Alpha- and beta-diversity statistical analysis

Supplementary table 1. Statistical analysis of the alpha- and beta-diversity results of the bacteriophage RNA-metagenome. Results with a p-value < 0.05 (bolded) were considered significant.

Pairwise comparison	Alpha-diversity p-value (Shannon index; Wilcoxon)	Beta-diversity Bray-Curtis p-value (PERMANOVA)	Beta-diversity Jaccard p-value (PERMANOVA)
SARS-CoV-2 positive *vs* SARS-CoV-2 negative	**0.03**	0.092	0.099
NPS *vs* TS	**<0.001**	**0.001**	**0.001**
Symptomatic *vs* asymptomatic	**0.04**	**0.001**	**0.001**
NPS-symptomatic-SARS-CoV-2 positive *vs* NPS-symptomatic SARS-CoV-2 negative	1.0	0.234	0.837
NPS-asymptomatic-SARS-CoV-2 positive *vs* NPS-asymptomatic-SARS-CoV-2 negative	0.13	**0.026**	**0.004**
TS-symptomatic-SARS-CoV-2 positive *vs* TS-symptomatic-SARS-CoV-2 negative	0.15	0.205	0.183

Volcano plot supplementary data

Supplementary table 2. Significant species (-Log10(q-val)>1.30)) observed between the fold change of NP symptomatic SARS-CoV-2 positive samples and healthy individuals.

Species	Log2(FC)	-Log10(q-val)
Actinomyces sp. 2129	-6.97144210727746	3.47998854
Candidatus Saccharibacteria oral taxon TM7x	-6.70612742905151	3.47998854
Capnocytophaga sputigena	-6.20647138540253	3.47998854
Candidatus Saccharibacteria bacterium FS15P	-6.12285647555089	3.47998854
Leptotrichia wadei	-6.01519772707412	3.47998854
Leptotrichia goodfellowii	-5.93626931836543	3.47998854
Alloprevotella sp. E39	-5.89713095103068	3.47998854
Porphyromonas asaccharolytica	-5.89509017851958	3.47998854
Prevotella enoeca	-5.81960075432376	3.47998854
Leptotrichia sp. oral taxon 498	-5.7940315689528	3.45070333
Prevotella denticola	-5.50533596550729	3.45070333
Atopobium parvulum	-5.48981202698121	3.45070333
Candidatus Saccharibacteria bacterium FS13P	-5.47825129694856	3.45070333
Actinomyces pacaensis	-5.47309776555904	3.45070333
Mogibacterium pumilum	-5.40907989582476	3.45070333
[Eubacterium] sulci	-5.39034312529511	3.45070333
Candidatus Saccharibacteria bacterium FS07P-B	-5.37981554857233	3.45070333
Leptotrichia buccalis	-5.36076216960278	3.45070333
Megasphaera stantonii	-5.31276388012429	3.45070333
Lachnoanaerobaculum umeaense	-5.30335557184688	3.45070333
Leptotrichia shahii	-5.17914229781393	3.45070333
Capnocytophaga sp. oral taxon 878	-5.06660334152056	3.45070333
Capnocytophaga leadbetteri	-4.96340744073626	3.45070333
Leptotrichia trevisanii	-4.93655421336354	3.45070333
Dialister pneumosintes	-4.8398654489815	3.45070333
Actinomyces sp. oral taxon 897	-4.83410605378407	3.45070333
Leptotrichia hofstadii	-4.82527976942094	3.45070333
Actinomyces hongkongensis	-4.72794972108188	3.45070333
Capnocytophaga sp. FDAARGOS_737	-4.65369071299944	3.45070333
Campylobacter showae	-4.63530883523954	3.45070333
Porphyromonas gingivalis	-4.62375197991705	3.45070333
Leptotrichia hongkongensis	-4.58238675976895	3.45070333
Aggregatibacter aphrophilus	-4.53213693250872	3.45070333
Capnocytophaga gingivalis	-4.48788660695844	3.45070333
Eikenella corrodens	-4.34099040431785	3.45070333
Aggregatibacter segnis	-4.24504452131846	3.45070333
Haemophilus haemolyticus	-4.23630293732862	3.45070333
Haemophilus sp. oral taxon 036	-4.22416433368701	3.45070333
Leptotrichia sp. oral taxon 212	-4.189810328986	3.45070333
Corynebacterium humireducens	-4.18030164571718	3.45070333

Negativicoccus massiliensis	-4.12262000013304	3.45070333
Actinomyces oris	-4.11710346674153	3.45070333
Neisseria mucosa	-3.9252969563647	3.45070333
Gemella sanguinis	-3.92138823389469	3.45070333
Fusobacterium pseudoperiodonticum	-3.81196581078016	3.45070333
Olsenella sp. oral taxon 807	-3.78998872182911	3.45070333
Actinomyces sp. oral taxon 171	-3.70889848452512	3.45070333
Haemophilus influenzae	-3.67686400441652	3.45070333
Actinomyces naeslundii	-3.60719659993851	3.45070333
Corynebacterium matruchotii	-3.59421327644431	3.45070333
Haemophilus pittmaniae	-3.49585557798169	3.45070333
Haemophilus parahaemolyticus	-3.47624720237673	3.45070333
Alloscardovia omnicolens	-3.4307289303342	3.45070333
Haemophilus parainfluenzae	-3.36979730047764	3.45070333
Fusobacterium periodonticum	-3.3051714153176	3.45070333
Bifidobacterium dentium	-3.00693821636284	3.45070333
Neisseria sp. oral taxon 014	-2.94823280134082	3.45070333
Campylobacter concisus	-2.94556716439363	3.45070333
Actinomyces sp. oral taxon 414	-2.93383750402153	3.45070333
Granulicatella adiacens	-2.8686569181649	3.45070333
Neisseria lactamica	-2.84158860704017	3.45070333
Fusobacterium hwasookii	-2.80380460694096	3.45070333
Actinomyces sp. oral clone CT068	-2.78396446766446	3.45070333
Aggregatibacter actinomycetemcomitans	-2.77224148824293	3.45070333
Parvimonas micra	-2.69489722456799	3.45070333
Gemella sp. oral taxon 928	-2.65637290312393	3.45070333
Fusobacterium sp. oral taxon 203	-2.53235950440809	3.45070333
Filifactor alocis	-2.50890200329412	3.45070333
Capnocytophaga sp. oral taxon 864	-2.48776724065971	3.45070333
Capnocytophaga sp. oral taxon 323	-2.4778318073797	3.45070333
Fusobacterium nucleatum	-2.45795937294954	3.45070333
Corynebacterium halotolerans	-2.441696196398	3.45070333
Neisseria elongata	-2.4188334385725	3.45070333
Capnocytophaga ochracea	-2.3215423094064	3.45070333
Bacillus thuringiensis	-2.29592270844087	3.45070333
Neisseria subflava	-2.28596703441463	3.45070333
Gemella haemolysans	-2.24755234193987	3.45070333
Lachnospiraceae bacterium oral taxon 500	-2.15853712970997	3.45070333
Neisseria gonorrhoeae	-2.06050512339854	3.45070333
Corynebacterium diphtheriae	-2.00106178768252	3.45070333
Neisseria flavescens	-1.80823738467065	3.45070333
Human betaherpesvirus 6	-1.80490481741298	3.45070333
Neisseria meningitidis	-1.8009469867746	3.45070333
Neisseria sp. KEM232	-1.78108492075284	3.45070333
Neisseria cinerea	-1.7649705373982	3.45070333

Neisseria polysaccharea	-1.66622862010011	3.45070333
Gemella morbillorum	-1.55323460703936	3.45070333
Aegilops speltoides	1.50078402030824	3.45070333
Glycine max	1.78454801936221	3.45070333
Glycine soja	2.04315593520215	3.45070333
Enterococcus cecorum	2.49327327268974	3.45070333
Dolosigranulum pigrum	2.79252744332395	3.45070333
Lactobacillus salivarius	2.99209715348235	3.45070333
Lactobacillus paragasseri	4.27005726906089	3.45070333
Lactobacillus paracasei	5.74866119892748	3.45070333
Avena strigosa	6.31759207050652	3.45070333
Lactobacillus vaginalis	6.42251451263149	3.45070333
Lactobacillus rhamnosus	7.28567106447585	3.45070333
Lactobacillus oris	10.418901156597	3.45070333
Prevotella fusca	-6.01450612023137	3.45070333
Prevotella intermedia	-5.79960585781393	3.36598985
Prevotella melaninogenica	-5.66543034017694	3.36598985
Prevotella oris	-5.18490026160827	3.36598985
Prevotella jejuni	-5.1027936804828	3.28378411
Prevotella histicola	-4.20566144989985	3.28378411
Prevotella sp. oral taxon 299	-7.20121828198195	3.28378411
Prevotella ruminicola	-7.03578702715013	3.28378411
Selenomonas sp. oral taxon 126	-5.46049435351101	3.28378411
Prevotella scopos	-5.19419967811139	3.20862397
Schaalia meyeri	-4.84306784526561	3.20862397
Schaalia odontolytica	-4.82068464851997	3.20862397
Schaalia cardiffensis	-4.48695028793937	3.20862397
Rothia sp. T40-1	-3.76618528784962	3.20862397
Rothia aeria	-3.38828123098386	3.20862397
Rothia mucilaginosa	-3.37875110709189	3.20862397
Rothia dentocariosa	-1.78692023970928	3.20862397
Prevotella salivae	-1.66609396593095	3.20862397
Selenomonas sp. oral taxon 920	-5.79854969244087	3.20862397
Selenomonas sp. oral taxon 478	-4.26232637194894	3.17604374
Selenomonas sp. oral taxon 136	-4.11342541778317	3.17604374
Streptococcus phage Javan374	-6.37430435091126	3.17604374
Streptococcus phage Javan371	-5.16074806832283	3.17604374
Streptococcus constellatus	-4.8667877416922	3.17604374
Streptococcus phage Javan377	-4.81163045579993	3.17604374
Streptococcus salivarius	-4.74344122456246	3.17604374
Streptococcus gordonii	-4.69100139335862	3.17604374
Streptococcus sp. oral taxon 064	-4.39189734686127	3.17604374
Streptococcus sp. NPS 308	-4.23854962978665	3.15935379
Streptococcus parasanguinis	-4.07628808657897	3.13919268
Streptococcus sp. LPB0220	-4.07065136900468	3.13919268

Streptococcus satellite phage Javan323	-4.0529065871514	3.13919268
Streptococcus sp. I-P16	-3.92212515150516	3.13919268
Streptococcus oralis	-3.85588275783764	3.13919268
Streptococcus phage Javan355	-3.81698171293315	3.13919268
Streptococcus sp. 1643	-3.59043277083971	3.13919268
Streptococcus infantis	-3.50596874740594	3.13919268
Streptococcus sanguinis	-3.29314418106152	3.13919268
Streptococcus mitis	-3.20951971793425	3.13919268
Streptococcus sp. I-G2	-3.19734270810211	3.13919268
Streptococcus sp. 116-D4	-3.18109881165719	3.13919268
Streptococcus gwangjuense	-3.17934482144221	3.13919268
Streptococcus pseudopneumoniae	-3.11662134607058	3.13919268
Streptococcus koreensis	-3.11538009387299	3.13919268
Streptococcus sp. A12	-2.99240379301716	3.13919268
Streptococcus phage Javan311	-2.92761284939209	3.13919268
Streptococcus pneumoniae	-2.83985801084027	3.13919268
Streptococcus intermedius	-2.76795687594035	3.13919268
Streptococcus sp. FDAARGOS_192	-2.62809911571559	3.13919268
Streptococcus phage Javan326	-2.54978052312163	3.13919268
Streptococcus australis	-2.44482742924891	3.13919268
Selenomonas sputigena	-2.37953002114023	3.13919268
Streptococcus equinus	-2.27490843840369	3.13919268
Streptococcus satellite phage Javan243	-1.84368420573697	3.13919268
Streptococcus phage Javan310	-1.78991594155303	3.13919268
Streptococcus cristatus	-1.7013210668819	3.13919268
Streptococcus sp. oral taxon 431	-4.84442952288394	3.13919268
Streptococcus sp. VT 162	-3.96391417038777	3.13919268
Tannerella sp. oral taxon HOT-286	-7.02076273899266	3.00348563
TM7 phylum sp. oral taxon 488	-5.61034091656462	3.00348563
Veillonella atypica	-5.2218856279505	3.00348563
uncultured Caudovirales phage	-5.17515896676709	3.00348563
Veillonella dispar	-3.98178798939368	3.00348563
Treponema sp. OMZ 804	-3.87827554075787	3.00348563
Veillonella parvula	-3.77690799585285	3.00348563
Thermoanaerobacterium xylanolyticum	-3.55987741870961	2.83115019
Treponema sp. OMZ 838	-3.40335858416535	2.83115019
Streptococcus virus MS1	-2.86157337236359	2.83115019
Tannerella forsythia	-2.82334586270996	2.83115019
Treponema denticola	-2.50603417804985	2.47553193
Streptococcus viridans	-2.08487238000787	2.47553193
Treponema putidum	-1.51569049179469	2.47553193

Supplementary table 3. Putative respiratory pathogens identified amongst the NPS samples in COVID-19 positive and COVID-19 negative patients. The proportion of identified organisms were compared using the Fisher's exact test with Benjamini-Hochberg correction.

Organism	**TS**			***P***	***q-val***
	All (n=24)	COVID-19 + (n=12)	COVID-19 – (n=12)		
Dolosigranulum pigrum	12.5% (3/24)	0% (0/12)	25% (3/12)	0.217	0.951
Enterobacter *Enterobacter asburiae* *Enterobacter cloacae* *Enterobacter hormaechei* *Enterobacter ludwigii* *Enterobacter sichuanensis* *Enterobacter sp. R4-368* *Enterobacter roggenkampii*	12.5% (3/24)	16.7% (2/12)	8.3% (1/12)	>0.999	1
Haemophilus influenzae	37.5% (9/24)	41.7% (5/12)	33.3% (4/12)	>0.999	1
Haemophilus parainfluenzae	79.2% (19/24)	66.6% (8/12)	91.7% (11/12)	0.317	0.951
Klebsiella pneumoniae	41.7% (10/24)	16.7% (2/12)	66.6% (8/12)	0.008	0.072
Serratia marcescens	4.2% (1/24)	0% (0/12)	8.3% (1/12)	>0.999	1
Staphylococcus aureus	25% (6/24)	16.7% (2/12)	33.3% (4/12)	0.640	1
Streptococcus pneumoniae	95.8% (23/24)	91.7% (11/12)	100% (12/12)	>0.999	1
Streptococcus pyogenes	4.2% (1/24)	0% (0/12)	8.3% (1/12)	>0.999	1

Supplementary table 4. Putative respiratory pathogens identified amongst the TS samples in COVID-19 positive and COVID-19 negative patients. The proportion of identified organisms were compared using the Fisher's exact test with Benjamini-Hochberg correction.

Organism	**NPS**			***P***	***q-val***
	All (n=47)	COVID-19 + (n=27)	COVID-19 – (n=20)		
Burkholderia multivorans	17.2 %(8/47)	11.1% (3/27)	25.0% (5/20)	0.258	0.401
Chlamydia pneumoniae	2.1% (1/47)	0% (0/27)	2.0% (1/20)	0.069	0.322
Dolosigranulum pigrum	23.4% (11/47)	40.7% (11/27)	0% (0/20)	0.001	**0.014**
Haemophilus influenzae	4.2% (2/47)	0% (0/27)	10.0% (2/20)	0.196	0.372
Haemophilus parainfluenzae	14.9% (7/47)	25.9% (7/27)	0% (0/20)	0.015	0.105
HHV-6	2.1% (1/47)	3.7 % (1/27)	0% (0/20)	>0.999	1
Human coronavirus HKU1	4.2% (2/47)	0% (0/27)	10% (2/20)	0.175	0.372
Human coronavirus NL63	2.1% (1/47)	0% (0/27)	5.0% (1/20)	0.425	0.507
Influenza A virus	2.1% (1/47)	0% (0/27)	5.0% (1/20)	0.435	0.507
Klebsiella pneumoniae	2.1% (1/47)	3.7 % (1/27)	0% (0/20)	>0.999	1
Moraxella catharralis	8.5% (4/47)	14.8% (4/27)	0% (0/20)	0.126	0.372
Mycoplasma pneumoniae	4.3% (2/47)	0% (0/27)	10% (2/20)	0.176	0.372
Rhinovirus A	2.1% (1/47)	0% (0/27)	5.0% (1/20)	0.435	0.507
Streptococcus pneumoniae	14.9% (7/47)	22.22% (6/27)	5.0% (1/20)	0.213	0.372

Genomic variants supplementary data

Supplementary table 5. SARS-CoV-2 SNPs among the NPS and TS clinical samples, along with its gene, and unique identified features.

SNP	SNP frequency	Mutation type	Gene/Protein	Unique characteristics and/or predicted changes from wild type	References
C100T	1/274	N/A	5'UTR	SNP associated with the P.2 (also known as B.1.1.28.2) Brazilian lineage.	[1]
A187G	6/274	N/A	5'UTR	Unresolved.	[2–4]
C241T	20/274	N/A	5'UTR	*In silico* prediction of a weaker interaction of mutant with the MADP1 host-transcription factor [5].	[5–10]
C379A	3/274	Synonymous: L64L	ORF1ab/Nsp1	Unresolved.	[11–14]
507-del	1/274	Deletion	ORF1ab/Nsp1	Unresolved.	Not found.
C913T	1/274	Synonymous: S36S	ORF1ab/Nsp2	Associated SNP with the VOC B.1.1.7	[15]
C1059T	12/274	Non-synonymous: T266I	ORF1ab/Nsp2	Unresolved.	[9,16,17]
C1616T	6/274	Non-synonymous: L271F	ORF1ab/Nsp2	Unresolved.	Not found
C2110T	1/274	Synonymous: N435N	ORF1ab/Nsp2	Unresolved	Not found
C2416T	1/274	Synonymous: Y537Y	ORF1ab/Nsp2	Unresolved.	[6]
T2423C	1/274	Non-synonymous: C540R	ORF1ab/Nsp2	Unresolved.	Not found
G2424A	2/274	Non-synonymous: C540Y	ORF1ab/Nsp2	Unresolved.	Not found
A2480G	1/274	Non-synonymous: I559V	ORF1ab/Nsp2	Unresolved.	[13,18,19]
G2528A	1/274	Non-synonymous: E575K	ORF1ab/Nsp2	Unresolved.	Not found.
C2558T	1/274	Non-synonymous: P585S	ORF1ab/Nsp2	Unresolved.	[13,20–22]
C3037T	21/274	Synonymous: F106F	ORF1ab/Nsp3	Unresolved.	3,12,,13
C3096T	1/274	Non-synonymous: S126L	ORF1ab/Nsp3	Unresolved.	[25,26]
C3267T	1/274	Non-synonymous: T183I	ORF1ab/Nsp3	Unresolved	[27]
A3405G	1/274	Non-synonymous: E229G	ORF1ab/Nsp3	Unresolved.	Not found.

T3477C	1/274	Non-synonymous: V253A	ORF1ab/Nsp3	Unresolved	Not found
T3766C	1/274	Synonymous: D349D	ORF1ab/Nsp3	Unresolved	[28]
T3905G	1/274	Non-synonymous: F396V	ORF1ab/Nsp3	Unresolved	Not found.
C4534T	1/274	Non-synonymous: Y605H	ORF1ab/Nsp3	Unresolved.	[29]
C5144T	2/274	Synonymous: L809L	ORF1ab/Nsp3	Unresolved.	Not found
C5694T	1/274	Non-synonymous: P992L	ORF1ab/Nsp3	Unresolved.	[30]
C5986T	1/274	Synonymous: F1089F	ORF1ab/Nsp3	Unresolved	[31]
G5992T	1/274	Non-synonymous: E1091D	ORF1ab/Nsp3	Unresolved.	Not found
C6310T	2/274	Synonymous: S1197S	ORF1ab/Nsp3 (NAR domain)	Unresolved.	[13,25]
A6425G	1/274	Non-synonymous: N1236D	ORF1ab/Nsp3	Unresolved.	Not found
C6664G	1/274	Synonymous: V1315V	ORF1ab/Nsp3	Unresolved.	Not found.
A6796G	1/274	Synonymous: L1359L	ORF1ab/Nsp3	Unresolved.	[30]
G7675T	2/274	Synonymous: A1652A	ORF1ab/Nsp3	Unresolved.	Not found.
G7745A	1/274	Non-synonymous: V1676I	ORF1ab/Nsp3	Unresolved.	Not found.
G8083A	1/274	Non-synonymous: M1788I	ORF1ab/Nsp3	Unresolved.	[32]
G8216A	1/274	Non-synonymous: V1833I	ORF1ab/Nsp3	Unresolved.	Not found
C9438T	1/274	Non-synonymous: T295I	ORF1ab/Nsp4	Unresolved.	[2,33,34]
C9711T	1/274	Non-synonymous: S387F	ORF1ab/Nsp4	Unresolved.	[30]
C10029T	1/274	Non-synonymous: T492I	ORF1ab/Nsp4	Unresolved. Often referred to as T3255I[35]. Potential signature mutation in the B.1 sub-clade[35].	[35]
C10319T	2/274	Non-synonymous: L89F	ORF1ab/3C-like proteinase	Unresolved.	[19,30,36–38]
C10422T	1/274	Non-synonymous: S123F	ORF1ab/3C-like proteinase	Unresolved.	Not found.
C10507T	1/274	Synonymous: N151N	ORF1ab/3C-like proteinase	Unresolved.	Not found.
T10667G	1/274	Non-synonymous: L205V	ORF1ab/3C-like proteinase	SNP associated with the B.1.1.28 Brazilian VOC[28,39]	[28,39]
A10770G	1/274	Non-synonymous: Y239C	ORF1ab/3C-like proteinase	Unresolved.	[30,37]
T11059G	2/274	Synonymous: T29T	ORF1ab/Nsp6	Unresolved.	Not found.

G11083T	1/274	Non-synonymous: L37F	ORF1ab/Nsp6	Unresolved.	[13,19,20,22,40,41]
C11289T	1/274	Non-synonymous: S106F	ORF1ab/Nsp6	Unresolved.	[30]
C11494T	1/274	Synonymous: N174N	ORF1ab/Nsp6	Unresolved.	Not found.
C11812A	1/274	Synonymous: G280G	ORF1ab/Nsp6	Unresolved.	Not found.
C11824T	1/274	Synonymous: I284I	ORF1ab/Nsp6	Associated SNP with the VOC B.1.1.28 Brazilian clade [42].	[42,43]
C11916T	1/274	Non-synonymous: S25L	ORF1ab/Nsp7	Unresolved.	[13,19,36]
C12053T	2/274	Non-synonymous: L71F	ORF1ab/Nsp7	Linked to increased severity outcome[44]. Odds ratio of 17.80 associated with increased death among Brazilian patients. Linked toG25088T (S: V11176F) SNP[45].	[44,45]
C12357T	1/274	Non-synonymous: T89I	ORF1ab/Nsp8	Unresolved.	[30,46]
C12412T	1/274	Synonymous: I107I	ORF1ab/Nsp8	Unresolved.	[30,47]
C12789T	1/274	Non-synonymous: T35I	ORF1ab/Nsp9	Unresolved.	[30]
G12794A	1/274	Non-synonymous: G37R	ORF1ab/Nsp9	Unresolved.	Not found.
A12964G	1/274	Synonymous: G93G	ORF1ab/Nsp9	SNP associated with the P.2 (also known as B.1.1.28.2) Brazilian lineage[1].	[1,48]
C13335T	1/274	Non-synonymous: A104V	ORF1ab/Nsp10	Unresolved.	[30]
T14313C	1/274	Synonymous: D291D	ORF1ab/RdRp	Unresolved.	[30]
C14408T	23/274	Non-synonymous: P323L	ORF1ab/RdRp	Unresolved modified activity. This mutant are more likely to have mutations in the membrane (M) and envelope proteins (E)[49]. Suggested to rigidify the RdRp protein structure[49].	[6,9,23,49–52]
G14559T	1/274	Synonymous: V373V	ORF1ab/RdRp	Unresolved.	Not found.
C14805T	1/274	Synonymous: Y455Y	ORF1ab/RdRp	Typically found with G11083T and G26114T[53]	[8,9,53]
T16176C	1/274	Synonymous: T912T	ORF1ab/RdRp	Unresolved.	[31]
C16806T	2/274	Synonymous: N190N	ORF1ab/Helicase	Unresolved.	[30]
G17014T	4/274	Non-synonymous: D260Y	ORF1ab/Helicase	Unresolved.	[46]
A17066G	1/274	Non-synonymous: Y277C	ORF1ab/Helicase	Unresolved	Not found.
G18155A	1/274	Non-synonymous: C39Y	ORF1ab/3’ to 5’ exonuclease	Unresolved.	Not found.
A18424G	1/274	Non-synonymous: N129D	ORF1ab/3’ to 5’ exonuclease	Unresolved.	[32]
G18651T	1/274	Non-synonymous: E204D	ORF1ab/3’ to 5’ exonuclease	Unresolved	Not found
C18998T	1/274	Non-synonymous: A320V	ORF1ab/3’ to 5’ exonuclease	Unresolved.	[13,45,54]
A19578G	2/274	Synonymous: Q513Q	ORF1ab/3’ to 5’ exonuclease	Unresolved.	Not found.
A19974G	1/274	Synonymous: P118P	ORF1b/ EndoRNAse	Unresolved.	[30]
C19983T	1/274	Synonymous: V121V	ORF1b/ EndoRNAse	Unresolved.	[55]
G20014T	2/274	Non-synonymous: D132Y	ORF1b/ EndoRNAse	Unresolved.	Not found

C20480T	1/274	Non-synonymous: S288L	ORF1b/ EndoRNAse	Unresolved.	Not found
G21088T	1/274	Non-synonymous: D144Y	ORF1ab/2'-O-methyltransferase	Unresolved.	[30]
C21302T	1/274	Non-synonymous P215L	ORF1ab/2'-O-methyltransferase	Unresolved.	[30]
C21304A	1/274	Non-synonymous: R216S	ORF1ab/2'-O-methyltransferase	Unresolved.	[23,30]
G21305A	1/274	Non-synonymous: R216H	ORF1ab/2'-O-methyltransferase	Unresolved.	[30]
C21306T	1/274	Synonymous: R216R	ORF1ab/2'-O-methyltransferase	Unresolved.	[30]
A21583G	1/274	Synonymous: L7L	S/Surface glycoprotein	Unresolved.	Not found
G23012A	1/274	Non-synonymous: E484K	S/Surface glycoprotein	Confers resistance to monoclonal antibodies[56]. Mutant associated with the P.1 (Brazil) and B.1.351 (South Africa) VOC[57].May improve binding affinity between RBD and the hACE2 receptor, potentially increasing transmission . Associated with cases of re-infection in Brazilian patients[57,59].	[45,56–59]
A23064C	1/274	Non-synonymous: N501T	S/Surface glycoprotein	This mutant could potentially increase the binding affinity between SARS-CoV-2 and the human receptor ACE2[60–64].	[60–65]
C23271A	1/274	Non-synonymous: A570T	S/Surface glycoprotein	SNP associated with the B.1.1.7 VOC[66]. Mutant correlated with 70% more transmissibility than the wild type[45].	[45,66,67]
A23403G	19/274	Non-synonymous: D614G	S/Surface glycoprotein	Unresolved modified activity. Found simultaneously with G25563T variation on ORF3a[9]. Potentially increased infectivity* [68,69]	[4,8,9,23,24,68,69]
G23608T	2/274	Synonymous: R682R	S/Surface glycoprotein	Unresolved.	Not found.
A23756G	1/274	Non-synonymous: T732A	S/Surface glycoprotein	Unresolved.	[30]
G25088T	1/274	Non-synonymous: V1176F	S/Surface glycoprotein	Correlated to increased mortality in Brazilian and Saudi-Arabian patients. Typically present with the D614G mutant[70].	[70]
A25097G	1/274	Non-synonymous: I1179V	S/Surface glycoprotein	Unresolved.	Not found.
G25234T	1/274	Non-synonymous: L1224F	S/Surface glycoprotein	Unresolved.	[71]
G25563T	17/274	Non-synonymous: Q57H	ORF3a/ORF3a protein	Unresolved modified activity. Typically found in C27964T mutants [72].	[9,13,30,47,50,72–74]
C25603T	1/274	Synonymous: L71L	ORF3a/ORF3a protein	Unresolved.	Not found.
C25688T	1/274	Non-synonymous: A99V	ORF3a/ORF3a protein	Unresolved.	[26]
A25718G	1/274	Non-synonymous: Y109C	OR3a/ORF3a protein	Unresolved.	[75]
C27213T	1/274	Synonymous: L4L	ORF6/ORF6 protein	Unresolved.	[26]
C27964T	7/274	Non-synonymous: S24L	ORF8/ORF8 protein	Unresolved modified activity. Typically found in G25563T mutants [72].	[13,30,72,76,77]
C27972T	1/274	Nonsense: Q27Stop	ORF8/ORF8 protein	Truncates the ORF8 protein and potentially inactivates the protein. This variant has been associated with the B.1.1.7 VOC[31]	[15,31,78]
G28048T	1/274	Non-synonymous: R52I	ORF8/ORF8 protein	SNP associated with the B.1.1.7 VOC [15,31,79].	[15,31,79]
A28095T	1/274	Nonsense: K68Stop	ORF8/ORF8 protein	This mutant has been found to be present along with the Q27Stop mutation[80].	[80]
A28111G	1/274	Non-synonymous: Y73C	ORF8/ORF8 protein	SNP typically found in B.1.1.7 VOC[81].	[81,82]
C28253T	1/274	Synonymous: F120F	ORF8/ORF8 protein	Unresolved.	[29,83]

28270 Δ	1/274	Deletion	N/A	Unresolved.	[30,84]
C28472T	1/274	Non-synonymous: P67S	N/Nucleocapsid phosphoprotein	Unresolved.	[30,32]
G28628T	1/274	Non-synonymous: A119S	N/Nucleocapsid phosphoprotein	SNP associated with the P.2 (also known as B.1.1.28.2) Brazilian lineage[1].	[1,39]
C28869T	1/274	Non-synonymous: P199L	N/Nucleocapsid phosphoprotein	Unresolved.	[85]
G28881A	2/274	Non-synonymous: R203K	N/Nucleocapsid phosphoprotein	*Unknown function. SNPs G28881A, G28882A, G28883C typically found together in a homopolymer mutant. Strong allelic pairwise association of the three SNPs clustered together.	[7,8,74,86]
G28882A	2/274	Non-synonymous: R203K	N/Nucelocapsid phosphoprotein	*Unknown function. SNPs G28881A, G28882A, G28883C typically found together in a homopolymer mutant. Strong allelic pairwise association of the three SNPs clustered together.	[7,8,86]
G28883C	2/274	Non-synonymous: G204R	N/Nucleocapsid phosphoprotein	*Unknown function. SNPs G28881A, G28882A, G28883C typically found together in a homopolymer mutant. Strong allelic pairwise association of the three SNPs clustered together.	[7,8,86]
T28921C	6/274	Synonymous: D206D	N/Nucleocapsid phosphoprotein	Unresolved.	Not found.
G28975T	6/274	Non-synonymous: M234I	N/Nucleocapsid phosphoprotein	Unresolved.	[85]
C28977T	6/274	Non-synonymous: S235F	N/Nucleocapsid phosphoprotein	Characteristic mutation of the B.1.1.7 VOC.[81] Predicted to provide a stabilization effect to the N protein. [82]	[15,81,82]
A29412T	1/274	Non-synonymous: Q380L	N/Nucleocapsid phosphoprotein	Unresolved.	Not found.
G29540A	1/274	Non-translatable SNP	3'UTR	Unresolved.	[87,88]
29743Δ	1/274	Deletion	3'UTR	Unresolved.	Not found
29746Δ	1/274	Deletion	3'UTR	Unresolved.	[89]
29749Δ	1/274	Deletion	3'UTR	Unresolved.	[90]
C29754T	1/274	Non-translatable SNP	3'UTR	SNP associated with the P.2 (also known as B.1.1.28.2) Brazilian lineage[1].	[1]
C29835T	1/274	Non-translatable SNP	3'UTR	Unresolved.	Not found.

Supplementary table 6. Identified SNPs per patient with its correspondent mutant, lineage and method used for lineage identification. Highlighted SNPs have been previously correlated to VOC/VOIs.

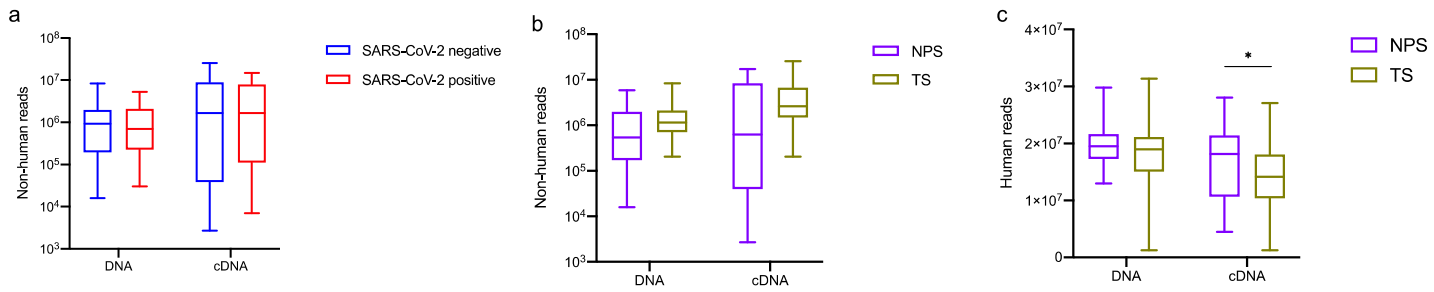
Patient ID	**Identified SNPs**	**Identified mutants**	**VOC lineage**
P004	C241T, C379A, C1059T, C3037T, C14408T, G25563T	Nsp1: L64L Nsp2: T266I; Nsp3: F106F; RdRp: P323L; ORF3a protein: Q57H	N/A
P019	C241T, C1059T, C3037T, G8216A, C14408T, G18155A, A23403G, G25563T	Nsp2: T266I; Nsp3: F106F;, V1833I; RdRp: P323L; 3' to 5' exonuclease: C39Y; Surface glycoprotein: D614G; ORF3a protein: Q57H	N/A
P022	C241T, C1059T, C3037T, C14408T, G17014T, A23403G, G25563T, C27964T	Nsp2: T266I; Nsp3: F106F; RdRp: P323L; Helicase: D260Y; Surface glycoprotein: D614G; ORF3a protein: Q57H; ORF8 protein: S24L	N/A
P35	A187G, C241T, C1616T, C3037T, C14408G, A17066G, C19983T, A23403G	Nsp2: L271F; Nsp3: F106F; RdRp: P323L; Helicase: Y277C; EndoRNAse: V121V; Surface glycoprotein: D614G	N/A
P066	A187G, C241T, C1616T, C3037T, C14408T	Nsp2: L271F; Nsp3: F106F; RdRp: P323L	N/A
P069	A187G, C241T, C1616T, C3037T, C14408T, A23403G, G23608T, A25097G	Nsp2: L271F; Nsp3: F106F; RdRp: P323L; Surface glycoprotein: D614G, R682R, I1179V	N/A
P082	C241T, C1059T, C3037T, C14408T, A23403G, G25563T, C25688T	Nsp2: T266I; Nsp3: F106F; RdRp: P323L; Surface glycoprotein: D614G; ORF3a protein: Q57H, A99V	N/A
P721	C241T, C1059T, C3037T, C6664G, A6796G, G8083A, C10319T, C12357T, C14408T, C14805T, A18424G, G21088T, C21302T, C21304A, G21305A, A23403G, G25563T, C27964T, C28472T, C28869T, A29412T	Nsp2: T266I; Nsp3: F106F, V1315V, L1359L,M1788I; 3C-like proteinase: L89F; Nsp8: T89I; RdRp: P323L, Y455Y; 3' to 5' exonuclease: N129D; 2-O-methyltransferase: 216H, P215L, R216S, R216H; Surface glycoprotein: D614G; ORF3a protein: Q57H; ORF8 protein: S24L; Nucleocapsid phosphoprotein: P67S, P199L, Q380L	N/A
P259	C1059T, C3037T, A23403G, G25563T	Nsp2: T266I; Nsp3: F106F; Surface glycoprotein: D614G; ORF3a protein: Q57H	N/A
P725	C1059T, C3037T, A23403G, G25563T	Nsp2: T266I; Nsp3: F106F; Surface glycoprotein:	N/A

		D614G; ORF3a protein: Q57H	
P264	C241T, C1059T, T2423C, C3037T, C3096T, C5144T, C14408T, G17014T, A23403G, G25563T, C27964T	Nsp2: T266I, C540R; Nsp3: F106F, S126L, L809L; RdRp: P323L; Helicase: D260Y; Surface glycoprotein: D614G; ORF3a protein: Q57H; ORF8 protein: S24L	N/A
P267	C3037T, C4534T, C5144T, C14408T, G17014T, A23403G, G25563T, C27964T	Nsp3: F106F, Y605H, L809L; RdRp: P323L; Helicase: D260Y; Surface glycoprotein: D614G; ORF3a protein: Q57H; ORF8 protein: S24L	N/A
P273	C241T, C1059T, C3037T, T3905G, C6310T, A10770G, T11059G, C14408T, A23064C, G23608T, G25563T, C27964T	Nsp2: T266I; Nsp3: F106F, F396V, S1197S, 3C-like proteinase: Y239C; Nsp6: T29T; RdRp: P323L; Surface glycoprotein: N501T, R682R; ORF3a protein: Q57H; ORF8 protein: S24L	N/A
P276	C241T, 507del, C1059T, C3037T, G5992T, C10422T, T11059G, C14408T, G17014T, A23403G, G25563T, C27964T	Nsp2: T266I; Nsp3: F106F, E1091D; 3C-like proteinase: S123F; Nsp6: T29T; RdRp: P323L; Helicase: D260Y; Surface glycoprotein: D614G; ORF3a protein: Q57H; ORF8 protein: S24L	N/A
P281	C241T, C3037T, C12412T, T14313C, C14408T, A23403G, G28881A, G28882A, G28883C	Nsp3: F106F; Nsp8: I107I; RdRp: D291D, P323 (or P314L in NSP12b)L; Surface glycoprotein: D614G; Nucleocapsid phosphoprotein: R203K, G204R	Potentially VOC P.2
P289	A2480G, C2558T, C6310T, G11083T, C13335T	Nsp2: I559V, P585S; Nsp3: S1197S; Nsp6:L37F; Nsp10: A104V	N/A
P297	C241T, C379A, C1059T, C3037T, C14408T, C20480T, A23403G, G25563T	Nsp1: L64L Nsp2: T266I; Nsp3: F106F; RdRp: P323L; EndoRNAse: S288L; Surface glycoprotein: D614G; ORF3a protein: Q57H	N/A
P303	C241T, C379A, C1059T, C3037T, C14408T, G25563T	Nsp1: L64L Nsp2: T266I; Nsp3: F106F; RdRp: P323L; ORF3a protein: Q57H	N/A
P306	A187G, C241T, C1616T, C3037T, A6425G, G7675T, C9348T, C14408T, A23403G	Nsp2: L271F; Nsp3: F106F, N1236, A1652A; Nsp4: T295I; RdRp: P323L; Surface glycoprotein: D614G	N/A
P311	A187G, C241T, C1616T, C3037T, A23403G	Nsp2: L271F; Nsp3: F106F; Surface glycoprotein: D614G	N/A
P312	A187G, C241T, C1616T, C3037T, G7675T, C14408T	Nsp2: L271F; Nsp3: F106F, A1652A; RdRp: P323L	N/A
P733	C10319T, C10507T, C11289T, A23403G, G25563T, C27964T, C28472T, C28869T	3C-like proteinase: L89F, N151N; Nsp6: S106F; Surface glycoprotein: D614G; ORF3a protein: Q57H; ORF8 protein: S24L; Nucleocapsid phosphoprotein: P67S, P199L	N/A
P344	C241T, A23403G, G25563T	Surface glycoprotein: D614G; ORF3a protein: Q57H	N/A
P348	C241T, C2416T, G2528A, A3405G, C11494T, C14408T, C16806T, G18651T, G20014T, A23403G, G25234T, G25563T, A25718G, C27213T, 29743del	Nsp2: Y537Y, E575K; Nsp3: E229G; Nsp6: N174N; RdRp: P323L; Helicase: N190N; 3' to 5' exonuclease: E204D; EndoRNAse: D132Y; Surface glycoprotein: D614,: L1224F; ORF3a protein: Q57H, Y109C; ORF6 protein: L4L	N/A
P734	C10029T, C12789T, G12794A, G14559T, A19974G, C21306T, A23756G, G28881A, G28882A, G28883C, T28912C	Nsp4: T492I; Nsp9: T35I; Nsp9: G37R; RdRp: V373V; EndoRNAse: P118P; 2-O-methyltransferase: R216R; Surface glycoprotein: T732A; Nucleocapsid phosphoprotein: R203K, G204R, D260D	N/A
P356	C11812A, C16806T, G20014T	Nsp6: G280G; Helicase: N190N; EndoRNAse: D132Y	N/A
P363	C11916T, C14408T, C18998T, A21583G, G25563T, G29540A, 29746del, 29749del	Nsp7: S25L; RdRp: P323L; 3' to 5' exonuclease: A320V; Surface glycoprotein: L7L; ORF3a protein: Q57H	N/A
P369	C14408T, G28881A, G28882A, G28883C	RdRp: P323L; Nucleocapsid phosphoprotein: R203K, G204R	N/A
P739**	C913T, C2110T, C3267T, C5986T, G7745A, C14408T, T16176C, C23271A, C25603T, C27972T, G28048T, A28095T, A28111G	Nsp2: S36S, N435N; Nsp3: T183I, A890D,F1089F, I1412T, V1676I; Nsp6: 104-107 del; RdRp: P323L (or P314L in NSP12b), T912T; Surface glycoprotein: 69/70 del, 144 del, N501Y, A570D, P681H, D1118H; ORF3a protein: L71L; ORF8 protein: Q27Stop, R52I, K68Stop, Y73C; Nucleocapsid phosphoprotein: 3DL, S253F	B.1.1.7 United Kingdom VOC
P743	G2424A, C12053T, C14408T, A19578G, G28628T, G28881A, G28882A, G28883C, C29754T	Nsp2: C540Y Nsp7: L71F; RdRp: P323L (or P314L in NSP12b); 3' to 5' exonuclease: Q513Q; Nucleocapsid phosphoprotein: A119S, R203K, G204R	P.2 Brazilian VOC
P744	C100T, C241T, G2424A, C3037T, T3477C, T3766C, C5694T, C9711T, T10667G, C11824T, C12053T, A12964G, C14408T, A19578G, G23012A, A23403G, G25088T, C28253T, G28628T, G28881A, G28882A, G28883C, G28975T, C29754T, C29835T	5'UTR: R81C, Nsp2: C540Y; Nsp3: F106F, V253A, D349D, P992L; Nsp4: S387F; 3C-like proteinase: L205V; Nsp6: I284I; Nsp7: L71F; Nsp9: G93G; RdRp: P323L (or P314L in NSP12b); 3' to 5' exonuclease: Q513Q; Surface glycoprotein: E484K, D614G, V1176F; ORF8 protein: F120F; Nucleocapsid phosphoprotein: A119S, R203K, G204R, M234I	P.2 Brazilian VOC

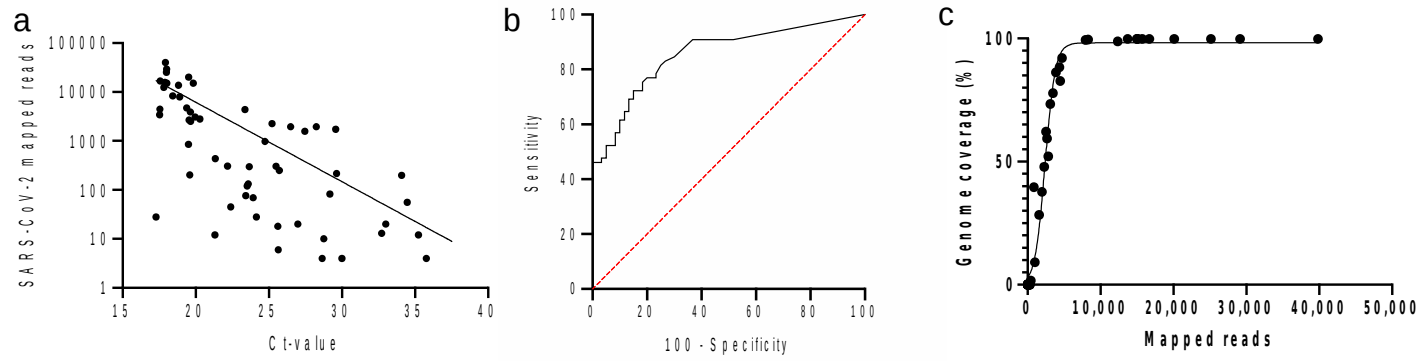
*: Homopolymer stretch of three SNPs occur in the two codons, the G2881A and G28882A translate together to lysine (K).

**: Majority of SNPs were detected under the default parameter threshold for SNP calling from IDseq.

Supplementary figures

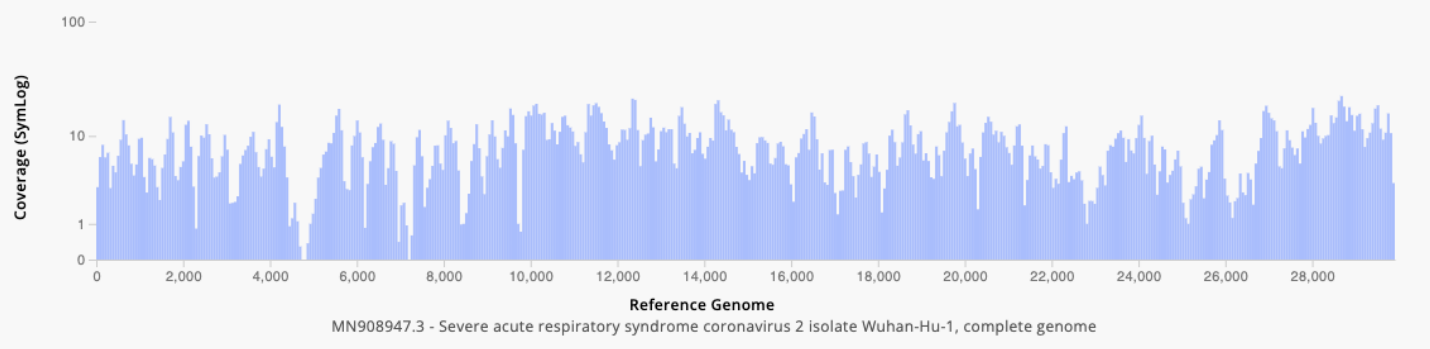


Supplementary figure 1. Number of human and non-human DNA and cDNA reads between SARS-CoV-2 diagnosis and anatomical swabbing site. (a) Number of non-human DNA and cDNA reads by SARS-CoV-2 diagnosis. Number of non-human (b) and human (c) DNA and cDNA reads by anatomical swabbing site.(*) indicates a significant p-value below 0.05.

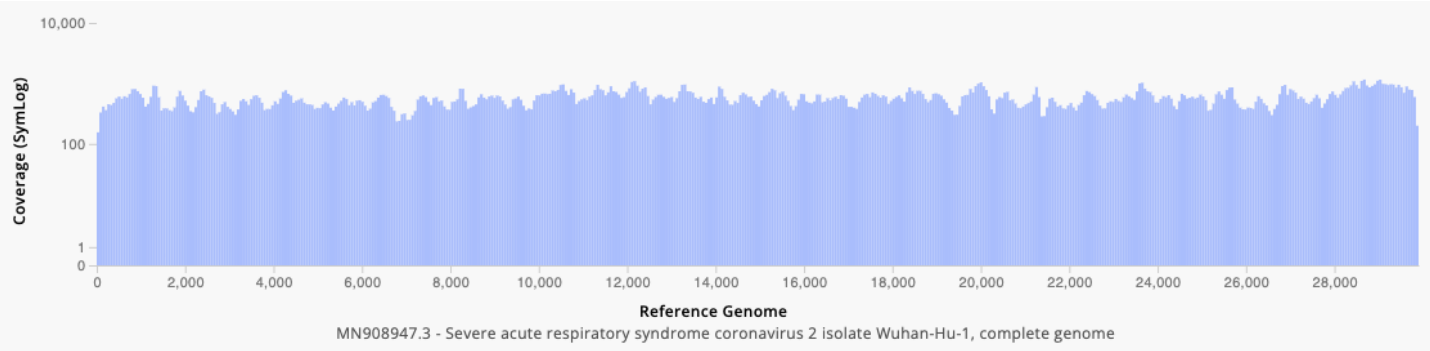


Supplementary figure 2. Regression analysis and ROC evaluation of all the clinical samples (n=125). (a) Exponential regression model of SARS-CoV-2 mapped reads vs RT-PCR Ct value. (b) ROC curve evaluation with a red dashed line depicting the 0.50 AUC. (c) Logistic regression model of genome coverage vs number of mapped reads.

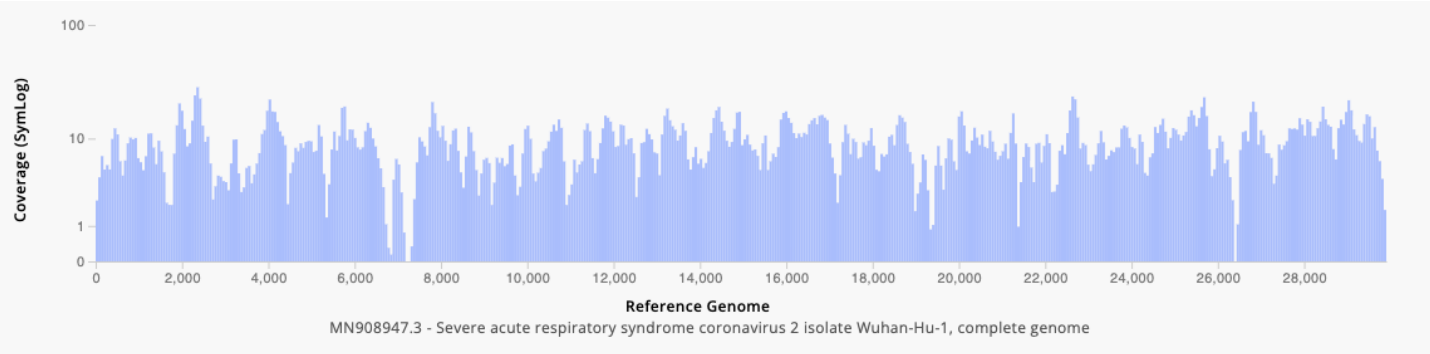
a



b



c



Supplementary figure 3. Genome coverage visualization of the (a,b) Zeta VOIs and the (c) Alpha VOC identified. The 'y' axis represents the depth of coverage of each nucleotide and the 'x' axis represents the nucleotide position of the MN908947.3 reference genome.

References:

1. Ferrareze, P. A. G. *et al.* E484K as an innovative phylogenetic event for viral evolution: Genomic analysis of the E484K spike mutation in SARS-CoV-2 lineages from Brazil. *bioRxiv* 2021.01.27.426895 (2021) doi:10.1101/2021.01.27.426895.
2. Yang, X., Dong, N., Chan, E. W.-C. & Chen, S. Genetic cluster analysis of SARS-CoV-2 and the identification of those responsible for the major outbreaks in various countries. *Emerg. Microbes Infect.* **9**, 1287–1299 (2020).

3. Large scale genomic analysis of 3067 SARS-CoV-2 genomes reveals a clonal geo-distribution and a rich genetic variations of hotspots mutations. 25.
4. Identification of super-transmitters of SARS-CoV-2. 38.
5. Chaudhari, A. *et al.* In-Silico analysis reveals lower transcription efficiency of C241T variant of SARS-CoV-2 with host replication factors MADP1 and HNRNP-1. *bioRxiv* 2020.11.22.393009 (2020) doi:10.1101/2020.11.22.393009.
6. Mercatelli, D. & Giorgi, F. M. Geographic and Genomic Distribution of SARS-CoV-2 Mutations. *Front. Microbiol.* **11**, (2020).
7. Mishra, A. *et al.* Mutation landscape of SARS-CoV-2 reveals three mutually exclusive clusters of leading and trailing single nucleotide substitutions. 19.
8. Yang, H.-C. *et al.* Analysis of genomic distributions of SARS-CoV-2 reveals a dominant strain type with strong allelic associations. *Proc. Natl. Acad. Sci.* **117**, 30679–30686 (2020).
9. UGUREL, O. M., ATA, O. & TURGUT-BALIK, D. An updated analysis of variations in SARS-CoV-2 genome. *Turk. J. Biol.* **44**, 157–167 (2020).
10. Bartolini, B. *et al.* SARS-CoV-2 Phylogenetic Analysis, Lazio Region, Italy, February–March 2020 - Volume 26, Number 8—August 2020 - Emerging Infectious Diseases journal - CDC. doi:10.3201/eid2608.201525.
11. Sashittal, P., Luo, Y., Peng, J. & El-Kebir, M. Characterization of SARS-CoV-2 viral diversity within and across hosts. *bioRxiv* 2020.05.07.083410 (2020) doi:10.1101/2020.05.07.083410.
12. Wan, J. *et al.* Human-IgG-Neutralizing Monoclonal Antibodies Block the SARS-CoV-2 Infection. *Cell Rep.* **32**, 107918 (2020).
13. Gomez-Carballa, A., Bello, X., Pardo-Seco, J., Martinon-Torres, F. & Salas, A. Mapping genome variation of SARS-CoV-2 worldwide highlights the impact of COVID-19 super-spreaders. *Genome Res.* gr.266221.120 (2020) doi:10.1101/gr.266221.120.
14. Guo, L. *et al.* Genomic epidemiology of the Los Angeles COVID-19 outbreak. *medRxiv* 2020.09.15.20194712 (2020) doi:10.1101/2020.09.15.20194712.

15. Public Health England. Investigation of SARS-CoV-2 variants of concern in England.
16. Liu, S. *et al.* Genetic Spectrum and Distinct Evolution Patterns of SARS-CoV-2. *Front. Microbiol.* **11**, (2020).
17. Zhang, W. *et al. Analysis of SARS-CoV-2 Genomes from Southern California Reveals Community Transmission Pathways in the Early Stage of the US COVID-19 Pandemic*. http://medrxiv.org/lookup/doi/10.1101/2020.06.12.20129999 (2020) doi:10.1101/2020.06.12.20129999.
18. Lemieux, J. E. *et al.* Phylogenetic analysis of SARS-CoV-2 in Boston highlights the impact of superspreading events. *Science* **371**, (2021).
19. Gámbaro, F. *et al.* Introductions and early spread of SARS-CoV-2 in France, 24 January to 23 March 2020. *Eurosurveillance* **25**, (2020).
20. Hanifehnezhad, A. *et al.* Characterization of local SARS-CoV-2 isolates and pathogenicity in IFNAR−/- mice. *Heliyon* **6**, e05116 (2020).
21. Martin, J. *et al.* Tracking SARS-CoV-2 in Sewage: Evidence of Changes in Virus Variant Predominance during COVID-19 Pandemic. *Viruses* **12**, (2020).
22. Li, J. *et al.* Rapid genomic characterization of SARS-CoV-2 viruses from clinical specimens using nanopore sequencing. *Sci. Rep.* **10**, 17492 (2020).
23. DEMİR, A. B., BENVENUTO, D., ABACIOĞLU, H., ANGELETTI, S. & CICCOZZI, M. Identification of the nucleotide substitutions in 62 SARS-CoV-2 sequences from Turkey. *Turk. J. Biol.* **44**, 178–184 (2020).
24. Nguyen, T. T. *et al.* Genetic diversity of SARS-CoV-2 and clinical, epidemiological characteristics of COVID-19 patients in Hanoi, Vietnam. *PLOS ONE* **15**, e0242537 (2020).
25. Salazar, C. *et al. Multiple introductions, regional spread and local differentiation during the first week of COVID-19 epidemic in Montevideo, Uruguay*. http://biorxiv.org/lookup/doi/10.1101/2020.05.09.086223 (2020) doi:10.1101/2020.05.09.086223.
26. Snell, L. B. *et al.* Combined epidemiological and genomic analysis of nosocomial SARS-CoV-2 transmission identifies community social distancing as the dominant intervention reducing outbreaks. *medRxiv* 2020.11.17.20232827 (2020) doi:10.1101/2020.11.17.20232827.

27. Loconsole, D. *et al.* Genome Sequence of a SARS-CoV-2 VUI 202012/01 Strain Identified from a Patient Returning from London, England, to the Apulia Region of Italy. *Microbiol. Resour. Announc.* **10**, (2021).
28. Franceschi, V. B. *et al.* Genomic Epidemiology of SARS-CoV-2 in Esteio, Rio Grande do Sul, Brazil. *medRxiv* 2021.01.21.21249906 (2021) doi:10.1101/2021.01.21.21249906.
29. Wen, S. *et al.* High-coverage SARS-CoV-2 genome sequences acquired by target capture sequencing. *J. Med. Virol.* **92**, 2221–2226 (2020).
30. Joshi, M. *et al.* Genomic variations in SARS-CoV-2 genomes from Gujarat: Underlying role of variants in disease epidemiology. *bioRxiv* 2020.07.10.197095 (2020) doi:10.1101/2020.07.10.197095.
31. Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations - SARS-CoV-2 coronavirus / nCoV-2019 Genomic Epidemiology. *Virological* https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563 (2020).
32. Pater, A. A. *et al.* Emergence and Evolution of a Prevalent New SARS-CoV-2 Variant in the United States. *bioRxiv* 2021.01.11.426287 (2021) doi:10.1101/2021.01.11.426287.
33. Leist, S. R. *et al.* A Mouse-Adapted SARS-CoV-2 Induces Acute Lung Injury and Mortality in Standard Laboratory Mice. *Cell* **183**, 1070-1085.e12 (2020).
34. SARS-CoV-2 has observably higher propensity to accept uracil as nucleotide substitution: Prevalence of amino acid substitutions and their predicted functional implications in circulating SARS-CoV-2 in India up to July, 2020 - Abstract - Europe PMC. https://europepmc.org/article/PPR/PPR223256.
35. Shen, L. *et al.* Emerging variants of concern in SARS-CoV-2 membrane protein: a highly conserved target with potential pathological and therapeutic implications. *bioRxiv* 2021.03.11.434758 (2021) doi:10.1101/2021.03.11.434758.
36. Zeng, L., Li, D., Tong, W., Shi, T. & Ning, B. Biochemical features and mutations of key proteins in SARS-CoV-2 and their impacts on RNA therapeutics. *Biochem. Pharmacol.* 114424 (2021) doi:10.1016/j.bcp.2021.114424.

37. Gimeno, A. *et al.* Prediction of Novel Inhibitors of the Main Protease (M-pro) of SARS-CoV-2 through Consensus Docking and Drug Reposition. *Int. J. Mol. Sci.* **21**, (2020).
38. Turakhia, Y. *et al.* Stability of SARS-CoV-2 phylogenies. *PLOS Genet.* **16**, e1009175 (2020).
39. Phylogenetic relationship of SARS-CoV-2 sequences from Amazonas with emerging Brazilian variants harboring mutations E484K and N501Y in the Spike protein - SARS-CoV-2 coronavirus / nCoV-2019 Genomic Epidemiology. *Virological* https://virological.org/t/phylogenetic-relationship-of-sars-cov-2-sequences-from-amazonas-with-emerging-brazilian-variants-harboring-mutations-e484k-and-n501y-in-the-spike-protein/585 (2021).
40. Sekizuka, T. *et al.* Haplotype networks of SARS-CoV-2 infections in the Diamond Princess cruise ship outbreak. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 20198–20201 (2020).
41. Capobianchi, M. R. *et al.* Molecular characterization of SARS-CoV-2 from the first case of COVID-19 in Italy. *Clin. Microbiol. Infect. Off. Publ. Eur. Soc. Clin. Microbiol. Infect. Dis.* **26**, 954–956 (2020).
42. Spike E484K mutation in the first SARS-CoV-2 reinfection case confirmed in Brazil, 2020 - SARS-CoV-2 coronavirus / nCoV-2019 Genomic Epidemiology. *Virological* https://virological.org/t/spike-e484k-mutation-in-the-first-sars-cov-2-reinfection-case-confirmed-in-brazil-2020/584 (2021).
43. Løvestad, A. H., Jørgensen, S. B., Handal, N., Ambur, O. H. & Aamot, H. V. Investigation of intra-hospital SARS-CoV-2 transmission using nanopore whole genome sequencing. *J. Hosp. Infect.* **0**, (2021).
44. Nagy, Á., Pongor, S. & Győrffy, B. Different mutations in SARS-CoV-2 associate with severe and mild outcome. *Int. J. Antimicrob. Agents* **57**, 106272 (2021).
45. Fang, S. *et al.* Updated SARS-CoV-2 Single Nucleotide Variants and Mortality Association. *medRxiv* 2021.01.29.21250757 (2021) doi:10.1101/2021.01.29.21250757.
46. Patro, L. P. P., Sathyaseelan, C., Uttamrao, P. P. & Rathinavelan, T. Global variation in the SARS-CoV-2 proteome reveals the mutational hotspots in the drug and vaccine candidates. *bioRxiv* 2020.07.31.230987 (2020) doi:10.1101/2020.07.31.230987.

47. Gniazdowski, V. *et al.* Repeat COVID-19 Molecular Testing: Correlation of SARS-CoV-2 Culture with Molecular Assays and Cycle Thresholds. *Clin. Infect. Dis. Off. Publ. Infect. Dis. Soc. Am.* (2020) doi:10.1093/cid/ciaa1616.

48. Francisco, R. da S. *et al.* Pervasive transmission of E484K and emergence of VUI-NP13L with evidence of SARS-CoV-2 co-infection events by two different lineages in Rio Grande do Sul, Brazil. *medRxiv* 2021.01.21.21249764 (2021) doi:10.1101/2021.01.21.21249764.

49. Eskier, D., Karakülah, G., Suner, A. & Oktay, Y. RdRp mutations are associated with SARS-CoV-2 genome evolution. *PeerJ* **8**, (2020).

50. Yuan, F., Wang, L., Fang, Y. & Wang, L. Global SNP analysis of 11,183 SARS-CoV-2 strains reveals high genetic diversity. *Transbound. Emerg. Dis.* (2020) doi:10.1111/tbed.13931.

51. Yu, J.-M. *et al.* Analysis of Continuous Mutation and Evolution on Circulating SARS-CoV-2. *Evol. Bioinforma.* **16**, 1176934320954870 (2020).

52. Pachetti, M. *et al.* Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant. *J. Transl. Med.* **18**, 179 (2020).

53. Kim, J.-M. *et al.* Genomic investigation of the coronavirus disease-2019 outbreak in the Republic of Korea. *Sci. Rep.* **11**, 6009 (2021).

54. Everett, J. *et al.* SARS-CoV-2 Genomic Variation in Space and Time in Hospitalized Patients in Philadelphia. *mBio* **12**, (2021).

55. Velasco, J. M. *et al.* Coding-Complete Genome Sequences of 23 SARS-CoV-2 Samples from the Philippines. *Microbiol. Resour. Announc.* **9**, (2020).

56. Xie, X. *et al.* Neutralization of SARS-CoV-2 spike 69/70 deletion, E484K and N501Y variants by BNT162b2 vaccine-elicited sera. *Nat. Med.* 1–2 (2021) doi:10.1038/s41591-021-01270-4.

57. Nonaka, C. K. V. *et al.* Early Release - Genomic Evidence of SARS-CoV-2 Reinfection Involving E484K Spike Mutation, Brazil - Volume 27, Number 5—May 2021 - Emerging Infectious Diseases journal - CDC. doi:10.3201/eid2705.210191.

58. Wang, W. B. *et al.* E484K mutation in SARS-CoV-2 RBD enhances binding affinity with hACE2 but reduces interactions with neutralizing antibodies and nanobodies: Binding free energy calculation studies. *bioRxiv* 2021.02.17.431566 (2021) doi:10.1101/2021.02.17.431566.

59. Resende, P. C. *et al.* Genomic surveillance of SARS-CoV-2 reveals community transmission of a major lineage during the early pandemic phase in Brazil. *bioRxiv* 2020.06.17.158006 (2020) doi:10.1101/2020.06.17.158006.

60. Wan, Y., Shang, J., Graham, R., Baric, R. S. & Li, F. Receptor Recognition by the Novel Coronavirus from Wuhan: an Analysis Based on Decade-Long Structural Studies of SARS Coronavirus. *J. Virol.* **94**, (2020).

61. Fiorentini, S. *et al.* First detection of SARS-CoV-2 spike protein N501 mutation in Italy in August, 2020. *Lancet Infect. Dis.* **0**, (2021).

62. Ahamad, S., Kanipakam, H. & Gupta, D. Insights into the structural and dynamical changes of spike glycoprotein mutations associated with SARS-CoV-2 host receptor binding. *J. Biomol. Struct. Dyn.* 1–13 doi:10.1080/07391102.2020.1811774.

63. Welkers, M. R. A., Han, A. X., Reusken, C. B. E. M. & Eggink, D. Possible host-adaptation of SARS-CoV-2 due to improved ACE2 receptor binding in mink. *Virus Evol.* **7**, (2021).

64. Sardar, R., Satish, D., Birla, S. & Gupta, D. Integrative analyses of SARS-CoV-2 genomes from different geographical locations reveal unique features potentially consequential to host-virus interaction, pathogenesis and clues for novel therapies. *Heliyon* **6**, e04658 (2020).

65. Shang, J. *et al.* Structural basis of receptor recognition by SARS-CoV-2. *Nature* **581**, 221–224 (2020).

66. Calcagnile, M., Forgez, P., Alifano, M. & Alifano, P. The lethal triad: SARS-CoV-2 Spike, ACE2 and TMPRSS2. Mutations in host and pathogen may affect the course of pandemic. *bioRxiv* 2021.01.12.426365 (2021) doi:10.1101/2021.01.12.426365.

67. Kemp, S. A. *et al.* Recurrent emergence and transmission of a SARS-CoV-2 Spike deletion ΔH69/ΔV70. *bioRxiv* 2020.12.14.422555 (2020) doi:10.1101/2020.12.14.422555.

68. Korber, B. *et al.* Spike mutation pipeline reveals the emergence of a more transmissible form of SARS-CoV-2. *bioRxiv* 2020.04.29.069054 (2020) doi:10.1101/2020.04.29.069054.

69. Zhang, L. *et al.* The D614G mutation in the SARS-CoV-2 spike protein reduces S1 shedding and increases infectivity. *bioRxiv* 2020.06.12.148726 (2020) doi:10.1101/2020.06.12.148726.

70. Farkas, C., Mella, A. & Haigh, J. J. Large-scale population analysis of SARS-CoV-2 whole genome sequences reveals host-mediated viral evolution with emergence of mutations in the viral Spike protein associated with elevated mortality rates. *medRxiv* 2020.10.23.20218511 (2020) doi:10.1101/2020.10.23.20218511.

71. Long, S. W. *et al.* Molecular Architecture of Early Dissemination and Massive Second Wave of the SARS-CoV-2 Virus in a Major Metropolitan Area. *medRxiv* 2020.09.22.20199125 (2020) doi:10.1101/2020.09.22.20199125.

72. Cai, H. Y., Cai, K. K. & Li, J. Identification of Novel Missense Mutations in a Large Number of Recent SARS-CoV-2 Genome Sequences. *J. Biotechnol. Biomed.* **3**, 93–103 (2020).

73. Hu, Y. & Riley, L. W. Dissemination and co-circulation of SARS-CoV2 subclades exhibiting enhanced transmission associated with increased mortality in Western Europe and the United States. *medRxiv* 2020.07.13.20152959 (2020) doi:10.1101/2020.07.13.20152959.

74. Paul, D. *et al.* Phylogenomic analysis of SARS-CoV-2 genomes from western India reveals unique linked mutations. (2020).

75. Majumdar, P. & Niyogi, S. ORF3a mutation associated with higher mortality rate in SARS-CoV-2 infection. *Epidemiol. Infect.* **148**,.

76. Yu, Y. *et al. CovProfile: profiling the viral genome and gene expressions of SARS-COV-2*. http://biorxiv.org/lookup/doi/10.1101/2020.04.05.026146 (2020) doi:10.1101/2020.04.05.026146.

77. Leuzinger, K. *et al.* Epidemiology and precision of SARS-CoV-2 detection following lockdown and relaxation measures. *J. Med. Virol.* **93**, 2374–2384 (2021).

78. Lopez-Rincon, A. *et al.* Design of Specific Primer Set for Detection of B.1.1.7 SARS-CoV-2 Variant using Deep Learning. *bioRxiv* 2020.12.29.424715 (2020) doi:10.1101/2020.12.29.424715.

79. Tu, H. *et al.* Distinct Patterns of Emergence of SARS-CoV-2 Spike Variants including N501Y in Clinical Samples in Columbus Ohio. *bioRxiv* 2021.01.12.426407 (2021) doi:10.1101/2021.01.12.426407.

80. Pereira, F. SARS-CoV-2 variants combining spike mutations and the absence of ORF8 may be more transmissible and require close monitoring. *Biochem. Biophys. Res. Commun.* **550**, 8–14 (2021).
81. Galloway, S. E. Emergence of SARS-CoV-2 B.1.1.7 Lineage — United States, December 29, 2020–January 12, 2021. *MMWR Morb. Mortal. Wkly. Rep.* **70**, (2021).
82. Singh, J., Ehtesham, N. Z., Rahman, S. A. & Hasnain, S. E. Structure-function investigation of a new VUI-202012/01 SARS-CoV-2 variant. *bioRxiv* 2021.01.01.425028 (2021) doi:10.1101/2021.01.01.425028.
83. Harper, H. *et al.* Detecting SARS-CoV-2 variants with SNP genotyping. *bioRxiv* 2020.11.18.388140 (2020) doi:10.1101/2020.11.18.388140.
84. Khatamzas, E. *et al.* Emergence of multiple SARS-CoV-2 mutations in an immunocompromised host. *medRxiv* 2021.01.10.20248871 (2021) doi:10.1101/2021.01.10.20248871.
85. Ward, D. *et al.* An integrated in silico immuno-genetic analytical platform provides insights into COVID-19 serological and vaccine targets. *Genome Med.* **13**, 4 (2021).
86. Caccuri, F. *et al.* A persistently replicating SARS-CoV-2 variant derived from an asymptomatic individual. *J. Transl. Med.* **18**, 362 (2020).
87. LitCovid. https://www.ncbi.nlm.nih.gov/research/coronavirus/.
88. Ryder, S. P., Morgan, B. R. & Massi, F. Analysis of Rapidly Emerging Variants in Structured Regions of the SARS-CoV-2 Genome. *bioRxiv* 2020.05.27.120105 (2020) doi:10.1101/2020.05.27.120105.
89. Miller, D. *et al.* Full genome viral sequences inform patterns of SARS-CoV-2 spread into and within Israel. *Nat. Commun.* **11**, 5518 (2020).
90. Khailany, R. A., Safdar, M. & Ozaslan, M. Genomic characterization of a novel SARS-CoV-2. *Gene Rep.* **19**, 100682 (2020).