

Supplementary Table S3.

Lineage counts for the training data set of samples collected from July 17 through December 25, 2021 used in this paper. The lineages are taken from GISAID metadata entries, which are based on Pango lineage assignments. (See O’Toole, Á., Scher, E., Underwood, A., Jackson, B., *et al.*, 2021. Assignment of epidemiological lineages in an emerging pandemic using the pangolin tool. *Virus evolution* 7, veab064. <https://doi.org/10.1093/ve/veab064>.)

Lineage	Count
B.1.617.2	8580
AY.43	7171
AY.122	5299
AY.20	4856
AY.4	4430
AY.26	2158
BA.1	1816
AY.100	1701
AY.3	1141
AY.98.1	1041
BA.1.1	971
None	920
AY.113	734
AY.5	692
AY.34	669
AY.9.2	643
AY.25.1	585
AY.103	567
B	560
AY.125	546
AY.33	500
B.1.621	459
AY.46.6	458
AY.34.2	429
P.1	394
AY.126	385
B.1.1.7	375
AY.9	372
AY.112	366
AY.23	354
AY.12	335
AY.42	331
AY.99.2	328

Lineage	Count
AY.43.3	321
AY.57	317
AY.44	312
AY.127	309
AY.102	308
AY.4.2	305
AY.25	298
B.1.640.1	294
BA.1.18	281
AY.95	270
AY.123	265
AY.34.1	253
AY.101	250
AY.121	248
AY.24.1	236
AY.109	207
AY.39	197
AY.59	197
AY.4.2.3	195
AY.5.4	188
AY.36	187
BA.1.4	182
AY.29	164
C.37	164
AY.129	162
AY.124	156
AY.92	152
BA.1.1.1	143
AY.120	130
AY.4.9	115
AY.28	111
AY.104	97
AY.24	91

Lineage	Count
BA.1.17	81
AY.46	76
AY.118	71
AY.122.4	68
AY.4.6	68
AY.98	65
AY.4.2.2	65
AY.7.2	63
BA.1.17.2	61
AY.94	60
AY.6	58
AY.4.4	58
AY.16	55
AY.79	53
AY.45	52
AY.119	52
AY.40	52
AY.76	50
AY.62	49
AY.133	48
AY.75	47
AY.37	47
AY.119.2	46
AY.13	46
AY.47	45
P.1.7	44
P.1.15	44
AY.116	43
AY.86	43
BA.1.14	42
AY.7.1	41
AY.123.1	41
AY.54	38

Predicting COVID-19 Disease Severity from SARS-CoV-2 Spike Protein Sequence by Mixed Effects Machine Learning – B.A. Sokhansanj & G.L. Rosen – Supplementary Table S3

Lineage	Count
AY.106	38
B.1.351	36
A.2.5	36
AY.73	34
AY.43.6	34
AY.61	34
B.1.628	34
AY.30	33
BA.1.13	33
B.1	32
AY.3.2	30
AY.105	29
AY.4.5	29
BA.1.1.7	29
AY.124.1	29
AY.88	29
AY.46.2	29
AY.85	28
AY.128	28
AY.117	28
AY.72	25
AY.4.11	24
AY.32	24
P.1.2	23
AY.84	22
AY.27	22
BA.1.16	22
BA.1.15	22
AY.91	21
AY.114	20
AY.51	20
B.1.621.1	19
AY.111	19
AY.110	19
AY.70	18
A.2.5.2	18
AY.14	17
B.1.1.519	16
AY.43.4	16
AY.4.2.1	16

Lineage	Count
AY.71	16
AY.39.2	15
AY.39.1	15
AY.5.2	15
B.1.1.318	15
B.1.632	14
AY.122.2	14
AY.29.1	14
AY.119.1	14
AY.60	13
AY.68	13
AY.4.3	13
AY.121.1	13
AY.10	13
P.1.12	13
AY.19	12
AY.33.1	12
AY.17	12
AY.15	12
P.1.9	12
AY.43.8	12
B.1.1	12
AY.3.1	11
AY.4.1	11
AY.87	10
AY.78	10
AY.25.3	10
B.1.631	10
P.1.8	9
BA.1.8	9
P.1.14	9
AY.53	9
AY.3.3	9
B.1.1.348	9
AY.1	9
AY.122.1	9
AY.77	9
AY.7	9
BA.1.19	8
AY.4.2.4	8

Lineage	Count
AY.65	8
AY.43.7	8
P.1.17	8
AY.58	7
AY.4.8	7
AY.16.1	7
AY.107	7
AY.66	7
AY.22	6
AY.125.1	6
AY.5.5	6
AY.120.2	5
AY.75.2	5
B.1.621.2	5
B.1.575	5
BA.1.1.4	5
B.1.640.2	5
AY.46.4	5
AY.35	5
C.36.3	5
AY.21	5
AY.99	5
AY.108	5
BA.2	5
B.1.609	5
AY.11	5
B.1.351.2	5
AY.2	5
B.1.177	4
AY.48	4
AY.127.1	4
AY.67	4
XB	4
BA.1.15.1	4
AY.83	4
AY.23.1	4
P.3	4
AY.82	4
BA.1.1.11	4
AY.120.1	4

Predicting COVID-19 Disease Severity from SARS-CoV-2 Spike Protein Sequence by Mixed Effects Machine Learning – B.A. Sokhansanj & G.L. Rosen – Supplementary Table S3

Lineage	Count
B.1.626	4
P.1.7.1	4
AY.56	4
AY.43.2	4
AY.34.1.1	4
AZ.5	4
AY.43.5	4
AY.25.2	3
AY.5.3	3
BA.3	3
B.1.160	3
AY.46.5	3
P.1.10	3
Unassigned	3
BA.1.1.2	3
AY.74	3
P.2	3
AY.4.7	3
AY.38	3
P.1.3	3
AY.41	3
AY.116.1	2
AY.46.1	2
B.1.620	2
B.1.1.523	2
P.1.1	2
A.1	2
B.1.160.25	2
Q.8	2
AY.64	2
AY.39.1.1	2
AY.36.1	2
B.1.351.5	2
B.1.243	2
B.1.462	2
P.1.4	2
B.1.243.2	2
AY.120.2.1	2
AZ.2	2
AY.50	2

Lineage	Count
B.1.36	2
AY.69	2
BA.1.1.15	2
B.1.1.207	2
C.37.1	2
P.1.12.1	1
AY.90	1
B.1.638	1
AY.63	1
C.36	1
C.26	1
A.2.5.1	1
AY.102.2	1
AY.103.1	1
B.1.36.16	1
BA.2.9	1
AY.39.1.3	1
AY.127.2	1
Q.4	1
BA.1.10	1
B.1.551	1
AY.99.1	1
B.1.411	1
AY.103.2	1
B.1.1.263	1
B.1.427	1
AY.20.1	1
P.1.6	1
B.1.625	1
B.1.526	1
B.1.637	1
P.1.11	1
AY.4.14	1
B.1.533	1
B.1.525	1
B.1.429	1
AY.52	1
B.1.1.317	1
B.1.1.525	1
C.40	1

Lineage	Count
Q.1	1
AY.5.1	1
B.1.1.117	1
AY.93	1
B.1.1.1	1
B.1.1.241	1
B.1.630	1
B.1.2	1
B.1.385	1
AY.46.6.1	1
AY.46.3	1
AZ.3	1
B.1.617.1	1
B.1.396	1
BA.1.13.1	1