

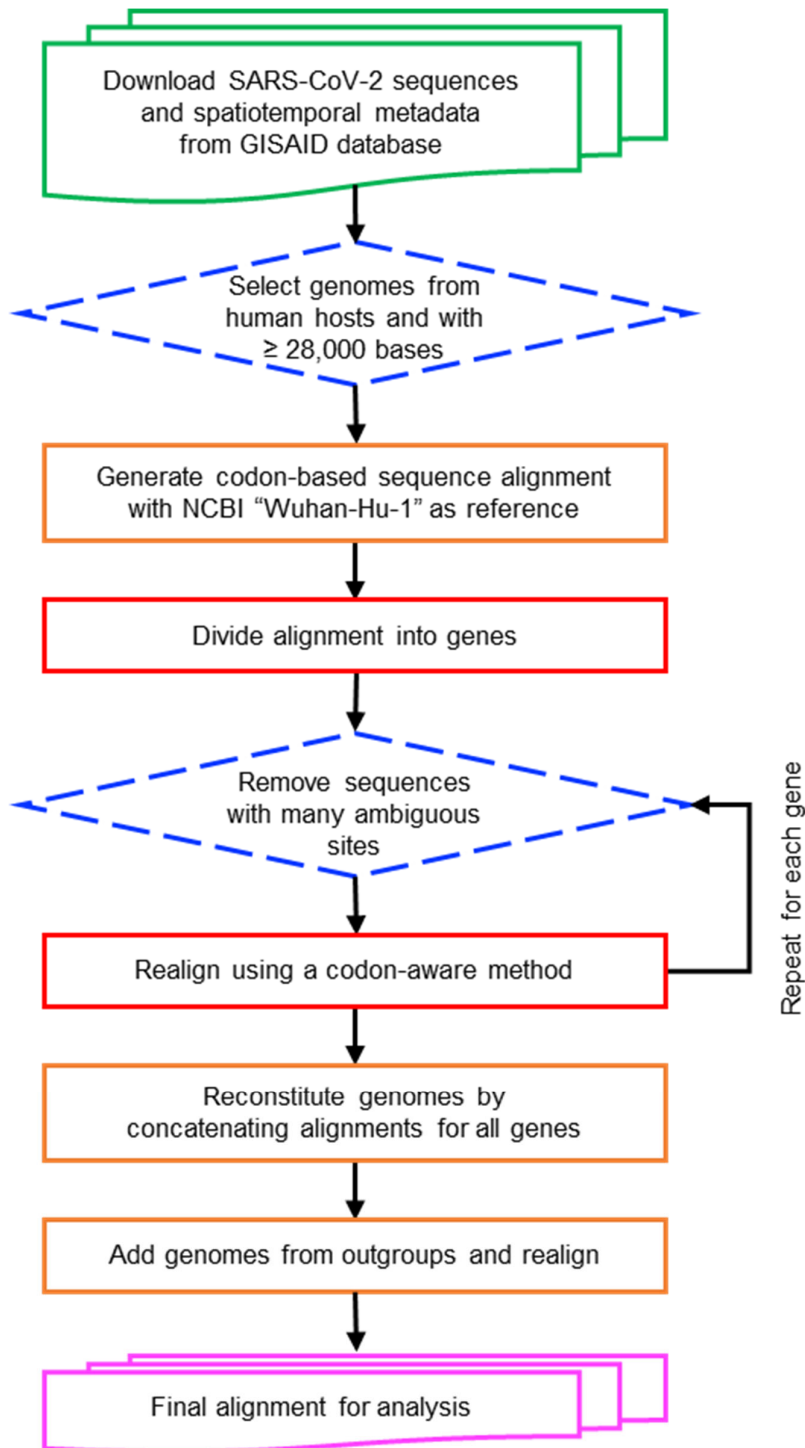
Supplementary Table 1. SARS-CoV-2 variants in the 68KG dataset.

Mutant (major)	Mutant (minor)	Gene	Genomic Position	Nucleotide change	Amino acid change	Time (days)	Variant Frequency	Genomes mapped	First location
$\mu_1$		ORF1ab	2416	U>C		0	98.4%	0	China, Asia
$\mu_2$		ORF1ab	19524	U>C		0	99.0%	18	China, Asia
$\mu_3$		S	23929	U>C		0	98.9%	0	China, Asia
$\mu_4$		ORF1ab	15933	U>C		0	98.8%	0	China, Asia
$\mu_5$		ORF8	27944	U>C		0	97.0%	0	China, Asia
$\mu_6$		ORF1ab	6286	U>C		0	95.6%	0	China, Asia
$\mu_7$		S	22444	U>C		0	98.7%	0	China, Asia
$\alpha_1$		ORF1ab	18060	U>C		0	97.3%	1114	China, Asia
	$\alpha_{1a}$	N	28657	C>U		63	1.0%	3	France, Europe
	$\alpha_{1b}$	ORF1ab	9477	U>A	F>Y	63	0.7%	3	France, Europe
	$\alpha_{1c}$	N	28863	C>U	S>L	63	0.7%	7	France, Europe
	$\alpha_{1d}$	ORF3a	25979	G>U	G>V	63	0.7%	451	France, Europe
$\alpha_2$		ORF1ab	8782	U>C		0	94.9%	51	China, Asia
$\alpha_3$		ORF8	28144	C>U	S>L	0	94.9%	1281	China, Asia
	$\alpha_{3a}$	ORF1ab	1606	U>C		43	0.9%	578	United Kingdom, Europe
	$\alpha_{3b}$	ORF1ab	11083	G>U	L>F	24	7.5%	417	China, Asia
	$\alpha_{3c}$	N	28311	C>U	P>L	64	1.4%	4	South Korea, Asia
	$\alpha_{3d}$	ORF1ab	13730	C>U	A>V	33	1.4%	5	China, Asia
	$\alpha_{3e}$	ORF1ab	6312	C>A	T>K	71	1.2%	767	Taiwan, Asia
	$\alpha_{3f}$	ORF3a	26144	G>U	G>V	28	3.0%	160	China, Asia
	$\alpha_{3g}$	ORF1ab	14805	C>U		54	3.7%	511	United Kingdom, Europe
	$\alpha_{3h}$	ORF1ab	17247	U>C		64	1.0%	682	Switzerland, Europe
	$\alpha_{3i}$	ORF1ab	2558	C>U	P>S	54	1.0%	44	United Kingdom, Europe
	$\alpha_{3j}$	ORF1ab	2480	A>G	I>V	54	1.0%	648	United Kingdom, Europe
$\beta_1$		ORF1ab	3037	C>U		31	87.2%	45	China, Asia
$\beta_2$		S	23403	A>G	D>G	31	87.2%	15	China, Asia
$\beta_3$		ORF1ab	14408	C>U	P>L	41	87.1%	4450	Saudi Arabia, Middle East
	$\beta_{3a}$	ORF1ab	20268	A>G		64	6.0%	2388	Italy, Europe
	$\beta_{3b}$	N	28854	C>U	S>L	29	4.5%	1782	China, Asia
	$\beta_{3c}$	ORF1ab	15324	C>U		29	2.2%	1463	China, Asia
	$\beta_{3d}$	ORF3a	25429	G>U	V>L	77	1.1%	719	United Kingdom, Europe
	$\beta_{3e}$	N	28836	C>U	S>L	74	0.8%	3	Switzerland, Europe
	$\beta_{3f}$	ORF1ab	13862	C>U	T>I	74	0.8%	85	Switzerland, Europe
	$\beta_{3g}$	ORF1ab	10798	C>A		86	0.6%	435	United Kingdom, Europe
$\gamma_1$		ORF3a	25563	G>U	Q>H	41	24.4%	1671	Saudi Arabia, Middle East
	$\gamma_{1a}$	ORF1ab	18877	C>U		41	4.2%	1201	Saudi Arabia, Middle East
	$\gamma_{1b}$	M	26735	C>U		41	2.7%	1784	Saudi Arabia, Middle East
$\delta_1$		ORF1ab	1059	C>U	T>I	54	17.6%	8284	Singapore, Asia
	$\delta_{1a}$	S	24368	G>U	D>Y	75	0.7%	466	Sweden, Europe
	$\delta_{1b}$	ORF8	27964	C>U	S>L	76	2.9%	1152	USA, North America
	$\delta_{1c}$	ORF1ab	11916	C>U	S>L	72	1.9%	807	USA, North America
	$\delta_{1d}$	ORF1ab	18998	C>U	A>V	72	0.7%	458	USA, North America
	$\delta_{1e}$	ORF1ab	10319	C>U	L>F	76	1.2%	799	USA, North America
$\zeta_1$		ORF1ab	445	U>C		179	4.4%	18	Netherlands, Europe
$\zeta_2$		M	26801	C>G		82	4.3%	7	Canada, North America
$\zeta_3$		S	22227	C>U	A>V	84	4.5%	1	Spain, Europe
$\zeta_4$		N	28932	C>U	A>V	96	4.4%	5	Portugal, Europe
$\zeta_5$		ORF10	29645	G>U	V>L	78	4.4%	2	Denmark, Europe
$\zeta_6$		ORF1ab	21255	G>C		80	4.4%	1557	USA, North America
$\zeta_7$		S	21614	C>U	L>F	79	2.5%	1442	United Kingdom, Europe

**Supplementary Table 1.** SARS-CoV-2 variants in the 68KG dataset (continued).

Mutant (major)	Mutant (minor)	Gene	Genomic Position	Nucleotide change	Amino acid change	Time (days)	Variant Frequency	Genomes mapped
ε <sub>1</sub>		N	28881	G>A	R>K	54	41.7%	5
ε <sub>2</sub>		N	28882	G>A	R>K	54	41.6%	0
ε <sub>3</sub>		N	28883	G>C	G>R	54	41.6%	13394
	ε <sub>3a</sub>	ORF1ab	313	C>U		64	2.4%	1630
	ε <sub>3b</sub>	ORF1ab	19839	U>C		64	2.9%	1227
	ε <sub>3c</sub>	M	27046	C>U	T>M	69	0.8%	548
	ε <sub>3d</sub>	ORF1ab	10097	G>A	G>S	69	3.2%	11
	ε <sub>3e</sub>	S	23731	C>U		69	3.2%	425
	ε <sub>3f</sub>	N	28580	G>U	D>Y	69	1.0%	678
	ε <sub>3g</sub>	ORF1ab	13536	C>U		69	1.6%	23
	ε <sub>3h</sub>	ORF1ab	4002	C>U	T>I	69	1.6%	1066
	ε <sub>3i</sub>	ORF1ab	10265	G>A	G>S	63	1.4%	879
	ε <sub>3j</sub>	S	21575	C>U	L>F	54	1.0%	248
	ε <sub>3k</sub>	S	21637	C>U		111	1.3%	873
	ε <sub>3l</sub>	ORF8	28169	A>G		103	1.3%	0
	ε <sub>3m</sub>	ORF1ab	16968	G>U		114	1.0%	702
η <sub>1</sub>		ORF1ab	1163	A>U	I>F	86	9.6%	339
	η <sub>1a</sub>	ORF1ab	14202	G>U		159	1.1%	7
	η <sub>1b</sub>	ORF1ab	19542	G>U	M>I	81	1.2%	23
	η <sub>1c</sub>	S	22388	C>U		90	1.2%	21
	η <sub>1d</sub>	N	29466	C>U	A>V	91	1.2%	4
	η <sub>1e</sub>	ORF1ab	19718	C>U	T>I	73	1.5%	23
	η <sub>1f</sub>	ORF3a	26060	C>U	T>I	92	1.2%	7
	η <sub>1g</sub>	N	29227	G>U		55	1.2%	24
	η <sub>1h</sub>	ORF1ab	3256	U>C		167	1.1%	0
	η <sub>1i</sub>	ORF1ab	5622	C>U	P>L	67	1.2%	775
η <sub>2</sub>		ORF1ab	18555	C>U		51	8.0%	25
η <sub>3</sub>		ORF1ab	16647	G>U		84	8.0%	8
η <sub>4</sub>		ORF1ab	7540	U>C		86	7.9%	0
η <sub>5</sub>		S	23401	G>A		86	7.9%	1
η <sub>6</sub>		S	22992	G>A	S>N	86	8.5%	4583
	η <sub>6a</sub>	S	22480	C>U		66	1.3%	878
v <sub>1</sub>		ORF1ab	17858	A>G	Y>C	59	2.6%	61
v <sub>2</sub>		ORF1ab	17747	C>U	P>L	59	2.5%	1677

Note.- Genomic locations correspond to those of the NCBI genome (GenBank ID: NC\_04551.2). Amino acid changes are shown for nonsynonymous variants.

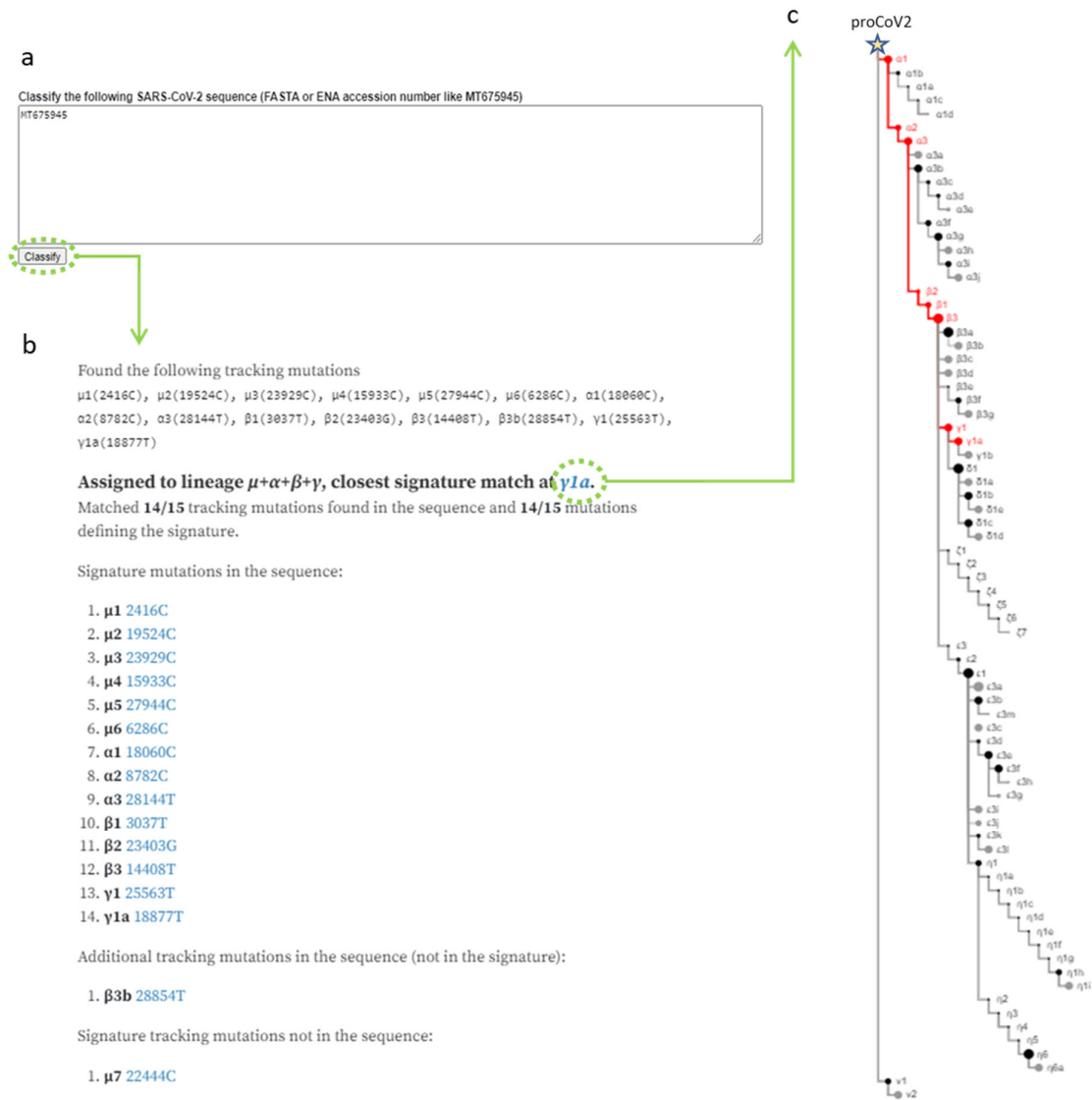


**Supplementary Figure 1.** Flowchart summarizing the procedure to generate two curated multiple sequence alignments used in this study (29KG and 68KG) from the SARS-CoV-2 genomes available in the GISAID database. Abbreviations are as follows: GISAID (Global Initiative on Sharing Avian Influenza Data [<https://www.gisaid.org>]); NCBI (National Center for Biotechnology Information [<https://www.ncbi.nlm.nih.gov>]).



**Supplementary Figure 2.** Flowchart summarizing the steps performed to construct the mutation order tree from the curated SARS-CoV-2 alignments (i.e., 29KG and 68KG). Abbreviations are as follows: GISAID (Global Initiative on Sharing Avian Influenza Data); SCITE (Single Cell Inference of Tumor Evolution); FPR (False Positive Rates); FNR (False Negative Rates); COI (Co-Occurrence Index).





**Supplementary Figure 4.** An example of sequence classification (ENA Accession MT675945) based on the 84 signature mutations (<http://sars2evo.datamonkey.org/>; “Classify your Sequence” option). (a) Input window to provide identifiers of sequences to be classified (e.g., MT675945). (b) The input sequence is classified into a mutational fingerprint. A list of mutations that are appeared in the input sequence is shown in the output window. (c) A waterfall phylogeny shows the input sequence's location in the phylogeny, which appears after clicking the closet signature matched mutation in panel b.