

Supplementary figures and tables list:

Fig S1. BHS and BTHS scores identified from our customized SARS-CoV-2 genome database.
Fig S2. BLAST hit ratios of bat and pangolin coronavirus genomes by the SARS-CoV-2 sequences.
Fig S3. The BHS comparison of virus samples between USA and China in the cluster of 4I.C1.
Fig S4. The BHS comparison of virus samples between USA and China in the cluster of 4I.C2.
Fig S5 Dynamic S/ACE2 binding identified for the alignment in the cluster of 4I.C1 by protein docking.

Table S1. Dataset_summary.xlsx

Table S2. Statistics of infection cases and death by 04222020.xlsx

Table S3. Sample collection date.xlsx

Table S4. BHS for 4I.C1.xlsx

Table S5. BHS for 4I.C2.xlsx

Table S6. USA genome MT334524 loci hit by 4I.C1.txt

Table S7. USA genome MT345850 loci hit by 4I.C1.txt

Table S8. USA genome MT345852 loci hit by 4I.C1.txt

Table S9. USA genome MT334562 loci hit by 4I.C1.txt

Table S10. NCBI betacoronavirus country list.txt

Table S11. Customized database.xlsx

Table S12. GISAID list in our customized database.xlsx

Supplementary Tables are available at: <https://github.com/guangxujin/COVID-19/tree/master/Supplementary%20Tables>

Fig S1_Blast_hits of raw sequences using customized betacoronavirus database.pdf

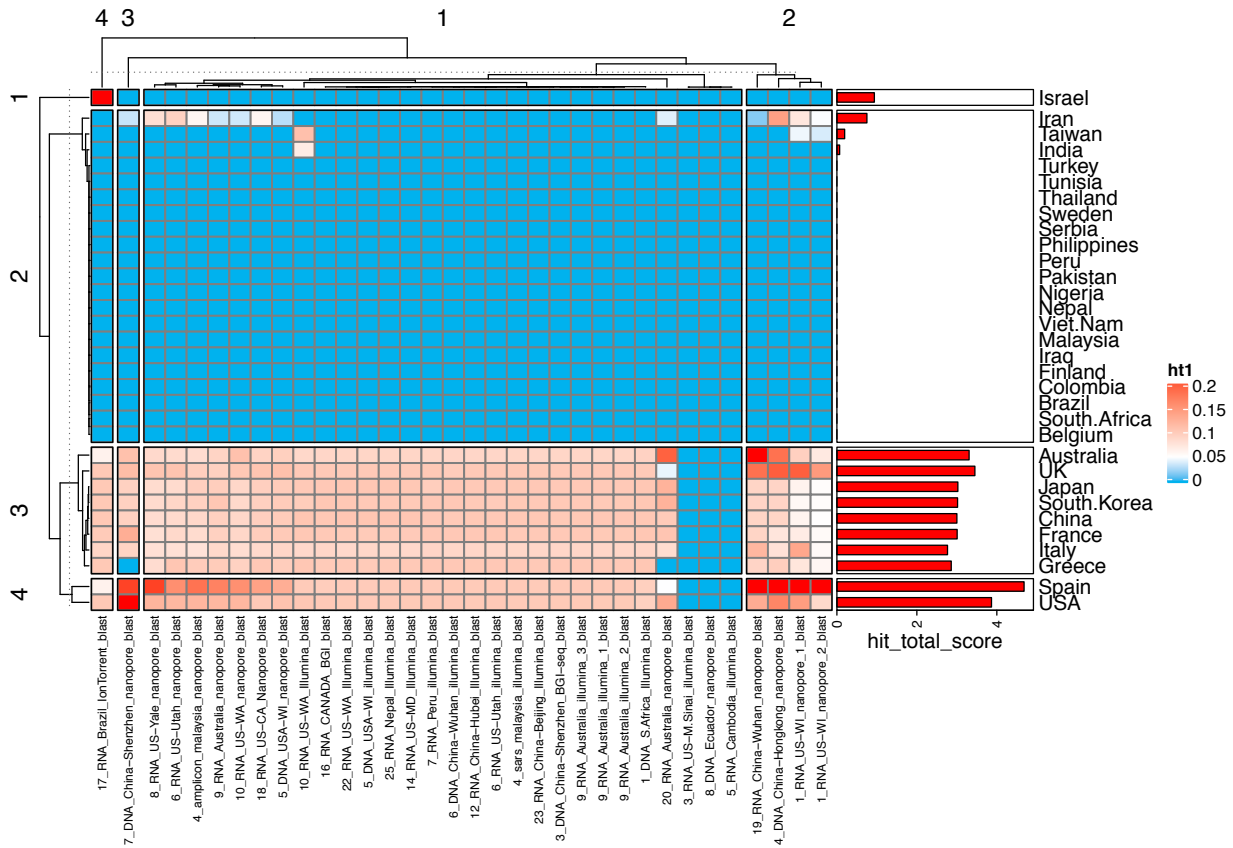


Fig S1 Hub viruses revealed by the local genomic similarity mapping. The local genomic similarity mapping by the BHS values of the BLAST alignments between the raw sequencing reads from each dataset and the genome sequence in the searching database of customized Betacoronavirus database.

Fig S2_the bat_pangolin_blast.pdf

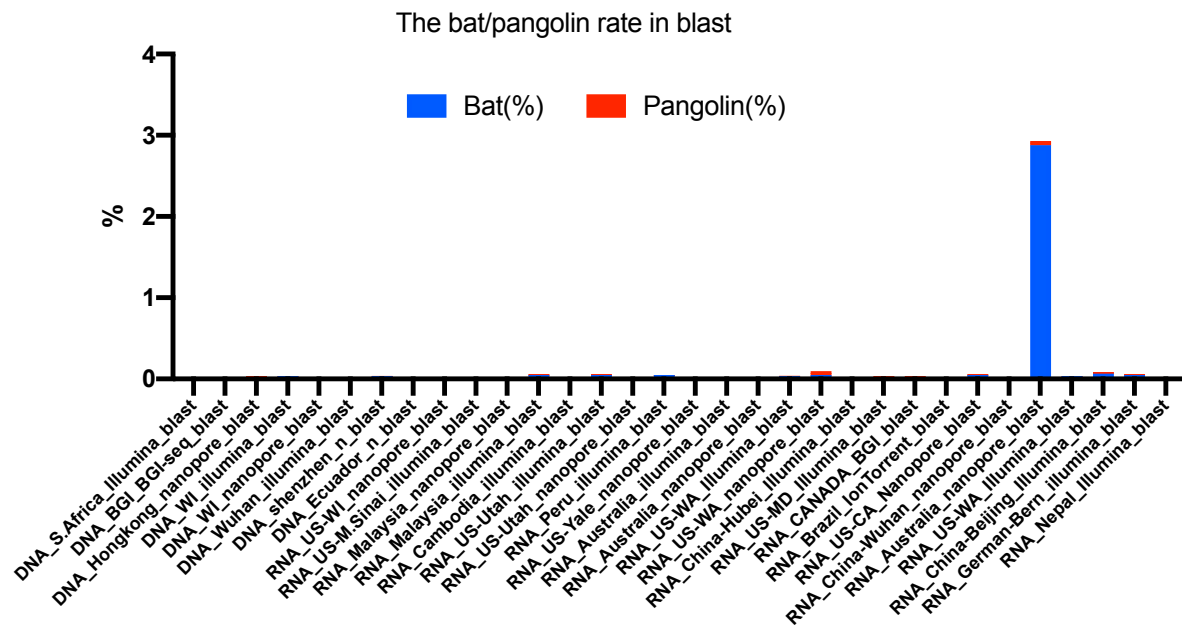


Fig. S2 The percentage of Bat and Pangolin BLAST hits. The sample sequences were aligned to the NCBI assembled Betacoronavirus database and the percentage of sequences matched to the Bat or Pangolin genome was calculated. The percentage of all but one sample (RNA_Australia_nanopore_blast) is low than 0.1 %.

Fig S3

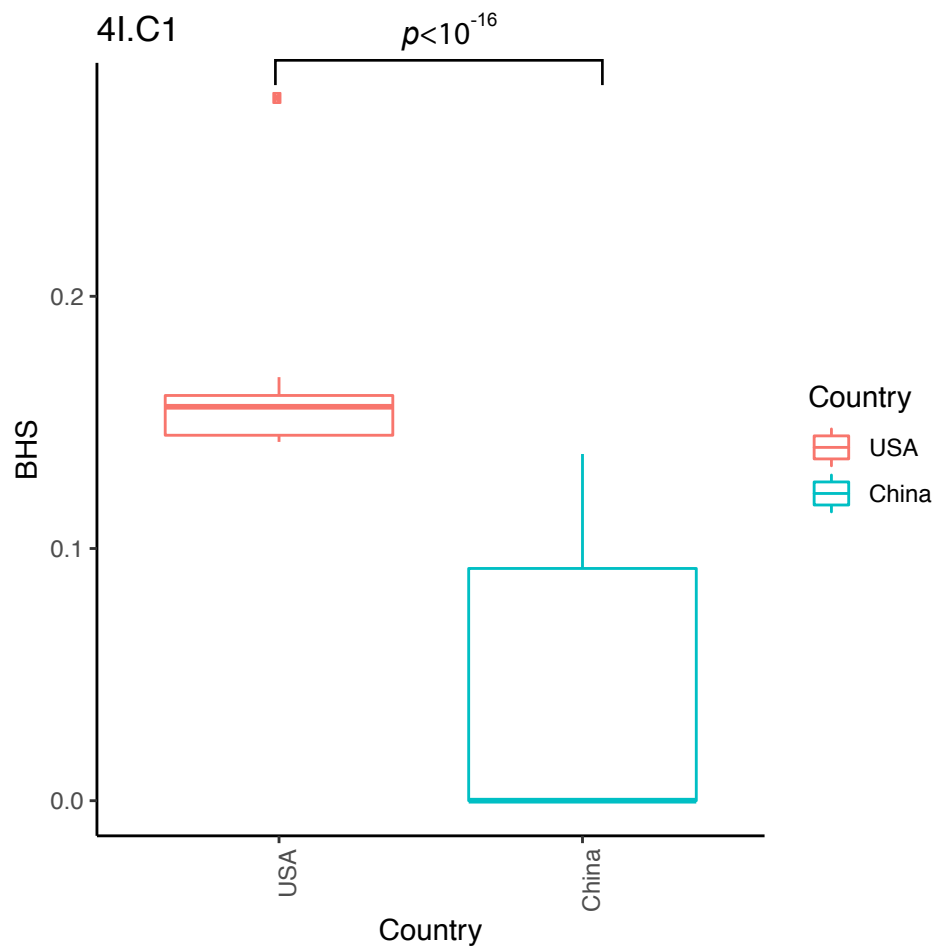


Fig S3 The BHS comparison of virus samples between USA and China in the cluster of 4I.C1. USA has higher BHS scores for early SARS-CoV-2 viruses from China. ($p < 2.3e-16$, Welch Two Sample t-test, two- sided, $t = -13.13$, $df = 41.55$).

Fig S4

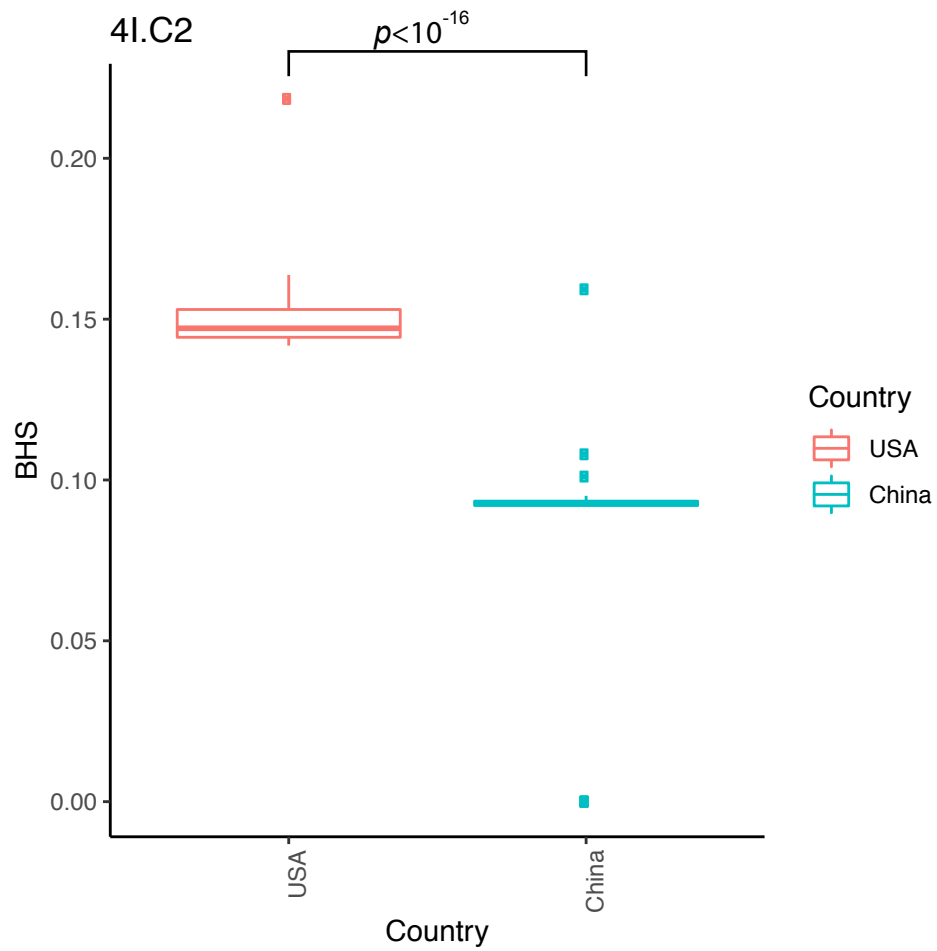
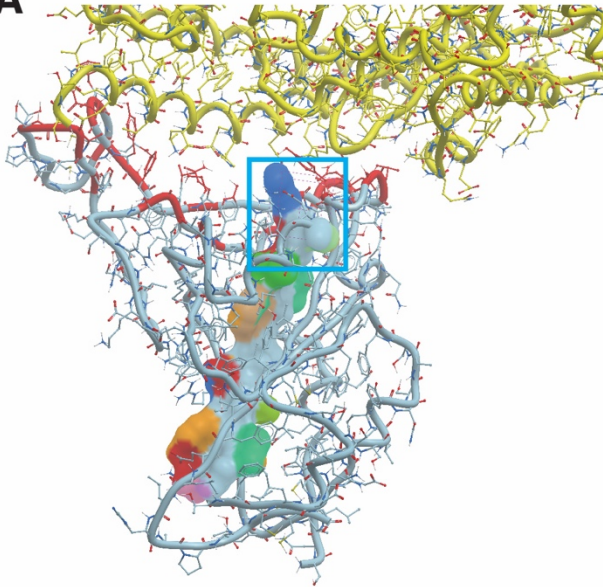


Fig S4 The BHS comparison of virus samples between USA and China in the cluster of 4I.C2. USA have higher BHS scores for early SARS-CoV-2 viruses from China and USA. ($p < 2.2e-16$, Welch Two Sample t-test, two-sided, $t = -12.05$, $df = 49.95$).

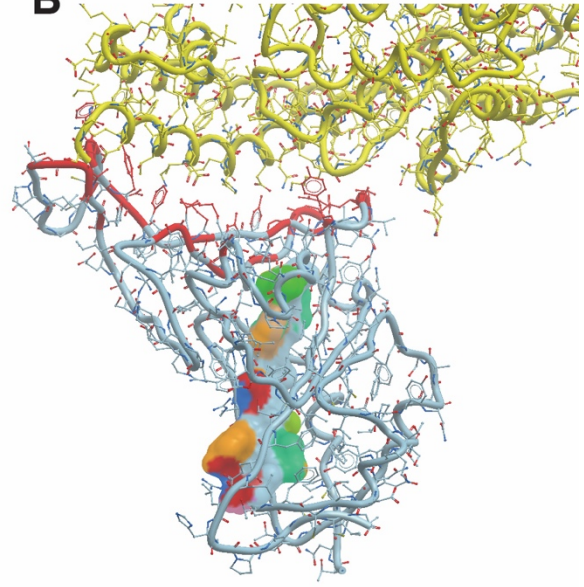
Fig S5

A



Ligand rmsd (Å) = 0.90

B



Ligand rmsd (Å) = 1.64

Fig S5 Dynamic S/ACE2 binding identified by protein docking. **A.** Predicted binding of S protein with ACE2. Meshes are for the backbone from N394 to G404. **B.** Predicted binding of S protein with a deletion and ACE2. Meshes are for the backbone from N394 to I402. Red color: the amino acids in RBM close to binding region.