

## Supplementary Information

### Genomic epidemiology reveals multiple introductions and spread of SARS-CoV-2 in the Indian state of Karnataka

Chitra Pattabiraman<sup>1</sup>, Farhat Habib<sup>2</sup>, Harsha PK<sup>1</sup>, Risha Rasheed<sup>1</sup>, Vijayalakshmi Reddy<sup>1</sup>, Prameela Dinesh<sup>3</sup>, Tina Damodar<sup>1</sup>, Kiran Hosallimath<sup>1</sup>, Anson K George<sup>1</sup>, Nakka Vijay Kiran Reddy<sup>1</sup>, Banerjee John<sup>1</sup>, Amrita Pattanaik<sup>1</sup>, Narendra Kumar<sup>1</sup>, Reeta S Mani<sup>1</sup>, Manjunatha M Venkataswamy<sup>1</sup>, Shafeeq K Shahul Hameed<sup>1</sup>, Prakash Kumar B.G<sup>3</sup>, Anita Desai<sup>1</sup>, Ravi Vasanthapuram<sup>1</sup>

## Contents

Supplementary Table 1

Supplementary Figure 1-2

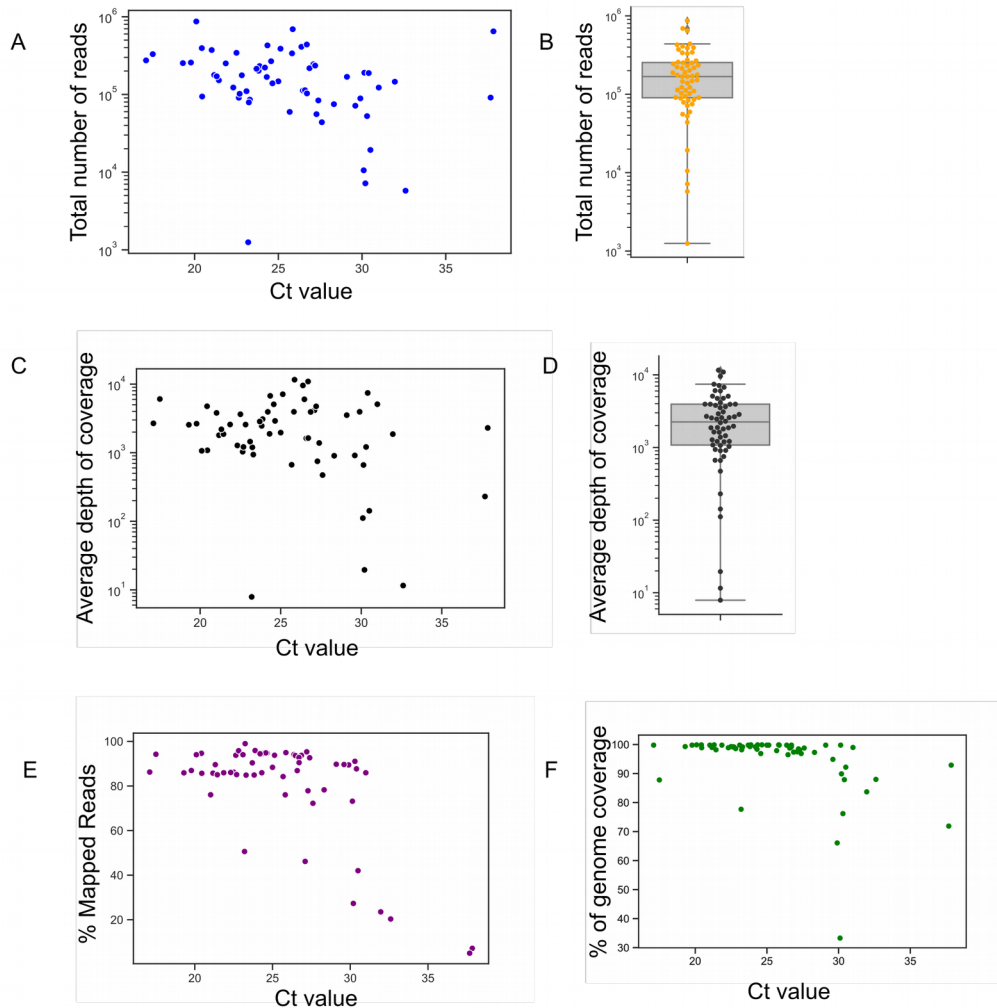
Supplementary Table 2-5

## Supplementary Table 1

Supplementary Table 1: Ct value, coverage depth and breadth of SARS-CoV-2 genomes						
Sr. No	Ct Value	kit	Total Reads	Mapped Reads	% of genome covered	Average Depth of coverage
1	17.1	ICMR-NIV	274243	236503	99.8	2681
2	19.78	ICMR-NIV	257231	223629	99.8	2651
3	30.14	A*Star	190178	139123	99.8	662
4	19.3	ICMR-NIV	252864	217195	99.3	2553
5	22.5	Mylab	343163	295756	99.3	3637
6	22.64	A*Star	90658	77036	99.2	1033
7	23.3	Mylab	85694	72731	99.2	939
8	21.17	A*Star	177961	152594	99.0	1798
9	22.31	A*Star	122679	105577	98.9	1276
10	27.28	ICMR-NIV	55714	43373	98.1	750
11	20.1	A*Star	868971	897046	99.9	1066.6
12	20.43	BGI	3,94,431	373374	99.9	4755
13	21.32	A*Star	171088	153235	99.9	2193
14	23.88	A*Star	2,31,800	222294	99.9	3092.3
15	24.35	A*Star	4,27,763	404509	99.9	6746
16	24.65	A*Star	139037	131752	99.9	2910.3
17	25.82	A*Star	337778	256837	99.9	3950
18	21.02	A*Star	372807	283520	99.8	3807.6
19	21.86	A*Star	251094	213959	99.8	2580
20	23.1	A*Star	110061	103481	99.8	1451.8
21	23.7	A*Star	213839	193264	99.8	2852.3
22	25	A*Star	147496	130336	99.8	1963.6
23	25.13	A*Star	387775	363643	99.8	7126
24	25.86	A*Star	691406	656607	99.8	11602
25	26.38	A*Star	4,10,206	385974	99.8	9568.5
26	26.71	A*Star	103092	93277	99.8	1634.5
27	29.1	A*Star	168480	151202	99.8	3528
28	23.82	A*Star	201647	171229	99.3	2463
29	22.69	A*Star	101884	86721	99.2	1214
30	24.2	A*Star	2,21,647	209359	99.2	3943.2
31	26.7	A*Star	439001	407218	99.2	10931
32	31	A*Star	1,22,573	105329	99	5092.5
33	20.45	A*Star	93992	80541	98.9	1084
34	26.59	A*Star	113388	98515	98.8	1620.3
35	27.6	A*Star	43884	31697	98.8	472
36	22.82	A*Star	176354	168960	98.6	2572
37	27.2	A*Star	235096	224120	98.5	4741.3
38	24.31	A*Star	167748	139447	98.4	1884
39	21.45	A*Star	152005	129254	98.2	1867
40	23.24	A*Star	78932	78131	98.2	1198.2
41	25.68	A*Star	59579	50176	97.9	667
42	27.1	BGI	244455	112843	97.5	4161.3
43	26.86	A*Star	2,17,498	203880	97.4	3930.5
44	28.32	A*Star	74958	58662	97.3	902.62
45	24.57	A*Star	2,67,466	253685	96.9	5078.6
46	27.39	A*Star	83692	77535	96.9	1385.5
47	26.47	Altona	112288	105163	96.5	6009.9
48	29.6	A*Star	71,530	64110	94.9	911.7
49	37.85	A*Star	648341	46502	92.9	2299.1
50	30.5	A*Star	19297	8168	92.2	142.17
51	30.2	A*Star	7166	1957	89.9	19.56
52	32.6	A*Star	5766	1172	88	11.56
53	30.4	A*Star	188366	165182	87.9	7434.6
54	17.5	BGI	330020	310965	87.8	6066
55	31.96	A*Star	145769	34295	83.7	1866.9
56	23.2	A*Star	1251	633	77.7	7.9
57	30.3	A*Star	52582	47882	76.2	1212
58	37.68	A*Star	91043	4535	71.9	229.9
59	29.9	A*Star	88861	79483	66.1	3949.7

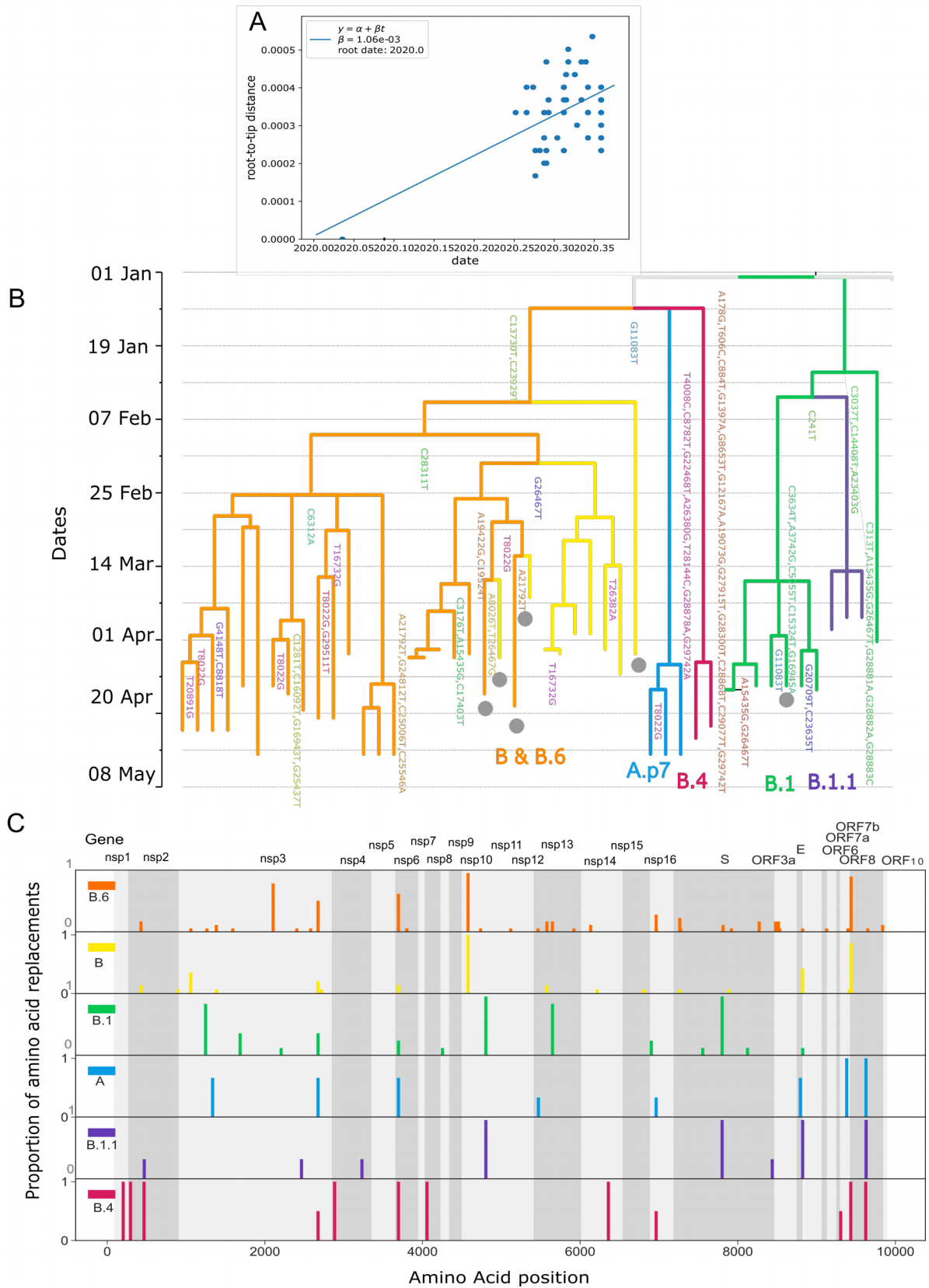
*Supplementary Table 1- Table describes data from sequencing of 60 COVID19 positive samples. Ct Values represent the results of RT-PCR with the indicated kits. Total reads is the total number of sequences obtained from each sample. Mapped reads represent the number of reads that aligned to the SARS-CoV-2 reference genome (NC\_045512). % of genome/breadth of coverage is the percentage of the reference genome covered by at least one read. Average depth of coverage is the average number of reads that cover each base of the genome. Samples for which genomes were not recovered are shaded in grey*

## Supplementary Figure 1



Supplementary Figure 1- Amplicon sequencing of SARS-CoV-2 genomes from PCR positive samples. Figure panels (A, C, E, F) show the relationship between Ct values from RT-PCR with total number of reads obtained from each sample, % of reads that mapped to the SARS-CoV-2 reference genome, average depth of sequencing across the genome and % of reference genome covered by at least one read (% genome coverage). Panel B and D show the spread of the total number of sequencing reads obtained from each sample and average number of reads per nucleotide base (average depth of coverage) respectively. Each point denotes a positive sample,  $n=60$ .

# Supplementary Figure 2



Supplementary Figure 2- **A**. Root to tip regression analysis of sequences used in this study. Mutation rate is estimated at  $1.062e-03$  mutations/site/year with  $r^2 = 0.28$ . Note: While the temporal signal in this data is weak, analysis of larger datasets by others is suggestive of a molecular clock. **B**. Time scaled maximum likelihood tree of genomes from this study. Coloured circles represent sequences from symptomatic individuals. Mutations/SNPs at each node are labelled in the tree. All changes are with respect to the reference genome. **C**. Frequency of amino acid replacements in different lineages. Details are provided in Supplementary Table 3

## Supplementary Table 2

Supplementary Table 2: Frequency of SNPs in different lineages							
Sr. No	SNP	A	B	B.1	B.1.1	B.4	B.6
48	C16092T						0.18
49	G16396A						0.59
50	A16408G	0.33					
51	T16732G		0.14				0.18
52	G16943T						0.18
53	G16945A			0.88			
54	C17403T		0.36				0.59
55	C17762G						0.59
56	G18397T						0.12
57	G18651T		0.71				
58	A19073G					1.00	
59	A19422G		0.50				0.59
60	C19524T		0.50				0.59
61	G20419T		0.71				
62	A20435G		0.71				
63	G20709T			0.25			
64	T20891G	0.33				0.50	0.29
65	A20895T						0.59
66	C20896G						0.59
67	A21792T		0.71				0.24
68	G21824C						0.59
69	G22468T	1.00					
70	T22665C			0.13			
71	C23320T		0.71				
72	A23403G			1.00	1.00		
73	T23431A						0.59
74	T23433C						0.59
75	G23441C						0.12
76	A23442C						0.59
77	C23635T			0.25			
78	G23663T		0.71				
79	G23755T						0.59
80	C23929T		0.86				0.82
81	G24368C			0.13			
82	G24812T						0.18
83	C25006T						0.18
84	G25311T				0.33		
85	G25437T						0.12
86	C25546A						0.18
87	C25584T			0.13			
88	G25598A						0.59
89	G26211T	0.33					
90	A26380G	0.67					
91	T26382A		0.14		0.33		0.59
92	T26397C						0.59
93	G26467T		0.43	0.13	1.00		0.59
94	C26753T		0.71				
95	C26801T		0.71				

Supplementary Table 2: Frequency of SNPs in different lineages

Sr. No	SNP	A	B	B.1	B.1.1	B.4	B.6
48	C16092T						0.18
49	G16396A						0.59
50	A16408G	0.33					
51	T16732G		0.14				0.18
52	G16943T						0.18
53	G16945A			0.88			
54	C17403T		0.36				0.59
55	C17762G						0.59
56	G18397T						0.12
57	G18651T		0.71				
58	A19073G					1.00	
59	A19422G		0.50				0.59
60	C19524T		0.50				0.59
61	G20419T		0.71				
62	A20435G		0.71				
63	G20709T			0.25			
64	T20891G	0.33				0.50	0.29
65	A20895T						0.59
66	C20896G						0.59
67	A21792T		0.71				0.24
68	G21824C						0.59
69	G22468T	1.00					
70	T22665C			0.13			
71	C23320T		0.71				
72	A23403G			1.00	1.00		
73	T23431A						0.59
74	T23433C						0.59
75	G23441C						0.12
76	A23442C						0.59
77	C23635T			0.25			
78	G23663T		0.71				
79	G23755T						0.59
80	C23929T		0.86				0.82
81	G24368C			0.13			
82	G24812T						0.18
83	C25006T						0.18
84	G25311T				0.33		
85	G25437T						0.12
86	C25546A						0.18
87	C25584T			0.13			
88	G25598A						0.59
89	G26211T	0.33					
90	A26380G	0.67					
91	T26382A		0.14		0.33		0.59
92	T26397C						0.59
93	G26467T		0.43	0.13	1.00		0.59
94	C26753T		0.71				
95	C26801T		0.71				

Supplementary Table 2: Frequency of SNPs in different lineages							
Sr. No	SNP	A	B	B.1	B.1.1	B.4	B.6
96	G27378T						0.59
97	G27915T					0.50	
98	T28144C	1.00					
99	G28188C						0.59
100	A28247C		0.71				
101	A28249G						0.59
102	G28300T					1.00	
103	C28311T		0.86				0.88
104	T28688C					0.50	
105	C28868T					1.00	
106	G28878A	1.00					
107	G28881A				1.00		
108	G28882A				1.00		
109	G28883C				1.00		
110	G28946T						0.59
111	C29077T					1.00	
112	G29511T						0.59
<p>Supplementary Table 2: Frequency of SNPs in different lineages            Single nucleotide polymorphisms were identified in the genomes with respect to the reference sequence. The frequency of SNPs within each lineage was calculated by the number of sequences with SNP/total number of sequences in the lineage. SNPs are ordered by their position on the genome.</p>							

Supplementary Table 3

Sr. No	Amino Acid Replacements (Gene – change)	Proportion of amino acid replacements by lineage					
		A	B	B.1	B.1.1	B.4	B.6
1	E-I46V	0.67					
2	E-V75F		0.43	0.13	1.00		0.59
3	N-D225Y						0.59
4	N-G204R				1.00		
5	N-P13L		0.86				0.94
6	N-P199S					1.00	
7	N-Q9H					1.00	
8	N-R203K				1.00		
9	N-S202N	1.00					
10	N-S413I						0.12
11	ORF3a-L15F						0.18
12	ORF3a-L52I						0.18
13	ORF3a-W69*						0.59
14	ORF6-E59D						0.59
15	ORF8-D119G						0.59
16	ORF8-G8*					0.50	
17	ORF8-L118F		0.71				
18	ORF8-L84S	1.00					
19	ORF8-V99L						0.59
20	S-A701S		0.71				
21	S-C1250F				0.33		
22	S-D1084Y						0.18
23	S-D614G			1.00	1.00		
24	S-D627H						0.59
25	S-D627P						0.59
26	S-D88H						0.59
27	S-D936H			0.13			
28	S-I624T						0.59
29	S-K77M		0.71				0.24
30	S-L368P			0.13			
31	S-M731I						0.59
32	nsp1-I114T					1.00	
33	nsp12-A97V		1.00				1.00
34	nsp12-H256P						0.59
35	nsp12-L638F						0.59
36	nsp12-P323L			1.00	1.00		
37	nsp13-A237T			0.88			
38	nsp13-A509G						0.59
39	nsp13-G54S						0.59
40	nsp13-S166A		0.14				0.18
41	nsp13-S236I						0.18
42	nsp13-T58A	0.33					
43	nsp14-D345G					1.00	
44	nsp14-E204D		0.71				
45	nsp14-V120L						0.12
46	nsp15-D267Y		0.71				
47	nsp15-D272G		0.71				
48	nsp16-M17I			0.25			



Sr. No	Amino Acid Replacements (Gene – change)	Proportion of amino acid replacements by lineage					
		A	B	B.1	B.1.1	B.4	B.6
49	nsp16-P80A						0.59
50	nsp16-V78G	0.33				0.50	0.29
51	nsp2-A159V		0.14				0.18
52	nsp2-E201D				0.33		
53	nsp2-R27C					1.00	
54	nsp2-T629T/I		0.71				
55	nsp2-V198I					1.00	
56	nsp3-A358S						0.59
57	nsp3-A480V		0.71				
58	nsp3-D1674N						0.59
59	nsp3-F430S	0.33					
60	nsp3-F430S/F	0.33					
61	nsp3-G1300C			0.13			
62	nsp3-I341M			0.88			
63	nsp3-P153S		0.36				0.59
64	nsp3-P1558P/L				0.33		
65	nsp3-T1198K						0.82
66	nsp3-T686I						0.59
67	nsp3-T779I			0.38			
68	nsp3-V1768G	0.67	0.21	0.38		0.50	0.53
69	nsp3-V1811A		0.71				
70	nsp3-V477F						0.12
71	nsp3-W1498R						0.59
72	nsp4-A380V				0.33		
73	nsp4-M33I					1.00	
74	nsp6-L142V						0.59
75	nsp6-L37F	0.67	0.14	0.25		1.00	0.65
76	nsp6-M143M/L						0.59
77	nsp6-V145L						0.59
78	nsp8-V26I					1.00	
79	nsp9-D25E			0.13			

## Supplementary Table 4

Supplementary Table 4: Description of contact clusters			
		Our Centre	
	State	tested	sequenced
Total No of contact clusters	104	38	12
Total No. of people connected	822	398	294
No of people in a cluster			
0–5	74	22	2
>5	30	16	10
Nature of contact/exposure of index case*			
Contact with COVID19 case	0	0	0
No known Contact			
i. Travel History (International)	14	0	0
ii. Travel History (Domestic)	20	10	5
iii.ILI (under investigation)	19	4	0
iv.SARI (under investigation)	29	9	6
v.Under Investigation	10	8	1
vi.Contact Unknown	12	7	0
Clinical state of index case*			
Asymptomatic(Asymp)	29	19	3
Symptomatic (Symp)	75	19	9
Ratio (Asymp: Symp)	0.39	1.00	0.33
Nature of contact/exposure of spreader**			
Contact with COVID19 case	12	4	3
No known Contact			
i. Travel History (International)	14	0	0
ii. Travel History (Domestic)	16	9	4
iii.ILI (under investigation)	19	4	0
iv.SARI (under investigation)	22	7	4
v.Under Investigation	10	8	1
vi.Contact Unknown	11	6	0
Clinical state of spreader**			
Asymptomatic(Asymp)	35	20	4
Symptomatic (Symp)	69	18	8
Ratio (Asymp: Symp)	0.51	1.11	0.50

\* earliest detected individual from a cluster, \*\* individual with maximum number of connections in a cluster, 'asymptomatic at the time of sampling

*Note: Nature of contact/exposure is classified as Contact with COVID19 case- where tested individual was in contact with a known positive case or No known contact, divided into six categories. i. Travel history (international) - travel history to other countries, ii. Travel history (domestic) - travel within the state or interstate, iii. ILI (under investigation) - individuals with Influenza like illness with no known source of infection, iv. SARI (under investigation) - individuals with severe acute respiratory infection where the source of infection is not known, v. Under investigation- source of infection is not yet known/ contact tracing has not been completed, vi. Contact Unknown- where the tested individual was from a location where there were cases (eg. A containment Zone) but a specific contact could not be identified.*

## Supplementary Table 5

Supplementary Table 5: Characteristics of sequenced clusters										
Clusters	# people	Lineages	Age Distribution					Clinical State		Districts
			0-20	20-40	40-60	60-80	80-100	Asym*	Symp	
C1	67	B.1.1 (3)	7	46	11	3	0	51	16	Mysuru (57),Vijayapura (8),Mandya(2)
C2	43	B.6(1), B(1)	10	18	13	2	0	39	4	Belagavi(41),Tumkur(2)
C3	50	B.4(2), B(1)	10	37	2	1	0	48	2	Bengaluru Urban (49), Davanagere (1)
C4	35	B.1(7)	4	23	7	0	1	32	3	Bengaluru Urban (35)
C5	21	B.6(4)	4	10	5	2	0	20	1	Bagalkot(21)
C6	27	B.6(3)	7	14	5	1	0	26	1	Mandya(20), Mysuru(5), Kalaburgi(2)
C7	18	B(6), B.6(3)	1	9	8	0	0	18	0	Bengaluru Urban(18)
C8	16	B.6(1)	3	6	5	1	1	14	2	Bidar(16)
C9	7	B(1)	4	0	2	1	0	6	1	Bengaluru urban(7)
C10	6	B(1)	1	4	1	0	0	5	1	Belagavi(6)
C11	2	B.6(1)	1	1	0	0	0	2	0	Bidar(2)
C12	2	B(1)	1	1	0	0	0	1	1	Belagavi(2)