**Virologica Sinica**

**Supplementary Data**

**RNA barcode segments for SARS-CoV-2 identification from HCoVs andSARSr-CoV-2 lineages**

Changqiao You a, 1, Shuai Jiang a, 1, Yunyun Ding a, 1, Shunxing Ye b, Xiaoxiao Zou a, Hongming Zhang a,

Zeqi Li a, Fenglin Chen a, Yongliang Li a, \*, Xingyi Ge a, \*, Xinhong Guo a, \*

a *College of Biology, Hunan University, Changsha, 410082, China*

b *College of Bioscience and Biotechnology, Hunan Agricultural University, Changsha, 410128, China*

\*Corresponding authors.

*E-mail addresses*: hnugxh@163.com (X. Guo), xyge@hnu.edu.cn (X. Ge), yongliangli1357@163.com (Y. Li).

ORCID: 0000-0002-8125-1906 (X. Guo), 0000-0003-3964-5140 (X. Ge), 0009-0009-7570-7880 (Y. Li).

1Changqiao You, Shuai Jiang and Yunyun Ding contributed equally to this work.

**Appendix A. Supplementary data**

**Supplementary Fig. S1** The length and relative position of conserved regions in SARS-CoV-2.

**Supplementary Fig. S2** Correlation of parameters (length, Blast scores and *P* value) of barcode segments.

**Supplementary Fig. S3** The visual dynamic 2D barcode for subordinate barcode segments. **A** The 2D barcode. People obtain 1D barcodes and basic information about barcode segments by scanning. **B** The basic information screenshot of barcode segments (the top five) with 2D barcode scanning. The barcode segments use the standard representation of DNA sequences, which is denoted as ACGT.

**Supplementary Table S1.xlsx** Accession and version numbers of sequences in training sets on NCBI.

**Supplementary Table S2.xlsx** Detailed test set results for barcode segments. The green-colored cells in the table represent the data of the main barcode segments.

**Supplementary Table S3.xlsx** The nucleotide polymorphism and genetic features of SDs.

**Supplementary Table S4.xlsx** The basic information of 75 barcode segments.

**Supplementary File 1.bat** The script for batching file format (file extension) conversion.

**Supplementary File 2.bat** The script for the merge of various files.

**Supplementary File 3.py** The Python code for the bulk export of feature string (simple version).

**Supplementary File 4.py** The Python code for the bulk export of feature string with an ambiguity range (extended version).

**Supplementary File 5.py** The Python code for the replacement of sequence names and the deletion of degenerate bases in SDs.

**Supplementary File 6.py** The Python code for batch amending conserved region names from the resulting files of DnaSP6 “Conserved Regions” function.

**Supplementary File 7.py** The Python code for analyzing the characteristics of conserved regions from the resulting files of DNAsp6 “Conserved Regions” function.