**Intra-host dynamic variations in SARS-CoV-2**

Jiarui Li[1,2#], Pengcheng Du[1,2#], Lijiang Yang[3,4#], Ju Zhang[1,2#], Chuan Song[1,2#], Danying Chen[1,2#], Yangzi Song[1,2], Nan Ding[1,2], Mingxi Hua[1,2], Kai Han[1,2], Rui Song[1], Wen Xie[1], Zhihai Chen[1], Xianbo Wang[1], Jingyuan Liu[1], Yanli Xu[1], Guiju Gao[1], Qi Wang[1], Lin Pu[1], Lin Di[4,5,6], Jie Li,[7] Jinglin Yue[8], Junyan Han[1,2], Xuesen Zhao[1,2],Yonghong Yan[1,2], Fengting Yu[1], Angela R. Wu[9], Fujie Zhang[1*], Yi Qin Gao[3,4,10*], Yanyi Huang[3,4,5,6,11*], Jianbin Wang[7,11*], Hui Zeng[1,2*], Chen Chen[1,2*]

1. Beijing Ditan Hospital, Capital Medical University, Beijing 100015, P. R. China
2. Beijing Key Laboratory of Emerging Infectious Diseases, Beijing 100015, P. R. China
3. Beijing National Laboratory for Molecular Sciences, College of Chemistry and Molecular Engineering, Peking University, Beijing 100871, China.
4. Beijing Advanced Innovation Center for Genomics, Biomedical Pioneering Innovation Center, Peking University, Beijing 100871, China.
5. School of Life Sciences, Peking-Tsinghua Center for Life Sciences, Peking University, Beijing 100871, China.
6. Institute for Cell Analysis, Shenzhen Bay Laboratory, Shenzhen 518055, China
7. School of Life Sciences, Tsinghua-Peking Center for Life Sciences, Beijing Advanced Innovation Center for Structural Biology, Tsinghua University, Beijing 100084, China
8. Peking University Ditan Teaching Hospital, Beijing 100015, P. R. China
9. Division of Life Science and Department of Chemical and Biological Engineering, Hong Kong University of Science and Technology, Hong Kong SAR, P.R. China
10. Chinese Center for Disease Control and Prevention, Beijing, P. R. China
11. Institute of Systems and Physical Biology, Shenzhen Bay Laboratory, Shenzhen 518055, China
12. Chinese Institute for Brain Research, Beijing 102206, China

[#]These authors contributed equally to this work.

*Corresponding authors:

Prof. Chen Chen,

Institute of Infectious Diseases, Beijing Ditan Hospital, Capital Medical University and Beijing Key Laboratory of Emerging Infectious Diseases

No.8 Jingshundongjie, Beijing 100015, China

E-mail: chenchen1@ccmu.edu.cn Tel.: 00861084322612 Fax: 00861084322606

Prof. Hui Zeng,

Institute of Infectious Diseases, Beijing Ditan Hospital, Capital Medical University and Beijing Key Laboratory of Emerging Infectious Diseases

No.8 Jingshundongjie, Beijing 100015, China

E-mail: zenghui@ccmu.edu.cn Tel.: 00861084322621 Fax: 00861084322606

Prof. Jianbin Wang

School of Life Sciences, Tsinghua-Peking Center for Life Sciences, Beijing Advanced Innovation Center for Structural Biology

Tsinghua University, Beijing 100084, P. R. China

Email: jianbinwang@tsinghua.edu.cn

Prof. Yanyi Huang

Biomedical Pioneering Innovation Center, Beijing Advanced Innovation Center for Genomics, College of Chemistry and Molecular Engineering, Peking-Tsinghua Center for Life Sciences

Peking University, Beijing 100871, P. R. China

Email: yanyi@pku.edu.cn

Prof. Yiqin Gao

Beijing National Laboratory for Molecular Sciences, College of Chemistry and Molecular Engineering, Beijing Advanced Innovation Center for Genomics, Biomedical Pioneering Innovation Center, Peking University, Beijing 100871, China.

Email: gaoyq@pku.edu.cn

Prof. Fujie Zhang

Institute of Infectious Diseases, Beijing Ditan Hospital, Capital Medical University

No.8 Jingshundongjie, Beijing 100015, China

E-mail: treatment@chinaaids.cn

**This PDF file includes:**

**Supplementary figures and tables**

Figures S1 to S6

Table S1 to S5

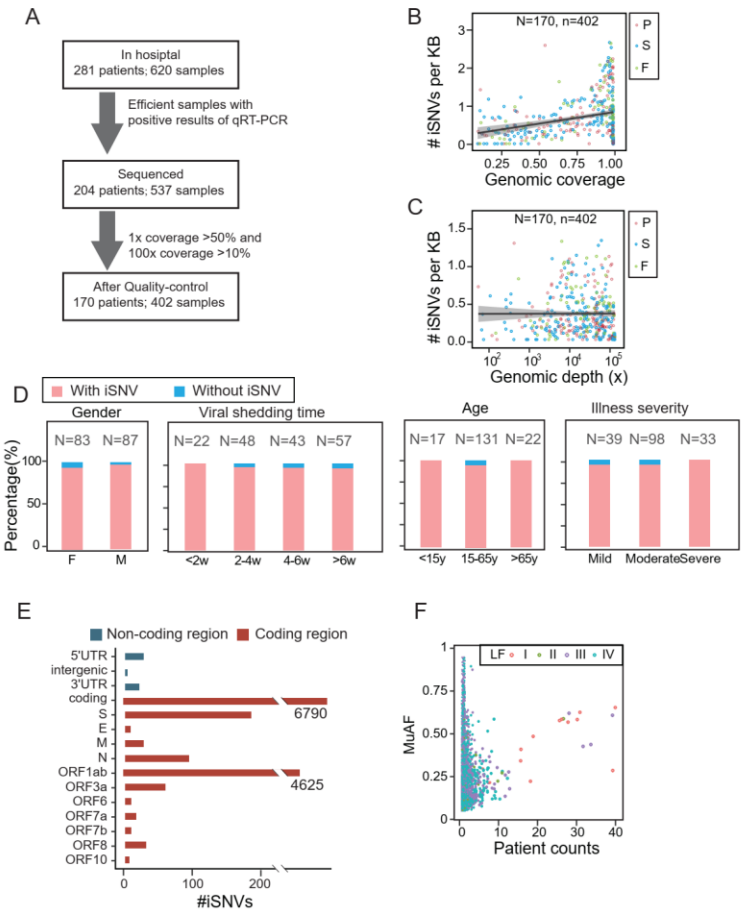


**Figure S1.** The pipeline of samples selection and iSNV distribution along SARS-Cov-2 genome in patients. (A) The pipeline of patients and sample selected for sequencing and quality control. (B) The number of iSNVs per kb against the genomic coverage with a linear regression. (C) The number of iSNVs per kb against the sequencing depth with a linear regression. (D) The numbers of patients grouped by gender, age, viral shedding time and illness severity, and the proportion of patients with or without iSNV in each group. In the gender group, F is short for female and M is short for male. (E) The iSNV counts in coding (red) and non-coding regions (dark blue). (F) The MuAF of iSNVs against the number of patients with the iSNVs. The average MuAF was used if the iSNV was shared in patients. The color of the point represents the level of SNP frequency in public database reported previously, Level I to IV, the SNP frequency from high to low.

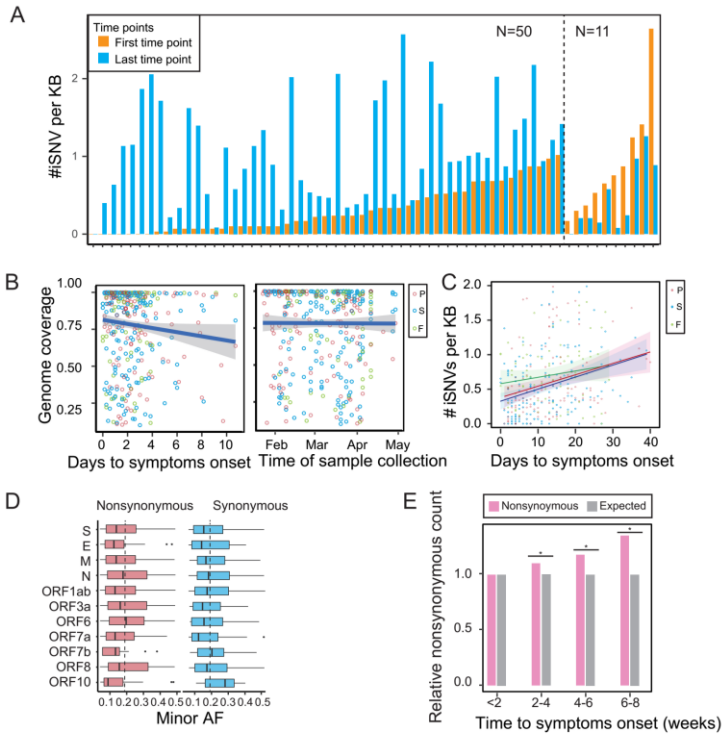


Figure S2. The iSNV distribution along the symptoms onset and genes. (A) The iSNV per KB of the same patients in the first and last time point. The duration of first and last time is more than 5 days. (B) The correlation of iSNV minor allele frequency (minor AF) and genomic coverage with the days post symptom onset and the time of sample collection, respectively, using the samples with more than 100x sequencing depth and linear regression were calculated. (C) The distribution of iSNV per KB against the days post symptom onset, and linear regression were calculated for three sample types. (D) The minor allele frequency (minor AF) of iSNVs causing nonsynonymous (left) and synonymous (right) mutations in each gene. The average frequency in each type of iSNVs was marked by dashed vertical line. (E) The relative nonsynonymous count in epitope regions against time to symptoms onset. The nonsynonymous mutation was colored by pink and expected was colored by grey. The relative nonsynonymous were normalized by the whole epitope region and the expected value was set to 1.

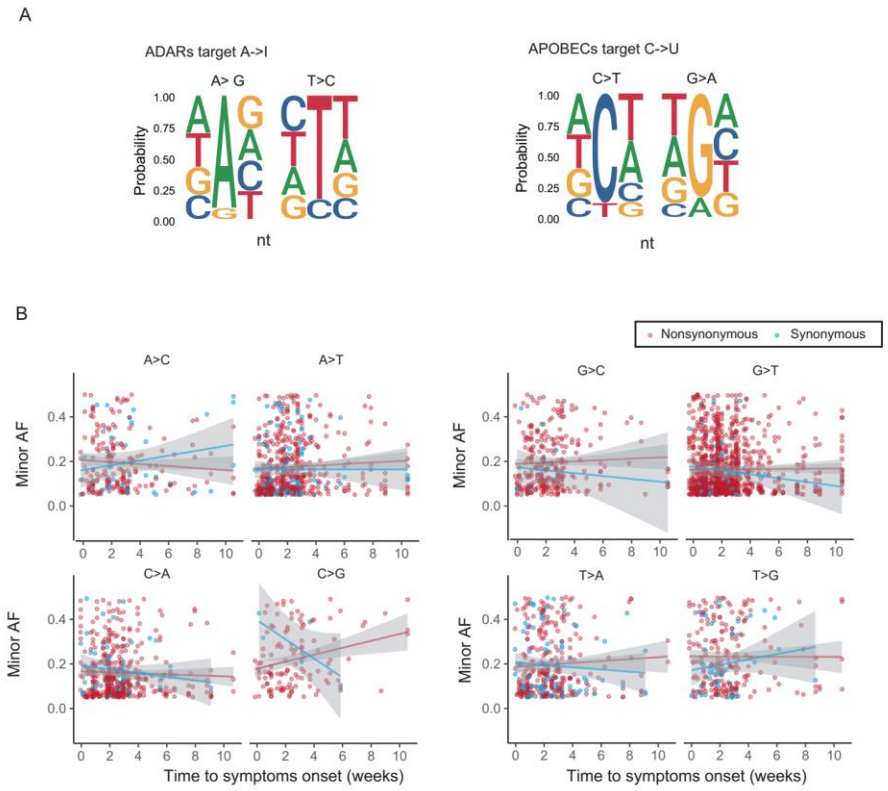


Figure S3. The nucleotide changes of iSNVs and distribution along symptoms onset time. (A) The nucleotide sequence context for the ADARs target (A -> I) and APOBECs target (C -> U). (B) The minor AF of different nucleotide change against the time to symptoms onset of patients. The mutations causing nonsynonymous and synonymous mutations were distinguished by color (red: nonsynonymous mutations, blue: synonymous mutations).

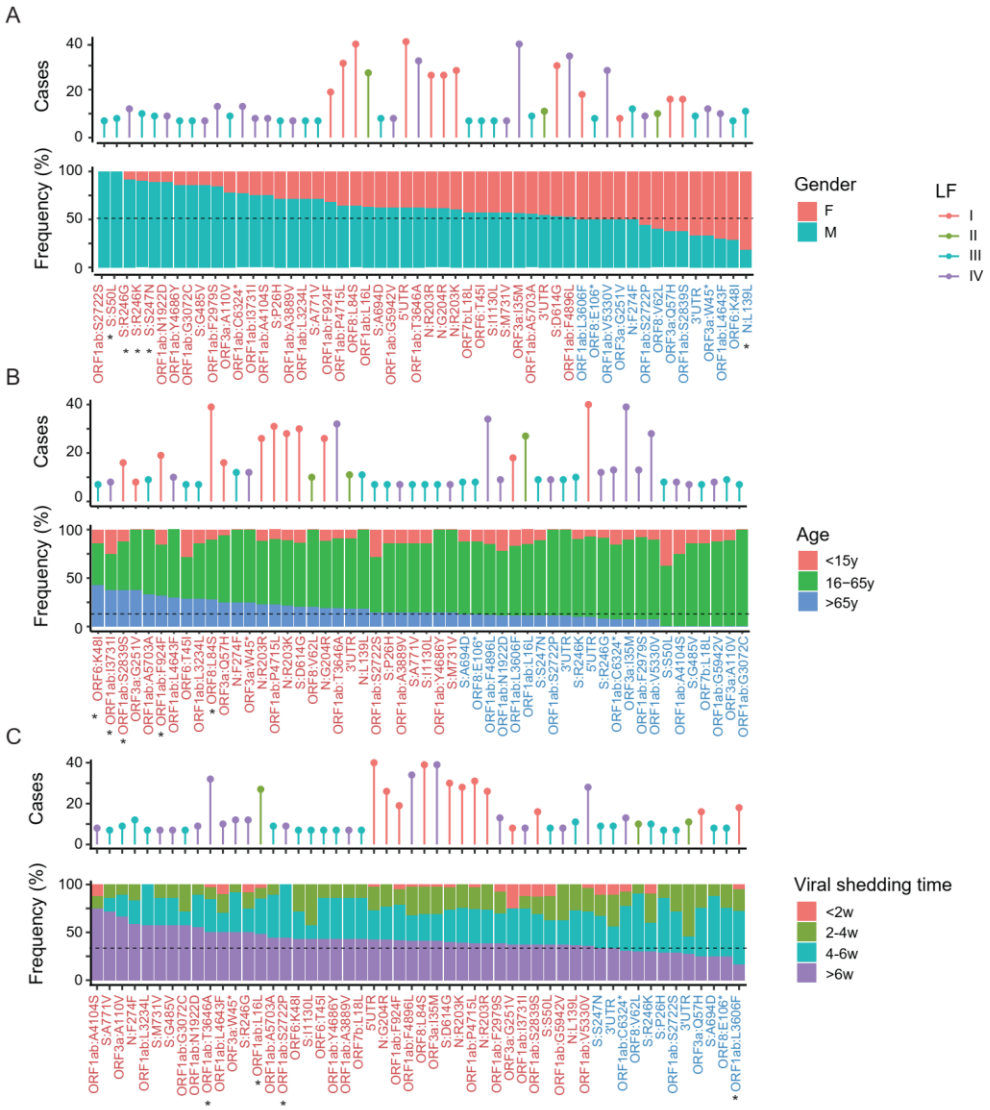


Figure S4. The association of iSNVs with gender, age and viral shedding time of patients. Distribution of iSNVs in patients grouped by gender, age and viral shedding time. The point plots on the top mark the level of SNPs in public database 2019nCovR corresponding to the iSNV site. The middle color titles represent the gene region that the iSNV located. The histogram at the bottom shows the proportion of patients with different clinical states that carrying the iSNV site. The iSNVs marked with star represent that the population carrying this iSNV was significantly differed from the whole patient population. The bars in the lower bar diagram were colored by gender (A), age (B), and viral shedding time (C), respectively.

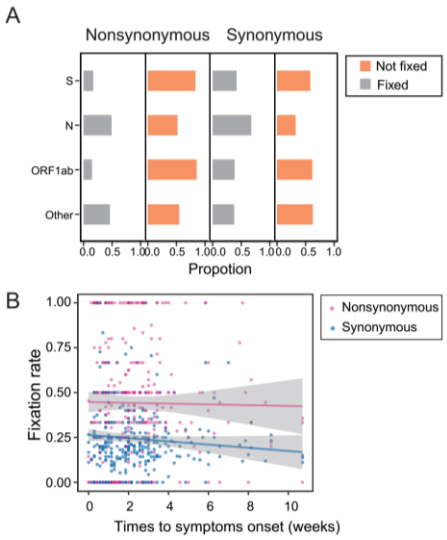


Figure S5. The fixation of iSNVs in different gene and symptoms onset time. (A) The proportion of fixed and not fixed iSNVs causing nonsynonymous and synonymous mutation in each gene. (B) The fixation rate of nonsynonymous and synonymous mutation against the time post symptom onset.

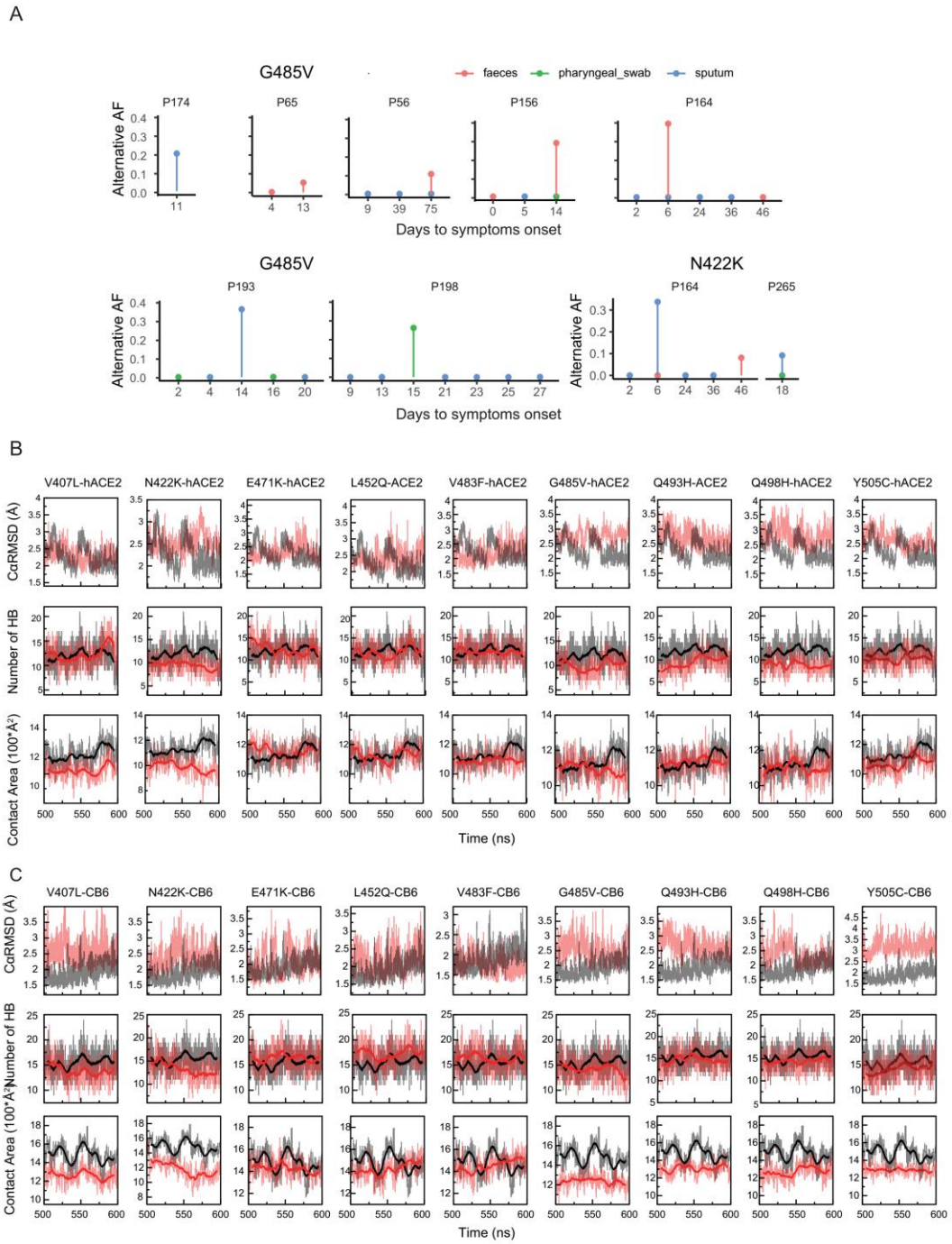


Figure S6. The iSNVs occurs in the S genes. (A) The iSNVs that occurred in different patients along the symptoms onset. The y-axis represents the mutated allele frequency. The color of the point represents samples from pharyngeal swabs, sputum and fecal. (B) The RMSD comparisons, number of hydrogen bond and contact area of the mutants in RBD region compared to WT RBD bound to hACE2. The mutants were colored by red and WT was colored by grey. (D) The RMSD

comparisons, number of hydrogen bond and contact area of the mutants in RBD region compared to WT RBD bound to CB6. The mutants were colored by red and WT was colored by grey.

**Table S1** Clinical information for patients included in the study

| **Characteristics** | **In hospital patients No.** | **Sequenced patients No.** | **QC patients No.(%)** |
|---|---|---|---|
| Total | 281 | 204 | 170(83.33%) |
| Age(years) | | | |
| <15 | 26 | 22 | 17(77.27%) |
| 15-65 | 226 | 158 | 131(82.91%) |
| >65 | 29 | 24 | 22(91.67%) |
| Gender | | | |
| Female | 143 | 102 | 83(81.37%) |
| Male | 138 | 102 | 87(85.29%) |
| Illness severity | | | |
| Mild (M) | 61 | 47 | 39(82.98%) |
| Moderate (N) | 167 | 117 | 98(83.76%) |
| Severe (S) | 53 | 40 | 33(82.50%) |
| Viral shedding time(week) | | | |
| 0-2 | 80 | 36 | 22(61.11%) |
| 2-4 | 70 | 54 | 48(88.89%) |
| 4-6 | 63 | 51 | 43(84.31%) |
| >6 | 68 | 63 | 57(90.48%) |

**Table S2** Single nucleotide substitution rate and iSNV identified in other RNA virus

| | **SNP ($\times10^{-3}$/site/year)** | **iSNV($\times10^{-3}$/site )** | **Reference** |
|---|---|---|---|
| SARS-CoV-2 | 0.35–4.67 | 0.53(this paper) | [1,2] |
| Ebola virus | 0.8–1.9 | ~0.5 | [3-6] |
| Influenza A virus | 1.43–11.62 | 0.43 | [7-9] |
| Yellow fever virus | ~0.42 | 0.44 | [10,11] |

**Table S3** Samples information that after quality control

| Characteristics | Sample Count No.(%) | iSNV count median (qu1-qu3) | Wilcoxon test P-value | # of normalized iSNV median (qu1-qu3) | Wilcoxon test P-value | #of samples with iSNV (%) | Fisher-exact test P-value |
|---|---|---|---|---|---|---|---|
| Total | 402 | | | | | | |
| Days to symptoms onset | | | | | | | |
| 0-7 | 128(31.84%) | 7(2-18.5) | - | 0.331(0.100-0.688) | - | 112(87.5%) | - |
| 8-14 | 104(25.87%) | 11(4-25) | **0.011** | 0.541(0.294-0.972) | **5.74e-4** | 99(95.19%) | 0.0637 |
| 15-21 | 81(20.15%) | 15(5-29) | **6.05e-5** | 0.684(0.334-1.133) | **1.79e-6** | 79(97.53%) | **0.0112** |
| 22-28 | 45(11.19%) | 11(5-31) | **0.006** | 0.513(0.343-1.111) | **1.23e-3** | 42(93.33%) | 0.408 |
| 29-42 | 22(5.47%) | 20.5(11.25-37.75) | **0.0001** | 0.933(0.645-1.423) | **4.00e-6** | 22(100%) | 0.130 |
| 43- | 22(5.47%) | 22.5(6-32.25) | **0.015** | 0.836(0.240-1.232) | **5.38e-3** | 20(90.91%) | 1 |
| Specimens | | | | | | | |
| Pharyngeal swab | 136(33.83%) | 10(4-22.25) | 0.791 | 0.439(0.235-0.888) | 0.982 | 128(94.12%) | 0.133 |
| sputum | 182(45.27%) | 11(3-24) | - | 0.514(0.197-0.905) | - | 165(90.66%) | - |
| faces | 84(20.89%) | 21(5.75-30.25) | **0.001** | 0.771(0.355-1.076) | **0.004** | 81(96.43%) | 0.297 |

**Table S4** Patients clinical information that carrying iSNVs located in S protein RBD domain

| Patient ID | Gender | Age | Viral shedding time | Illness severity | Clustering onset groups | iSNV (Days to symptoms onset) |
|---|---|---|---|---|---|---|
| P13 | M | 33 | 63 | Moderate | - | V483F (62) |
| P56 | M | 59 | 75 | Moderate | F4 | G485V (75) |
| P65 | M | 6 | 42 | Mild | F6 | G485V (13) |
| P103 | M | 78 | 35 | Severe | - | E471K (27) |
| P156 | M | 42 | 48 | Mild | - | G485V (14) |
| P164 | M | 50 | 52 | Moderate | F1 | V483F (6, 36, 46); G485V (6) |
| P174 | M | 30 | 23 | Moderate | F1 | G485V (11); Q493H (11) |
| P165 | M | 33 | 55 | Moderate | F1 | L452Q (6) |
| P183 | M | 49 | 43 | Moderate | - | Q498H (24) |
| P193 | M | 41 | 48 | Moderate | - | G485V (14) |
| P198 | F | 21 | 37 | Moderate | - | V483F (23); G485V (15); Q498H (21) |
| P201 | F | 76 | 23 | Severe | - | Q493H (5) |
| P215 | F | 20 | 31 | Moderate | - | E471K (13); Q493H (16) |
| P230 | M | 24 | 58 | Moderate | - | Y505C (6, 16) |
| P235 | F | 25 | 10 | Moderate | F2 | L452Q (3) |
| P247 | F | 23 | 51 | Mild | - | Y505C (21) |
| P279 | F | 44 | 33 | Moderate | F3 | Q498H (16) |

**Table S5. Neutralizaing susceptibility of SARS-CoV-2 reference strains and RBD mutants to CB6 mAb**

| RBD mutant | CB6 $IC_{50}$ (μg/mL) | Fold change vs. WT |
|---|---|---|
| WT | 0.009521 | 1.00 |
| D614G | 0.009122 | 0.96 |
| V407L | 0.01194 | 1.25 |
| L452Q | 0.00853 | 0.90 |
| V483F | 0.00561 | 0.59 |
| Q493H | 0.00412 | 0.43 |
| Q498H | 0.00371 | 0.39 |

## Reference


1 Tang, X. *et al.* On the origin and continuing evolution of SARS-CoV-2. (2020).

2 Zhang, X. *et al.* Viral and host factors related to the clinical outcome of COVID-19. *Nature* **583**, 437-440, doi:10.1038/s41586-020-2355-0 (2020).

3 Park, D. J. *et al.* Ebola Virus Epidemiology, Transmission, and Evolution during Seven Months in Sierra Leone. *Cell* **161**, 1516-1526, doi:10.1016/j.cell.2015.06.007 (2015).

4 Gire, S. K. *et al.* Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak. *Science* **345**, 1369-1372, doi:10.1126/science.1259657 (2014).

5 Holmes, E. C., Dudas, G., Rambaut, A. & Andersen, K. G. The evolution of Ebola virus: Insights from the 2013-2016 epidemic. *Nature* **538**, 193-200, doi:10.1038/nature19790 (2016).

6 Ni, M. *et al.* Intra-host dynamics of Ebola virus during 2014. *Nat Microbiol* **1**, 16151, doi:10.1038/nmicrobiol.2016.151 (2016).

7 Rejmanek, D., Hosseini, P. R., Mazet, J. A., Daszak, P. & Goldstein, T. Evolutionary Dynamics and Global Diversity of Influenza A Virus. *J Virol* **89**, 10993-11001, doi:10.1128/JVI.01573-15 (2015).

8 Bedford, T., Cobey, S., Beerli, P. & Pascual, M. Global migration dynamics underlie evolution and persistence of human influenza A (H3N2). *PLoS Pathog* **6**, e1000918, doi:10.1371/journal.ppat.1000918 (2010).

9 Debbink, K. *et al.* Vaccination has minimal impact on the intrahost diversity of H3N2 influenza viruses. *PLoS Pathog* **13**, e1006194, doi:10.1371/journal.ppat.1006194 (2017).

10 Liang, H., Lee, M. & Jin, X. Guiding dengue vaccine development using knowledge gained from the success of the yellow fever vaccine. *Cell Mol Immunol* **13**, 36-46, doi:10.1038/cmi.2015.76 (2016).

11 Chen, C. *et al.* Phylogenomic analysis unravels evolution of yellow fever virus within hosts. *PLoS Negl Trop Dis* **12**, e0006738, doi:10.1371/journal.pntd.0006738 (2018).