

Data Information

1. Detailed information of big sequences in Practical Use Case 1

We collected nearly thirty thousand publicly available complete genome sequences of COVID-19 virus from 79 different countries from December 2019 to May 2020 (shown in Table 1), and related flu-causing coronaviruses/viruses, e.g., HCoV-229E (shown in Table 2), HCoV-HKU1 (shown in Table 3), HCoV-NL63 (shown in Table 4), HCoV-OC43 (shown in Table 5), etc. These viruses are big sequences with a length of about 30,000. These sequences are all from NCBI GenBank database. You can search and download them by using their serial number (No.). Some collected sequences are shown in the following tables 1-5 (see following references [1-2] for details).

Table 1 Detailed information of the COVID-19 strains

No.	Name of SARS-CoV-2 strain	Serial No.	Data source	Sampling time	Sampling location	Sequence length
1	BetaCoV/Wuhan/HUICDC-HB-01/2019	EPI_ISL_412899	GISAI	2019.12.30	China/Hubei/Wuhan	29848
2	BetaCoV/Wuhan/HUICDC-HB-02/2019	EPI_ISL_412898	GISAI	2019.12.30	China/Hubei/Wuhan	29848
3	BetaCoV/Wuhan/HUICDC-HB-01/2019	EPI_ISL_402132	GISAI	2019.12.30	China/Hubei	29848
4	BetaCoV/Wuhan/IVDC-HH-05/2019	EPI_ISL_402121	GISAI	2019.12.30	China/Hubei/Wuhan	29891
5	BetaCoV/Wuhan/IVDC-HH-01/2019	EPI_ISL_402119	GISAI	2019.12.30	China/Hubei/Wuhan	29891
6	BetaCoV/Wuhan/WH1903/2019	NMDC60013002-06	NMDC	2019.12.30	China/Hubei/Wuhan	29891
7	BetaCoV/Wuhan/WH1901/2019	NMDC60013002-08	NMDC	2019.12.30	China/Hubei/Wuhan	29891
8	BetaCoV/Wuhan/WH1903/2019	NMDC60013002-10	NMDC	2019.12.30	China/Hubei/Wuhan	29891
9	BetaCoV/Wuhan/WH1901/2019	CNA0007332	CNGBLs	2019.12.26	China/Hubei/Wuhan	29866
10	WIV02	GWIAHKK000000000	Genome Warehouse	2019.12.30	China/Hubei/Wuhan	29825
11	WIV04	GWIAHKL000000000	Genome Warehouse	2019.12.30	China/Hubei/Wuhan	29891
12	WIV05	GWIAHKM000000000	Genome Warehouse	2019.12.30	China/Hubei/Wuhan	29852
13	WIV06	GWIAHKN000000000	Genome Warehouse	2019.12.30	China/Hubei/Wuhan	29854
14	WIV07	GWIAHKO000000000	Genome Warehouse	2019.12.30	China/Hubei/Wuhan	29857
15	BetaCoV/Wuhan/PPICAMS-WH-01/2019	GWIAHKT000000000	Genome Warehouse	2019.12.23	China/Hubei/Wuhan	29899
16	BetaCoV/Wuhan/PPICAMS-WH-02/2019	GWIAHKG000000000	Genome Warehouse	2019.12.30	China/Hubei/Wuhan	29889
17	BetaCoV/Wuhan/PPICAMS-WH-03/2019	GWIAHKL000000000	Genome Warehouse	2019.12.30	China/Hubei/Wuhan	29899
18	BetaCoV/Wuhan/PPICAMS-WH-04/2019	GWIAHKL000000000	Genome Warehouse	2019.12.30	China/Hubei/Wuhan	29890
19	BetaCoV/Australia/VIC01/2020	MT097544	GenBank	2020.01.25	Australia/Victoria/Clayton	29893
20	BetaCoV/Sydney/2/2020	EPI_ISL_408976	GISAI	2020.01.22	Australia/New South Wales	29741
21	BetaCoV/Australia/NSW01/2020	EPI_ISL_407893	GISAI	2020.01.24	Australia/New South Wales	29782
22	BetaCoV/Sydney/3/2020	EPI_ISL_408977	GISAI	2020.01.25	Australia/New South Wales	29782
23	BetaCoV/Australia/VIC01/2020	EPI_ISL_406844	GISAI	2020.01.25	Australia/Victoria/Clayton	29893
24	BetaCoV/Cambodia/01/2/2020	EPI_ISL_411902	GISAI	2020.01.27	Cambodia/Sihanoukville	29873
25	BetaCoV/Canada/ON-VIC01/2020	EPI_ISL_411015	GISAI	2020.01.23	Canada/Ontario	29879
26	BetaCoV/Canada/ON-PIH1244/2020	EPI_ISL_411014	GISAI	2020.01.25	Canada/Ontario	29851
27	SARS-CoV-2/WH-09-human/2020/CHN	MT091631	GenBank	2020.01.08	China	29911
28	BetaCoV/Anhui/SZ005/2020	EPI_ISL_413485	GISAI	2020.01.24	China/Anhui/Suzhou	29860
29	SARS-CoV-2/105-human/2020/CHN	MT135041	GenBank	2020.01.26	China/Beijing	29903
30	SARS-CoV-2/233-human/2020/CHN	MT135044	GenBank	2020.01.28	China/Beijing	29903
31	SARS-CoV-2/233-human/2020/CHN	MT135043	GenBank	2020.01.28	China/Beijing	29903
32	SARS-CoV-2/231-human/2020/CHN	MT135042	GenBank	2020.01.28	China/Beijing	29903
33	Chongqing/IVDC-CQ-001/2020	EPI_ISL_404481	GISAI	2020.01.19	China/Chongqing	29891
34	BetaCoV/Taiwan/9/2020	EPI_ISL_411060	GISAI	2020.01.21	China/Taiwan	29871
35	BetaCoV/Taiwan/13/2020	EPI_ISL_411066	GISAI	2020.01.22	China/Taiwan	29891
36	SARS-CoV-2/QTC02-human/2020/CHN	MT123291	GenBank	2020.01.29	China/Guangdong / Guangzhou	29927
37	2019-nCoV_HKU1-SZ-009b_2020	MN8975262	GenBank	2020.01	China/Guangdong / Shenzhen	29891
38	2019-nCoV_HKU1-SZ-002a_2020	MN8918184	GenBank	2020.01.01	China/Guangdong / Shenzhen	29838
39	BetaCoV/Wuhan/20SF211/2020	EPI_ISL_406536	GISAI	2020.01.22	China/Guangdong Province	29803
40	BetaCoV/Wuhan/20SF210/2020	EPI_ISL_406535	GISAI	2020.01.22	China/Guangdong Province	29839
41	BetaCoV/Wuhan/20SF207/2020	EPI_ISL_406534	GISAI	2020.01.22	China/Guangdong Province	29819
42	BetaCoV/Guangdong/20SF174/2020	EPI_ISL_406531	GISAI	2020.01.22	China/Guangdong Province	29808
43	BetaCoV/Guangdong/20SF201/2020	EPI_ISL_406538	GISAI	2020.01.23	China/Guangdong Province	29839
44	BetaCoV/Guangdong/20SF206/2020	EPI_ISL_406533	GISAI	2020.01.22	China/Guangdong Province / Guangzhou City	29839
45	BetaCoV/Guangdong/20SF012/2020	EPI_ISL_403932	GISAI	2020.01.14	China/Guangdong Province / Shenzhen City	29845
46	BetaCoV/Guangdong/20SF025/2020	EPI_ISL_403935	GISAI	2020.01.15	China/Guangdong Province / Shenzhen City	29845
47	BetaCoV/Guangdong/20SF014/2020	EPI_ISL_403934	GISAI	2020.01.15	China/Guangdong Province / Shenzhen City	29845
48	BetaCoV/Guangdong/20SF011/2020	EPI_ISL_403933	GISAI	2020.01.15	China/Guangdong Province / Shenzhen City	29845

49	BetaCoV/Guangdong/20S1028/2020	EPI_ISL_401936	GISAID	2020-01-17	China/Guangdong Province / Zhuhai City	29845
50	BetaCoV/Guangdong/20S1040/2020	EPI_ISL_401937	GISAID	2020-01-18	China/Guangdong Province / Zhuhai City	29845
51	BetaCoV/Hong Kong/VM20001061/2020	EPI_ISL_412028	GISAID	2020-01-22	China/Hong Kong	29851
52	BetaCoV/Hong Kong/VM20001988/2020	EPI_ISL_412029	GISAID	2020-01-30	China/Hong Kong	29793
53	BetaCoV/Hangzhou/HHDC-HH-01/2020	EPI_ISL_412459	GISAID	2020-01-08	China/Hubei/Hangzhou	29848
54	BetaCoV/Wuhan/VIDC-HH-04/2020	EPI_ISL_402120	GISAID	2020-01-01	China/Hubei/Wuhan	29896
55	BetaCoV/Wuhan/HHDC-HH-02/2020	EPI_ISL_412978	GISAID	2020-01-17	China/Hubei/Wuhan	29846
56	BetaCoV/Wuhan/HHDC-HH-05/2020	EPI_ISL_412981	GISAID	2020-01-18	China/Hubei/Wuhan	29843
57	BetaCoV/Wuhan/HHDC-HH-04/2020	EPI_ISL_412980	GISAID	2020-01-18	China/Hubei/Wuhan	29845
58	BetaCoV/Wuhan/HHDC-HH-03/2020	EPI_ISL_412979	GISAID	2020-01-18	China/Hubei/Wuhan	29848
59	Jiangxi/VIDC-JX-002/2020	EPI_ISL_408486	GISAID	2020-01-12	China/Jiangxi/Pingxiang	29891
60	BetaCoV/Shandong/VIDC-SD-001/2020	EPI_ISL_408482	GISAID	2020-01-19	China/Shandong/Qingdao	29891
61	Sichuan/VIDC-SC-001/2020	EPI_ISL_408484	GISAID	2020-01-16	China/Sichuan/Chengdu	29891
62	Yunnan/VIDC-YN-003/2020	EPI_ISL_408480	GISAID	2020-01-18	China/Yunnan/Kunming	29891
63	BetaCoV/Zhejiang/WZ-01/2020	EPI_ISL_404227	GISAID	2020-01-16	China/Zhejiang	29839
64	BetaCoV/Zhejiang/WZ-02/2020	EPI_ISL_404228	GISAID	2020-01-17	China/Zhejiang	29859
65	BetaCoV/Hangzhou/HZCDC001/2020	EPI_ISL_407313	GISAID	2020-01-19	China/Zhejiang/Hangzhou	29836
66	2019-nCoV-HL	GWH1AHK500000000	Genome Warehouse	2020-01-20	China/Zhejiang/Hangzhou	29833
67	BetaCoV/Shenzhen/SZTH-002/2020	EPI_ISL_406593	GISAID	2020-01-13	China/Guangdong Province / Shenzhen City	29893
68	BetaCoV/Shenzhen/SZTH-003/2020	EPI_ISL_406594	GISAID	2020-01-16	China/Guangdong Province / Shenzhen City	29898
69	2019-nCoV-WHU02	MS900669	GenBank	2020-01-02	China/Hubei/Wuhan	29881
70	HU-1	MT019873	GenBank	2020-01-20	China/Hangzhou	29833
71	BetaCoV/Wuhan/WHICAMS-WH-03/2020	MT019833	Genome Warehouse	2020-01-01	China/Hubei/Wuhan	29883
72	BetaCoV/Wuhan/YS0011/2020	NMDC/601/3002-07	NMDC	2020-01-07	China/Hubei/Wuhan	29890
73	BetaCoV/Wuhan/WH19004/2020	NMDC/601/3002-09	NMDC	2020-01-01	China/Hubei/Wuhan	29896
74	BetaCoV/Wuhan/WH-03/2019	CNA0007334	CNGHLdb	2020-01-01	China/Hubei/Wuhan	29868
75	BetaCoV/Wuhan/WH-04/2019	CNA0007335	CNGHLdb	2020-01-03	China/Hubei/Wuhan	29872
76	2019-nCoV-WHU01	MS900668	GenBank	2020-01-02	China/Hubei/Wuhan	29881
77	Wuhan-Hu-1	MS900947	GenBank	2020-01-02	China/Hubei/Wuhan	29903
78	BetaCoV/Wuhan/WHICAMS-WH-05/2020	GWH1AHKJ000000000	GenBank	2020-01-01	China/Hubei/Wuhan	29883
79	BetaCoV/Jiangsu/JS01/2020	EPI_ISL_411953	GISAID	2020-01-24	China/Jiangsu	29891
80	BetaCoV/Jiangsu/JS02/2020	EPI_ISL_411952	GISAID	2020-01-24	China/Jiangsu	29891
81	BetaCoV/Jiangsu/JS01/2020	EPI_ISL_411950	GISAID	2020-01-23	China/Jiangsu	29891
82	SARS-CoV-2/NTU01/2020/TWN	MT066175	GenBank	2020-01-31	China/Taiwan/Taipei	29870
83	BetaCoV/Taiwan/4/2020	EPI_ISL_411927	GISAID	2020-01-28	China/Taiwan/Taipei	29890
84	BetaCoV/Taiwan/3/2020	EPI_ISL_411926	GISAID	2020-01-24	China/Taiwan/Taipei	29891
85	BetaCoV/Taiwan/CGMH-CGL-01/2020	EPI_ISL_411915	GISAID	2020-01-25	China/Taiwan/Taiwan	29862
86	SARS-CoV-2/Yunnan-01/human/2020/CHN	MT049951	GenBank	2020-01-17	China/Yunnan	29903
87	FranceIDH0626/2020	EPI_ISL_408431	GISAID	2020-01-30	France/Ile De France/Paris	29851
88	FranceIDH0515/2020	EPI_ISL_408430	GISAID	2020-01-30	France/Ile De France/Paris	29863
89	BetaCoV/FranceIDH072-id/2020	EPI_ISL_410720	GISAID	2020-01-23	France/Ile-de-France/Paris	29874
90	BetaCoV/FranceIDH0386/2020	EPI_ISL_411220	GISAID	2020-01-28	France/Ile-de-France/Paris	29878
91	BetaCoV/FranceIDH0173/2020	EPI_ISL_406597	GISAID	2020-01-23	France/Ile-de-France/Paris	29899
92	BetaCoV/FranceIDH0172/2020	EPI_ISL_406596	GISAID	2020-01-23	France/Ile-de-France/Paris	29851
93	BetaCoV/FranceIDH0515-id/2020	EPI_ISL_410984	GISAID	2020-01-29	France/Ile-de-France/Paris	29903
94	BetaCoV/FranceIDH0386/2020	EPI_ISL_411220	GISAID	2020-01-28	France/Ile-de-France/Paris	29872
95	BetaCoV/Germany/BavPat1/2020	EPI_ISL_406862	GISAID	2020-01-28	Germany/Bavaria/Munich	29782
96	BetaCoV/Italy/SPL1/2020	EPI_ISL_412974	GISAID	2020-01-29	Italy / Rome	29903
97	BetaCoV/Italy/INMI1-id/2020	EPI_ISL_410545	GISAID	2020-01-29	Italy/Rome	29867
98	BetaCoV/Italy/INMI1-cv/2020	EPI_ISL_410546	GISAID	2020-01-31	Italy/Rome	29785
99	2019-nCoV/Japan/KYN-029/2020	LC322972	GenBank	2020-01-29	Japan/Kyoto	29876
100	BetaCoV/Japan/NA-20-05-1/2020	EPI_ISL_410531	GISAID	2020-01-25	Japan/Nara	29878
101	BetaCoV/Japan/OS-20-07-1/2020	EPI_ISL_410532	GISAID	2020-01-23	Japan/Osaka	29878
102	2019-nCoV/Japan/TY/AWK-521/2020	LC322975	GenBank	2020-01-31	Japan/Tokyo	29878
103	2019-nCoV/Japan/TY/AWK-501/2020	LC322974	GenBank	2020-01-31	Japan/Tokyo	29878
104	2019-nCoV/Japan/TY/AWK-012/2020/RNA	LC322973	GenBank	2020-01-29	Japan/Tokyo	29878
105	BetaCoV/Korea/KCDC03/2020	EPI_ISL_407193	GISAID	2020-01-25	Korea/Gyeonggi-do	29818
106	SARS-CoV-2/61-TW/human/2020/NPI	EPI_ISL_410301	GISAID	2020-01-13	Nepal/Kathmandu	29811
107	BetaCoV/Singapore/1/2020	EPI_ISL_406973	GISAID	2020-01-23	Singapore	29845
108	BetaCoV/Singapore/7/2020	EPI_ISL_410713	GISAID	2020-01-27	Singapore	29891
109	BetaCoV/Singapore/2/2020	EPI_ISL_407987	GISAID	2020-01-25	Singapore	29844
110	SNU01	MT039890	GenBank	2020-01-01	South Korea	29903
111	BetaCoV/Korea/KCDC07/2020	EPI_ISL_412871	GISAID	2020-01-31	South Korea / Seoul	29789
112	BetaCoV/Korea/KCDC05/2020	EPI_ISL_412869	GISAID	2020-01-30	South Korea/Seoul	29789
113	BetaCoV/Korea/KCDC06/2020	EPI_ISL_412870	GISAID	2020-01-30	South Korea/ Seoul	29789
114	BetaCoV/Nonthaburi/61/2020	EPI_ISL_401962	GISAID	2020-01-08	Thailand/Nonthaburi Province	29848
115	BetaCoV/Nonthaburi/74/2020	EPI_ISL_401963	GISAID	2020-01-03	Thailand/Nonthaburi Province	29839
116	BetaCoV/England/02/2020	EPI_ISL_407073	GISAID	2020-01-29	United Kingdom	29843
117	2019-nCoV/USA-AZ1/2020	MS997409	GenBank	2020-01-22	United States / Arizona / Phoenix	29882
118	2019-nCoV/USA-CA1/2020	MS994467	GenBank	2020-01-23	United States / California / Los Angeles	29882

119	2019-nCoV/USA-CA2/2020	MN994468	GenBank	2020.01.22	United States / California / Orange County	29883
120	2019-nCoV/USA-WA1/2020	MN985325	GenBank	2020.01.19	United States / Washington / Snohomish County	29882
121	2019-nCoV/USA-AZ1/2020	EPI_ISL_406223	GISAI	2020.01.22	United States/Arizona	29882
122	2019-nCoV/USA-CA2/2020	EPI_ISL_406036	GISAI	2020.01.22	United States/California	29883
123	2019-nCoV/USA-CA1/2020	EPI_ISL_406034	GISAI	2020.01.23	United States/California	29882
124	2019-nCoV/USA-WA1/2020	EPI_ISL_404895	GISAI	2020.01.19	United States/Washington	29882
125	2019-nCoV/USA-CA6/2020	MT044258	GenBank	2020.01.27	USA/CA	29858
126	2019-nCoV/USA-CA5/2020	MT027064	GenBank	2020.01.29	USA/California	29882
127	2019-nCoV/USA-CA4/2020	MT027063	GenBank	2020.01.29	USA/California	29882
128	2019-nCoV/USA-CA3/2020	MT027062	GenBank	2020.01.29	USA/California	29882
129	2019-nCoV/USA-IL2/2020	MT044257	GenBank	2020.01.28	USA/IL	29882
130	2019-nCoV/USA-MA1/2020	MT019088	GenBank	2020.01.29	USA/MA	29882
131	BetaCoV/USA/WA1-1/6/2020	MT020881	GenBank	2020.01.25	USA/Washington	29882
132	BetaCoV/USA/WA1-1/12/2020	MT020880	GenBank	2020.01.25	USA/Washington	29882
133	2019-nCoV/USA-WI1/2020	MT019087	GenBank	2020.01.31	USA/Wisconsin	29879
134	BetaCoV/Vietnam/VR01-18142/2020	EPI_ISL_404668	GISAI	2020.01.24	Vietnam/Thanh Hoa	29839
135	BetaCoV/Australia/NSW03/2020	EPI_ISL_412975	GISAI	2020.02.28	Australia / New South Wales / Sydney	29782
136	BetaCoV/Australia/NSW07/2020	EPI_ISL_413214	GISAI	2020.02.29	Australia / New South Wales / Sydney	29782
137	BetaCoV/Australia/NSW06/2020	EPI_ISL_413213	GISAI	2020.02.29	Australia / New South Wales / Sydney	29782
138	BetaCoV/Australia/QLD03/2020	EPI_ISL_410717	GISAI	2020.02.05	Australia/Queensland	29665
139	BetaCoV/Australia/QLD04/2020	EPI_ISL_410718	GISAI	2020.02.05	Australia/Queensland	29665
140	BetaCoV/Belgium/GHB-03021/2020	EPI_ISL_407076	GISAI	2020.02.03	Belgium / Leuven	29882
141	SARS-CoV-2/SPI02/Inman/2020/IRA	MT126078	GenBank	2020.02.28	Brazil	29876
142	BetaCoV/Brazil/SPI02/2020	EPI_ISL_413016	GISAI	2020.02.28	Brazil / Sao Paulo/ Sao Paulo	29903
143	BetaCoV/Hefei/2/2020	EPI_ISL_412026	GISAI	2020.02.23	China / Anhui / Hefei	29870
144	SARS-CoV-2/QTC01/Inman/2020/CHN	MT123290	GenBank	2020.02.05	China / Guangdong / Guangzhou	29891
145	BetaCoV/Hong Kong/VH2002666/2020	EPI_ISL_412030	GISAI	2020.02.01	China / Hong Kong	29843
146	BetaCoV/Taiwan/HKDC-HB-07/2020	EPI_ISL_412983	GISAI	2020.02.08	China / Hubei / Taiwan	29697
147	BetaCoV/Wuhan/HKDC-HB-06/2020	EPI_ISL_412982	GISAI	2020.02.07	China / Hubei / Wuhan	29838
148	SARS-CoV-2/NT103/2020/TWN	MT066176	GenBank	2020.02.04	China/Taiwan	29870
149	Taiwan/NTU01/2020	MT066175	GenBank	2020.02.01	China/Taiwan/Taipei	29903
150	BetaCoV/France/ID10371/2020	EPI_ISL_411218	GISAI	2020.02.02	France/Paris	29873
151	BetaCoV/Germany/Baden-Wuerttemberg-1/2020	EPI_ISL_412912	GISAI	2020.02.25	Germany/Baden-Wuerttemberg	29756
152	BetaCoV/Germany/NRW-01/2020	EPI_ISL_413488	GISAI	2020.02.28	Germany/North Rhine Westphalia/Heinberg District	29903
153	BetaCoV/Italy/CDG1/2020	EPI_ISL_412973	GISAI	2020.02.20	Italy/Lombardy	29903
154	SARS-CoV-2/HuDP/Knp/19-027	LC528233	GenBank	2020.02.10	Japan/Kanagawa	29902
155	SARS-CoV-2/HuDP/Knp/19-020	LC528232	GenBank	2020.02.10	Japan/Kanagawa	29902
156	BetaCoV/Japan/TK/20-11-1/2020	EPI_ISL_413459	GISAI	2020.02.20	Japan/Tokyo	29835
157	BetaCoV/Mexico/CDMX/InfIRU_01/2020	EPI_ISL_412972	GISAI	2020.02.27	Mexico/Mexico City	29865
158	BetaCoV/New Zealand/01/2020	EPI_ISL_413490	GISAI	2020.02.27	New Zealand/Auckland	29903
159	BetaCoV/Singapore/11/2020	EPI_ISL_410719	GISAI	2020.02.02	Singapore	29845
160	BetaCoV/Singapore/6/2020	EPI_ISL_410537	GISAI	2020.02.09	Singapore	29829
161	BetaCoV/Singapore/4/2020	EPI_ISL_410535	GISAI	2020.02.03	Singapore	29809
162	BetaCoV/Singapore/5/2020	EPI_ISL_410536	GISAI	2020.02.06	Singapore	29841
163	BetaCoV/Singapore/3/2020	EPI_ISL_407988	GISAI	2020.02.01	Singapore	29843
164	BetaCoV/Singapore/10/2020	EPI_ISL_410716	GISAI	2020.02.04	Singapore	29839
165	BetaCoV/Singapore/9/2020	EPI_ISL_410715	GISAI	2020.02.04	Singapore	29891
166	BetaCoV/Singapore/8/2020	EPI_ISL_410714	GISAI	2020.02.03	Singapore	29758
167	BetaCoV/South Korea/KUMC02/2020	EPI_ISL_413018	GISAI	2020.02.06	South Korea	29838
168	BetaCoV/South Korea/KUMC01/2020	EPI_ISL_413017	GISAI	2020.02.06	South Korea	29903
169	BetaCoV/Korea/KCDC24/2020	EPI_ISL_412873	GISAI	2020.02.06	South Korea/Chungcheongnam-do	29789
170	BetaCoV/Korea/KCDC12/2020	EPI_ISL_412872	GISAI	2020.02.01	South Korea/Gyeonggi-do	29789
171	SARS-CoV-2/01/Inman/2020/SWE	MT093571	GenBank	2020.02.07	Sweden	29836
172	SARS-CoV-2/01/Inman/2020/SWE	EPI_ISL_411951	GISAI	2020.02.07	Sweden	29836
173	BetaCoV/Switzerland/1000477377/2020	EPI_ISL_413020	GISAI	2020.02.27	Switzerland/Zurich	29864
174	BetaCoV/Switzerland/1000477806/2020	EPI_ISL_413024	GISAI	2020.02.29	Switzerland/Zurich	29871
175	BetaCoV/Switzerland/1000477797/2020	EPI_ISL_413023	GISAI	2020.02.29	Switzerland/Zurich	29788
176	BetaCoV/Switzerland/1000477796/2020	EPI_ISL_413022	GISAI	2020.02.29	Switzerland/Zurich	29845
177	BetaCoV/Switzerland/1000477757/2020	EPI_ISL_413021	GISAI	2020.02.29	Switzerland/Zurich	29875
178	BetaCoV/Finland/03/2020	EPI_ISL_412116	GISAI	2020.02.09	United Kingdom	29903
179	2019-nCoV/USA-CA7/2020	MT109052	GenBank	2020.02.06	United States/California	29882
180	2019-nCoV/USA-CA8/2020	MT109053	GenBank	2020.02.10	United States/California	29882
181	2019-nCoV/USA-CA9/2020	MT118815	GenBank	2020.02.23	United States/California / Solano	29882
182	2019-nCoV/USA-TX1/2020	MT109054	GenBank	2020.02.11	United States/Texas	29882
183	BetaCoV/USA/WA3-UW1/2020	EPI_ISL_413025	GISAI	2020.02.27	United States/Washington	29903
184	BetaCoV/USA/WA4-UW2/2020	EPI_ISL_413455	GISAI	2020.02.28	United States/Washington	29897
185	BetaCoV/USA/WA6-UW3/2020	EPI_ISL_413457	GISAI	2020.02.29	United States/Washington	29903
186	BetaCoV/USA/WA-S2/2020	EPI_ISL_413456	GISAI	2020.02.20	United States/Washington / King County	29903
187	BetaCoV/USA/WA2/2020	MT153824	GenBank	2020.02.24	United States/Washington / Snohomish County	29878
188	2019-nCoV/USA-CA8/2020	EPI_ISL_411955	GISAI	2020.02.10	United States/California	29882

189	2019-nCoV/USA-CA7/2020	EPI_ISL_411955	GISAID	2020.02.06	United States/California	29882
190	2019-nCoV/USA-TX1/2020	EPI_ISL_411956	GISAID	2020.02.11	United States/Texas	29882
191	BetaCoV/Italy/UniSR1/2020	EPI_ISL_413489	GISAID	2020.03.03	Italy/Lombardy/Milan	29887
192	BetaCoV/USA/WA7-UW4/2020	EPI_ISL_413458	GISAID	2020.03.01	United States/Washington	29903
193	BetaCoV/USA/WA9-UW6/2020	EPI_ISL_413487	GISAID	2020.03.01	United States/Washington	29325
194	BetaCoV/USA/WA8-UW5/2020	EPI_ISL_413486	GISAID	2020.03.01	United States/Washington	29732

Table 2 Detailed information of the HCov-229E strains

No.	The virus strains	Serial No.	Data source	Sampling time	Sampling location	Sequence length
1	HCov-229E/BN1/GER/2015	KU291448	GenBank	2015	Germany	27022
2	J0304	JX503061	GenBank	2009	Italy	27208
3	0349	JX503060	GenBank	2010	Netherlands	27263

Table 3 Detailed information of the HCov-HKU1 strains

No.	The virus strains	Serial No.	Data source	Sampling time	Sampling location	Sequence length
1	SI17244	MH940245	GenBank	2017.6.4	Thailand	29811
2	BJ01-p9	KT779556	GenBank	2009.4.23	China	29887
3	BJ01-p3	KT779555	GenBank	2009.4.23	China	29887
4	Caen1	HM034837	GenBank	2005.3.4	France	29862

Table 4 Detailed information of the HCov-NL63 strains

No.	The virus strains	Serial No.	Data source	Sampling time	Sampling location	Sequence length
1	496	DQ445912	GenBank	2015	Netherlands	27535
2	057	DQ445911	GenBank	2004	Netherlands	27550
3	HCov NL63/Haiti-1/2015	KT266906	GenBank	2015.1.16	Haiti	27553
4	Kilifi HH 3808 24-May-2010	MG428706	GenBank	2010.5.24	Kenya	27637
5	Kilifi HH 0522 21-May-2010	MG428705	GenBank	2010.5.21	Kenya	27587
6	Kilifi HH 5402 20-May-2010	MG428704	GenBank	2010.5.20	Kenya	27832
7	Kilifi HH 0511 01-Jun-2010	MG428703	GenBank	2010.6.1	Kenya	27685
8	Kilifi HH 3807 11-May-2010	MG428702	GenBank	2010.5.11	Kenya	27781
9	Kilifi HH 0512 04-Jun-2010	MG428701	GenBank	2010.6.4	Kenya	27623
10	Kilifi HH 5709 19-May-2010	MG428699	GenBank	2010.5.19	Kenya	27513
11	CN0601/14	MG772808	GenBank	2014.11.25	South Korea	27535
12	NL63/UF-2/2015	KU521535	GenBank	2015.9.1	USA	27553
13	NL63/UF-1/2015	KT381875	GenBank	--	USA	27553
14	NL63/RPTEC/2004	JX504050	GenBank	--	USA	27553
15	CBJ123	JX524171	GenBank	2009.1.24	China	27538
16	CBJ037	JX104161	GenBank	2008.11.4	China / Beijing	27538

Note: "--" indicates that the information is not given, which is the same for the rest of the tables.

Table 5 Detailed information of the HCov-OC43 strains

No.	The virus strains	Serial No.	Data source	Sampling time	Sampling location	Sequence length
1	MDS16	MK303625	GenBank	--	--	30668
2	MDS14	MK303624	GenBank	--	--	30664
3	MDS12	MK303623	GenBank	--	--	30419
4	MDS11	MK303622	GenBank	--	--	30665
5	MDS4	MK303621	GenBank	--	--	30664
6	MDS2	MK303620	GenBank	--	--	30665
7	MDS6	MK303619	GenBank	--	--	30664
8	HCov-OC43/USA/ACRI 0213/2016	MH121121	GenBank	2016.12.19	USA/Little Rock	30722
9	HCov-OC43/USA/ACRI 0052/2016	MF314143	GenBank	2016.3.7	USA/Little Rock	30737
10	HCov-OC43/USA/TCNP 00212/2017	MF374985	GenBank	2017.1.17	USA/Albuquerque	30705
11	HCov-OC43/USA/TCNP 0070/2016	MF374983	GenBank	2016.2.1	USA/Albuquerque	30721
12	HCov-OC43/USA/TCNP 00204/2017	MF374984	GenBank	2017.1.3	USA/Albuquerque	30713
13	OC43_KLF_01_2018	MN026164	GenBank	2018.1.18	Kenya	30777
14	TNP 12643	MG977452	GenBank	2016.12.10	Cote d'Ivoire	30704
15	TNP 12636	MG977451	GenBank	2016.12.10	Cote d'Ivoire	30722
16	TNP F1834_2	MG977449	GenBank	2016.12.28	Cote d'Ivoire	30721
17	TNP F1832_2	MG977447	GenBank	2016.12.26	Cote d'Ivoire	30705

18	TNP F1790_2	MG977445	GenBank	2016.12.31	Cote d'Ivoire	30677
19	TNP F1778_2	MG977444	GenBank	2017.1.4	Cote d'Ivoire	30571
20	2007-09	KY014282	GenBank	2007	France	30659
21	2002-04	KY014281	GenBank	2002	France	30602
22	MY-U1975/13	KX538979	GenBank	2013.2.15	Malaysia	30713
23	MY-U1758/13	KX538978	GenBank	2013.1.2	Malaysia	30713
24	MY-U1140/12	KX538977	GenBank	2012.9.10	Malaysia	30716
25	MY-U1057/12	KX538976	GenBank	2012.8.27	Malaysia	30713
26	MY-U1024/12	KX538975	GenBank	2012.8.24	Malaysia	30716
27	MY-U945/12	KX538974	GenBank	2012.8.1	Malaysia	30716
28	MY-U868/12	KX538973	GenBank	2016.7.12	Malaysia	30716
29	MY-U774/12	KX538972	GenBank	2012.7.4	Malaysia	30713
30	MY-U732/12	KX538971	GenBank	2012.6.25	Malaysia	30716
31	MY-U710/12	KX538970	GenBank	2012.6.20	Malaysia	30713
32	MY-U523/12	KX538969	GenBank	2012.5.18	Malaysia	30716
33	MY-U464/12	KX538968	GenBank	2012.5.9	Malaysia	30716
34	MY-U413/12	KX538967	GenBank	2012.5.2	Malaysia	30716
35	MY-U236/12	KX538966	GenBank	2012.4.2	Malaysia	30713
36	MY-U208/12	KX538965	GenBank	2012.3.28	Malaysia	30716
37	MY-U002/12	KX538964	GenBank	2012.2.22	Malaysia	30713
38	LRTI_238	KX344031	GenBank	2011.2.9	Mexico	30713
39	10574/2010	KF923925	GenBank	2010.9	China/Beijing	30713
40	10290/2010	KF923924	GenBank	2010.7	China/Beijing	30713
41	892A/2008	KF923923	GenBank	2008.1	China/Beijing	30713
42	8164/2009	KF923922	GenBank	2009.3	China/Beijing	30713
43	69A/2007	KF923921	GenBank	2007.5	China/Beijing	30713
44	5617/2007	KF923920	GenBank	2007.7	China/Beijing	30713
45	5595/2007	KF923919	GenBank	2007.7	China/Beijing	30713
46	10108/2010	KF923918	GenBank	2010.5	China/Beijing	30713
47	5566/2007	KF923917	GenBank	2007.6	China/Beijing	30713
48	5519/2007	KF923916	GenBank	2007.6	China/Beijing	30713
49	5517/2017	KF923915	GenBank	2007.6	China/Beijing	30713
50	5508/2007	KF923914	GenBank	2007.6	China/Beijing	30713
51	5485/2007	KF923913	GenBank	2007.6	China/Beijing	30713
52	5484/2007	KF923912	GenBank	2007.6	China/Beijing	30713
53	5479/2007	KF923911	GenBank	2007.6	China/Beijing	30713
54	5445/2007	KF923910	GenBank	2007.6	China/Beijing	30713
55	5442/2007	KF923909	GenBank	2007.6	China/Beijing	30713
56	5414/2007	KF923908	GenBank	2007.6	China/Beijing	30713
57	5370/2007	KF923907	GenBank	2007.5	China/Beijing	30713
58	229/2005	KF923905	GenBank	2005.6	China/Beijing	30722
59	12694/2012	KF923904	GenBank	2012.5	China/Beijing	30713
60	12691/2012	KF923903	GenBank	2012.5	China/Beijing	30713
61	12689/2012	KF923902	GenBank	2012.5	China/Beijing	30713
62	5472/2007	KF923901	GenBank	2007.6	China/Beijing	30713
63	3582/2006	KF923899	GenBank	2006.9	China/Beijing	30722
64	3184A/2012	KF923898	GenBank	2012.3	China/Beijing	30716
65	3269A/2012	KF923897	GenBank	2012.6	China/Beijing	30711
66	10285/2010	KF923895	GenBank	2010.7	China/Beijing	30731
67	5352/2007	KF923894	GenBank	2007.5	China/Beijing	30713
68	2151A/2010	KF923893	GenBank	2010.7	China/Beijing	30713
69	5345/2007	KF923892	GenBank	2007.5	China/Beijing	30713
70	39A/2007	KF923890	GenBank	2007.4	China/Beijing	30713
71	1926/2006	KF923889	GenBank	2006.3	China/Beijing	30719
72	1997A/2010	KF923887	GenBank	2010.4	China/Beijing	30719
73	1908A/2010	KF923886	GenBank	2010.3	China/Beijing	30719
74	3194A/2012	KF923906	GenBank	2012.3	China/Beijing	30737
75	3647/2006	KF923900	GenBank	2006.1	China/Beijing	30722
76	3074A/2012	KF923896	GenBank	2012.2	China/Beijing	30737
77	5240/2007	KF923891	GenBank	2007.5	China/Beijing	30713
78	2145A/2010	KF923888	GenBank	2010.7	China/Beijing	30719

References

- [1] Li Y, Liu B, Cui J, et al. Similarities and evolutionary relationships of COVID-19 and related viruses[J]. arXiv preprint arXiv:2003.05580, 2020.
- [2] Yanni Li, Bing Liu, Zhi Wang, Jiangtao Cui, et al. 2020. COVID-19 Evolves in Human Hosts[C]. In Proceedings of the 26st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 1-10.

2. Detailed information of big sequences in Practical Use Case 2

We selected 11 publicly available complete genome sequences (length $\geq 10,000$) obtained through the latest sequencing technology from liver cancer patients. These sequences are all from the NCBI GenBank database, which can be searched and downloaded using their corresponding serial number (No.). The collected 11 sequences are presented in Table 6 below.

Table 6 Detailed information of the liver cancer whole genome sequencing

No.	Whole genome sequencing name	Serial No.	Data source	Sampling time	Sequence length
1	primary tumor of HCC patient 1	SRR2157816	GeneBank	2016-05-05	10000
2	primary tumor of HCC patient 2	SRR2157821	GeneBank	2016-05-05	10000
3	primary tumor of HCC patient 3	SRR2157826	GeneBank	2016-05-05	10000
4	primary tumor of HCC patient 4	SRR2157832	GeneBank	2016-05-05	10000
5	primary tumor of HCC patient 5	SRR2157838	GeneBank	2016-05-05	10000
6	primary tumor of HCC patient 6	SRR2157843	GeneBank	2016-05-05	10000
7	primary tumor of HCC patient 7	SRR2157849	GeneBank	2016-05-05	10000
8	primary tumor of HCC patient 8	SRR2157858	GeneBank	2016-05-05	10000
9	primary tumor of HCC patient 9	SRR2157863	GeneBank	2016-05-05	10000
10	primary tumor of HCC patient 10	SRR2157866	GeneBank	2016-05-05	10000
11	portal vein tumor thrombosis NO.1 of HCC patient 6	SRR2157846	GeneBank	2016-05-05	10000