

- 1 Generation and transmission of inter-lineage recombinants in the SARS-CoV-2
- 2 pandemic
- 3
- 4 **Jackson et al**
- 5
- 6 **Supplementary Information**

7 **Supplementary Figure Legends**

8

9 ***Figure S1***

10 The distribution of minor allele frequencies from the read data for the 16 putative recombinants (red  
11 bars) and 20 samples suspected of being sequenced mixtures, either due to co-infection or laboratory  
12 contamination (grey bars). For each recombinant, the minor allele frequency is the mean across all  
13 sites that differ by a nucleotide change from the reference (MN908947.3) in it or either of its putative  
14 parents by genetic similarity. For the mixtures, the minor allele frequency is the mean across the  
15 sites that differ by a nucleotide change from the reference at genomic positions where mutations occur  
16 in B.1.1.7.

17

18 ***Figure S2***

19 The nucleotide variation with respect to the reference sequence (MN908947.3; grey genome far  
20 bottom) for each of the recombinant genomes (middle coloured genomes in each panel) and their  
21 closest neighbours by genetic similarity among all UK sequences from the same time period, for the  
22 B.1.1.7-like and non-B.1.1.7-like regions of their genomes (top and bottom coloured genomes in each  
23 panel). Panels: A) Group B; B) Group C; C) Group D; D) CAMC-CBA018; E) CAMC-CB7AB3; F)  
24 MILK-103C712; G) QEUH-1067DEF.

25

26 ***Figure S3***

27 The distribution of the most frequent SARS-CoV-2 lineages in the NUTS1 location of each set  
28 (Groups A-D and the four singletons) of recombinants for the four weeks immediately preceding each  
29 set's (earliest) sample date. Here, B.1.177 refers to B.1.177 itself and all its descendant lineages (e.g.  
30 B.1.177.9); the same is true for B.1.36.

31

32 ***Figure S4***

33 The distribution of nucleotide variation in the original members of Group A (top four coloured rows)  
34 and the 41 additional sequences that are derived from it (bottom 41 coloured rows), with respect to the  
35 reference sequence (MN908947.3; very bottom grey sequence).

36

37 **Table S1**

38

39 Nucleotide, amino acid, or deletion states at the 22 sites where mutations define lineage B.1.1.7 from  
 40 its immediate ancestor, for each putative recombinant. The genotype column names are in one of the  
 41 following formats: (i) “snp”, followed by the nucleotide in MN908947.3 (reference), the position in  
 42 the MN908947.3 genome, then the nucleotide in B.1.1.7; (ii) “aa” (amino acid), followed by the name  
 43 of the coding sequence, the residue in MN908947.3, the position of the residue in the coding  
 44 sequence, then the residue in B.1.1.7; (iii) “del” (deletion) followed by the first position of the  
 45 deletion, then the length of the deletion. For the contents of the amino acid columns, an asterisk (\*)  
 46 denotes a stop codon. For the contents of the deletion genotype columns, del=deletion present;  
 47 ref=deletion absent. Lineage B.1.1.7 states are coloured red, and MN908947.3 alleles are coloured  
 48 white. The mosaic structures visible here informed the first iteration of recombination detection.  
 49

Sample Name	Group	Sample Date	snp:C913T	aa:ORF1ab:T1001I	aa:ORF1ab:A1708D	snp:C5986T	aa:ORF1ab:I2230T	del:11288:9	snp:C14676T	snp:C15279T	snp:T16176C	del:21765:6	del:21991:3	aa:N501Y	aa:A570D	aa:P681H	aa:T716I	aa:S982A	aa:D1118H	aa:ORF8:Q27*	aa:ORF8:R52I	aa:ORF8:Y73C	aa:N:D3L	aa:N:S235F
CAMC-CBA018	-	18/12/2020	C	T	A	C	I	ref	C	C	T	del	del	Y	D	H	I	A	H	*	I	C	L	F
ALDP-11CF93B	A	30/01/2021	C	T	A	C	I	ref	C	C	T	del	del	Y	D	H	I	A	H	*	I	C	L	F
ALDP-125C4D7	A	06/02/2021	C	T	A	C	I	ref	C	C	T	del	del	Y	D	H	I	A	H	*	I	C	L	F
LIVE-DFCFE	A	14/02/2021	C	T	A	C	I	ref	C	C	T	del	del	Y	D	H	I	A	H	*	I	C	L	F
ALDP-130BB95	A	21/02/2021	C	T	A	C	I	ref	C	C	T	del	del	Y	D	H	I	A	H	*	I	C	L	F
QEUH-CCCB30	B	23/12/2020	C	T	A	C	T	del	T	T	C	del	del	Y	D	H	I	A	H	*	I	C	L	F
QEUH-CD0F1F	B	24/12/2020	C	T	A	C	T	del	T	T	C	del	del	Y	D	H	I	A	H	*	I	C	L	F
MILK-1166F52	C	24/01/2021	T	I	D	T	T	del	T	T	C	del	del	Y	D	H	I	A	H	Q	R	Y	D	S
MILK-11C95A6	C	30/01/2021	T	I	D	T	T	del	T	T	C	del	del	Y	D	H	I	A	H	Q	R	Y	D	S
QEUH-109B25C	C	18/01/2021	T	I	D	T	T	del	T	T	C	del	del	Y	D	H	I	A	H	Q	R	Y	D	S
MILK-126FE1F	D	07/02/2021	C	T	A	C	I	ref	C	C	T	ref	ref	Y	D	H	I	A	H	*	I	C	L	F
RAND-12671E1	D	02/02/2021	C	T	A	C	I	ref	C	C	T	ref	ref	Y	D	H	I	A	H	*	I	C	L	F
RAND-128FA33	D	02/02/2021	C	T	A	C	I	ref	C	C	T	ref	ref	Y	D	H	I	A	H	*	I	C	L	F
CAMC-CB7AB3	-	18/12/2020	T	I	A	C	I	ref	C	C	T	del	del	Y	D	H	I	A	H	*	I	C	L	F
MILK-103C712	-	12/01/2021	C	T	A	C	I	ref	C	C	T	ref	ref	N	A	P	T	S	D	*	I	C	L	F
QEUH-1067DEF	-	17/01/2021	T	I	D	T	T	ref	C	C	T	ref	ref	N	A	P	T	S	D	Q	R	Y	D	S

50  
51  
52  
53

54 **Table S2**

55

56 GISAID and ENA accession numbers for the 16 putative recombinant sequences.

57

Sample	Group	GISAID accession	ENA sample accession	ENA sample secondary accession	ENA run accession
ALDP-11CF93B	A	EPI_ISL_989697	ERS5764793	SAMEA8077663	ERR5308556
ALDP-125C4D7	A	EPI_ISL_1019487	ERS5792357	SAMEA8105282	ERR5323237
ALDP-130BB95	A	EPI_ISL_1122630	ERS5883487	SAMEA8196877	ERR5414941
LIVE-DFCFE	A	EPI_ISL_1104468	ERS5872630		ERR5404883
QEUH-CCCB30	B	EPI_ISL_782203	ERS5522602	SAMEA7775362	ERR5058070
QEUH-CD0F1F	B	EPI_ISL_782442	ERS5522673	SAMEA7775433	ERR5058141
MILK-1166F52	C	EPI_ISL_938901	ERS5697106	SAMEA8009867	ERR5272107
MILK-11C95A6	C	EPI_ISL_1050799	ERS5874115		ERR5406307
QEUH-109B25C	C	EPI_ISL_909004	ERS5669824	SAMEA7982538	ERR5232711
MILK-126FE1F	D	EPI_ISL_1057512	ERS5812369	SAMEA8125397	ERR5349458
RAND-12671E1	D	EPI_ISL_1045772	ERS5805519	SAMEA8118538	ERR5335088
RAND-128FA33	D	EPI_ISL_1042110	ERS5809401	SAMEA8122425	ERR5340986
CAMC-CBA018	-	EPI_ISL_777766	ERS5517714	SAMEA7770468	ERR5054123
CAMC-CB7AB3	-	EPI_ISL_777974	ERS5517673	SAMEA7770427	ERR5054082
MILK-103C712	-	EPI_ISL_994038	ERS5760589	SAMEA8073454	ERR5304348
QEUH-1067DEF	-	EPI_ISL_917983	ERS5674718	SAMEA7987439	ERR5238288

58

59

60

61 **Table S3**

62

63 Phylogenetic placement and UF bootstrap support for the two parental tracts of each recombinant's  
 64 genome, among the whole UK epidemic. For CAMC-CB7AB3, "Left" refers to the inner part of its  
 65 genome, and "Right" refers to the outer part of its genome. The full phylogenies with bootstrap  
 66 support are available at <https://github.com/COG-UK/UK-recombination-analysis>

67

68

Sample	Group	Left placement	Left support	Right placement	Right support
ALDP-11CF93B	A	B.1.177	81	B.1.1.7	100
ALDP-125C4D7	A	B.1.177	81	B.1.1.7	100
LIVE-DFCFE	A	B.1.177	81	B.1.1.7	100
ALDP-130BB95	A	B.1.177	81	B.1.1.7	100
QEUH-CCCB30	B	B.1.36.28	97	B.1.1.7	100
QEUH-CD0F1F	B	B.1.36.28	97	B.1.1.7	100
MILK-1166F52	C	B.1.1.7	100	B.1.221.1	98
MILK-11C95A6	C	B.1.1.7	100	B.1.221.1	98
QEUH-109B25C	C	B.1.1.7	100	B.1.221.1	98
MILK-126FE1F	D	B.1.36.17/B.1.36.39	100	B.1.1.7	100
RAND-12671E1	D	B.1.36.17/B.1.36.39	100	B.1.1.7	100
RAND-128FA33	D	B.1.36.17/B.1.36.39	100	B.1.1.7	100
CAMC-CB7AB3	-	B.1.177	39	B.1.1.7	100
CAMC-CBA018	-	B.1.177	94	B.1.1.7	100
MILK-103C712	-	B.1.177	100	B.1.1.7	100
QEUH-1067DEF	-	B.1.1.7	100	B.1.177	100

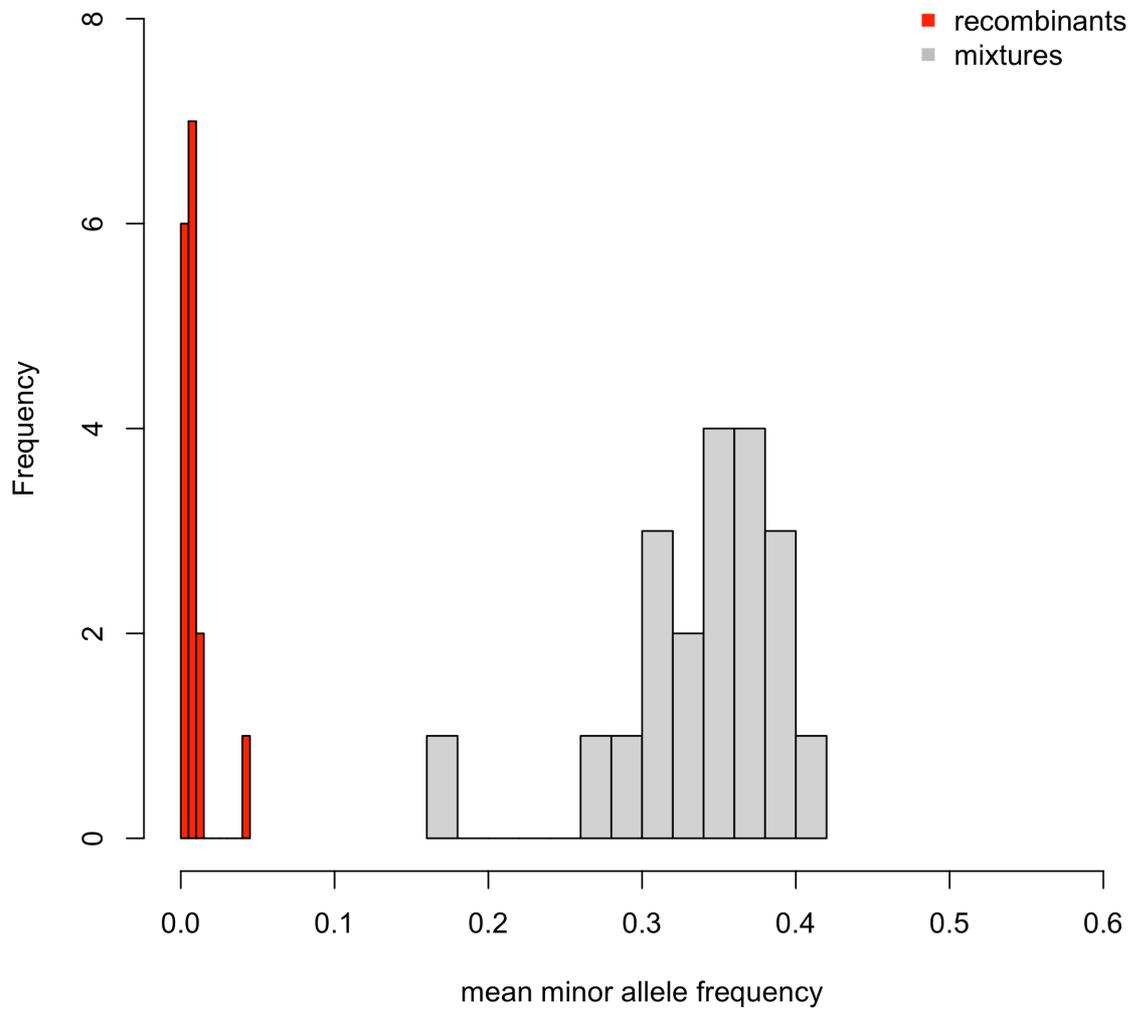
69

70

71

72 **Figure S1**  
73

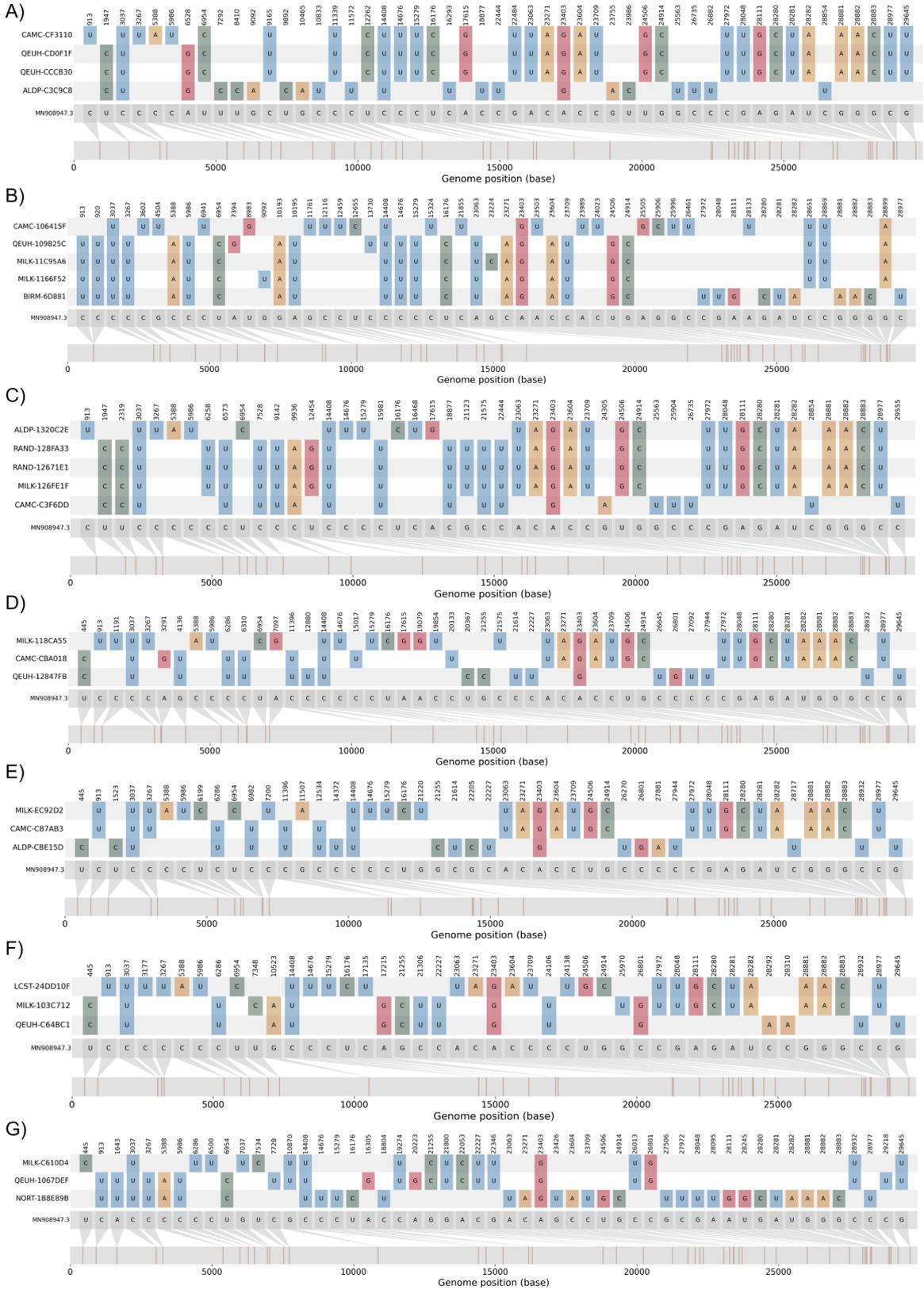
### raw read minor allele frequencies



74  
75  
76

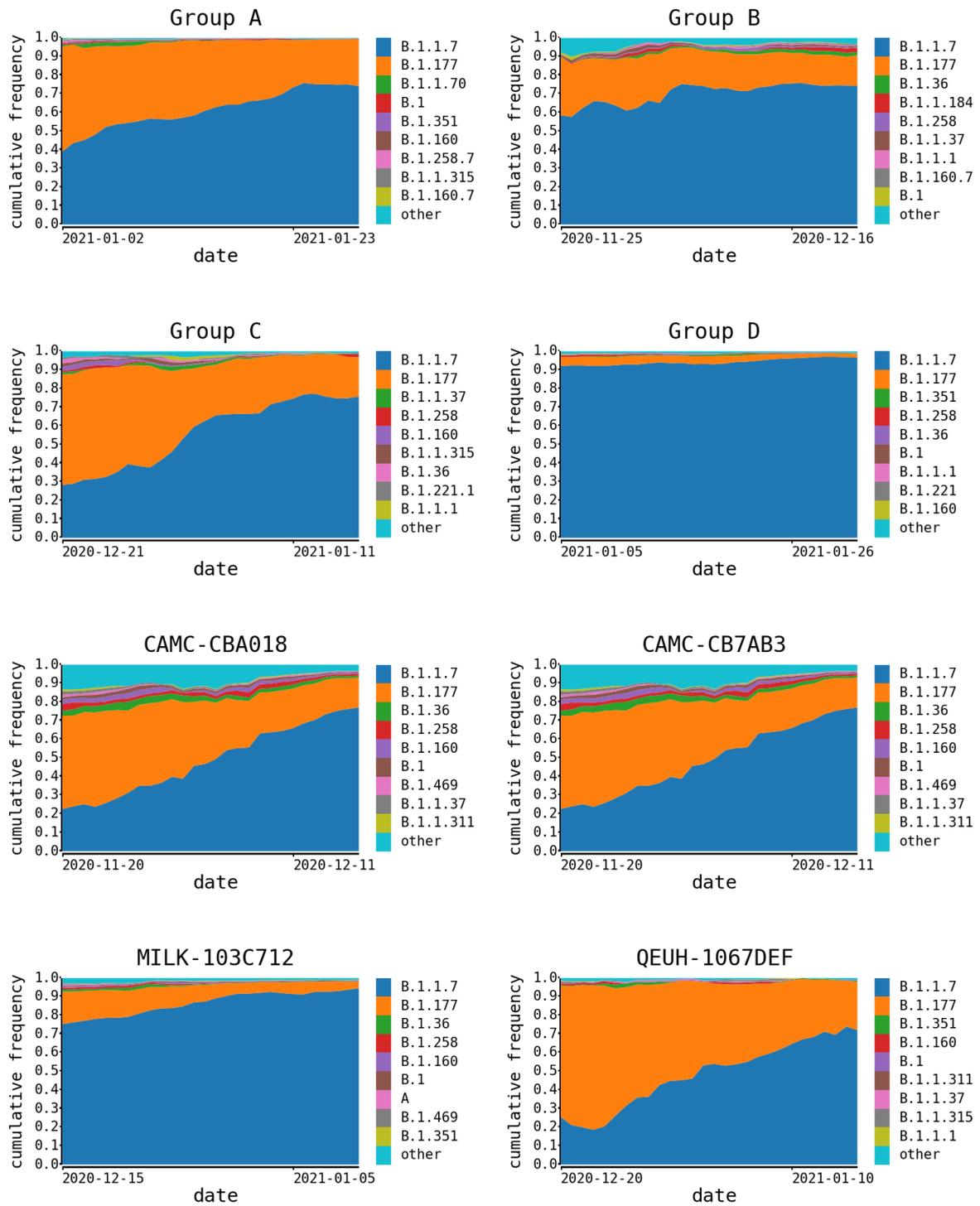
---

77 **Figure S2**  
78



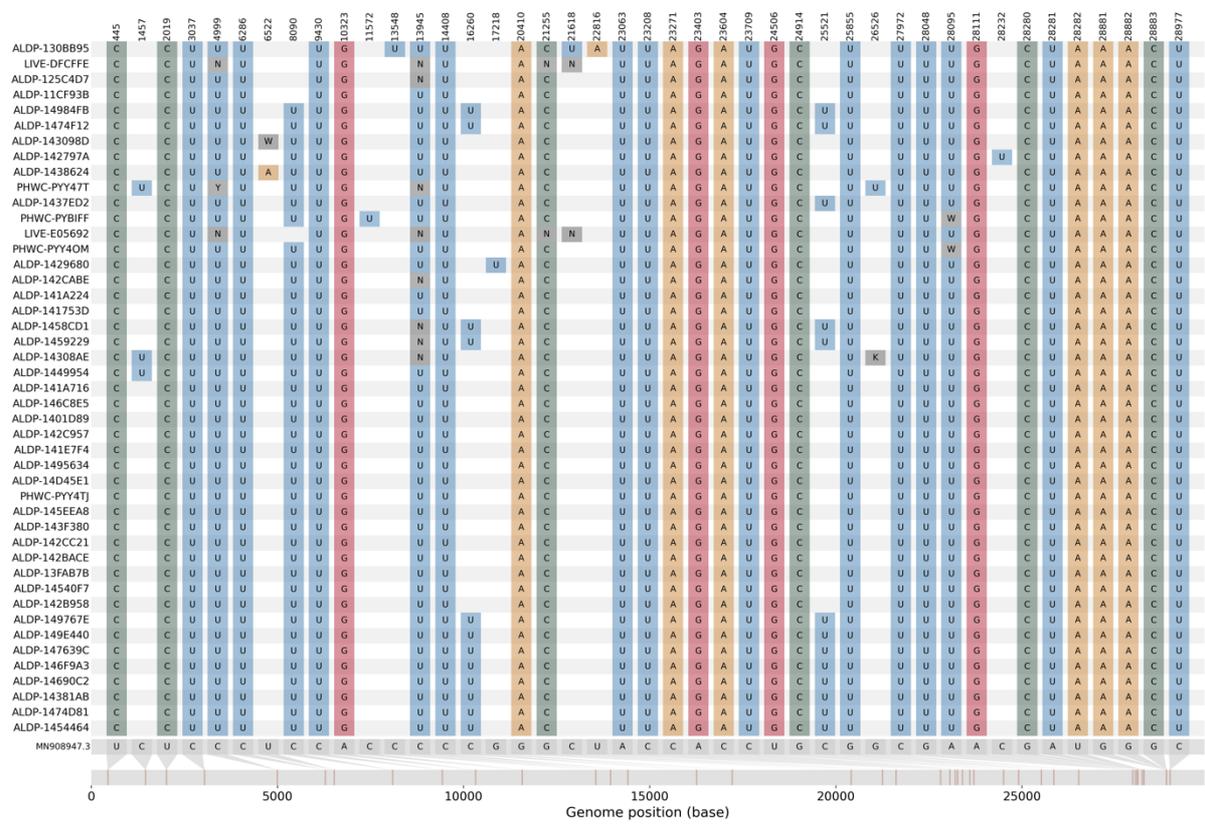
79  
80  
81

82 **Figure S3**  
 83



84  
 85  
 86

87 Figure S4  
88



89  
90  
91  
92