

Supplementary Information

Analysis of SARS-CoV-2 mutations in the United States suggests presence of four substrains and novel variants

Rui Wang¹, Jiahui Chen¹, Kaifu Gao¹, Yuta Hozumi¹, Changchuan Yin², and Guo-Wei Wei^{1,3,4*}

¹ Department of Mathematics,

Michigan State University, MI 48824, USA.

² Department of Mathematics, Statistics, and Computer Science,
University of Illinois at Chicago, Chicago, IL 60607, USA

³ Department of Electrical and Computer Engineering,
Michigan State University, MI 48824, USA.

⁴ Department of Biochemistry and Molecular Biology,
Michigan State University, MI 48824, USA.

October 17, 2020

Contents

S1 Supplementary Tables	1
S2 Supplementary Data	1
S3 Supplementary Figures	2
S3.1 Mutation Tracker	2
S3.2 <i>K</i> -Means clustering	2
S3.3 Proteoforms	3
S3.4 Protein structures, network analysis, and alignment	4

*Corresponding author. E-mail: weig@msu.edu

S1 Supplementary Tables

Table S1 shows the cluster distributions of samples (N_{NS}), total SNP counts (N_{TF}), and the ratios of SNPs ($R_{TF} = N_{TF}/7123$) in 20 states of the U.S. The 20 states are Alaska (AK), Arizona (AZ), California (CA), Connecticut (CT), Washington, D.C. (DC), Florida (FL), Louisiana (LA), Maine (ME), Maryland (MD), Michigan (MI), Minnesota (MN), Nevada (NV), New Mexico (NM), New York (NY), Oregon (OR), Texas (TX), Utah (UT), Virginia (VA), Washington (WA), and Wisconsin (WI).

Table S2 lists the number of samples (N_{NS}) of a particular mutation on 5'UTR in the United States and worldwide. Here, only the mutations that occurred more than 9 times are listed.

Table S1: The cluster distributions of samples (N_{NS}), total SNP counts (N_{TF}), and the ratios of SNPs ($R_{TF} = N_{TF}/7123$) in 20 states of the U.S. The 20 states are Alaska (AK), Arizona (AZ), California (CA), Connecticut (CT), Washington, D.C. (DC), Florida (FL), Louisiana (LA), Maine (ME), Maryland (MD), Michigan (MI), Minnesota (MN), Nevada (NV), New Mexico (NM), New York (NY), Oregon (OR), Texas (TX), Utah (UT), Virginia (VA), Washington (WA), and Wisconsin (WI).

State	Cluster A			Cluster B			Cluster C			Cluster D		
	N_{NS}	N_{TF}	R_{TF}	N_{NS}	N_{TF}	R_{TF}	N_{NS}	N_{TF}	R_{TF}	N_{NS}	N_{TF}	R_{TF}
AK	13	108	1.5%	15	183	2.6%	4	27	0.4%	19	154	2.2%
AZ	36	273	3.8%	4	34	0.5%	12	95	1.3%	25	224	3.1%
CA	461	3574	50.2%	544	6147	86.3%	258	1507	21.2%	566	5440	76.4%
CT	29	226	3.2%	17	166	2.3%	3	35	0.5%	67	756	10.6%
DC	13	94	1.3%	2	54	0.8%	4	29	0.4%	4	36	0.5%
FL	122	907	12.7%	132	1531	21.5%	47	522	7.3%	299	3492	49.0%
LA	179	1451	20.4%	7	188	2.6%	1	15	0.2%	163	1753	24.6%
ME	26	175	2.5%	4	31	0.4%	13	101	1.4%	47	304	4.3%
MD	45	359	5.0%	17	320	4.5%	29	365	5.1%	78	899	12.6%
MI	253	1868	26.2%	10	100	1.4%	16	157	2.2%	90	900	12.6%
MN	209	1590	22.3%	52	622	8.7%	159	1260	17.7%	312	3576	50.2%
NV	52	395	5.5%	8	79	1.1%	19	107	1.5%	69	728	10.2%
NM	17	115	1.6%	15	138	1.9%	0	0	0.0%	8	71	1.0%
NY	925	6700	94.1%	68	651	9.1%	96	675	9.5%	300	2846	40.0%
OR	41	337	4.7%	153	1753	24.6%	70	401	5.6%	255	2503	35.1%
TX	10	79	1.1%	9	114	1.6%	8	95	1.3%	15	160	2.2%
UT	39	293	4.1%	49	505	7.1%	21	159	2.2%	252	2819	39.6%
VA	199	1561	21.9%	63	803	11.3%	35	274	3.8%	291	3146	44.2%
WA	595	4593	64.5%	1213	15630	219.4%	1155	9005	126.4%	680	7925	111.3%
WI	111	806	11.3%	322	4219	59.2%	187	934	13.1%	414	4324	60.7%

Table S2: The number of samples (N_{NS}) of a particular mutation on 5'UTR in the United States and worldwide. Here, only the mutations that occurred more than 9 times are listed.

Mutation	US N_{NS}	World N_{NS}	Mutation	US N_{NS}	World N_{NS}
241C>T	9628	36786	187A>G	36	207
199G>T	28	33	222C>T	17	50
208G>T	13	15	218C>T	13	29
242G>T	11	17	169A>G	10	11

S2 Supplementary Data

Total 13 spreadsheets are merged in the Supplementary Data.xlsx:

- Supplementary Data 1: The GISAID IDs we use in the world. (Up to September 11, 2020).
- Supplementary Data 2: The GISAID IDs we use in the United States. (Up to September 11, 2020).
- Supplementary Data 3: The world clusters information.
- Supplementary Data 4: The US clusters information.
- Supplementary Data 5: Acknowledgment table provided by GISAID in Jan 2020.
- Supplementary Data 6: Acknowledgment table provided by GISAID in Feb 2020.
- Supplementary Data 7: Acknowledgment table provided by GISAID in March 2020.
- Supplementary Data 8: Acknowledgment table provided by GISAID in April 2020.
- Supplementary Data 9: Acknowledgment table provided by GISAID in May 2020.
- Supplementary Data 10: Acknowledgment table provided by GISAID in June 2020.
- Supplementary Data 11: Acknowledgment table provided by GISAID in July 2020.
- Supplementary Data 12: Acknowledgment table provided by GISAID in August 2020.
- Supplementary Data 13: Acknowledgment table provided by GISAID in September 2020.

S3 Supplementary Figures

S3.1 Mutation Tracker

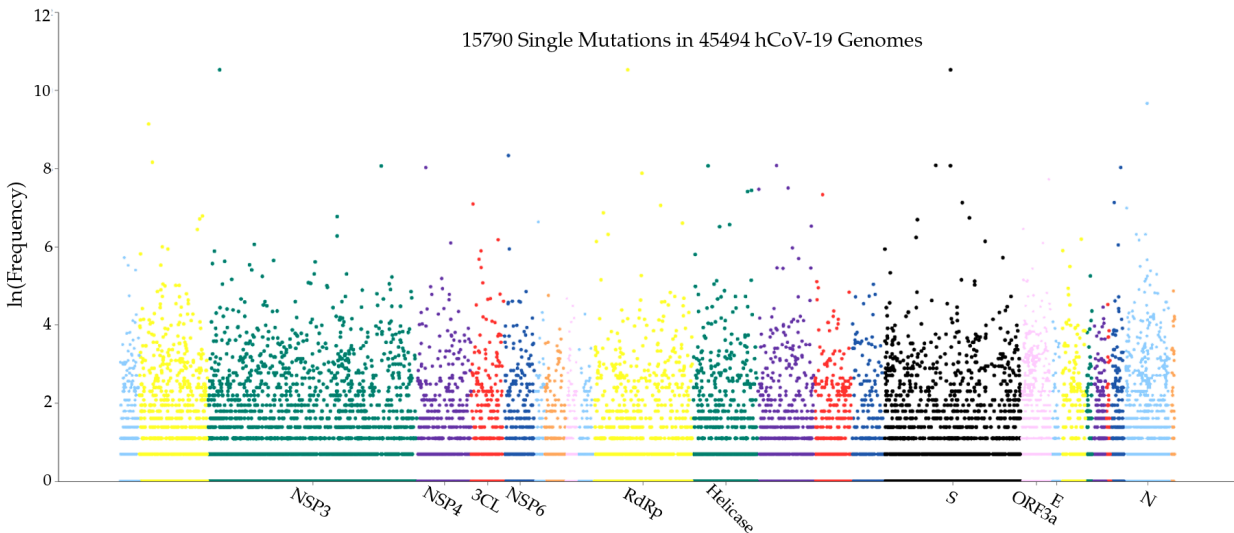


Figure S1: Genome-wide SARS-CoV-2 mutation distribution. The y -axis represents the natural log frequencies of mutations on specific positions of the complete SARS-CoV-2 genome. A few landmark positions are labeled with gene (protein) names.

S3.2 K -Means clustering

The k -means clustering is used to classify the SARS-CoV-2 the SNP variants. The Elbow method is used to determine the optimal number of clusters. Our results demonstrate four main clusters in the United States (US) as shown in Figure S2, which plots the within-cluster sum of squares according to the number

of clusters k for the SNP variants in the United States based on Jaccard distance metric. The optimal values of k -mean clusters is shown as the turning point in the in the elbow plots.

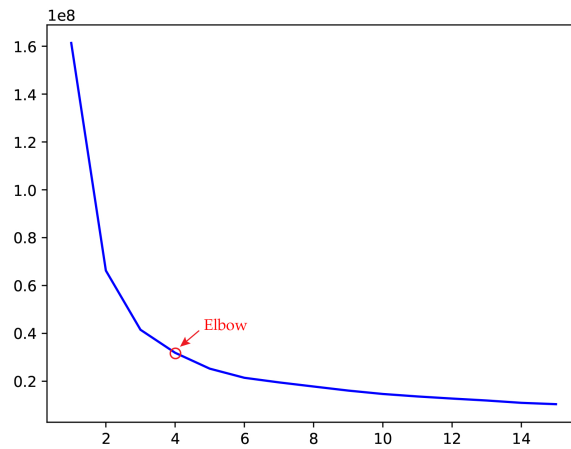


Figure S2: The plot of WCSS according to the number of clusters based on Jaccard distance metric. Here, Jaccard distance-based representation is taken as the input feature. The arrows point out the optimal number of clusters. The within-cluster sum of squares against the number of clusters for the SNP variants in the United States. The optimal number of clusters in the United States is four.

S3.3 Proteoforms

Figure S3 shows the visualization of the proteoforms of SARS-CoV-2 NSP2, NSP13, NSP12, spike protein, ORF8, and ORF3a. The top 8 mutations are marked in color. The red color represents the wild type residue and the yellow color represents the mutant type.

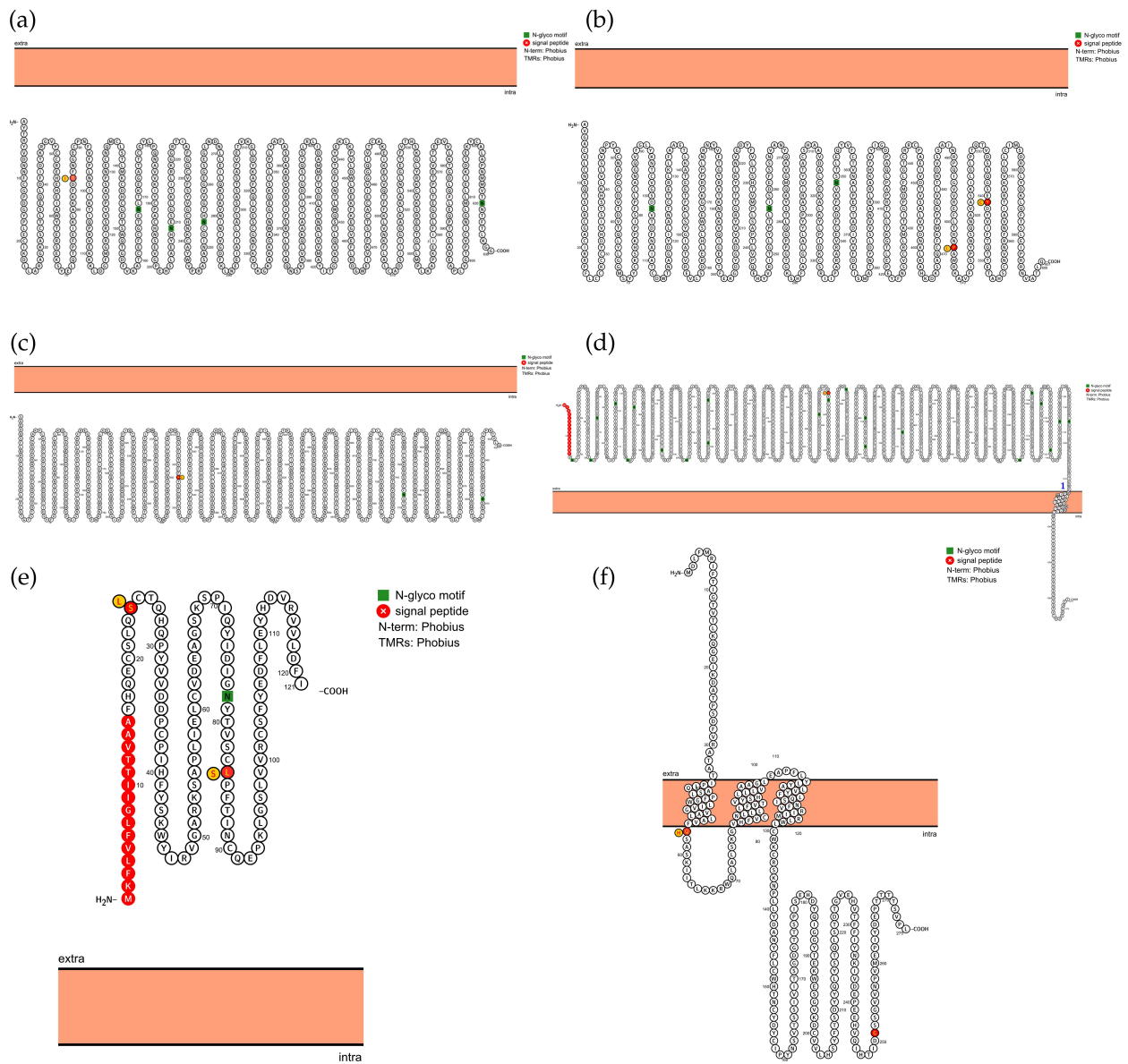


Figure S3: The visualization of six SARS-CoV-2 proteoforms. (a) Proteoform of NSP2. (b) Proteoform of NSP13. (c) Proteoform of NSP12. (d) Proteoform of spike protein. (e) Proteoform of ORF8. (f) Proteoform of ORF3a. The red color represents the wild type and the yellow represents the wild type.

S3.4 Protein structures, network analysis, and alignment

Figure S6 illustrates the 3D structure of SARS-CoV-2 NSP2 protein and the difference of FRI rigidity index, the subgraph centrality between the network with wild type and the network with mutant type.

Figure S7 illustrates sequence alignments for the NSP2 of SARS-CoV-2, SARS-CoV, bat coronavirus RaTG13, bat coronavirus CoVZC45, bat coronavirus BM48-31. Detailed numbering is given according to SARS-CoV-2. One high-frequency mutation 1059C>T(T85I) locates on the NSP2 protein. Here, the red rectangle marks the T85I position with its neighborhoods.

Figure S8 illustrates the sequence alignments for the NSP13 protein of SARS-CoV-2, SARS-CoV, bat coronavirus RaTG13, bat coronavirus CoVZC45, bat coronavirus BM48-31. Detailed numbering is given according to SARS-CoV-2. Two high-frequency mutations 17858A>G-(Y541C) and 17747C>T-(P504L) locate on NSP13. Here, the red rectangles mark the Y541C and P504: mutations with their neighborhoods.

Figure S9 illustrate the 3D structure of SARS-CoV-2 NSP13 protein and the difference of FRI rigidity index, the subgraph centrality between the network with wild type and the network with mutant type.

Figure S10 illustrates the sequence alignments for the ORF8 protein of SARS-CoV-2, SARS-CoV, bat coronavirus RaTG13, bat coronavirus CoVZC45, bat coronavirus BM48-31. Detailed numbering is given according to SARS-CoV-2. Two high-frequency mutations 28144T>C-(L84S) and 27964C>T-(S24L) locate on the ORF8. Here, the red rectangles mark the S24L and L84S mutations with their neighborhoods.

Figure S11 illustrate the 3D structure of SARS-CoV-2 ORF8 protein and the difference of FRI rigidity index, the subgraph centrality between the network with wild type and the network with mutant type.

Figure S12 illustrate the 3D structure of SARS-CoV-2 N protein and the difference of FRI rigidity index, the subgraph centrality between the network with wild type and the network with mutant type.

Figure S13 illustrates the sequence alignments for the N protein of SARS-CoV-2, SARS-CoV, bat coronavirus RaTG13, bat coronavirus CoVZC45, bat coronavirus BM48-31. Detailed numbering is given according to SARS-CoV-2. Three high-frequency mutations 28881G>A-(R203K), 28882G>A-(R203K), and 28883G>C-(G204R) locate on the N protein. Here, the red rectangles mark mutations R203K and G204R with their neighborhoods.

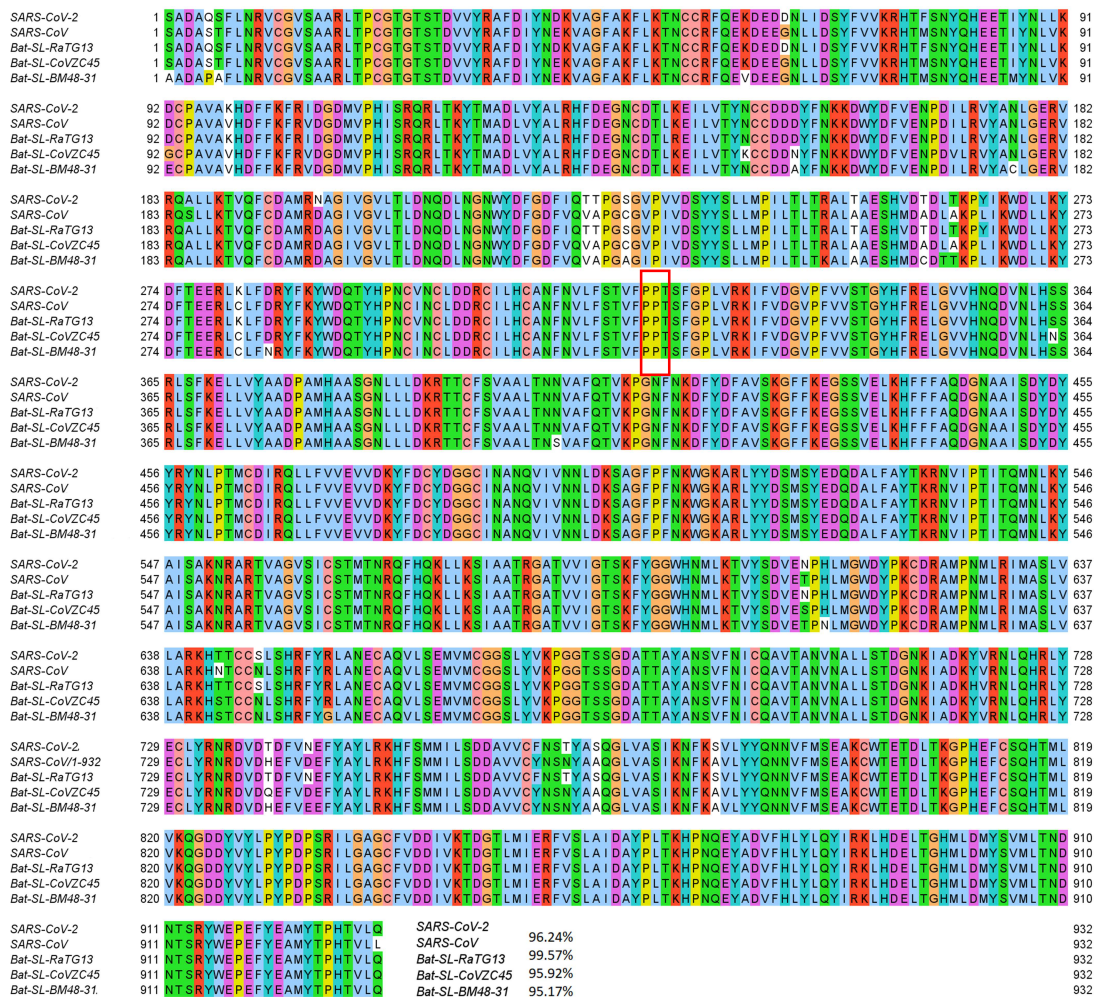


Figure S4: Sequence alignments for the NSP12 of SARS-CoV-2, SARS-CoV, bat coronavirus RaTG13, bat coronavirus CoVZC45, bat coronavirus BM48-31. Detailed numbering is given according to the SARS-CoV-2. One high-frequency mutation 14408C>T (P323L) is detected on the NSP12 protein. Here, the red rectangle marks the P323L mutations with its neighborhoods.

SARS-CoV-2	1	-----MFVFLVLLPLVSSGCVNLTTRTQLPPAYNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFSNVWFHAIHVSQTNGTKRFDNPVYL	84	
SARS-CoV	1	MFIFLLFLTLTSGDLDRCTTDDVQAPNRYTQHTSSMRGVVYYPDEIFRSDDLTLTQDLFLPFFSNVTFGHTIN-----HTFGNPVI	81	
Bat-SL-RaTG13	1	-----MFVFLVLLPLVSSGCVNLTTRTQLPPAYNSFTRGVYYPDKVFRSSVLHSTQDLFLPFFSNVWFHAIHVSQTNGTKRFDNPVYL	84	
Bat-SL-CoVZC45	1	-----MLFFLFLOFALVNSGCVNLTGRTPLNPNYTNSSQRGVYYPDTIYRSDTLVLSQGYFLPFFSNVSWYSLTTN-NAATKRITDNPIL	84	
Bat-SL-BM48-31	1	MKFLAFLLCLLGFANAOQDKGCTLSNKSPPSKLQTPSSRRGVYYPDDIFRSSLRVLLTGHFLPFFNLTWVLLTKSN-GKQRIYDNPNI	88	
SARS-CoV-2	85	PFNDGVYFASIEKSNIIIRGWIIFGTTLDKSTQSLLIIVNNAIVVIVKVEFQFQNDPFLGVYVYKNNKMWSEFRVYSSANNCTFEYVVSQ	173	
SARS-CoV	82	PFKDDGIYFAAEKSNVVRGWVFGSTMNKKSQSVIINNSTNVVIRACNFELGDNPFVAVSKPMG-----TQHTMIFDNAFNCFTFEYVVSQ	166	
Bat-SL-RaTG13	85	PFNDGVYFASIEKSNIIIRGWIIFGTTLDKSTQSLLIIVNNAIVVIVKVEFQFQNDPFLGVYVYKNNKMWSEFRVYSSANNCTFEYVVSQ	173	
Bat-SL-CoVZC45	85	DFKDDGIYFAAEKSNIIIRGWIIFGTTLDKSTQSLLIIVNNAIVVIVKVEFQFQNDPFLGVYVYKNNKMWSEFRVYSSANNCTFEYVVSQ	172	
Bat-SL-BM48-31	89	NFGDGVYFGLIEKSNVFRGWIIFGTTLDNTQSAVLFNNGHILVIVDGNFNFQADPMFAVNS-----GDPYKWTWYISAAANCTYHRAH	170	
SARS-CoV-2	174	FFLMDLIEKQGNFKNLRFEVFKKIDGYFKIYVSKHTPIINLRDLQPGFSALEPLVDLPIGINIRFQILLALHRVYLTGGDSSGMITAGA	262	
SARS-CoV	167	AFLSDVSEKSGNFKHLRFVFKKIDGFLVYKQYQIDVVDLPLSGFNILKRPFKLPLGINITNFRALIAFSP-----AQDIWIGLSA	249	
Bat-SL-RaTG13	174	FFLMDLIEKQGNFKNLRFEVFKKIDGYFKIYVSKHTPIINLRDLQPGFSALEPLVDLPIGINIRFQILLALHRVYLTGGDSSGMITAGA	262	
Bat-SL-CoVZC45	173	FFLMDLIEKQGNFKNLRFEVFKKIDGYFKIYVSKHTPIINLRDLQPGFSALEPLVDLPIGINIRFQILLALHRVYLTGGDSSGMITAGA	258	
Bat-SL-BM48-31	171	AFNLISNNMPPGKHFHREHLFKNVDGFLVYVHNVEPILDLSGFPSPGFSVLKRLILKLPFGLNITTYKAIMLLESS-----TQSNFEDA	253	
SARS-CoV-2	263	AAYVGVYLPRTFLLKYNEGTITDAVDCALDPLSEKCTKLKSFVEKGIYQTSNFRVPTESIVRFPNITNLCPFEVFNATRFASVYV	351	
SARS-CoV	250	AAYVGVYLPRTFLLKYNEGTITDAVDCSONPLAELKCTKSKSEIDKGIYQTSNFRVPTSDVVRFPNITNLCPFEVFNATRFASVYV	338	
Bat-SL-RaTG13	263	AAYVGVYLPRTFLLKYNEGTITDAVDCALDPLSEKCTKLKSFVEKGIYQTSNFRVPTESIVRFPNITNLCPFEVFNATRFASVYV	351	
Bat-SL-CoVZC45	259	AAYVGVYLPRTFLLKYNEGTITDAVDCALDPLSEKCTKLKSFVEKGIYQTSNFRVPTSDVVRFPNITNLCPFEVFNATRFASVYV	347	
Bat-SL-BM48-31	254	AYFVGHKPLITMLVDFDENGTITDAIDCSODPLSEKCTKLKSFVEKGIYQTSNFRVPTESIVRFPNITNLCPFEVFNATRFASVYV	342	
SARS-CoV-2	352	AWNRKRIISNADVSVLYN-ASFSFTFCYGVSPKLNLDLCTFNVAADFVIRGDEVRIAPGQTGKIADYNYKLPDDFGCVIAWNSN	439	
SARS-CoV	339	AWERKRIISNADVSVLYN-STFSTFCYGVSPKLNLDLCTFNVAADFVIRGDEVRIAPGQTGKIADYNYKLPDDFGCVIAWNSN	426	
Bat-SL-RaTG13	352	AWNRKRIISNADVSVLYN-STFSTFCYGVSPKLNLDLCTFNVAADFVIRGDEVRIAPGQTGKIADYNYKLPDDFGCVIAWNSN	439	
Bat-SL-CoVZC45	348	AWERTKIDCADIYVLYN-STFSTFCYGVSPKLNLDLCTFNVAADFVIRGDEVRIAPGQTGKIADYNYKLPDDFGCVIAWNSN	435	
Bat-SL-BM48-31	343	AWERMRIISNADVSVLYNSASFSFTFCYGVSPKLNLDLCTFNVAADFVIRGDEVRIAPGQTGKIADYNYKLPDDFGCVIAWNSN	431	
SARS-CoV-2	440	NLSKVGNNYLYRFLKSNLKPFFERDITSEIYQAGSTPCNGVEGFNCYFPLQSYGFOPTNGVGYQPYRVVLLSFELLNAPATVCGPK	528	
SARS-CoV	427	NIDATSTGNYYKYRYLRHGKLRPFERDIN-VPFSPDGKFCPPALNCYWRPNDYGFYTTIGIYQPYRVVLLSFELLNAPATVCGPK	514	
Bat-SL-RaTG13	440	HIDAKEGNNYLYRFLKSNLKPFFERDITSEIYQAGSTPCNGVEGFNCYFPLQSYGFOPTNGVGYQPYRVVLLSFELLNAPATVCGPK	528	
Bat-SL-CoVZC45	443	KQDVG-----NYFYRSHRSTLKPFFERDLSDENGVR-----TISTYDFNPNVPEYQATRVVLLSFELLNAPATVCGPK	505	
Bat-SL-BM48-31	432	LDSSNE-----FFYRFRHGKIKPYGRDLSN-VLFNPSGGTCSAEGLCYKPLASVYGFQTSGGIGFPPYRVVLLSFELLNAPATVCGPK	515	
SARS-CoV-2	529	KSTLNLVKKNKCVNFFNGLTGTGVLTESENKFLPFQDFGRDIDTDAVRDPTLEILDIITPCSGFGGVSIVTPTNTSNVAVLYQDVNC	617	
SARS-CoV	515	LSTDLIKKQCVNFFNGLTGTGVLTPSSKRFQPFQDFGRDIDTDAVRDPTLEILDIITPCSGFGGVSIVTPTNTSNVAVLYQDVNC	603	
Bat-SL-RaTG13	529	KSTLNLVKKNKCVNFFNGLTGTGVLTESENKFLPFQDFGRDIDTDAVRDPTLEILDIITPCSGFGGVSIVTPTNTSNVAVLYQDVNC	617	
Bat-SL-CoVZC45	506	LSTLNLVKKNKCVNFFNGLKGTGVLTPSSKRFQPFQDFGRDIDTDAVRDPTLEILDIITPCSGFGGVSIVTPTNTSNVAVLYQDVNC	594	
Bat-SL-BM48-31	516	CSLNLVKKKCVNFFNGLTGTGVLTESENKFLPFQDFGRDIDTDAVRDPTLEILDIITPCSGFGGVSIVTPTNTSNVAVLYQDVNC	604	
SARS-CoV-2	618	TEVPVAIHADQLPTWRYVYVSTGS--NVFDTRAGGLIGAETHVNSYECDDIPIGAGICASYTOTNPPRRARSVASQSIAYTMSLGAENS	704	
SARS-CoV	604	TVSTAIHADQLPTAWRIYVYVSTGN--NVFDTRAGGLIGAETHVNSYECDDIPIGAGICASYTVLSLLR---STQKSIYVAYTMSLGAENS	696	
Bat-SL-RaTG13	618	TEVPVAIHADQLPTWRYVYVSTGS--NVFDTRAGGLIGAETHVNSYECDDIPIGAGICASYTOTNPPRRARSVASQSIAYTMSLGAENS	700	
Bat-SL-CoVZC45	595	DVPTLIIHADQLPTAWRIYVYVSTGS--NVFDTRAGGLIGAETHVNSYECDDIPIGAGICASYTASILR---STSQKAIYVAYTMSLGAENS	677	
Bat-SL-BM48-31	605	TVPTMLHADQISHDWRVYAFRNDGIFQTAGGLIGAAYVNSYECDDIPIGAGICAKYTNVSSITL---VRGGHSIYVAYTMSLGDNDG	690	
SARS-CoV-2	705	VAYSNNSIAIPTNFTISVTTTEIIPVSMKTSDVCDTMYICGDSIECSNLLLYGSGFCTQLNRALTGIAVEDDKNTQEVFAVQKVIYKTPP	793	
SARS-CoV	687	IAYSNNSIAIPTNFTISVTTTEIIPVSMKTSDVCDTMYICGDSIECANLLLYGSGFCTQLNRALSGIAAEDDKNTQEVFAVQKVIYKTPP	775	
Bat-SL-RaTG13	705	VAYSNNSIAIPTNFTISVTTTEIIPVSMKTSDVCDTMYICGDSIECSNLLLYGSGFCTQLNRALTGIAVEDDKNTQEVFAVQKVIYKTPP	789	
Bat-SL-CoVZC45	678	IAYSNNSIAIPTNFTISVTTTEIIPVSMKTSDVCDTMYICGDSIECSNLLLYGSGFCTQLNRALSGIAAEDDKNTQEVFAVQKVIYKTPP	766	
Bat-SL-BM48-31	691	IYVSNNSIAIPTNFTISVTTTEIIPVSMKTSDVCDTMYICGDSIECSNLLLYGSGFCTQLNRALGIAAEDDKNTQEVFAVQKVIYKTPP	779	
SARS-CoV-2	794	KDFGGFNFSQILPDPSPKSRFSIEDLLFNKVLADAGFIKQVGDCLGDIIAARDLICAQKFNGLTVLPPLLTDEMIAYTAALASGIA	882	
SARS-CoV	776	KYFGGFNFSQILPDLKPKTKRFSIEDLLFNKVLADAGFMKQVGECLGDIINARDLICAQKFNGLTVLPPLLTDEMIAYTAALASGIA	864	
Bat-SL-RaTG13	790	KDFGGFNFSQILPDPSPKSRFSIEDLLFNKVLADAGFIKQVGDCLGDIIAARDLICAQKFNGLTVLPPLLTDEMIAYTAALASGIA	878	
Bat-SL-CoVZC45	767	KDFGGFNFSQILPDPSPKSRFSIEDLLFNKVLADAGFIKQVGDCLGDIIAARDLICAQKFNGLTVLPPLLTDEMIAYTAALASGIA	855	
Bat-SL-BM48-31	780	KDFGGFNFSQILPDRAPKPSRFSIEDLLFNKVLADAGFMKQVGDCLGGVNAARDLICAQKFNGLTVLPPLLTDEMIAYTAALASGIA	868	
SARS-CoV-2	883	SGWTFGAGAALDIPFAMQMAYRFNGIIVTQNVLYENQRLIANGFNKAIISQIDSLSTTASALGLQDDVVNQNAGALNTLVKQLSSNFG	971	
SARS-CoV	865	TAQWTFGAGAALDIPFAMQMAYRFNGIIVTQNVLYENQRLIANGFNKAIISQIQEISLTTTALGLQDDVVNQNAGALNTLVKQLSSNFG	953	
Bat-SL-RaTG13	879	SGWTFGAGAALDIPFAMQMAYRFNGIIVTQNVLYENQRLIANGFNKAIISQIDSLSTTASALGLQDDVVNQNAGALNTLVKQLSSNFG	967	
Bat-SL-CoVZC45	856	TAQWTFGAGAALDIPFAMQMAYRFNGIIVTQNVLYENQRLIANGFNKAIISQIQEISLTTASALGLQDDVVNQNAGALNTLVKQLSSNFG	944	
Bat-SL-BM48-31	869	TAQWTFGAGAALDIPFAMQMAYRFNGIIVTQNVLYENQRLIANGFNKAIISQIQEISLTTTALGLQDDVIVNQNATLNTLVKQLSSNFG	957	
SARS-CoV-2	972	AISSVLNDILSRDLKVEAEVDIIRLITGRLDLQTYVTOQLIRAAEIRASANLAATKMSCEVLDGSKRVDFCGKGYHLMSPDQAPHG	1060	
SARS-CoV	954	AISSVLNDILSRDLKVEAEVDIIRLITGRLDLQTYVTOQLIRAAEIRASANLAATKMSCEVLDGSKRVDFCGKGYHLMSPDQAPHG	1042	
Bat-SL-RaTG13	969	AISSVLNDILSRDLKVEAEVDIIRLITGRLDLQTYVTOQLIRAAEIRASANLAATKMSCEVLDGSKRVDFCGKGYHLMSPDQAPHG	1056	
Bat-SL-CoVZC45	945	AISSVLNDILSRDLKVEAEVDIIRLITGRLDLQTYVTOQLIRAAEIRASANLAATKMSCEVLDGSKRVDFCGKGYHLMSPDQAPHG	1033	
Bat-SL-BM48-31	958	AISSVLNDILSRDLKVEAEVDIIRLITGRLDLQTYVTOQLIRAAEIRASANLAATKMSCEVLDGSKRVDFCGKGYHLMSPDQAPHG	1046	
SARS-CoV-2	1061	VFLHVTYYPADQENKFTTAPAIICHGKAHFPREGVFSNGTHWFVTRNFYEPQIITDNTFVSGNCDVVIIGIINNVTYDPLQPELDSFK	1149	
SARS-CoV	1043	VFLHVTYYPADQENKFTTAPAIICHGKAHFPREGVFSNGTHWFVTRNFYEPQIITDNTFVSGNCDVVIIGIINNVTYDPLQPELDSFK	1131	
Bat-SL-RaTG13	1057	VFLHVTYYPADQENKFTTAPAIICHGKAHFPREGVFSNGTHWFVTRNFYEPQIITDNTFVSGNCDVVIIGIINNVTYDPLQPELDSFK	1145	
Bat-SL-CoVZC45	1034	VFLHVTYYPADQENKFTTAPAIICHGKAHFPREGVFSNGTHWFVTRNFYEPQIITDNTFVSGNCDVVIIGIINNVTYDPLQPELDSFK	1122	
Bat-SL-BM48-31	1047	VFLHVTYYPADQENKFTTAPAIICHGKAHFPREGVFSNGTHWFVTRNFYEPQIITDNTFVSGNCDVVIIGIINNVTYDPLQPELDSFK	1135	
SARS-CoV-2	1150	EELDKYFNKHTSPDVLGDIGSINASVVDIENKEDRLNEVAKNLESILDLQELGKYEQYIKWPWYIWLGFAGLIAIIVMVTIMLCOMT	1238	
SARS-CoV	1132	EELDKYFNKHTSPDVLGDIGSINASVVDIENKEDRLNEVAKNLESILDLQELGKYEQYIKWPWYIWLGFAGLIAIIVMVTIMLCOMT	1220	
Bat-SL-RaTG13	1146	EELDKYFNKHTSPDVLGDIGSINASVVDIENKEDRLNEVAKNLESILDLQELGKYEQYIKWPWYIWLGFAGLIAIIVMVTIMLCOMT	1234	
Bat-SL-CoVZC45	1123	EELDKYFNKHTSPDVLGDIGSINASVVDIENKEDRLNEVAKNLESILDLQELGKYEQYIKWPWYIWLGFAGLIAIIVMVTIMLCOMT	1211	
Bat-SL-BM48-31	1136	EELDKYFNKHTSQNVSLDGLNININASVVDIENKEDRLNEVAKNLESILDLQELGKYEQYIKWPWYIWLGFAGLIAIIVMVTIMLCOMT	1224	
SARS-CoV-2	1239	CCSCCKLKGCCSCGSCCKFDEDDSEPLVKGVKLHYT	SARS-CoV-2	1273
SARS-CoV	1221	CCSCCKLKGACSCGSCCKFDEDDSEPLVKGVKLHYT	SARS-CoV	73.86%
Bat-SL-RaTG13	1235	CCSCCKLKGCCSCGSCCKFDEDDSEPLVKGVKLHYT	Bat-SL-RaTG13	97.47%
Bat-SL-CoVZC45	1212	CCSCCKLKGCCSCGSCCKFDEDDSEPLVKGVKLHYT	Bat-SL-CoVZC45	80.49%
Bat-SL-BM48-31	1225	CCSCCKLKGVCSCGSCCKFDEDDSEPLVKGVKLHYT	Bat-SL-BM48-31	67.11%

Figure S5: Sequence alignments for the S proteins of SARS-CoV-2, SARS-CoV, bat coronavirus RaTG13, bat coronavirus CoVZC45, bat coronavirus BM48-31. Detailed numbering is given according to SARS-CoV-2 S protein. One high-frequency mutation 23403A>G- (D614G) is detected on the S protein. Here, the red rectangle marks the D614G mutations with its neighborhoods

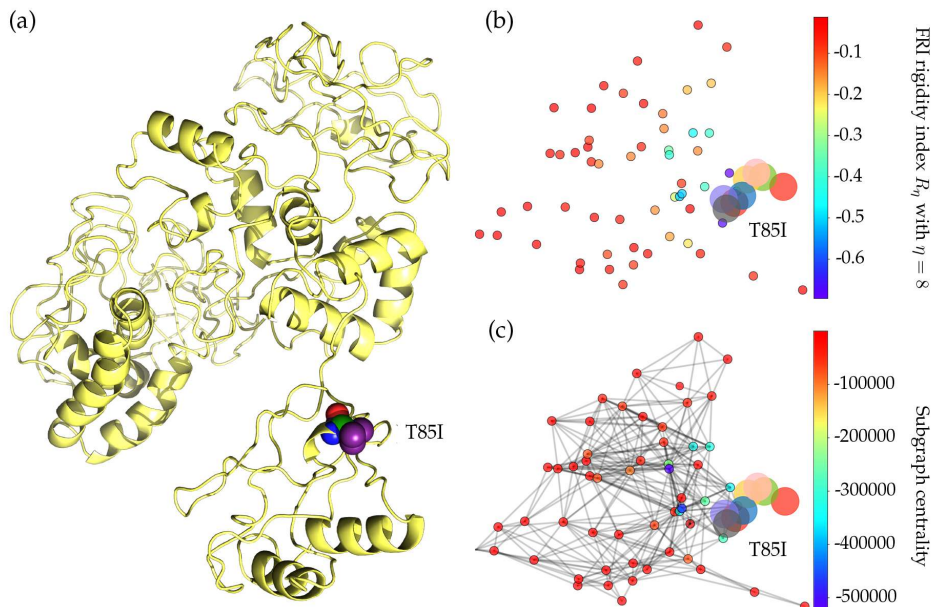


Figure S6: (a) The 3D structure of SARS-CoV-2 NSP2 protein. The mutant residue is marked with color balls. (b) The difference of FRI rigidity index between the network with wild type and the network with mutant type. (c) The difference of the subgraph centrality between the network with wild type and the network with mutant type.

SARS-CoV-2	1	AYTRVYVDNNFCGPDGVP	LECI	KDLLARAGKASCTLSEQLDFIDTKRGGVYCCREHEHEIAWYTERSEKSYELQTPFEIKLAKKFDITFKGCECP	91																												
SARS-CoV	1	AVTRVYVDNNFCGPDGVP	LDCI	KDFLARAGKSMCTLSEQLDYIESKRGVYCCRDHEHEIAWFTERSDKSYEHQTPFEIKSAKKFDITFKGCECP	91																												
Bat-SL-RaTG13	1	AYTRVYVDNNFCGPDGVP	LECI	KDLLARAGKASCTLSEQLDFIDTKRGGVYCCREHEHEIAWYTERSEKSYELQTPFEIKLAKKFDITFKGCECP	91																												
Bat-SL-CoVZC45	1	AYTRVYVDNNFCGPDGVP	LECI	KDLLARAGKASYALEQLDFIDTKRGGVYCCREHEHEIAWYTERSEKSYELQTPFEIKLAKKFDITFKGCECP	91																												
Bat-SL-BM48-31	1	VYTRVYVDNNFCGPDGVP	LECI	KDLLARAGKASAPLAELDFLESKRGVYCCREHEHEIAWYTERSDKSYELQTPFEITNAKKFDITFKGCECP	91																												
SARS-CoV-2	92	VFVPLNSTIKTIDPRVEKKLDGFMGR	RSVYVVASPMECNMCLSTLMKCDHCGE	TSWQTDGFVKAICFCFCGTELLKKEGATTCGYLPS	182																												
SARS-CoV	92	KVFPLNSTIKVQIDPRVEKKTEGFMGR	RSVYVVASPMECNMHLSTLMKCNHCDVSWQTDGFVKAICFCFCGTEENLVLEGPTTCGYLPT	182																													
Bat-SL-RaTG13	92	VFVPLNSTIKTIDPRVEKKLDGFMGR	RSVYVVASPMECNMCLSTLMKCDHCGE	TSWQTDGFVKAICFCFCGTEENLVLEGPTTCGYLPT	182																												
Bat-SL-CoVZC45	92	VFVPLNSTIKTIDPRVEKKLDGFMGR	RSVYVVASPMECNMCLSTLMKCDHCGE	TSWQTDGFVKAICFCFCGTEENLVLEGPTTCGYLPT	182																												
Bat-SL-BM48-31	92	KVFPLNSTIKVQIDPRVEKKTEGFMGR	IRTVYQVASPECNMHLSTYMNGNHCGE	TSWQTDGFVKAICFCFCGTEENLVLEGPTTCGYLPS	182																												
SARS-CoV-2	183	NAVVKLYCPACHNSEVGP	PEHSLAEYHNESGLKTI	LRKGGRTIAFGGCVFSYVGCYNKCAIYVWP	PRASANI	GCNHTGVVGGES	EGLNDNLLLEI	273																									
SARS-CoV	183	NAVVKMPACQDPEIGPEH	SVADYHNHNSIETRLR	KGGRTIRFCGGCVFSYVGCYNKRAYVWP	PRASADI	GSHTGII	IGDNVETL	NEDLLEI	273																								
Bat-SL-RaTG13	183	NAVVKLYCPACHNSEVGP	PEHSLAEYHNESGLKPI	LRKGGRTIAFGGCVFSYVGCYNKCAIYWI	PRASANI	GCNHTGVVGGES	EGLNDNLLLEI	273																									
Bat-SL-CoVZC45	183	NAVVKLYCPACHNSEVGP	PEHSLAEYHNESGLKPI	LRKGGRTIAFGGCVFSYVGCYNKCAIYWI	PRASANI	GCNHTGVVGGES	EGLNDNLLLEI	273																									
Bat-SL-BM48-31	183	NAVVKMVPACQDPEIGPDH	SVADYHNHNSKIETRLR	KGGRTIKSFGGCVFSYVGCYNKRAFWVWP	PRASANI	GSNHTGVVGGEV	ETMNDLLEI	273																									
SARS-CoV-2	274	LRKRVNINIVGDFKLN	EEIAILLASFRAS	TSFAFVETVKGIDYKAFKGI	VESCGNFVTKGAKK	GAWNIGEOKSI	LSPL	YAFAS	EAARVV	364																							
SARS-CoV	274	LSRERVNINIVGDFHLE	EEVAIILLASFRAS	TSFAFIDITIKSLDYKSFKTI	VESCGNFVTKGAPV	KGAWNIGBDR	SVLTP	LCGFP	SSAAGV	364																							
Bat-SL-RaTG13	274	LRKRVNINIVGDFKLN	EEIAILLASFRAS	TSFAFVETVKGIDYKTFKGI	VESCGNFVTKGAKK	GAWNIGEOKSI	LSPL	YAFAS	EAARVV	364																							
Bat-SL-CoVZC45	274	LRKRVNINIVGDFKLN	EEIAILLASFRAS	TSFAFVETVKGIDYKTFKGI	VESCGNFVTKGAKK	GAWNIGEOKSI	LSPL	YAFAS	EAARVV	364																							
Bat-SL-BM48-31	274	LSRERVNINIVGDFCL	NEEIAILLASL	ASFAFVETVKNLDFKTFKGI	IESCGNYKVT	KGAKFGVWN	IGTSK	SLTLP	LHCF	SSAAGVV	364																						
SARS-CoV-2	365	RSIFSR	LTAAQNSVR	LQKAAITLDGI	SQYSRLIDAMMFT	SDLATN	NLVVMAYITGGVV	QLTSQWLTNI	FGTV	EKLKPVLDWLEEK	455																						
SARS-CoV	365	RSIFART	LDAAHNSIDLQRA	AAVTLDGI	SEQSLR	LDVAMVYTS	DLTLN	SVIIMAYVT	GGLV	QOTSQWLS	NLLGTTVEKLRPI	FEWIEAKL	455																				
Bat-SL-RaTG13	365	RSIFSR	LTAAQNSVRA	LQKAAITLDGI	SQYSRLIDAMMFT	SDLV	TNNLVVMAYITGGVV	QLTSQWLTNI	FGTV	EKLKPVLDWLEEK	455																						
Bat-SL-CoVZC45	365	RSIFSR	LTAAHHSVH	LQKAAITLDGI	SQYSRLIDAMMFT	SDLV	TNNLVVMAYITGGVV	QMSQWLTNI	FGTV	EKLKPVLDWLEEK	455																						
Bat-SL-BM48-31	365	RSIFSR	LTAAHNSIV	DLHRAAMIIFSDI	SDQANR	LDVAMVYTS	DLVT	ESVVM	MAYITGGVV	QCVSTWLS	QLLN	SVDFK	SAVLRWLEOKL	455																			
SARS-CoV-2	456	KEGVEFL	RDGWEIVKFI	STCACEIVGGIVT	CAKEIKESVDT	FFKLVN	KFLALCADSII	IGGAKL	KALNLGET	FVTHSKGLYR	KCVK	KREE	546																				
SARS-CoV	456	SAGVEFL	DAWEILKFLIT	GVFDIVKGDIVQASDNI	KDCVKCFIDV	NKALEMCI	DOVTI	AGAKLR	SNLNLGEV	FI	AQSKGLYR	QCIR	KGEC	546																			
Bat-SL-RaTG13	456	KEGVEFL	RDGWEIVKFI	STCACEIVGGIVT	CAKEIKESVDT	FFKLVN	KFLALCADSII	IGGAKL	KALNLGET	FVTHSKGLYR	KCVK	KREE	546																				
Bat-SL-CoVZC45	456	KEGVEFL	RDGWEIVKFI	STCACEIVGGIVT	CAKEIKESVDT	FFKLVN	KFLALCADSII	IGGAKL	KALNLGET	FVTHSKGLYR	KCVR	KREE	546																				
Bat-SL-BM48-31	456	QGGIDFL	QAWGLL	LLVIGAYVMIR	GKI	IQVNTS	SEIECVT	FEVDV	NKVFELCTD	YITV	AGAR	RAIN	FG	EV	LIA	QSRGLYR	QCV	RND	546														
SARS-CoV-2	547	TGLLMP	LKAPKEIIF	LEGETLP	EVLT	EEVV	LKGLD	LP	LEPT	SEAVEAP	LVGT	PVCI	NG	LMLLEIKD	TEKY	CALAP	NMMVT	NTFT	FL	KG	637												
SARS-CoV	547	LQLLMP	LKAPKEVIF	LEGDSH	IVLT	SEVV	LKNG	LEAL	ET	PVDS	FN	NGAI	VG	TPV	CV	NG	LMLLEIKD	TEKY	QAL	SP	GL	LAT	NN	VF	RL	KG	637						
Bat-SL-RaTG13	547	TGLLMP	LKAPKEIIF	LEGETLP	EVLT	EEVV	LKGLD	LP	LEPT	SEAVEAP	LVGT	PVCI	NG	LMLLEIKD	TEKY	CALAP	NMMVT	NTFT	FL	KG	637												
Bat-SL-CoVZC45	547	TGLLMP	LKAPKEIIF	LEGETLP	EVLT	EEVV	LKGLD	LP	LEPT	SEAVEAP	LVGT	PVCI	NG	LMLLEIKD	TEKY	CALAP	NMMVT	NTFT	FL	KG	637												
Bat-SL-BM48-31	547	LQLLMP	LKSPK	DVDF	DGDAD	DL	LT	SEEV	VKNG	TL	EA	LD	LE	LD	SV	VT	GV	AE	GP	V	CV	NG	LMLLEIKD	TEKY	QAL	SP	LL	AT	NN	VF	RL	KG	637
SARS-CoV-2	638	G	SARS-CoV-2	68.33%	638																												
SARS-CoV	638	G	SARS-CoV	98.27%	638																												
Bat-SL-RaTG13	638	G	Bat-SL-RaTG13	95.29%	638																												
Bat-SL-CoVZC45	638	G	Bat-SL-CoVZC45	62.38%	638																												
Bat-SL-BM48-31	638	G	Bat-SL-BM48-31		638																												

Figure S7: Sequence alignments for the NSP2 of SARS-CoV-2, SARS-CoV, bat coronavirus RaTG13, bat coronavirus CoVZC45, bat coronavirus BM48-31. Detailed numbering is given according to SARS-CoV-2. One high-frequency mutation 1059C>T (T85I) locates on the NSP2 protein. Here, the red rectangle marks the T85I position with its neighborhoods.

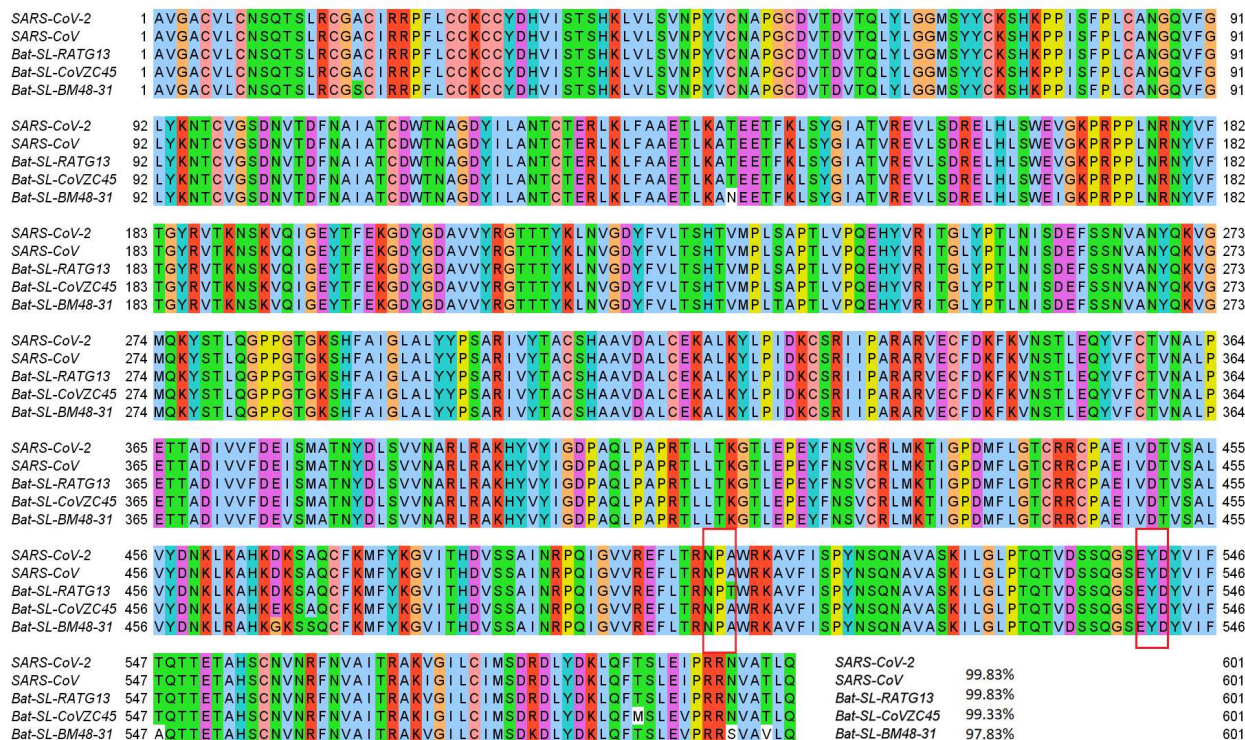


Figure S8: Sequence alignments for the NSP13 protein of SARS-CoV-2, SARS-CoV, bat coronavirus RaTG13, bat coronavirus CoVZC45, bat coronavirus BM48-31. Detailed numbering is given according to SARS-CoV-2. Two high-frequency mutations 17858A>G-(Y541C) and 17747C>T-(P504L) locate on NSP13. Here, the red rectangles mark the Y541C and P504L mutations with their neighborhoods.

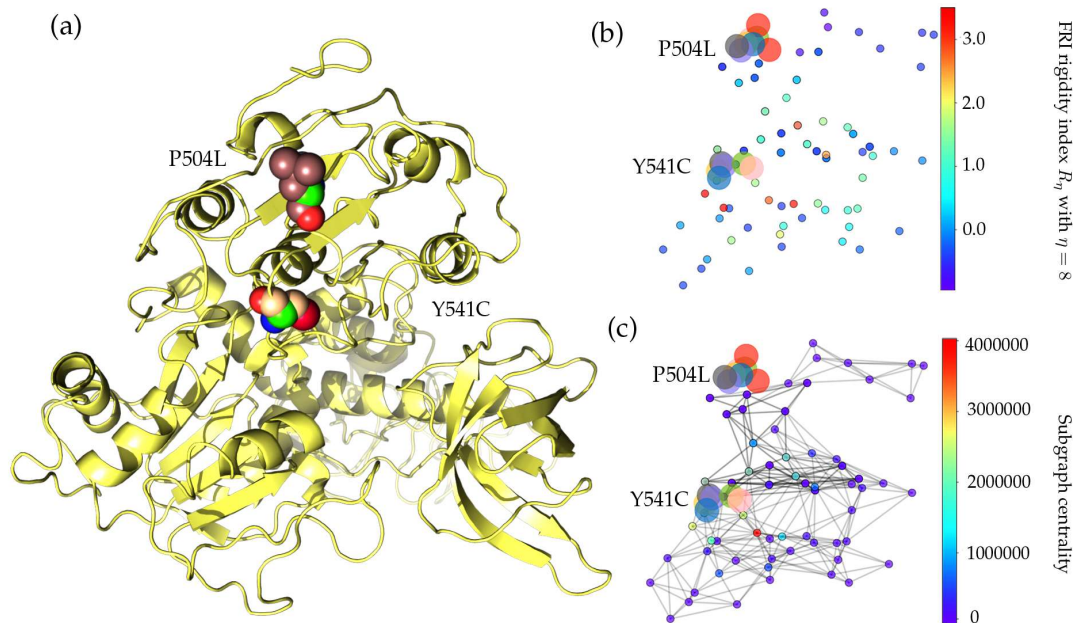


Figure S9: (a) The 3D structure of SARS-CoV-2 NSP13 protein. The mutant residue is marked with color balls. (b) The difference of FRI rigidity index between the network with wild type and the network with mutant type. (c) The difference of the subgraph centrality between the network with wild type and the network with mutant type.

SARS-CoV-2 1 M K F L V F L G I I T T V A A F H Q E C S L Q S C T C H Q P Y V V D D P C P I H F Y S - K W Y I R V G A R K S A P L I E L C V D E A G S K S P I Q Y I D I G N Y T V S C L P F T I N C 90
 SARS-CoV 1 M K I I L F L T L I V F T S C E L Y H - - Y Q E C V R G T T V L L K E P C P S G T Y E G N S P F H P L A D N K F A L T C T S T H F A F A C A D G T R H T Y Q L R A R S V S P K L F I R 89
 Bat-SL-RaTG13 1 M K I I L F L V L V T L A T C E L Y H - - Y Q E C V R G T T V L L K E P C S S G T Y E G N S P F H P L A D N K F A L T C F S T Q F A F A C P D G V R H T F Q L R A R S V S P K L F I R 89
 Bat-SL-CoVZC45 1 M K I I F F L V L I T L V T G E L Y H - - Y Q E C I K G T T V L L K E P C S S G T Y E G N S P F H P L A D N K F A L A C F S T Q F A F A C P D G V R H T F Q L R A R S V S P K L F T R 89
 Bat-SL-BM48-31 1 M K F L L L V A I V S I A S A E L Y H - - Y Q E C A R G T T V L L K E P C Q P N T Y E G N S P Y H P L A D N K F A I T C T N I K F S F V G O D E T R H V F Q L R A R S I S P R L F A S 89

SARS-CoV-2	91	Q E P K L G S L V V R C S F Y E D F L E Y H D V R V V L D F I - -	SARS-CoV-2	12.39%	121
SARS-CoV	90	Q E E V Q Q E L Y S P L F L I V A A L V F L I L C F T I K R K T E	SARS-CoV	10.74%	122
Bat-SL-RaTG13	90	Q E - E V Q E L Y S P I F L I A A I V F I T L C F T L K R K T E	Bat-SL-RaTG13	13.22%	121
Bat-SL-CoVZC45	90	Q E - E V Q E L Y S P V F L I V A A I V F I I L C F T F K R K I E	Bat-SL-CoVZC45	12.71%	118
Bat-SL-BM48-31	90	P K - H H S D D F T P V I L I I V T L L E V I Y C C M K R Q - - -			

Figure S10: Sequence alignments for the ORF8 protein of SARS-CoV-2, SARS-CoV, bat coronavirus RaTG13, bat coronavirus CoVZC45, bat coronavirus BM48-31. Detailed numbering is given according to SARS-CoV-2. Two high-frequency mutations 28144T>C-(L84S) and 27964C>T-(S24L) locate on the ORF8. Here, the red rectangles mark the S24L and L84S mutations with their neighborhoods.

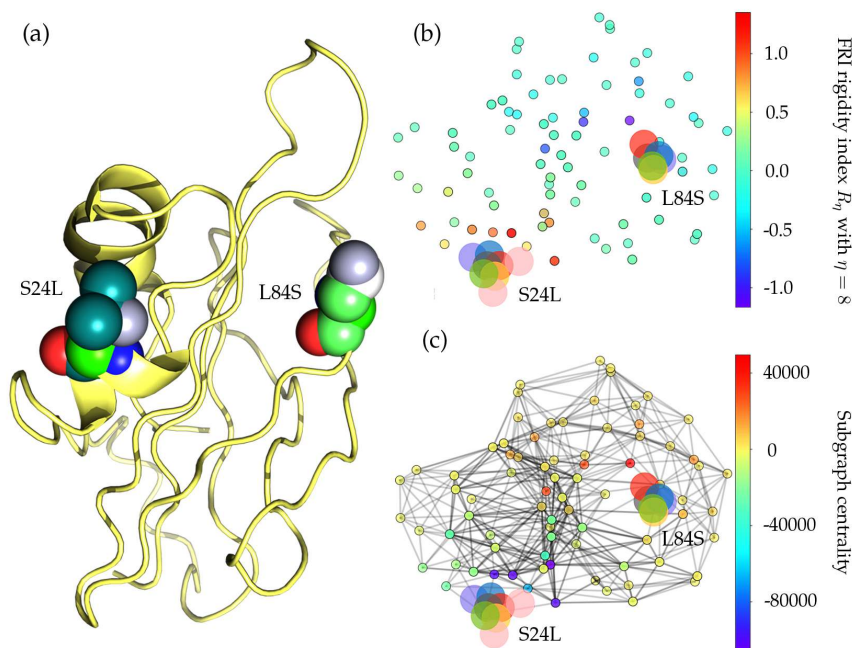


Figure S11: (a) The 3D structure of SARS-CoV-2 ORF8 protein. The mutant residue is marked with color balls. (b) The difference of FRI rigidity index between the network with wild type and the network with mutant type. (c) The difference of the subgraph centrality between the network with wild type and the network with mutant type.

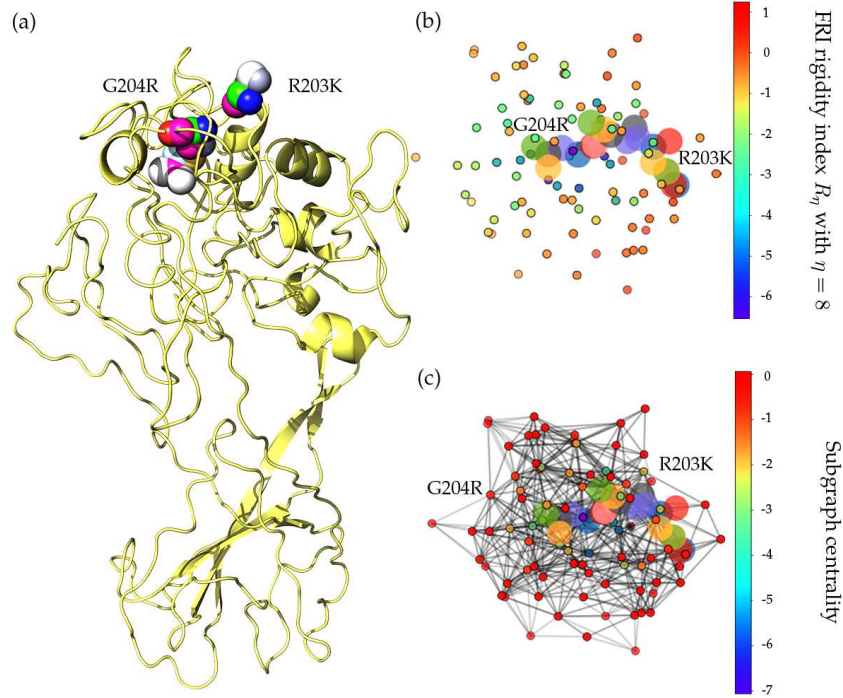


Figure S12: (a) The 3D structure of SARS-CoV-2 N protein. The mutant residue is marked with color balls. (b) The difference of FRI rigidity index between the network with wild type and the network with mutant type. (c) The difference of the subgraph centrality between the network with wild type and the network with mutant type.

SARS-CoV-2	1	MSDNGPQ-NORNAPRIITFGGPDSTGSNQNGERSGARSKORRPOGLPNNTASWFTALTQHGKEDLKFPFRGGQVPIINTNSSPDDQIGYRRAT	91		
SARS-CoV	1	MSDNGPQSNORSAPRIITFGGPTDSTDNQNGGRNGARPQORRPOGLPNNTASWFTALTQHGKEELRFPFRGGQVPIINTNSGPDDQIGYRRAT	92		
Bat-SL-RaTG13	1	MSDNGPQ-NORNAPRIITFGGPDSTGSNQNGERSGARPKORRPOGLPNNTASWFTALTQHGKEDLKFPFRGGQVPIINTNSSPDDQIGYRRAT	91		
Bat-SL-CoVZC45	1	MSDNGPQ-NORSAPRIITFGGPDSSDNSKNGERNGARPKORRPOGLPNNTASWFTALTQHGKENLTFPRGGQVPIINTNSSKDDQIGYRRAT	91		
Bat-SL-BM48-31	1	MTDNGQSNRNAPRIITFGVSDTSDNNQNAERAGARPKORRPOGPPNNTASWFTALTQHGKELSLFPFRGGQVPIINTNSTRDDQIGYRRAT	90		
SARS-CoV-2	92	RRIRGGDGMKDLSPRWYFYLLGTGPEAGLPYGANKDGIWVATEGALNTPKDHIGTRNPANNAIVLQLPQGTTLPKGFYAEGRGGSSQAS	183		
SARS-CoV	93	RRVRGGDGMKELSPRWYFYLLGTGPEASLPYGANKKEGIWVATEGALNTPKDHIGTRNPANNAIVLQLPQGTTLPKGFYAEGRGGSSQAS	184		
Bat-SL-RaTG13	92	RRIRGGDGMKDLSPRWYFYLLGTGPEAGLPYGANKDGIWVATEGALNTPKDHIGTRNPANNAIVLQLPQGTTLPKGFYAEGRGGSSQAS	183		
Bat-SL-CoVZC45	92	RRIRGGDGMKELSPRWYFYLLGTGPEAGLPYGANKKEGIWVATEGALNTPKDHIGTRNPANNAIVLQLPQGTTLPKGFYAEGRGGSSQAS	183		
Bat-SL-BM48-31	91	RRVRGGDGMKELSPRWYFYLLGTGPEAALPYGANKDGIWVATEGALNTPKDHIGTRNPANNAIVLQLPQGTTLPKGFYAEGRGGSSQAS	182		
SARS-CoV-2	184	SRSSSRNRSRNPSTPGSSRGTSPARMAGGGDAALALLLDRLNQLESKMSGKGOQQQGTVTKKSAAEASKKPRQKRATKAYNVTQAFG	275		
SARS-CoV	185	SRSSSRNRSRNPSTPGSSRGTSPARMAGGGETAALALLLDRLNQLESKMSGKGOQQQGTVTKKSAAEASKKPRQKRATKQYNVTQAFG	276		
Bat-SL-RaTG13	184	SRSSSRNRSRNPSTPGSSRGTSPARMAGGGDAALALLLDRLNQLESKMSGKGOQQQGTVTKKSAAEASKKPRQKRATKQYNVTQAFG	275		
Bat-SL-CoVZC45	184	SRSSSRNRSRNPSTPGSSRGTSPARMAGGGDTALALLLDRLNQLENKMSGKGOQQQGTVTKKSAAEASKKPRQKRATKQYNVTQAFG	275		
Bat-SL-BM48-31	183	SRSSSRNRSRNPSTPGSSRGTSPARMAAGG-DTALALLLDRLNQLESKMSGKTPQQ-SQVVTKKTAAEASKKPRQKRATKAYNVTQAFG	272		
SARS-CoV-2	276	RRGPEQTQGNFGDQELIRQGTDYKHWPQIAQFAPASAFFGMSRIGMEVTPSGTWLTYTGAIKLDDKDPNFKDQVILLNKHIDAYKTFPPT	367		
SARS-CoV	277	RRGPEQTQGNFGDQDLIRQGTDYKHWPQIAQFAPASAFFGMSRIGMEVTPSGTWLTYHGAIKLDDKDPQFKDQVILLNKHIDAYKTFPPT	368		
Bat-SL-RaTG13	276	RRGPEQTQGNFGDQELIRQGTDYKHWPQIAQFAPASAFFGMSRIGMEVTPSGTWLTYTGAIKLDDKDPNFKDQVILLNKHIDAYKTFPPT	367		
Bat-SL-CoVZC45	276	RRGPEQTQGNFGDQELIRQGTDYKHWPQIAQFAPASAFFGMSRIGMEVTPSGTWLTYHGAIKLDDKDPQFKDQVILLNKHIDAYKTFPPT	367		
Bat-SL-BM48-31	273	RRGPEPTQGNFGDQELIRLGTDYKNWPQIAQFAPASAFFGMSRIGMEVTPGTWLTYNGAIKLDDKDPNFKDQVILLNKHIDAYKTFPPT	364		
SARS-CoV-2	368	PKKDKKKKADETQALPQRQKKQQTVTLLPAADLDDFSKQLQSSMSG--ADSTQA	SARS-CoV-2	419	
SARS-CoV	369	PKKDKKKKTDEAQLPQRQKKQPTVTLLPAADMDDFSRQLQNSMSGASADSTQA	SARS-CoV	94.27%	422
Bat-SL-RaTG13	368	PKKDKKKKADETQALPQRQKKQQTVTLLPAADLDDFSKQLQSSMSG--ADSTQA	Bat-SL-RaTG13	89.74%	419
Bat-SL-CoVZC45	368	PKKDKKKKADETQALPQRQKKQQTVTLLPAADLDEFKQLQSSMSG--TDDSTQA	Bat-SL-CoVZC45	99.05%	419
Bat-SL-BM48-31	365	PKKDKKKKADENVSLPQRQKKQATVTLLPAADLDDFSKQLQNSMNA-ASPDSTQA	Bat-SL-BM48-31	85.27%	417

Figure S13: Sequence alignments for the N protein of SARS-CoV-2, SARS-CoV, bat coronavirus RaTG13, bat coronavirus CoVZC45, bat coronavirus BM48-31. Detailed numbering is given according to SARS-CoV-2. Three high-frequency mutations 28881G>A-(R203K), 28882G>A-(R203K), and 28883G>C-(G204R) locate on the N protein. Here, the red rectangles mark mutations R203K and G204R with their neighborhoods.