Supplementary Material:

Table I: Summarizes details on the 13 coronaviruses studied here. In this work, we chose the papain-like protease domains (20 in all) of those coronaviruses analogous to the SARS-CoV PLpro minimal core domain (1540-1854) following the basic strategy of Barretto and co-workers (1). Seven of these nine coronaviruses, namely MHV strain JHM, bovine CoV (BCoV), HCoV-HKU1, TGEV, HCoV-NL63, HCoV-229E and HCoV-OC43 encode two papain like protease domains, termed PL1 and PL2. The SARS-CoV, aIBV and the four Bat coronaviruses, namely Bat-SARS-CoV (NCBI Accession: AAZ67050) BtCoV/273/2005 (ABG47059), BtCoV/133/2005 (Q0Q4F3), and BtCoV-HKU9 (A3EXG5) encode only one papain-like protease domain termed PLpro (2, 3)

Full name of coronaviruses (Primary species infected)	Abbreviations	PLPs	Domain of PLP sequences analyzed	Serological groups
Transmissible Gastroenteritis virus[4] (Pigs)	TGEV-CoV	PL1 PL2	992-1248 1479-1781	G1a
Human Coronavirus NL or New Haven Coronavirus[5, 6] (Human)	HcoV-NL63	PL1 PL2	952-1250 1578-1876	G1b
Human Coronavirus 229E [7, 8] (Human)	HcoV-229E	PL1 PL2	949-1245 1599-1905	G1b
Bovine Coronavirus [9] (Cows/cattle)	BcoV	PL1 PL2	965-1250 1562-1870	G2a
Human Coronavirus Hong Kong University1[10, 11] (Human)	HcoV-HKU1	PL1 PL2	1028-1329 1648-1955	G2a
Murine Hepatitis virus strain JHM[12-14] (Mouse)	MHV-JHM-CoV	PL1 PL2	1013-1309 1606-1915	G2a
Human Coronavirus OC43[15] (Human)	HcoV-OC43	PL1 PL2	965-1250 1562-1870	G2a
Severe Acute Respiratory Syndrome Coronavirus (CoV)[2, 16] (Human)	HCoV-SARS	PLP	1540-1854	G2b
Bat-SARS-Coronavirus[22,23] (Bat)	Bt-SARS-CoV	PLP	1539-1854	G2b
Bat Coronavirus 273[17] (Bat)	BtCoV-273	PLP	1537-1855	G2b
Bat Coronavirus 133[17] (Bat)	BtCoV-133	PLP	1533-1855	G2c
Bat coronavirus HKU9[18] (Bat)	BtCoV-HKU9	PLP	1426-1737	G2d
Avian infectious Bronchitis virus[24] (Birds)	a-IBV	PLP	1174-1536	G3

The PLP1s of the coronaviruses were first aligned with the SARS-CoV-PLpro amino acid sequence using ClustalW [19]. The result was manually checked and edited to make sure that the key residues identified from the structure of the 315aa region of the minimal core domain (actual amino acid sequence numbers) of SARS-CoV-PLpro, namely W107 (W1646), N110 (N1649), C112 (C1651), Y113 (Y1652), L163 (L1702), G164 (G1703), D165 (D1704), E168 (E1707), Y265 (Y1804), G267 (G1806), G272 (G1811), H273 (H1812), Y274 (Y1813) and D287 (D1826), were aligned correctly with the other seven coronaviral PLP1s. This multiple sequence alignment was named Profile1. Next, the PLP2s of the coronaviruses, along with aIBV and the BtCoV-PLPs were aligned to the SARS-CoV-PLpro sequence (with special attention given to the 14 residues noted above) to obtain a second alignment termed Profile2. ClustalX [19] was then used to perform a profile alignment between Profile1 and Profile2 to conserve the alignments within the individual profiles and yet align them with each other. Profile alignment works much better among the PLpro amino acid sequences since the low sequence identity barely preserves the alignment of the important residues among the family, especially among the PLP1s and PLP2s. To our knowledge, this type of profile aligning approach has not been used in any prior work. The resulting multiple sequence alignment shows that even though the overall sequence identity in the coronaviral protease domain is very low (highest being ~30%), the amino acids likely to participate in the catalytic core are highly conserved.

Table II: The multiple sequence alignment of coronavirus papain-like protease (PLP) core domains shows the active site conserved residues highlighted in yellow starting with W107 of SARS-CoV-PLpro that stabilizes the oxyanion hole. The catalytic cysteine, histidine and aspartic acid are completely conserved among all PLPs. The residues highlighted in yellow are the 14 residues that are studied in detail among all the coronaviruses for cluster generation and classification studies. Besides these, the four Zn-chelating cysteines intermittently substituted by a histidine in 229E-PLP2, NL63-PLP2 and TGEV-PLP2 are highlighted in magenta. Additional residues such as N111 and Q122 that are also totally conserved among all the PLPs, but are not present in the active site; the residues corresponding to P249 or T302 that participates in the formation of the hydrophobic crevice that is balanced by leucine in the S4 substrate position of LXGG are highlighted in green. The arrows point to the key SARS-CoV-PLpro residues (all residue numbering is referenced to the 316-amino-acid catalytic core domain of PLpro, Ratia et al. (21)

```
SARS     PLP1540      -----MEVKTIKVFTTVDNTNLHTQLVDMSMTYGQQFGPTYLDG-ADVTKIKPHVNHEGK
BtSARS   PLP1539      ------EVKTIKVFTTVDNTNLHTQLVDMSMTYGQQLGPTYLEG-ADVTKIKPHVNHEGK
Bt273    PLP1537      ------EVKTIKVFTTVDNTNLHTQLVDMSMTYGQHFGPTYLDG-ADVTKVKPHVNHEGK
Bt133    PLP1533      ---------NIEVLVTIDGVNFRTVILNDATTFRKQLGATFYKG-VDISDALPTVKMGGE
BtHKU9   PLP1426      ETKLCAPEPLIKVLTTVDGINYSTVLVSTAQSYRAQIGTVFCDG-HDWSNKNPMPTDEGT
MHVJ     P2 1606      ------LANKVDVLCTVDGVNFRSCCVAEGEVFGKTLGSVFCDG-INVTKVRCSAIYKGK
HKU1     P2 1648      ------LAKKIDVLLTVDGVNFKSISLTVGEVFGKILGNVFCDG-IDVTKLKCSDFYADK
BCoV     P2 1562      -------LDKVDILLTVDGVNFTNRFVPVGESFGKSLGNVFCDG-VNVTKHKCDINYKGK
OC43     P2 1562      -------LDKVDILLTVDGVNFTNRFVPVGESFGKSLGNVFCDG-VNVTKHKCDINYKGK
229E     P2 1599      ------EAKVITIKVTEDGVNVHDVTVTTDKSFEQQVG-VIADKDKDLSGAVPSDLNTSE
NL63     P2 1578      ------KNDNVVLKITEDGINVKDVVVESSKSLGKQLG-VVSDGVDSFEGVLP--INTDT
TGEV     P2 1479      QTIENFFSCSIPVNVTEDNVNHERVSVSFDKTYGEQLKGTVVIKDKDVTNQLPSAFDVGQ
aIBV     PLP1174      ------TCKQKTIYLTEDGVKYRSIVLKPGDSLG-QFGQVYA-KNKIVFTAD-DVEDKEI
229E     P1 949       ----FCKTIQSALSVVSCYVNLPTYYIYDEEGGNDLSLPVMISEWPLSVQQ----AQQEA
NL63     P1 952       ------EVLTSAMNVIGQHIKLPQFYIYDEEGGYDVSKPVMISQWPISNDSNGCVVEAST
BCoV     P1 965       -ED--DDFLEESGVEEDDVEGEETDLTVTSAGEPCVASEQEESSEILEDTLDDGPCVETS
OC43     P1 965       -ED--DDFLEESDVEEDDVEGEETDLTVTSAGQPCVASEQEESSEVLEDTLDDGPSVETS
HKU1     P1 1028      -EDVVTGDNDDEDVVTGDNDDEDVVTGDNDDEDVVTGDNDDEDVVTGDNDDQIVVTGDDV
MHVJ     P1 1013      ---PIASAEETEVGEACDREG-IAEVKATVCADALDACPDQVEAFDIEKVEDSILSELQT
TGEV     P1 992       -----EVSASSEEEEVESSVEEDPENEIVEASEGAEGTS----SQEEVETVEVADITSTEE
```

```
                                                 (W107) W1646  C1651 (C112)
                                                      ↓     ↓Y1652
SARS     PLP1594      TFFVLPSDDTLRSEAFEYYHTL-DESFLGRYMSALNHTKKWKFPQVGGLTSIKWADNNCY
BtSARS   PLP1592      TFFVLPSDDTLRSEAFEYYHTL-DESFLGRYMSALNHTKKWKFPQVGGLTSIKWADNNCY
Bt273    PLP1590      TFFVLPSDDTLRSEAFEYYHTL-DESFLGRYMSALNHTKKWKFPQVGGLTSIKWADNNCY
Bt133    PLP1580      SLFVADNLSESEEVVLKEYYGTSDVTFLQRYYSLQPLVQQWKFVVHDGVKSLKLSNYNCY
BtHKU9   PLP1485      HLYKQDNFSSAEVTAIREYYGVDDSNIIARAMSIRKTVQTWPYTVVDGRVLLAQRDSNCY
MHVJ     P2 1660      VFFQYSDLSEADLAAVKDAFGF-DEPQLLQYYSMLGMC-KWPVVVCGNYFAFKQSNNNCY
HKU1     P2 1701      ILYQYENLSLADISAVQSSFGF-DQQQLLAYYNFLTVC-KWSVVVNGPFFSFEQSHNNCY
BCoV     P2 1614      VFFQFDNLSSEDLKAVRSSFNF-DQKELLAYYNMLVNCKWQVVVNGKYFTFKQANNNCF
OC43     P2 1614      VFFQFDNLSSEDLKAVRSSFNF-DQKELLAYYNMLVNCFKWQVVVNGKYFTFKQANNNCF
229E     P2 1652      LLTKAIDVDWVEFYGFKDAVTF-ATVDHSAFA-------YESAVVNGIRVLKTSDNNCW
NL63     P2 1629      VLSVAPEVDWVAFYGFEKAALF-ASLDVKPYG-------YPNDFVGGFRVLGTTDNNCW
TGEV     P2 1539      KVIKAIDIDWQAHYGFRDAAAF-SASSHDAYK-------FEVVTHSNFIVHKQTDNNCW
aIBV     PLP1225      LYVPTTDKSILEYYGLDAQKYVIYLQTLAQK--------WNVQYRDNFLILEWRDGNCW
229E     P1 1001      TLPDIAEDVVDQVEEVNSIFDIETVDVKHDVS---PFEMPFE--ELNGLKILKQLDNNCW
NL63     P1 1006      DFHQLECIVDDSVREEVDIIEQPFEEVEHVLSI--KQPFSFSFRDELGVRVLDQSDNNCW
BCoV     P1 1022      DSQVEEDVEMSDFADLESVIQDYENVCFEFYT----TEPEFV--KVLDLYVPKATRNNCW
OC43     P1 1022      DSQVEEDVEMSDFVDLESVIQDYENVCFEFYT----TEPEFV--KVLGLYVPKATRNNCW
HKU1     P1 1087      DD-IESIYDFDTYKALLVFNDVYNDALFVSYGSSVETETYF---KVNGLWSPTITHTNCW
MHVJ     P1 1069      ELNAPADKTYEDVLAFDAIYSETLSAFYAVPS----DETHF---KVCGFYSPAIERTNCW
TGEV     P1 1043      DVDIVEVSAKDDPWAAAVDVQEAEQFNPSLPP--------FKTTNLNGKIILKQGDNNCW
                                                                          :   :**:
```

```
                     Q1661 (Q122)                                    L1702 (L163)
                         ↓                                          ↓ D1704 (D165)
SARS     PLP1653      LSSVLLALQQL-EVKFN----APALQEAYYRARAGDAANFCALILAYS--NKTVGELGDVR
BtSARS   PLP1651      LSSVLLALQQI-EVKFN----APALQEAYYRARAGDAANFCALILAYS--NKTVGELGDVR
Bt273    PLP1649      LSSVLLALQQI-EVKFN----APALQEAYYRARAGDAANFCALILAYS--NKTVGELGDVR
Bt133    PLP1640      INATIMMIDMLHDIKFVVP--A--LQNAYLRYKGGDPYDFLALIMAYG--DCTFDNPDD--
BtHKU9   PLP1545      LNVAISLLQDI-DVSFSTP--WVCRA--YDALKGGNPLPMAEVLIALG--KATPGVSDDAH
MHVJ     P2 1718      INVACLMLQHL-SLKFP----KWQWRRPGNEFRSGKPLRFVSLVLAKG--SFKFNEPSDST
HKU1     P2 1759      VNVACLMLQHI-NLKFN----KWQWQEAWYEFRAGRPHRLVALVLAKG--HFKFDEPSDAT
BCoV     P2 1673      VNVSCLMLQSL-NLKFK----IVQWQEAWLEFRSGRPARFVSLVLAKG--GFKFGDPADSR
OC43     P2 1673      VNVSCLMLQSL-HLTFK----IVQWQEAWLEFRSGRPARFVALVLAKG--GFKFGDPADSR
229E     p2 1703      VNAVCIALQYS-KPHFI----SQGLDAAWNKFVLGDVEIFVAFVYVA--RLMKGDKGDA-
NL63     P2 1680      VNATCIILQYL-KPTFK----SKGLNVLWNKFVTGDVGPFVSFIYFIT--MSSKGQKGDA-
TGEV     P2 1590      INAICLALQRL-KPQWK----FPGVRGLWNEFLERKTQGFVHMLYHIS--GVKKGEPGDA-
aIBV     PLP1276      ISSAIVLLQAA-KIRFK----G-FLTEAWAKLLGGDPTDFVAWCYASC--TAKVGDFSDAN
229E     P1 1056      VNSVMLQIQLT-GI-------LDGDYAMQFFKMGRVAKMIERCYTAE--QCIRGAMGDVG
NL63     P1 1061      ISTTLVQLQLT-KL-------LDDSIEMQLFKVGKVDSIVQKCYELS--HLISGSLGDSG
BCoV     P1 1076      LRSVLAVMQKL-PCQF----KDKNLQDLWVLYKQQYSQLFVDTLVNKIPANIVVPQGGYVA
OC43     P1 1076      LRSVLAVMQKL-PCQF----KDKNLQDLWVLYKQQYSQLFVDTLVNKIPANIVLPQGGYVA
HKU1     P1 1143      LRSVLLVMQKL-PFKF----KDLAIENMWLSYKVGYNQSFVDYLLTTIPKAIVLPQGGFVA
MHVJ     P1 1122      LRSTLIVMQSL-PLEF----KDLGMQKLWLSYKAGYDQCFVDKLVKSAPKSIILPQGGYVA
TGEV     P1 1095      INACCYQLQAF---DF-------FNNEAWEKFKKGDVMDFVNLCYAAT--TLARGHSGDA-
                         :
```

```
                    E1707(E168)              C1729(C190)
                    ↓                        ↓
SARS    PLP1707     ETMTHLLQH---ANLESAKRVLNVVCKHCGQKT-TTLTGVE-AVMYM-GTLS--Y-DNLKTGVS--
BtSARS  PLP1705     ETMTHLLQH---ANLESAKRVLNVVCKHCGQKT-TTLTGVE-AVMYM-GTLS--Y-DNLKMGVS--
Bt273   PLP1703     ETMTHLLQH---ANLESAKRVLNLVCKHCGQKT-TTLMGVE-AVMYM-GTLS--Y-DELKAGVS--
Bt133   PLP1693     EAKLLHTLLAKAELTVSAKMVWREWCTVCGIRDIEYT-GMR-ACVYA-GVNS--M-EELQSVFN--
BtHKU9  PLP1599     MVLSAVLNHG--TVTA--RRVMQTVCEHCGVSQMVFT-GTD-ACTFY-GSVV--L-DDLYAPVS-V
MHVJ    P2 1772     DF-IRVELRE--ADLSGATCDLEFICK-CGVKQ-EQRKGVD-AVMHF-GTLD--KSGLVK-GYN--
HKU1    P2 1813     DF-IRVVLKQ--ADLSGAICELELICD-CGIKQ-ESRVGVD-AVMHF-GTLA--KTDLFN-GYK--
BCoV    P2 1727     DF-LRVVFSQ--VDLTGAICDFEIACK-CGVKQ-EQRTGVD-AVMHF-GTLS--REDLEI-GYT-V
OC43    P2 1727     DF-LRVVFSQ--VDLTGAICDFEIACK-CGVKQ-EQRTGLD-AVMHF-GTLS--REDLEI-GYT-V
229E    p2 1756     EDTL-TKLSKYLANEAQVQLEHYSSCVECDAKFKNSVASINSAIVCASVK----R-D----GVQ-V
NL63    P2 1733     EEAL-SKLSEYLISDSIVTLEQYSTCDICKSTVVEVKSAIVCASVL-------K-D----GCD-V
TGEV    P2 1643     ELML-HKLGDLMDNDCEIIVTHTTACDKCAKVE-KFVGPVVAAPLAI--------H----GTD--
aIBV    PLP1329     WL-LANLAEHFDADYTNAFLKKRVSCN-CGIKSYELRGLE--ACIQPVRAT-NLLHF--KTQYS--
229E    P1 1106     LCMYRLLK-----DLHTGFMVMDYKC-SCTSGR--LEESG--AVLFCTPTK--------KAFPY--
NL63    P1 1111     KL-LSELLK----EKYTCSITFEMSC-DCGKKF--DDQ---VGCLFWIMPYT-------KLFQK--
BCoV    P1 1132     DFAYWFLTL---CDWQ---CVAYWKCIKCDLAL--KLKGLD-A-MFFYG-------DVV----SHV
OC43    P1 1132     DFAYWFLTL---CDWQ---CVAYWKCIKCDLAL--KLKGLD-A-MFFYG-------DVV----SH-
HKU1    P1 1199     DFAYWFLNQ---FDIN---AYANWCCLKCGFSF--DLNGLD-A-LFFYG-------DIV----SHV
MHVJ    P1 1178     DFAYFFLSQ---CSFK--AYANWRCLKCGMEL--KLQGLD-AV-FFYG-------DVV----SH-
TGEV    P1 1143     EYLLELMLN----DYSTAKIVLAAKCG-CGEKE--IVLER--AV--FKLTPLK---ESFNY-G-V


                    C1766(C227)         P1788(P249)    Y1804(Y265)   HIS1812(H273)
                                                                     ↓Y1813(Y274)
                    ↓                   ↓              ↓
SARS    PLP1762     IPCV-CG--RDATQYLVQQESSFVMMS---APPA-EYKL---QQGTFLCANEYTGNYQ----CGHYTHI
BtSARS  PLP1760     IPCV-CG--RDATQYLVQQESSFVMMS---APPA-EYKL---QQGTFLCANEYTGNYQ----CGHYTHI
Bt273   PLP1758     IPCV-CG--RDATQYLVQQESSFVMMS---APPA-EYKL---QQGTFLCANEYTGNYQ----CGHYTHI
Bt133   PLP1751     ETCV-CG--SVKHRQLVEHSTPWLLV-S---GLN--EVKVSTSTDPVYRAFNVFQGVET---SVGHYVHV
BtHKU9  PLP1654     -VCQ-CG--RPAIRYVSEQKSPWLLMS---CTPTQVPLDTSG---IWKTAIVFRGPVT----AGHYMYA
MHVJ    P2 1826     IACT-CG--DKLVHCTQFNVPFL-ICS---NTP--EGK---KLPDDVVAANIFTGGS-----VGHYTHV
HKU1    P2 1867     IGCN-CA--GRIVHCTKLNVPFL-ICS---NTP--LSK---DLPDDVVAANMFMGVG-----VGHYTHL
BCoV    P2 1782     -DCS-CG--KKLIHCVRFDVPFL-ICS---NTPASV-----KLPKGVGSANIFKGDK-----VGHYVHV
OC43    P2 1782     -DCS-CG--KKLIHCVRFDVPFL-ICS---NTPASV-----KLPKGVGSANIFIGDK-----VGHYVHV
229E    p2 1810     GYCV-HGIKYYSRVRSVRGRAIIVSVE--QLEPCAQ-----SRLLSGVAYTAFSGPVD----KGHYT-V
NL63    P2 1783     GFCP-HRHKLRSRVKFVNGRVVITNV----GEPIISQP---SKLLNGIAYTTFSGSFD----NGHYV-V
TGEV    P2 1784     ETCV-HGVSVNVKVTQIKGTVAITSL----IGPIIG------EVLEATGYICYSGSNR----NGHYTY-
aIBV    PLP1386     -NCPTCGANNTDEVIEASLPYLLLFATD---GPATVD-----CDEDAVGTVVFVGSTN---SGHCYT-
229E    P1 1152     GTCLNCNAPRMCTIRQLQGTIIFVQQK---PEPVN----PVSFVVKPVCSSIFRGAVS----CGHYQT-
NL63    P1 1160     GECCICHKMQTYKLVSMKGTGVFVQD----PAPID----IDAFPVKPICSSVYLGVKG----SGHYQT-
BCoV    P1 1176     --CK-CGESMVLIDVDVPFTAHFA-------LKDK-LFCAF-----ITKRSVYKAACVVDVNDSHSMAV
OC43    P1 1176     -ICK-CGESMVLIDVDVPFTAHFA-------LKDK-LFCAF-----ITKRIVYKAACVVDVNDSHSMAV
HKU1    P1 1244     CKCG-HNMTLIAADLPCTLHFS---------LFDDN-FCAFCTP---KK-I-FIAACAVDVNVCHSVAV
MHVJ    P1 1222     -MCK-CGNSMTLLSADIPYTFDFG-------VRDD-KFCAF-----YTPRKVFRAACAVDVNDCHSMAV
TGEV    P1 1192     --CGDCMQVNTCRFLSVEGSGVFVHDILSKQTPEAMFVV---KP--VMHA-VYTGTTQ----NGHYM-V
                    *  *                   :               :           :*:


                    D1826(D287)        T1841(T302)
                    ↓                   ↓
SARS    PLP1817     TAKE-TLYRIDGAHLTKMS-EYKGP--VTDVFYKETSYTTTI-
BtSARS  PLP1816     TAKE-TLYRIDGAHLTKMS-EYKGP--VTDVFYKETSYTTTI-
Bt273   PLP1814     TAKE-TLYRIDGAHLTKMS-EYKGP--VTDVFYKETSYTTTI-
Bt133   PLP1809     RVKDGLFYKYDSGSLTKTS-DMKCK--MTSVWYPKVRYTADCN
BtHKU9  PLP1709     VNGT-LISVYDANTRRRTS-DLKLP--AIDILYGPTSFTSDSK
MHVJ    P2 1879     KCKP-KYQLYDACNVSKVS-EAKGN--FTDCLYLKNLKQTF--
HKU1    P2 1919     KCGS-PYQHYDACSVKKYT-GVSGC--LTDCLYLKNLTQTF--
BCoV    P2 1834     KCEQ-SYQLYDASNVKKVT-DVTGN--LSDCLYLKNLKQTF--
OC43    P2 1834     KCEQ-SYQLYDASNVKKVT-DVTGK--LSDCLYLKNLKQTF--
229E    p2 1869     YDTA-KKSMYDGDRFVKHD-LSLLS--VTSVVMVGGYVAPV--
NL63    P2 1840     YDAA-NNAVYDGARLFSSD-LSTLA--VTAIVVVGGCVTSN--
TGEV    P2 1748     YDNR-NGLVVDAEKAYHFN-RDLLQ--VTTAIASNFVV-----
aIBV    PLP1502     QAA---GQAFDNLAKDRKFGKKSPY--ITAMYTRFAFKNE---
229E    P1 1209     NIYS-QNLCVDGFGVNKIQ-PWTNDALNTICIKDADYNA----
NL63    P1 1216     NLYS-FNKAIDGFGVFDIK----NSSVNTVCFVDVDFHSV---
BCoV    P1 1220     VD----GKQIDDHRVTSIT----SDKFDFIIGHGMSFSM----
OC43    P1 1220     VD----GKQIDDHRITSIT----SDKFDFIIGHGMSFSM----
HKU1    P1 1298     IGDEQ----IDGKFVTKFS----GDKFDFIVGYGMSFSMS---
MHVJ    P1 1276     VD----GKQIDGKVVTKFN----GDKFDFMVGHGMTFSMSPF-
TGEV    P1 1249     DD-IEHGYCVDGMGIKPLK-----KRCYTSTLFINANVMTRA-
```

References:

1. 1. Barretto N, Jukneliene D, Ratia K, Chen Z, Mesecar AD, Baker SC. The Papain-Like Protease of Severe Acute Respiratory Syndrome Coronavirus has Deubiquitinating Activity. Journal of virology 2005; 79.24: 15189-98.
2. Snijder EJ, Bredenbeek PJ, Dobbe JC, Thiel V, Ziebuhr J, Poon LLM, Guan Y, Rozanov M, Spaan WJM, Gorbalenya AE. Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage. J Mol Biol 2003; 331: 991-1004
3. Sulea T, Lindner HA, Purisima EO, Ménard R. Binding Site-Based Classification of Coronaviral Papain-Like Proteases. Proteins. 2006; 62.3: 760-75.
4. Ortega J, Escors D, Laude H & Enjuanes L. J.Virol. 2002; 76: 11518-11529.
5. 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. Identification of a new human coronavirus. Nat Med 2004; 10: 368-373.
6. Fouchier RA, Hartwig NG, Bestebroer TM. A previously undescribed coronavirus associated with respiratory disease in humans. Proc Natl Acad Sci USA 2004; 101:6212–6.
7. Thiel V, Herold J., Schelle B and Siddell SG. Infectious RNA transcribed in vitro from a cDNA copy of the human coronavirus genome cloned in vaccinia virus. J. Gen. Virol. 2001; 82: PT 6, 1273-1281.
8. Herold J, Siddell SG and Gorbalenya AE. A human RNA viral cysteine proteinase that depends upon a unique Zn2+-binding finger connecting the two domains of a papain-like fold. J. Biol. Chem. 1999; 274:21, 14918-14925.
9. Current Opinion in Infectious Diseases. 2007; 20.2: 204-236
10. Woo PCY, Lau SKP, Chu C, Chan K, Tsoi H, Huang Y, Wong BHL, Poon RWS, Cai JJ, Luk W, Poon LLM, Wong SSY, Guan Y, Peiris JSM, and Yuen K. Characterization and Complete Genome Sequence of a Novel Coronavirus, Coronavirus HKU1, from Patients with Pneumonia. Journal of Virology, January 2005; 79.2: 884-895
11. Lau SKP, Woo PCY, Yip CCY, Tse H, Tsoi H, Cheng VCC, Lee P, Tang BSF, Cheung CHY, Lee RA, So L, Lau Y, Chan K, and Yuen K. Coronavirus HKU1 and Other Coronavirus Infections in Hong Kong. Journal of clinical microbiology 2006; 44.6: 2063-71.
12. Lee HJ, Shieh CK, Gorbalenya AE, Koonin EV, La Monica N, Tuler J, Bagdzhadzhyan A, and Lai MM. The complete sequence (22 kilobases) of murine coronavirus gene 1 encoding the putative proteases and RNA polymerase. Virology. 1991; 180 (2), 567-582.
13. Schiller JJ., Kanjanahaluethai A. and Baker SC. Processing of the coronavirus MHV-JHM polymerase polyprotein: identification of precursors and proteolytic products spanning 400 kilodaltons of ORF1a. Virology. 1998; 242:2, 288-302.
14. Mouse Hepatitis Virus (MHV) Division of Animal Resources University of Illinois, Urbana, http://www.dar.uiuc.edu/FACT%20Sheets/MHV.pdf#search=%22Mouse%20hepatitis%20virus%22
15. Vijgen L., Keyaerts E., Moes E., Thoelen I., Wollants E., Lemey P., Vandamme AM, and Van Ranst M. Complete genomic sequence of human coronavirus OC43: molecular clock analysis suggests a relatively recent zoonotic coronavirus transmission event. J. Virol. 2005; 79:3, 1595-1604.
16. Harcourt BH, Jukneliene D, Kanjanahaluethai A, Bechill J, Severson KM, Smith CM, Rota PA, Baker SC. Identification of Severe Acute Respiratory Syndrome Coronavirus

Replicase Products and Characterization of Papain-Like Protease Activity. Journal of virology 2004, 78.24: 1360-12.

17. 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 LLM., Peiris JSM., Smith GJD., Chen H., Guan Y. Prevalence and genetic diversity of coronaviruses in bats from China. J. Virol. 2006; 80:15, 7481-7490.

18. 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 and Yuen KY. Comparative analysis of twelve genomes of three novel group 2c and group 2d coronaviruses reveals unique group and subgroup features. J. Virol. 2007; 81:4, 1574-1585.

19. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 1997; 25.24: 4876-82.

20. Lau SK, Woo PC, Li KS, Huang Y, Tsoi HW, Wong BH, Wong SS, Leung SY, Chan KH, Yuen KY. Severe acute respiratory syndrome coronavirus-like virus in Chinese horseshoe bats. Proc Natl Acad Sci U S A. 2005; 102(39):14040-5. Epub 2005 Sep 16.

21. Ratia K, Saikatendu KS, Santarsiero BD, Barretto Naina, Baker SC, Stevens RC, Mesecar AD, Severe Acute Respiratory Syndrome Coronavirus Papain-Like Protease: Structure of a Viral Deubiquitinating Enzyme. Proceedings of the National Academy of Sciences of the United States of America 2006; 103.15: 5717-22.

22. 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 and Wang LF. Bats are natural reservoirs of SARS-like coronaviruses. Science.2005; 310 (5748), 676-679.

23. Vijaykrishna D, Smith GJ, Zhang JX, Peiris JS, Chen H, Guan Y. Evolutionary insights into the ecology of coronaviruses. J Virol. 2007; 81(8):4012-20. Epub 2007 Jan 31.

24. Liu DX, Tibbles KW, Cavanagh D, Brown TD, and Brierley I. Identification, expression, and processing of an 87-kDa polypeptide encoded by ORF 1a of the coronavirus infectious bronchitis virus. Virology. 1995; 208 (1), 48-57.