bioRxiv preprint doi: https://doi.org/10.1101/2020.07.27.223495. this version posted July 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

Supplementary Figures (1-4)



Supplementary Fig. 1. Genome coverage plots for the three SARS-CoV-2 negative samples.

Coverage is localized despite the 45-91 M reads that these samples obtained post-capture.

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.27.223495. this version posted July 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.

Poference	1	1,963	3,963	5,96	3 7,963	9,961	11,961	13,961	15,961	17,961	19,961	21,961	23,961	25,961	27,961	29,903			
NC_045512					orf1ab gene							S gene		orf3 M N ene					
Samples and lin	eage	•												E	0rt6-8	orf10			
192000119B A.1	_					_													
192000135B A.3	-			-		_													
192000051B A.3																			
192000052B <i>B.1</i>			1 1																
192000088B B.1				-												_			
192000251B B.1						_													
192000159B B.1																			
192000440B B.1																			
192000207B B.7																			
192000331B B 1																			
19200003B B 1						_										-			
192000021B B.1				_															
192000124B B.1																-			
192000296B B.1								-											
192000354B B.1				1															
192000072B B.1.1	1		-	-															
									Ambiguous base (N) Difference to reference										

Supplementary Fig. 2. A multiple sequence alignment (using MAFFT) of 17 reconstructed SARS-CoV-2 genomes and Wuhan-Hu-1 reference genome (NC_045512). Grey indicates agreement with the reference, black is a disagreement, and pink marks areas in the reconstruction with an ambiguous nucleotide, "N". The pangolin lineage assignment is listed next to the sample name. The extra length of the 192000251B seen here is an assembly artifact and was excluded from analysis.



Supplementary Fig. 3. Stop codon variants in sampled SARS-CoV-2 genomic assemblies. A snapshot of full length SARS-CoV-2 genome assemblies from GISAID and NCBI on 27 May 2020 was downloaded (comprising 39246 entries), and processed to detect single nucleotide variant alleles that introduced a stop codon. Introduced stop codons were detected in 270 entries, and the frequency of these alleles are plotted along the SARS-CoV-2 reference genome position. Introduced stop codons are rare but are distributed throughout the genomic sequences. Multiple loci harbor stop codons in unrelated assemblies.

bioRxiv preprint doi: https://doi.org/10.1101/2020.07.27.223495. this version posted July 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. It is made available under a CC-BY-NC-ND 4.0 International license.



Supplementary Fig. 4. Junctions reads to support expression of ORF10 192000052B,

192000251B and 192000440B. Expression values were calculated as 0.13, 0.13 and 0.02

reads/million. Few examples of those junction reads are shown in the figure (purple arrows).

Supplementary Tables (1-4)

Supplementary Table 1. Sample information, capture pools and sequencing metrics details.

Supplementary. Table 2. Lineage analysis of the 17 full-length genomes.

Supplementary Table 3. Junction read counts is reads/million identified in the post capture

data of 17 samples with full-length genomes.

Supplementary Table 4. Junction read counts in reads/million identified in the nine samples

sequenced before (IDxxxxB-2) and after capture (IDxxxxB) enrichment.