tr>
<tr>
<td>ORF3</td>
<td>25198–25857</td>
<td>657</td>
<td>25366–26025</td>
<td>657</td>
<td>25083–25751</td>
<td>666</td>
</tr>
<tr>
<td>E</td>
<td>25851–26075</td>
<td>222</td>
<td>26019–26243</td>
<td>222</td>
<td>25835–25959</td>
<td>222</td>
</tr>
<tr>
<td>M</td>
<td>26082–26840</td>
<td>756</td>
<td>26250–27017</td>
<td>765</td>
<td>25966–26721</td>
<td>753</td>
</tr>
<tr>
<td>N</td>
<td>26861–28030</td>
<td>1167</td>
<td>27038–28189</td>
<td>1149</td>
<td>26770–28038</td>
<td>1266</td>
</tr>
<tr>
<td>ORF7</td>
<td>NA</td>
<td>–</td>
<td>NA</td>
<td>–</td>
<td>27788–28534</td>
<td>744</td>
</tr>
</tbody>
</table>
OF3 of 1A and 1B have the same size of 660 nt and that of HKU8 is 669 nt. In terms of amino acid sequence similarity, ORF3 is the least conserved gene amongst these three viruses (Supplementary Table S3 available in JGV Online). Signal anchors in these ORF3 proteins were predicted with the probability of 0.545–0.689 by SignalP. Multiple transmembrane motifs were also predicted in these ORF3 proteins, suggesting these proteins might be surface proteins.

The E genes of these viruses are both the same size of 225 nt. One transmembrane signal was predicted in each of these E proteins (P>0.990), with the N terminus exposed to the surface.

Signal peptides were predicted in these M proteins (P>0.933). Cleavage sites were located between aa 17 and 18 in these M proteins. The N-terminal signal peptide was not observed in other BtCoV M proteins, including BtCoV/512/2005 and HKU2. Three strong transmembrane helices and one marginal transmembrane helix were predicted with the N terminus exposed to the surface. It has been reported that the M protein of the transmissible gastro-enteritis virus (TGEV) adopts two different topologies in the virus envelope with three transmembrane helices and the C terminus located inside the virus particle, or with four transmembrane helices and the C terminus exposed to the virion surface (Escors et al., 2001). This may also be the case for M proteins of these group 1 BtCoVs.

ORF7 is present only in HKU8, and it is located downstream of the N protein. A BLAST search of the ORF7 amino acid sequence revealed no homologous gene in GenBank. Prediction of an ORF (or ORFs) downstream of the N gene was also reported in the genomes of some group 1 CoVs, including TGEV (GenBank accession no. NC_002306), feline infectious peritonitis virus (FIPV) (GenBank accession no. AY994055), BtCoV/512/2005 (Tang et al., 2006) and BtCoV HKU2 (GenBank accession no. NC_009988). ORF7 in TGEV was shown to produce an endoplasmic reticulum associated protein, suggesting it plays a role in the association of replication complexes or in virion assembly (Tung et al., 1992). The ORF7ab of FIPV has been suggested to affect virulence (Haijema et al., 2004). However, there is no significant sequence similarity between these ORFs and it is likely that these viruses have acquired their unique ORFs from different sources.

The putative core sequence 5’-CUAAAC-3’ in the transcription regulatory sequence (TRS), which was identified in other group 1 CoVs (Dye & Siddell, 2005; Hiscox et al., 1995; Lau et al., 2007; Pyrc et al., 2004), is present in the leader TRSs of the studied genomes. They start at position 67 in BtCoV 1A and position 68 in both BtCoV 1B and HKU8. All the core TRS sequences are highly similar to its corresponding leader TRS (Supplementary Fig. S2) except the TRSs for all ORF3 and the E gene of HKU8. It has been reported that the leader TRS can also join the body segment at an unusual consensus sequence preceding the ORFs without the presence of the core sequence (Hofmann et al., 1993). However, we were not able to demonstrate this in these bat viruses because none of these viruses could be cultured in vitro and we have no means of obtaining sufficient viral RNA for detailed investigation. Thus, the authentic TRSs of these bat viruses, in particular those for ORF3, are yet to be confirmed.

Phylogenetic analyses of RdRp, helicase, S, E, M and N proteins from these BtCoVs were performed. Mouse hepatitis virus was included as an out-group in the analyses. In all cases, the three viruses studied formed a clade that is distinct to group 1 members isolated from other animals (Fig. 1). The BtCoVs 1A, 1B and HKU8 viruses phylogenetically cluster with BtCoVs Shandong/977/2006 and Fujian/773/2005 detected in mainland China (Fig. 1). Sequence analysis on partial RdRp sequences revealed that the Fujian/773/2005 found in Miniopterus schreibersii could be a HKU7-like virus (not shown). The previously published S gene and helicase gene sequences of BtCoV Shandong/977/2006 (Vijaykrishna et al., 2007), which was detected in bats of the species Rhinolophus ferrumequinum, show a close genetic relationship to the three CoVs described in this report. The S protein amino acid sequence similarity between BtCoV Shandong/977/2006 and BtCoV 1A is 46.5 %, while BtCoV 1A to HKU8 is 46.4 % (Supplementary Table S3). A similar relationship was also observed for the helicase gene (Shandong/977/2006 to 1A: 74.2 %; Shandong/977/2006 to HKU8: 75.5 %; 1A to HKU8: 76.5 %). Therefore, CoVs related to that found in the genus Miniopterus could also be found in other genera of bats. Considering the relationship between these CoV S proteins, which is one of the major determinants of host species specificity of infection (Dveksler et al., 1991), and the relatively low detection rates of BtCoVs HKU8 and HKU7 in Miniopterus (Chu et al., 2006), it is possible that the two viruses are naturally harboured in species of bats other than Miniopterus bats. Overall, the above findings highlight that some BtCoVs might have the ability for interspecies transmission, an event that is relevant to the genesis of SARS CoVs in humans.

It has been previously shown that recombination is not a rare event in CoVs (Makino et al., 1986). To detect possible recombination between the bat viruses we studied, which were all detected in Miniopterus bats found at a single habitat, genome sequences of BtCoVs 1A, 1B and HKU8 were aligned and examined by SimPlot and BootScan analyses. In addition, these genomes were also aligned with other full-length group 1 BtCoV genomes, BtCoV/512/2005 and BtCoV HKU2, for comparison. Bat-SARS-CoV was included in the analyses as this virus had been detected once in our study (Chu et al., 2006). No evidence of recombination between these genomes was detected (data not shown).

The limitation of direct sequencing of PCR amplicons is that the presence of co-infections may be masked by the
Fig. 1. Phylogenetic analyses of RdRp, Hel, S, E, M and N protein amino acid sequences. S protein of BtCoV HKU2 was not included in the analysis as the sequence similarity of this protein to that of all other group 1 CoVs was extremely low (Lau et al., 2007). MHV was included as an out-group in the analyses. Numbers at nodes indicate bootstrap values as a percentage. Bars show the estimated genetic distance of these viruses. GenBank accession numbers of the reference sequences and definition of abbreviations are listed in Supplementary Material (available in JGV Online).
sequence of the dominant virus. Given the high detection rates of BtCoVs 1A and 1B in our specimens and by the apparent species restriction of these two viruses to two different bat species that co-inhabit the same cave, we designed species-specific PCR assays for detecting BtCoVs 1A, 1B and HKU8 in our specimens. We applied these assays to all specimens that were positive in our CoV consensus screening RT-PCR in order to detect the presence of co-infections. These viruses were selected for screening as they were repeatedly detected in *Miniopterus* bats found within the same cave in Hong Kong. A total of 38 specimens collected in a period from 2004 to 2006 in Hong Kong, which were known to be positive for CoV (Chu et al., 2006), were tested by six different species-specific RT-PCR assays (RdRp and N genes of BtCoV 1A, 1B and HKU8). The identities of the amplified products were all confirmed by sequencing. In summary, infection with 1A alone was detected in 16 specimens collected from *M. magnater*, infection with 1B alone was detected in 14 specimens collected from *M. pusillus* and infection with BtCoV HKU8 alone was detected in one specimen collected from *M. magnater*. There was no evidence of co-infections with the more genetically closely related viruses 1A and 1B, even though these two viruses co-habited the same cave and were more frequently detected in our specimens from their corresponding hosts. These results suggest that there is a strong host restriction for these two viruses that apparently hinders interspecies transmission between these two species of bat. Interestingly, co-infections of 1B and HKU8 were detected in seven *M. pusillus* specimens collected in 2004 (*n*=2) and 2006 (*n*=5) (Table 2).

The partial N gene sequences (>500 nt) derived from the co-infected samples were aligned and a phylogenetic tree was constructed (Supplementary Fig. S3 available in JGV Online). The analysis revealed that the N gene sequences of the 1B viruses were not identical (similarities ranging from 96.8 to 98.8 %), confirming that our findings were not due to cross-contamination. On the other hand, three different HKU8 N gene sequences were detected (similarities ranging from 96.0 to 98.1 %) (Supplementary Fig. S3 available in JGV Online). HKU8 N gene sequences deduced from four of five specimens collected in the same sampling trip in 2006 were found to be identical. The detection of a genetically identical virus in these four specimens could represent a recent spread of a single infecting strain within this bat population, but the alternative possibility of cross-contamination at sampling or subsequent procedures cannot be conclusively excluded. However, the finding of HKU8 and 1B co-infections in samples AFC337 and WCF4 with unique sequences in each of these specimens (Table 2) and the fact that such co-infections have been detected in multiple sampling occasions (2004 and 2006) strongly suggest that co-infections of these two viruses in *M. pusillus* occur in nature and are not isolated events.

In this report, we describe the full genomes of three group 1 BtCoVs. Co-infection with different CoVs in their host provides chances for producing recombinants between these CoVs. In our study, there was no evidence of recombination events detected between the reported genomes of BtCoV 1A, 1B and HKU8. But the diversity of CoVs found in bats and the detection of BtCoV co-infections suggest bats might play a key role in shaping the evolution and ecology of CoV. The availability of genetic information for these viruses helps us better understand the taxonomy and evolution of CoVs in general. But more importantly, since SARS and a number of other new human diseases have emerged through interspecies transmission of viruses harboured by bats, the biological basis of host restriction of bat CoVs remains an important subject for study for global public health.

### Table 2. Specimens co-infected with BtCoV 1B and HKU8

The host used was *M. pusillus*. –, Negative; +, positive; ND, not done.

<table>
<thead>
<tr>
<th>Date of collection (mm/yy)</th>
<th>Sample ID</th>
<th>RdRp</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1A</td>
<td>1B</td>
</tr>
<tr>
<td>05/04</td>
<td>WCF4</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>05/04</td>
<td>WCF14</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>03/06</td>
<td>AFC307†</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>03/06</td>
<td>AFC309</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>03/06</td>
<td>AFC323</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>03/06</td>
<td>AFC325</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>03/06</td>
<td>AFC337</td>
<td>–</td>
<td>+</td>
</tr>
</tbody>
</table>

*Insufficient RNA samples to perform the assay.
†This sample was further tested by RT-PCR assay specific for the nsp3 of HKU8 and was positive in the assay.
Republic of China. We thank S. F. Sia for her technical support and K. S. Cheung, C. T. Shek and C. S. M. Chan for facilitating this study.

References


