**Supplementary Material**

**Recurrent mutations in SARS-CoV-2 genomes isolated from mink point to rapid host-adaptation**

Lucy van Dorp1, Cedric CS Tan1, Su Datt Lam2,3, Damien Richard1,4, Christopher Owen1, Dorothea Berchtold1, Christine Orengo3, François Balloux1

1. UCL Genetics Institute, University College London, United Kingdom.
2. Department of Applied Physics, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, Malaysia.
3. Institute of Structural and Molecular Biology, University College London, United Kingdom.
4. UCL Institute of Child Health, University College London, United Kingdom.

Correspondence: [lucy.dorp.12@ucl.ac.uk](mailto:lucy.dorp.12@ucl.ac.uk) (Lucy van Dorp), [f.balloux@ucl.ac.uk](mailto:f.balloux@ucl.ac.uk) (François Balloux)

Contents

Figures S1 – S13

Tables S1-S4 are available as downloadable excel files.

**Chart, histogram

Description automatically generated**

**Figure S1:** (top) Density plot of SNP and homoplasy distributions across the mink SARS-CoV-2 alignment. (bottom) Count (y-axis) of specific mutations identified across the SARS-CoV-2 alignment having masked out sites flagged as putative sequencing errors.

Chart, scatter chart

Description automatically generated

**Figure S2:** (left) SARS-CoV-2 phylogenetic tree of included mink genomes. (right) Root-to-tip (y-axis) phylogenetic distance against sampling date (x-axis) provides a significant temporal regression. *P*-value at top is provided following 10,000 random permutations of the sampling date.

**Chart

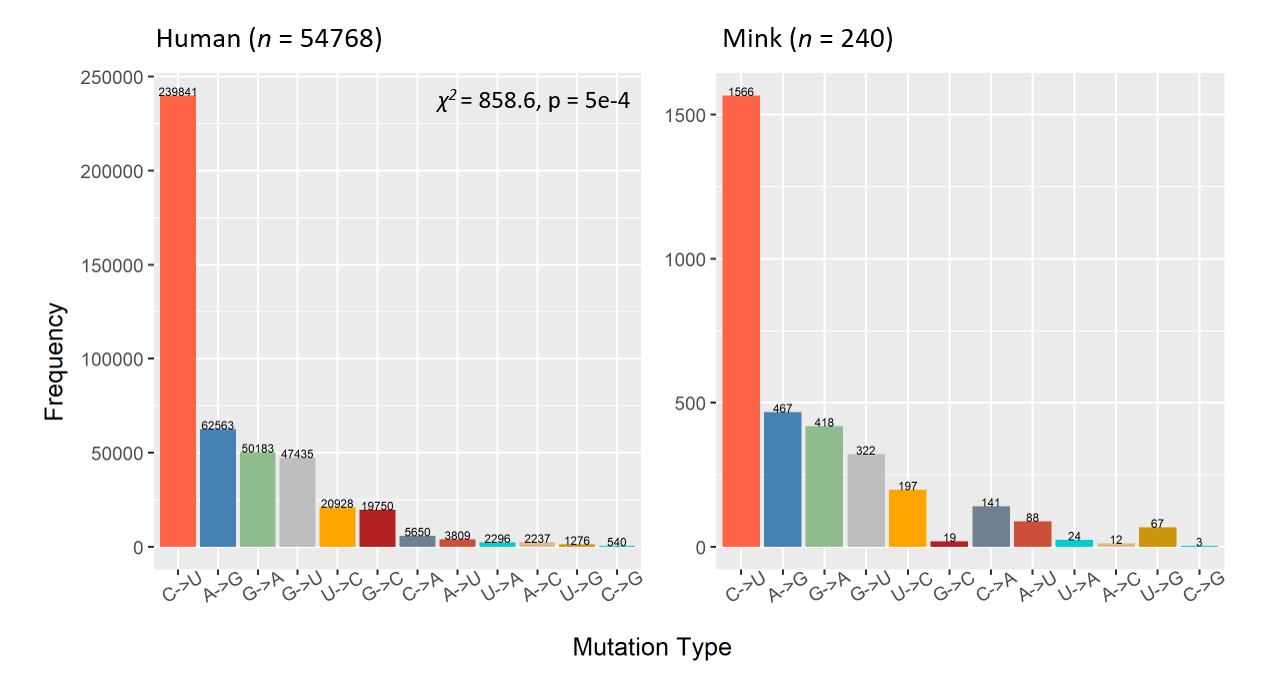
Description automatically generated**

**Figure S3:** Exhaustive list of nonsynonymous mutations and deletions observed in SARS-CoV-2 genomes circulating in minks identified by CoVSurver. Presence is denoted by colour with the colour providing the genomic region as given by the legend (NSP – non-structural protein, NS – non-structural/accessory). Rows are ordered by the phylogeny at left with tips coloured by SARS-CoV-2 lineage assignment.

Chart

Description automatically generated

**Figure S4:** (left) Maximum likelihood phylogenetic tree over the SARS-CoV-2 mink alignment. (right) Heatmap providing all non-synonymous mutations identified in the SARS-CoV-2 spike protein in the masked SARS-CoV-2 mink alignment. The RBD is highlighted together with the presence of D614G which has been claimed to modulate transmissibility in human SARS-CoV-2. Those sites in the RBD identified as highly homoplasic are denoted with an ‘\*’ at top.



**Figure S5**: Proportion of each mutation observed for possible transition states across the SARS-CoV-2 human alignment (left) and mink alignment (right) including Wuhan-Hu-1 and global phylogenetic outliers. Bars of mutations and their corresponding revertant mutation have the same fill colour. A χ2 test with 11 degrees of freedom was used to determine whether the mutational frequencies differed between the two datasets.

Chart

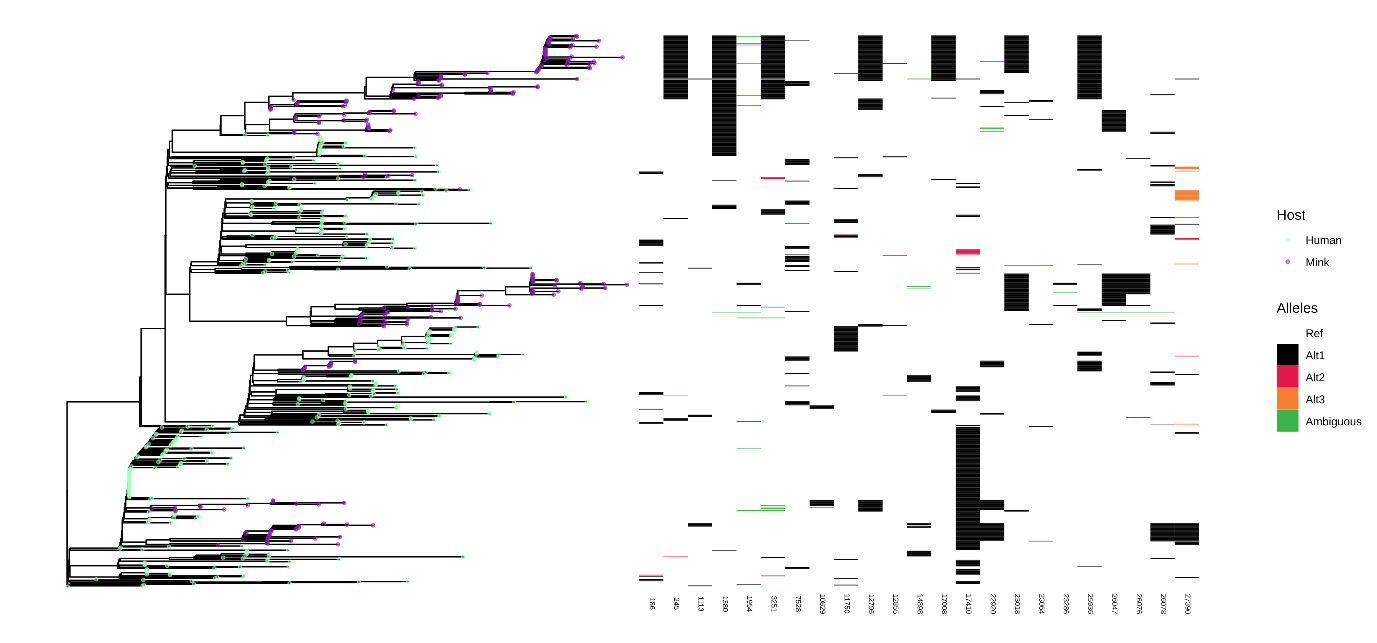
Description automatically generated

**Figure S6**: Raincloud plot (Ref: <https://doi.org/10.12688/wellcomeopenres.15191.1>) (half-violin + scatter plot + boxplot) visualisingCG dinucleotide frequencies of SARS-CoV-2 in mink (top; blue) and human (bottom; orange) hosts.A Wilcoxon rank sum test was used to determine whether the two distributions differed.

Chart, timeline

Description automatically generated

**Figure S7:** Distribution of homoplasic positions over the SARS-CoV-2 mink alignment. Mutations responsible for more than two changes across the maximum likelihood phylogeny are annotated by nucleotide position. Coloured bars at top provide genome annotation relative to Wuahan-Hu-1. Reported consistency indices and associated annotations are provided in **Table S4**.



**Figure S8:** Left panel: phylogenetic tree of SARS-CoV-2 genomes circulating in minks (n=237) together with human SARS-CoV-2 strains carrying derived alleles at nucleotide sites homoplasic in minks (n=426, as of 25/8/2020). Right panel: allelic variants at nucleotide sites identified as homoplasic in the SARS-CoV-2 mink genome dataset. Genome alignment and phylogenetic reconstruction was performed as explained in the *Alignment of mink SARS-CoV-2* paragraph of the *Materials and Methods* main text section.

Table

Description automatically generated

**Figure S9**: HADDOCK predicted energies for the S-protein:ACE2 complex. More negative values indicate stronger binding energy of the complex.

Table

Description automatically generated

**Figure S10:** mCSM-PPI2 predicted changes in complex stabilities. Negative DDG values are associated with destabilisation of the complex following mutation of the residue and positive values with stabilisation of the complex.

Diagram

Description automatically generated

**Figure S11:** Structural features of the S-protein from SARS-CoV-2 and SARS-CoV-2 bound to ACE2. (a) human ACE2 with Wuhan-Hu-1 reference (b) mink ACE2 with a mink associated SARS-CoV-2 S-protein (Y453F mutation in RBD). Dotted lines indicate residues <= 4A apart.

Diagram

Description automatically generated

**Figure S12:** Structural features of the S-protein from SARS-CoV-2 and SARS-CoV-2 bound to ACE2. (a) human ACE2 with Wuhan-Hu-1 reference (b) mink ACE2 with a mink associated SARS-CoV-2 S-protein (F486L mutation in RBD). Dotted lines indicate residues <= 4A apart.

A picture containing text

Description automatically generated

**Figure S13:** Structural features of the S-protein from SARS-CoV-2 and SARS-CoV-2 bound to ACE2. (a) human ACE2 with Wuhan-Hu-1 reference (b) mink ACE2 with a mink associated SARS-CoV-2 S-protein (N501T mutation in RBD). Dotted lines indicate residues <= 4A apart.