

1           **Analysis of the ARTIC V4 and V4.1 SARS-CoV-2 primers and their impact on the**  
2                           **detection of Omicron BA.1 and BA.2 lineage defining mutations**

3  
4       Fatima R. Ulhuq<sup>1\*</sup>, Madhuri Barge<sup>1</sup>, Kerry Falconer<sup>1</sup>, Jonathan Wild<sup>1</sup>, Goncalo Fernandes<sup>1</sup>,  
5       Abbie Gallagher<sup>1</sup>, Suzie McGinley<sup>1</sup>, Ahmad Sugadol<sup>1</sup>, Muhammad Tariq<sup>1</sup>, Daniel Maloney<sup>1,2</sup>,  
6                           Juliet Kenicer<sup>1</sup>, Rebecca Dewar<sup>1</sup>, Kate Templeton<sup>1</sup>, Martin McHugh<sup>1,3\*</sup>

7  
8       <sup>1</sup>Viral Genotyping Reference Laboratory, Royal Infirmary of Edinburgh, NHS Lothian, EH16  
9       4SA, Edinburgh, United Kingdom.

10       <sup>2</sup>Institute of Evolutionary Biology, University of Edinburgh, EH9 3JR, Edinburgh, United  
11       Kingdom

12       <sup>3</sup>School of Medicine, University of St Andrews, KY16 9TF, St Andrews, United Kingdom

13  
14       \*To whom correspondence should be addressed.

15       e-mail: [fatima.ulhuq@nhslothian.scot.nhs.uk](mailto:fatima.ulhuq@nhslothian.scot.nhs.uk), [mpm20@st-andrews.ac.uk](mailto:mpm20@st-andrews.ac.uk)

16  
17       **Abstract**

18       The ARTIC protocol uses a multiplexed PCR approach with two primer pools tiling the entire  
19       SARS-CoV-2 genome. Primer pool updates are necessary for accurate amplicon  
20       sequencing of evolving SARS-CoV-2 variants with novel mutations. The suitability of the  
21       ARTIC V4 and updated V4.1 primer scheme was assessed using whole genome sequencing  
22       of Omicron from clinical samples using Oxford Nanopore Technology. Analysis of Omicron  
23       BA.1 genomes revealed that 93.22% of clinical samples generated improved genome  
24       coverage at 50x read depth with V4.1 primers when compared to V4 primers. Additionally,  
25       the V4.1 primers improved coverage of BA.1 across amplicons 76 and 88, which resulted in  
26       the detection of the variant defining mutations G22898A, A26530G and C26577G. The  
27       Omicron BA.2 sub-variant (VUI-22JAN-01) replaced BA.1 as the dominant variant by March  
28       2022, and analysis of 168 clinical samples showed reduced coverage across amplicons 15

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

29 and 75. Upon further interrogation of primer binding sites, a mutation at C4321T (present in  
30 163/168, 97% of samples) was identified as a possible cause of complete dropout of  
31 amplicon 15. Furthermore, two mutations were identified within the primer binding regions for  
32 amplicon 75: A22786C (present in 90% of samples) and C22792T (present in 12.5% of  
33 samples). Together, these mutations may result in reduced coverage of amplicon 75 and  
34 further primer updates would allow the identification of the two BA.2 defining mutations  
35 present in amplicon 75; A22688G and T22679C. This work highlights the need for ongoing  
36 surveillance of primer matches as circulating variants evolve and change.

37

### 38 **Keywords**

39 SARS-CoV-2, Nanopore, Whole Genome Sequencing, Omicron, BA.1, BA.2

40

## 41 Introduction

42

43 On the 11<sup>th</sup> March 2020, the World Health Organisation (WHO) declared a global pandemic  
44 following the emergence of the severe acute respiratory syndrome coronavirus 2 (SARS-  
45 CoV-2). To date, 5<sup>th</sup> September 2022, there have been over 600 million confirmed cases and  
46 over 6.5 million deaths worldwide [1].

47

48 Worldwide genomic surveillance of SARS-CoV-2 has been employed to better understand  
49 the viral mutation rate and transmission dynamics of emerging SARS-CoV-2 variants.  
50 Identifying the genetic diversity of circulating SARS-CoV-2 variants aids research into  
51 understanding the pathogenicity of variants, identifying sites on the genome that may reduce  
52 the effectiveness of vaccines and drug therapies, and ultimately inform the design of future  
53 therapeutics [2–8]. Additionally, real-time surveillance of SARS-CoV-2 outbreaks informs the  
54 implementation of control and prevention measures for SARS-CoV-2 transmission. As of 6<sup>th</sup>  
55 September 2022, 12.9 million SARS-CoV-2 genomes have been submitted to the global  
56 initiative on sharing all influenza data (GISAID) database (GISAID - [gisaid.org](https://gisaid.org)).

57

58 The most widely used protocol to sequence SARS-CoV-2 is the ARTIC amplicon-based  
59 targeted whole-genome sequencing approach. The ARTIC protocol is a freely available  
60 protocol allowing for rapid responses to changes in the SARS-CoV-2 viral genome [9–14].  
61 This is a multiplex PCR approach, which consists of 98 pairs of primers that spans the  
62 ~30kb genome. However, one of the major limitations of the amplicon approach is the static  
63 nature of the primer scheme as the virus evolves and develops mutations and structural  
64 variants, which can lead to primer mismatches and an inability to construct near-complete  
65 viral genomes. Since January 2020 the primer scheme has been updated four times in  
66 response to changes in the prevailing virus lineages [9–11, 13, 14]. Inability of a given  
67 primer to bind to the complementary sequence results in amplicon dropouts and mutations in

68 the Beta, Delta and Gamma variants resulted in less efficient amplification of amplicons 72,  
69 74 and 76 [14, 15].

70

71 The ARTIC V4.1 primers were introduced to improve primer binding and eliminate amplicon  
72 dropouts for BA.1. In this work the suitability of the ARTIC V4.1 primer scheme by  
73 comparison with the V4 primer scheme will be assessed and reported for the Omicron BA.1  
74 lineage. To do this, Nanopore sequencing was used to sequence 70 SARS-CoV-2 samples,  
75 with both V4 and V4.1 primers, obtained from RT-PCR confirmed COVID-19 positive  
76 patients. Subsequently, BA.2 was assigned as a variant under investigation (VUI-22JAN-01)  
77 and has increased in prevalence. The ability of the V4.1 primer scheme to generate near  
78 complete genome coverage of BA.2 was also investigated in this work.

79

## 80 **Methods**

81 **Patient samples and RNA extraction.** 344 SARS-CoV-2 positive clinical samples were  
82 collected between Dec-2021 and Mar-2022. There were 325 nose and throat swabs  
83 (94.48%), 18 throat swabs (5.23%) and 1 tracheal aspirate (0.29%). Samples were  
84 confirmed positive for at least two SARS-CoV-2 real-time RT-PCR targets using a range of  
85 diagnostic assays in NHS laboratories across the east of Scotland and referred to the Viral  
86 Sequencing Service, Royal Infirmary of Edinburgh, for sequencing. RNA extraction was  
87 performed on confirmed SARS-CoV-2 samples using two automated extraction platforms;  
88 the Biomérieux NucliSENS EMAG<sup>®</sup> and Abbot *m2000sp*. First, SARS-CoV-2 positive  
89 samples were inactivated and lysed using NucliSENS<sup>®</sup> easyMAG<sup>®</sup> Lysis Buffer (Biomérieux,  
90 ref 280134). When using the *m2000sp*, VTM and lysis buffer were mixed at a 1:1 ratio and  
91 equal volumes of diluted lysis buffer was added to each sample for lysis (500 µl lysis buffer +  
92 500 µl sample). For the Biomérieux NucliSens EMAG<sup>®</sup>, neat lysis buffer was added to each  
93 sample (200 µl sample + 2 mL lysis buffer). The procedure was followed as outlined in the  
94 manufacturer's instructions and the RNA eluted in 110 µl. VTM only was used as an internal  
95 control for each extraction run.

96 **Ethical approval.** This work was performed as a diagnostic lab validation and ethical  
97 approval was obtained from the NHS Lothian NRS BioResource and Tissue Governance  
98 Unit (reference 10/S1402/33).

99 **ARTIC LoCost protocol Nanopore library preparation.** ONT libraries for SARS-CoV-2  
100 sequencing was performed according to the ARTIC Lo-Cost protocol [11, 12], with both V4  
101 and V4.1 primers. Slight modifications to the published method were made during the DNA  
102 barcode ligation step. For a single reaction, the ligation mastermix was prepared by adding  
103 nuclease-free water (2.1 µl), Ultra II Ligation Mastermix (5.0 µl), Ligation Enhancer (0.2µl)  
104 (NEBNext<sup>®</sup> Ultra™ II Ligation Module, E7595S) and the unique barcode (1.2 µl) (ONT Native  
105 Barcoding Expansion 96, EXP-NBD196). The end prep reaction mixture (1.5 µl), from the

106 previous step, was added to the ligation mastermix (8.50 µl) containing the unique barcodes.  
107 The protocol was then followed as outlined in the ARTIC Lo-Cost protocol [11].

108 **Nanopore sequence analysis.** High accuracy basecalling was performed using Guppy  
109 (v4.4.0) (Oxford Nanopore Technologies) and demultiplexed using guppy\_barcode with the  
110 option 'require\_barcodes\_both\_ends' and a minimum score of 60. The downstream analysis  
111 was performed using the ARTIC fieldbioinformatics pipeline (v1.2.1) ([https://github.com/artic-](https://github.com/artic-network/fieldbioinformatics)  
112 [network/fieldbioinformatics](https://github.com/artic-network/fieldbioinformatics)) and nanopolish variant calling (<https://github.com/jts/nanopolish>)  
113 to generate a consensus sequence for each sample in FASTA format. RAMPART (v1.0.6)  
114 (<https://github.com/artic-network/rampart>) was used to calculate genome coverage at 20x,  
115 100x and 200x read depth. Nextclade (<https://clades.nextstrain.org>) was used to determine  
116 the number of SNPs between the reference genome (NC\_045512.2) and sequenced clinical  
117 isolates. PANGOLIN (v3.1.20) (<https://github.com/cov-lineages/pangolin>) [16, 17] was used  
118 to assign a lineage to all SARS-CoV-2 samples, using the pangoleARN algorithm. Aln2type  
119 (v0.0.3) (<https://github.com/connor-lab/aln2type>) (analysis performed on 8th Feb 2022) was  
120 used to identify variant defining mutations for variants of concern (VOC) and variants under  
121 investigation (VUI) as curated by Public Health England ([https://github.com/phe-](https://github.com/phe-genomics/variant_definitions)  
122 [genomics/variant\\_definitions](https://github.com/phe-genomics/variant_definitions)).

123 The metrics and results of all experiments are available in the supplementary data (Datasets  
124 1, 2 and 3). For direct V4 to V4.1 primer comparison, two separate 48-plex libraries were  
125 loaded (Run1 and Run2, Dataset 1). For the analysis of BA.1 and BA.2 samples with V4.1  
126 primers only, a total of 7, 48-plex libraries were loaded and sequenced.

127 **Sequence alignment.** Sequences were aligned to the reference SARS\_CoV-2 isolate  
128 NC\_045512.2 using MAFFT (v7.505) (<https://anaconda.org/bioconda/mafft>) [18] and  
129 visualised using Jalview2 (v2.11.2.2) [19]. Depth at each nucleotide position was extracted  
130 using SAMtools (v1.15) (<https://github.com/samtools/samtools>).

131 **Quality Control.** Samples were run as 48-plex libraries, each included a negative and  
132 positive control (SARS-CoV-2 B.1 lineage, collected June 2020), and two extraction negative  
133 controls (Datasets 1, 2 and 3). For all sequencing runs, the negative control and extraction  
134 negative controls generated <5 mapped reads and 0% genome coverage at 20x. The viral  
135 lysate of the positive control generated 99.90% genome coverage at 20x read depth. All  
136 seven expected mutations were detected for the positive control and assigned the PANGO  
137 lineage B.1. Clinical samples that generated <70% whole genome coverage were excluded  
138 from further analysis.

139 **Data availability statement.** The data have been deposited with links to BioProject  
140 accession number PRJNA902683 in the NCBI BioProject database  
141 (<http://www.ncbi.nlm.nih.gov/bioproject/902683>).

142 **Conflict of interest statement.** The authors have declared no conflict of interest.

143

## 144 **Results**

145

### 146 **Genome coverage of Omicron BA.1 with ARTIC V4 and V4.1 primers**

147 In total, nine (13%) clinical samples were identified as Delta (VUI-21OCT-01 or VOC-  
148 21APR-02), 59 (84%) clinical samples were identified as Omicron lineage BA.1 (VOC-  
149 21NOV-01) and two (3%) clinical samples were identified as Omicron BA.2 (VUI-22JAN-01)  
150 with ARTIC V4 and V4.1 primers (Dataset 1). For the Delta lineage samples, the median  
151 genome coverage at 20x read depth improved from 99.40% with V4 primers to 99.90% with  
152 V4.1 primers, respectively. All nine samples were confirmed as either VUI-21OCT-01 or  
153 VOC-21APR-02 with both V4 and V4.1 primers. The two Omicron BA.2 samples increased  
154 from 95.20% to 99.30% and 97.20% to 99.90% genome coverage, with V4 and V4.1  
155 primers, respectively. Both samples had probable aln2type VUI-22JAN-01 calls with V4  
156 primers, however one sample improved from probable to confirmed with V4.1 primers.

157

158 Initial in-depth analysis focussed on Omicron BA.1, as the increasingly dominant variant and  
159 the main component of the runs during sequencing. The genome coverage at >50x read  
160 depth was investigated to identify regions of the genome with reduced amplification.  
161 Nucleotide positions that generated <50x read depth in 15% or more of the samples were  
162 considered as low coverage regions. Coverage was compared between the V4 and V4.1  
163 primer scheme. The V4.1 primer scheme included a new batch of primers added to the V4  
164 primers to improve the coverage across amplicons 10, 23, 27, 76, 79, 88, 89 and 90 [13].

165

166 Use of the V4 primers generated reduced amplification of nucleotides across 18 different  
167 amplicons, with less than 50x read depth in more than 15% of sequenced samples (Figure  
168 1A, Table S1). The use of the V4 primers resulted in dropout of amplicons 76 and 90  
169 (Figures 1A and 2A, Table S1) and poor coverage across amplicons 10, 23, 79, 88 and 89  
170 (Figures 1A and 2A, Table S1). Reduced coverage of these amplicons was due to primer  
171 mismatches between the V4 primer scheme and the SARS-CoV-2 BA.1 sequence. Reduced  
172 coverage of nucleotide positions across amplicons 22, 37, 72, 73, 74 and 95 were due to  
173 small deletions within the genome (3-9bp) (Table S1). Reduced amplification of amplicons  
174 21, 29, 31, 51 and 60 (Figure 1A, Table S1) were detected with the V4 primers in samples  
175 that generated <60,000 mapped reads. Due to the low number of mapped reads for these  
176 samples, the reduced coverage within these amplicons is likely to be due to the quality and  
177 quantity of the extracted RNA.

178 Utilisation of the V4.1 primer scheme [13] resulted in an overall improvement in genome  
179 coverage. The addition of the V4.1 alt primers increased the median sequencing depth  
180 across all eight amplicons with the modified primers in the updated V4.1 primer scheme  
181 (Figure 2A). Strikingly, the additional primers rescued the dropout of amplicons 76 and 90  
182 and increased sequencing coverage across amplicons 10, 23, 78, 88 and 89 (Figure 2A).  
183 The number of amplicons with nucleotide positions that generated <50x coverage in 15% or



184 more samples decreased from 18 amplicons with V4, to 10 amplicons with V4.1 primers  
185 (Figure 1B, Table S2). Despite the addition of modified primers to better amplify amplicon  
186 88, the coverage defect was not completely restored. However, the percentage of  
187 sequenced samples <50x coverage within this region decreased, from 40.68 - 66.1% to  
188 22.03 - 33.9%, with V4 and V4.1 primers, respectively (Tables S1 and S2). Regions with  
189 lower coverage with the V4.1 primers were investigated to identify the root cause of lower  
190 read depth at certain nucleotide positions. The reduced coverage across amplicons 22, 37,  
191 72, 73, 74 and 95 were due to small deletions within the genome (3-9 bp). Reduced  
192 amplification of amplicons 21, 51 and 66 was detected in samples that generated <60,000  
193 mapped reads, suggesting overall lower genome quality.

194

195 Analysis of all BA.1 sequences highlighted that in total, 2668 nucleotides generated reduced  
196 coverage (<50x read depth in 15% or more samples) with the V4 primers compared to 1085  
197 nucleotides with the V4.1 primers (Tables S1 and S2). The median number of reads mapped  
198 per sample was 98,609 and 114,476 with V4 and V4.1 primers, respectively (Dataset 1).  
199 Additionally, the median number of masked bases (N's) decreased from 1050 N's to 126 N's  
200 with the V4 and V4.1 primers, respectively (Dataset 1). In addition, the median number of  
201 SNPs increased from 52 to 58 with the V4 and V4.1 primers, respectively. This highlights the  
202 overall improvement detected with the V4.1 primers when compared to V4. The percentage  
203 of samples that generated improved genome coverage with the V4.1 primers when  
204 compared to V4 was 93.22% (55/59) at 50x read depth (Figure 2B). Additionally, 37 out of  
205 59 (62.71%) samples generated increased number of mapped reads with the V4.1 primers  
206 compared to V4 (Dataset 1). The median genome coverage per sample for V4 primers at  
207 200x, 100x and 20x was 95.59%, 96.30% and 98.00%, respectively (Dataset 1). The median  
208 coverage per genome with V4.1 primers at 200x, 100x and 20x was 98.80%, 99.60%, and  
209 99.90%, respectively.

210

211

212 **PANGO Lineage and VOC/VUI calls of Omicron BA.1 with ARTIC V4 and V4.1 primers**

213 PANGOLIN was used to assign a lineage to all SARS-CoV-2 consensus sequences and  
214 58/59 samples received identical PANGO lineages, either BA.1 (37/59) or BA.1.1 (21/59)  
215 with V4 and V4.1 primers (Dataset 1). The specificity of the lineage call for vssfru\_032  
216 altered from BA.1 to BA.1.1 with the V4.1 primers (vssfru\_032b) when compared to V4  
217 (vssfru\_032b) (Dataset 1). Based on PHE variant definitions, aln2type requires the detection  
218 of 17 variant defining mutations to be assigned as SARS-CoV-2 BA.1 (VOC-21NOV-01). All  
219 17 variant defining mutations must be detected for a 'confirmed' VOC/VUI status call.  
220 Interestingly, more accurate VOC-21NOV-01 status calls were obtained in 55/59 samples  
221 (93.22%), improving from probable to confirmed VOC-21NOV-01 when using V4.1  
222 compared to V4 primers (Dataset 1). The remaining 4/59 samples (6.78%) had probable  
223 VOC-21NOC-01 calls with both V4 and V4.1 primers (Dataset 1).

224 Further analysis highlighted that the median number of variant defining mutations detected  
225 for VOC-21NOV-01, increased from 14 to 17 with the V4 and V4.1 primers, respectively. The  
226 increased coverage across amplicons 76 and 88 with the V4.1 primers (Figure 2A) resulted  
227 in increased coverage at position 22898, 26530 and 26577 (Figure 3A) and the detection of  
228 the three additional variant defining SNPs, G22898A, A26530G and C26577G, present in  
229 amplicons 76 and 88 (Figure 3B). In addition, increased coverage of amplicon 60 allowed  
230 the identification of the variant defining mutation A18163G (Figure 3A and 3B). In total, 138  
231 variant defining bases were masked (N's) with the V4 primers and this was reduced to 12  
232 bases with V4.1 primers (Figure 3B). The 12 masked variant defining bases at position  
233 18163 (n=2), 22898 (n=2), 26530 (n=4) and 26577 (n=4) were only detected in 4 BA.1  
234 samples (Figure 3B) that generated <30,000 mapped reads, therefore overall reduced  
235 genome quality.

236 A larger dataset of 106 positive SAR-CoV-2 BA.1 samples collected between 21-02-2022  
237 and 07-03-2022 were investigated to compare the aln2type results obtained from the side-  
238 by-side V4 and V4.1 comparison above. Comparable results were obtained with reduced  
239 coverage at positions 18163, 22898, 26530 and 26577 (present in amplicons 60, 76 and 88)  
240 with the V4.1 primers (Figure S1A and S1B). Reduced amplification of these regions  
241 contributed to only 16/106 (15.1%) of samples with probable VOC-21NOV-01 calls (Dataset  
242 2) and a total of 48 masked variant defining mutations (Figure S1B, Dataset 2). Altogether,  
243 these results highlight the overall improvement in the amplification of Omicron BA.1 with the  
244 V4.1 primers when compared to V4.

245

#### 246 **Genome coverage of Omicron BA.2 with ARTIC V4.1 primers**

247 Genome coverage of Omicron BA.2 (VUI-22JAN-01) was further assessed with the ARTIC  
248 V4.1 primers by analysing 168 SARS-CoV-2 positive BA.2 samples collected from 2022-02-  
249 21 – 2022-03-06. Overall, the median genome coverage of 168 BA.2 samples with V4.1  
250 primers at 200x, 100x and 20x were 96.15%, 98.60% and 99.30%, respectively (Dataset 3).  
251 The median number of mapped reads was 82,598 (Dataset 3). The median number of  
252 masked bases (N's) was 387 with a median number of 68 SNPs across the BA.2 genomes.  
253 In total, 16 amplicons were identified with <50x read depth in more than 15% of the samples  
254 sequenced (Figure 4, Table S3). Reduced coverage of nucleotides across amplicons 37, 72,  
255 95 and 99 were due to deletions within BA.2 (8-26 bp deletions). It was not possible to  
256 identify a cause for the reduced read depth of nucleotides present in amplicons 1, 21, 22, 51,  
257 60, 66, 74, 76, 88 or 90, however, BA.2 samples with reduced nucleotide coverage in these  
258 regions generated a total of <60,000 mapped reads per genome. This suggests the low  
259 coverage may be due to pre-analytical issues with the amount and quality of RNA. Dropout  
260 of amplicon 15 and poor coverage across amplicon 75 was detected (Figure 4, Table 3).  
261 BA.2 samples with reduced nucleotide coverage within these regions generated >60,000

262 mapped reads. These samples were further investigated to identify the root cause of  
263 reduced coverage.

264 With regards to amplicon 15, a SNP (C4321T) was identified in 163/168 samples (97%)  
265 (Figure 5A). The primer 'SARS-CoV-2\_15\_LEFT' binds within this region of the genome and  
266 the C4321T SNP sits 9 bp into the forward primer (Figure 5A). In amplicon 75 two SNPs  
267 were identified within the primer binding region of 'SARS-CoV-2\_75\_RIGHT', A22786C and  
268 C22792T (Figure 5B). The variant defining mutation, A22786C is present in 90% of samples  
269 (151/168) while C22792T is present in 12.5% of samples (21/168). The variant defining  
270 mutation, A22786C sits at the end of the 'SARS-CoV-2\_75\_RIGHT' primer (3' region),  
271 therefore does not have a significant impact on primer binding. However, samples with the  
272 C22792T SNP exhibit dropouts in amplicon 75, highlighting the importance of this site for  
273 primer binding (Figure 5B). Updates to these primers may provide improved coverage of  
274 amplicons 15 and 75.

275 Reduced amplification of amplicons 75, 76 and 88 resulted in reduced sequencing coverage  
276 of four variant defining SNPs (T22679C, A22688G, A22786C and C26577G) (Figure S2,  
277 Table S3) required for *aln2type* Omicron BA.2 variant under investigation (VUI-22JAN-01)  
278 call criteria (Table S3). BA.2 (VUI-22JAN-01) required detection of 20 variant defining  
279 mutations for a confirmed status call and 80 probable VUI-22JAN-01 calls were obtained  
280 with a total of 197 masked bases (N's) at variant defining positions across all BA.2 samples  
281 sequenced (Figure S2B, Dataset 3). Of these, 71 BA.2 samples generated over 90%  
282 genome coverage at 20x read depth and were then further investigated. The 71 probable  
283 VUI-22JAN-01 samples accumulated a total of 131 masked bases (N's) at the following  
284 variant defining positions: 2790 (n=3), 9424 (n=29), 18163 (n=7), 22679 (n=34), 22688  
285 (n=34), 22775 (n=5), 22786 (n=9), 26577 (n=10) (Dataset 3). As expected, masked bases  
286 (N's) were detected at position 22679, 22688, 22786 and 26577 due to lower coverage  
287 across amplicons 75, 76 and 88. However, the number of samples with masked bases at  
288 position 9424 (n=29) was unexpected, as 27 out of the 29 samples generated more than 20x

289 coverage at position 9424 (Table S4). Further analysis revealed that the discrepancy in the  
290 detection of A9424G was due to the presence of mixed bases, with variant base calls often  
291 having low quality. This resulted in masking of the variant defining SNP A9424G in our  
292 consensus calling pipeline due to low support for the variant call.

293 In addition, reference base calls at variant defining sites, at either position 670 (6/168, 4%)  
294 or 9866 (1/168, 0.6%) (Table S5) contributed to probable VUI-22JAN-01 calls in seven BA.2  
295 samples. This might be due to diversity within the BA.2 population, where reversion  
296 mutations may have occurred. Furthermore, a small cluster of 10 BA.2 samples were  
297 defined as having mixed calls for the variant defining mutation at position 29510 (Table S5).  
298 An alignment comparing vssfru\_253 (confirmed A29510C SNP), the SARS-CoV-2 reference  
299 genome (NC\_045512.2) and the 10 BA.2 samples with aln2type mixed call status (Figure  
300 S3) was generated (Figure S3). The alignment highlights the presence of A29510C within all  
301 10 BA.2 samples, however an additional SNP, G29511T was present downstream (Figure  
302 S3). The expected codon is CGT, however, a subset of BA.2 samples have CTT within their  
303 consensus sequence, this is then defined as a mixed call by aln2type. Strikingly, the  
304 presence of this unique mutation, not identified within any other sequences analysed within  
305 this work, may indicate transmission during a small outbreak.

306

## 307 **Conclusion**

308 In summary, use of the ARTIC V4.1 primers, when compared to V4 primers, improved  
309 overall genome coverage of Omicron BA.1. The addition of the alt primers improved the  
310 sequencing depth across all eight amplicons which resulted in more accurate PANGO  
311 lineage and VOC/VUI calls. Further analysis of the BA.2 sub-variant revealed complete drop-  
312 out of amplicon 15 and reduced coverage of amplicon 75. Additionally, lower read depth of  
313 amplicon 75 resulted in the masking of two variant defining mutations. This work highlights  
314 the need for ongoing surveillance of primer matches as circulating variants evolve and  
315 change. The use of different primer schemes should be an important consideration when

316 identifying and assigning variants of concern and the genome completeness of SARS-CoV-2  
317 genomes should be routinely monitored with updates to primer schemes when required.

318

319

### 320 **Acknowledgements**

321 We would like to thank diagnostic laboratory teams at NHS Borders, NHS Fife, NHS  
322 Grampian, NHS Highland, NHS Lothian, NHS Orkney, NHS Shetland, and NHS Tayside for  
323 providing samples used in this work. We are also grateful to colleagues in Public Health  
324 Microbiology at Public Health Scotland, NHS Greater Glasgow and Clyde, National Services  
325 Division of NHS National Services Scotland, and Scottish Government with whom the  
326 sequencing service was established.

## 327 References

- 328  
329 1. WHO Coronavirus (COVID-19) Dashboard. <https://covid19.who.int>.
- 330  
331 2. Nadeau S, Beckmann C, Topolsky I, Vaughan T, Hodcroft E, *et al*. Quantifying SARS-  
332 CoV-2 spread in Switzerland based on genomic sequencing data. *medRxiv* 2020.
- 333  
334 3. Martin J, Klapsa D, Wilton T, Zambon M, Bentley E, *et al*. Tracking SARS-CoV-2 in  
335 Sewage: Evidence of Changes in Virus Variant Predominance during COVID-19  
336 Pandemic. *Viruses*;12.
- 337  
338 4. Volz E, Mishra S, Chand M, Barrett JC, Johnson R, *et al*. Transmission of SARS-CoV-  
339 2 Lineage B.1.1.7 in England: Insights from linking epidemiological and genetic data.  
340 *medRxiv* 2021.
- 341  
342 5. Chen C, Nadeau S, Yared M, Voinov P, Xie N, *et al*. CoV-Spectrum: Analysis of  
343 Globally Shared SARS-CoV-2 Data to Identify and Characterize New Variants.  
344 *Bioinformatics* 2021;38:1735–1737.
- 345  
346 6. Gram MA, Emborg H-D, Schelde AB, Friis NU, Nielsen KF, *et al*. Vaccine  
347 effectiveness against SARS-CoV-2 infection or COVID-19 hospitalization with the  
348 Alpha, Delta, or Omicron SARS-CoV-2 variant: A nationwide Danish cohort study.  
349 *PLOS Med*;19.
- 350  
351 7. Andrews N, Stowe J, Kirsebom F, Toffa S, Rikeard T, *et al*. Covid-19 Vaccine  
352 Effectiveness against the Omicron (B.1.1.529) Variant. *N Engl J Med* 2022;386:1532–  
353 1546.
- 354  
355 8. Kirsebom FCM, Andrews N, Stowe J, Toffa S, Sachdeva R, *et al*. COVID-19 vaccine  
356 effectiveness against the omicron (BA.2) variant in England. *Lancet Infect Dis*  
357 2022;22:931–933.
- 358  
359 9. Quick J. nCoV-2019 sequencing protocol V.1. *protocols.io*.
- 360  
361 10. Quick J. nCoV-2019 sequencing protocol v2 (GunIt) V.2. *protocols.io*.
- 362  
363 11. Quick J. nCoV-2019 sequencing protocol v3 (LoCost). *protocols.io*.
- 364  
365 12. Tyson JR, James P, Stoddart D, Sparks N, Wickenhagen A, *et al*. Improvements to  
366 the ARTIC multiplex PCR method for SARS-CoV-2 genome sequencing using

- 367 nanopore. *bioRxiv* 2020.
- 368
- 369 13. ARTIC Network. SARS-CoV-2 V4.1 update for Omicron variant.  
370 <https://community.artic.network/t/sars-cov-2-v4-1-update-for-omicron-variant/342>  
371 (2021).
- 372
- 373 14. ARTIC Network. SARS-CoV-2 version 4 scheme release.  
374 <https://community.artic.network/t/sars-cov-2-version-4-scheme-release/312>.
- 375
- 376 15. Davis JJ, Long SW, Christensen PA, Olsen RJ, Olson R, *et al.* Analysis of the ARTIC  
377 Version 3 and Version 4 SARS-CoV-2 Primers and Their Impact on the Detection of  
378 the G142D Amino Acid Substitution in the Spike Protein. *Microbiol Spectr*;9.
- 379
- 380 16. O'Toole Á, Scher E, Underwood A, Jackson B, Hill V, *et al.* Assignment of  
381 epidemiological lineages in an emerging pandemic using the pangolin tool. *Virus*  
382 *Evol*;7.
- 383
- 384 17. Rambaut A, Holmes EC, O'Toole Á, Hill V, McCrone JT, *et al.* A dynamic  
385 nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat*  
386 *Microbiol* 2020;5:1403–1407.
- 387
- 388 18. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple  
389 sequence alignment based on fast Fourier transform. *Nucleic Acids Res*  
390 2002;30:3059–3066.
- 391
- 392 19. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview Version 2--a  
393 multiple sequence alignment editor and analysis workbench. *Bioinformatics*  
394 2009;25:1189–1191.

395

## 396 **Figure legends**

397 **Figure 1. Coverage profile of SARS-CoV-2 BA.1 using V4 and V4.1 primers.** Percentage  
398 of samples with >50x coverage at each nucleotide position for the assembly of 59 SARS-  
399 CoV-2 BA.1 samples with **A.** V4 primers (in blue) or **B.** V4.1 primers (in orange). The depth  
400 of each nucleotide position was generated using SAMtools on the primer trimmed alignment  
401 files.



402 **Figure 2. Increased sequencing depth across amplicons and increased genome**  
403 **coverage in SARS-CoV-2 BA.1 samples amplified with V4.1 primers. A.** Median  
404 coverage across amplicons 10, 23, 27, 76, 79, 88, 89 and 90 with V4 (blue) and V4.1  
405 (orange). **B.** Percentage genome coverage at >50x coverage of all 59 SARS-CoV-2 BA.1  
406 samples. Assemblies with V4 primers in blue, and V4.1 primers in orange.

407 **Figure 3. Increased sequencing depth at BA.1 variant defining positions 22898, 26530**  
408 **and 26577.** Analysis of 59 BA.1 samples with V4 and V4.1 primers. **A.** Sequencing depth of  
409 variant defining mutations at all 17 nucleotide positions as defined by aln2type for BA.1 with  
410 V4 (blue) and V4.1 (orange) primers. **B.** Number of samples with masked bases (N's) at  
411 variant defining nucleotide positions in BA.1 samples with V4 (blue) and V4.1 (orange)  
412 primers.

413 **Figure 4. Coverage profile of SARS-CoV-2 BA.2 using V4.1 primers.** Percentage of  
414 samples with >50x coverage at each nucleotide position for the assembly of 168 SARS-  
415 CoV-2 BA.2 samples with V4.1 primers (in green). The depth of each nucleotide position  
416 was generated using SAMtools on the primer trimmed alignment files.

417 **Figure 5. Identification of SNPs responsible for low coverage of amplicons 15 and 75**  
418 **in BA.2 samples.** Multiple sequence alignment using MAFFT (v7.505) of NC\_045512.2  
419 reference with selected BA.2 samples with **A.** reduced coverage of amplicon 15 or **B.**  
420 reduced coverage of amplicon 75. Alignments viewed using Jalview (v2.11.2.2).

421 **Figure S1. Sequencing coverage and aln2type VOC-21NOV-01 defining mutations.**  
422 Analysis of 106 BA.1 samples with V4.1 primers. **A.** Sequencing depth of variant defining  
423 mutations at all 17 nucleotide positions as defined by aln2type for BA.1. **B.** Number of  
424 samples with masked bases (N's) at variant defining nucleotide positions in BA.1 samples.

425 **Figure S2. Sequencing coverage and aln2type VUI-22JAN-01 defining mutations.**  
426 Analysis of 168 BA.2 samples. **A.** Sequencing depth of variant defining mutations at all 20

427 nucleotide positions as defined by aln2type for BA.2. **B.** Number of samples with masked  
428 bases (N's) at variant defining nucleotide positions in BA.2 samples.

429 **Figure S3. A29510C aln2type mixed-call status due to downstream mutation present**  
430 **in subset of samples.** Multiple sequence alignment using MAFFT (v7.505) of NC\_045512.2  
431 reference with 10 BA.2 samples with mixed-call status for A29510C variant defining mutation  
432 and vssfru\_253 (A29510C detected). Alignment viewed using Jalview (v2.11.2.2).

433

#### 434 **Supplementary Datasets**

435 **Dataset 1. Overview of SARS-COV-2 genome coverage with V4 and V4.1 primers.** Total  
436 number of SNPs and masked bases (N's) generated using NextClade. Mapped reads, 20x,  
437 100x and 200x coverage data generated using RAMPART (v1.0.6). Lineage and conflict  
438 generated using PANGOLIN (v3.1.20). PHE-label, status and mutation calls generated using  
439 aln2type (v0.0.3). Data from two 48-plex libraries (Run1 and 2).

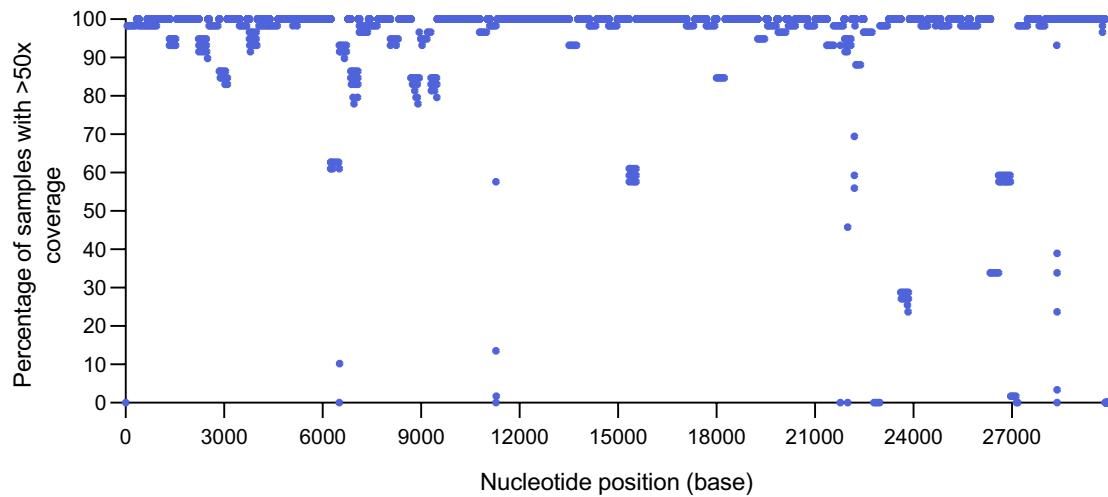
440 **Dataset 2. Overview of SARS-COV-2 BA.1 genome coverage with V4.1 primers.** Total  
441 number of SNPs and masked bases (N's) generated using NextClade. Mapped reads, 20x,  
442 100x and 200x coverage data generated using RAMPART (v1.0.6). Lineage and conflict  
443 generated using PANGOLIN (v3.1.20). PHE-label, status and mutation calls generated using  
444 aln2type (v0.0.3). Data from seven 48-plex libraries (Run3 – Run9) displaying only the BA.1  
445 samples.

446 **Dataset 3. Overview of SARS-COV-2 BA.2 genome coverage with V4.1 primers.** Total  
447 number of SNPs and masked bases (N's) generated using NextClade. Mapped reads, 20x,  
448 100x and 200x coverage data generated using RAMPART (v1.0.6). Lineage and conflict  
449 generated using PANGOLIN (v3.1.20). PHE-label, status and mutation calls generated using  
450 aln2type (v0.0.3). Data from seven 48-plex libraries (Run3 – Run9) displaying only the BA.2  
451 samples.

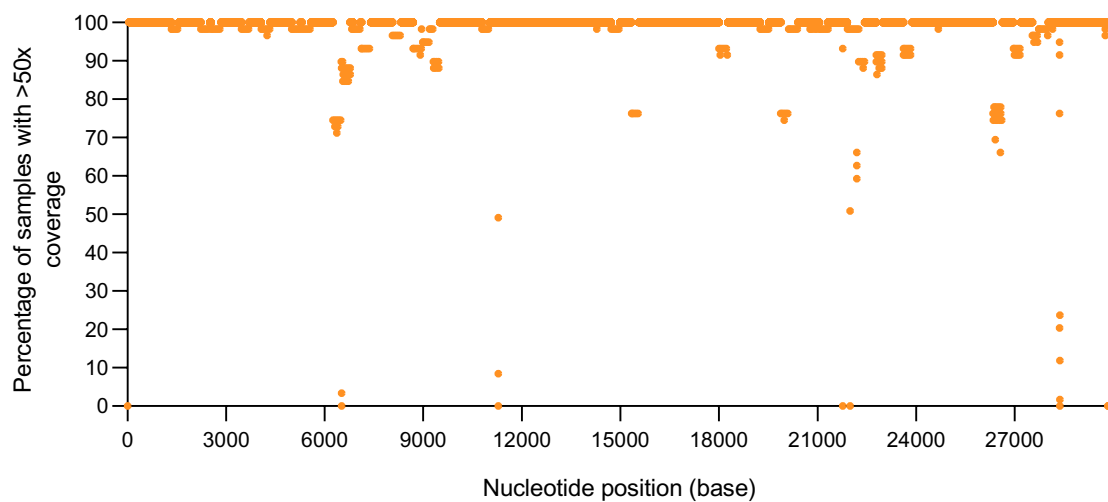
452

453

A

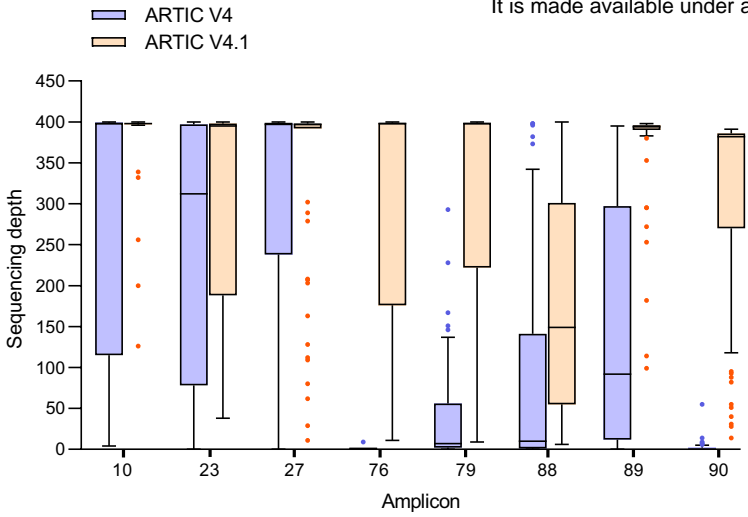


B

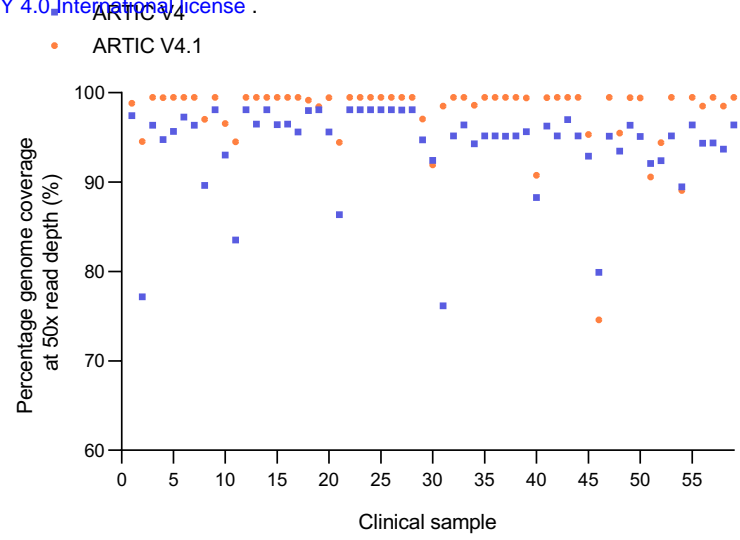


**Figure 1. Coverage profile of SARS-CoV-2 BA.1 using V4 and V4.1 primers.** Percentage of samples with >50x coverage at each nucleotide position for the assembly of 59 SARS-CoV-2 BA.1 samples with **A.** V4 primers (in blue) or **B.** V4.1 primers (in orange). The depth of each nucleotide position was generated using SAMtools on the primer trimmed alignment files.

A



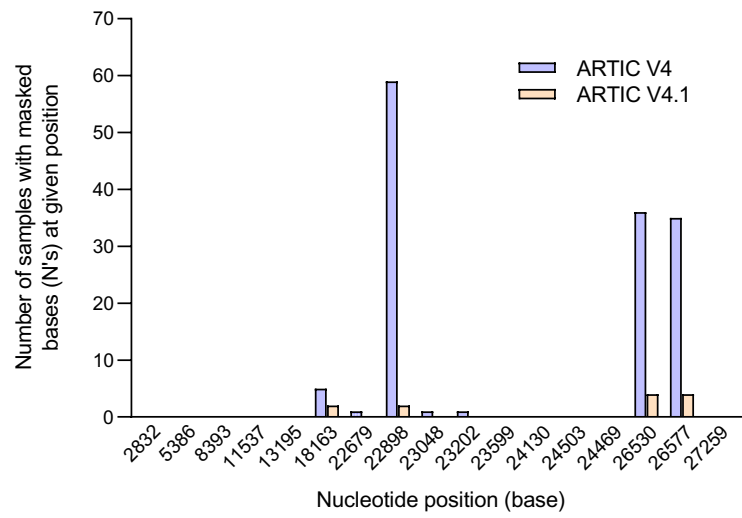
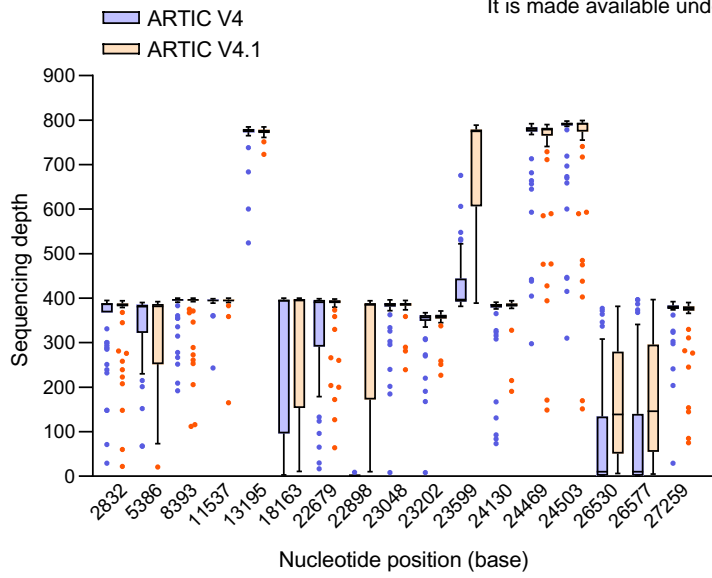
B



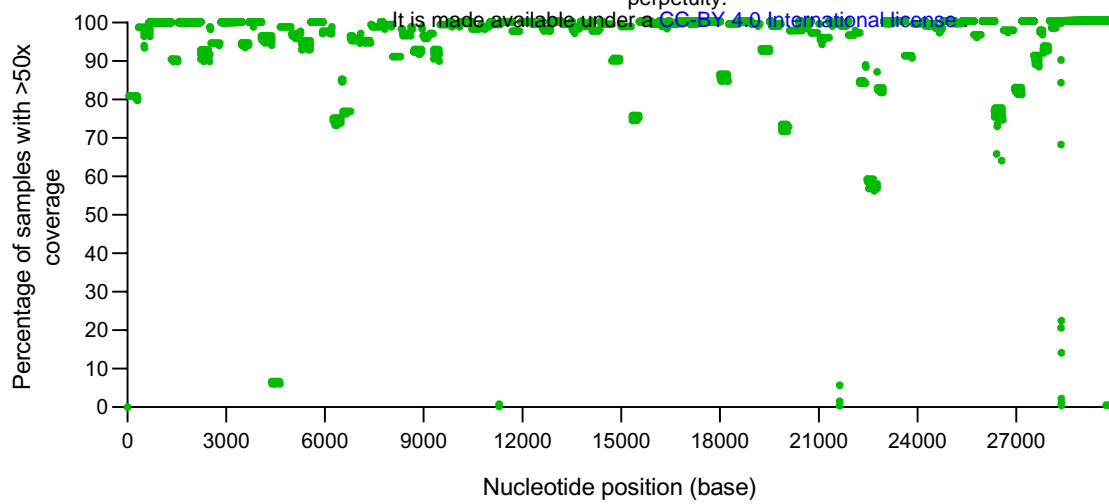
**Figure 2. Increased sequencing depth across amplicons and increased genome coverage in SARS-CoV-2**

**BA.1 samples amplified with V4.1 primers. A.** Median coverage across amplicons 10, 23, 27, 76, 79, 88, 89 and 90 with V4 (blue) and V4.1 (orange). **B.** Percentage genome coverage at >50x coverage of all 59 SARS-CoV-2 BA.1 samples. Assemblies with V4 primers in blue, and V4.1 primers in orange.

A

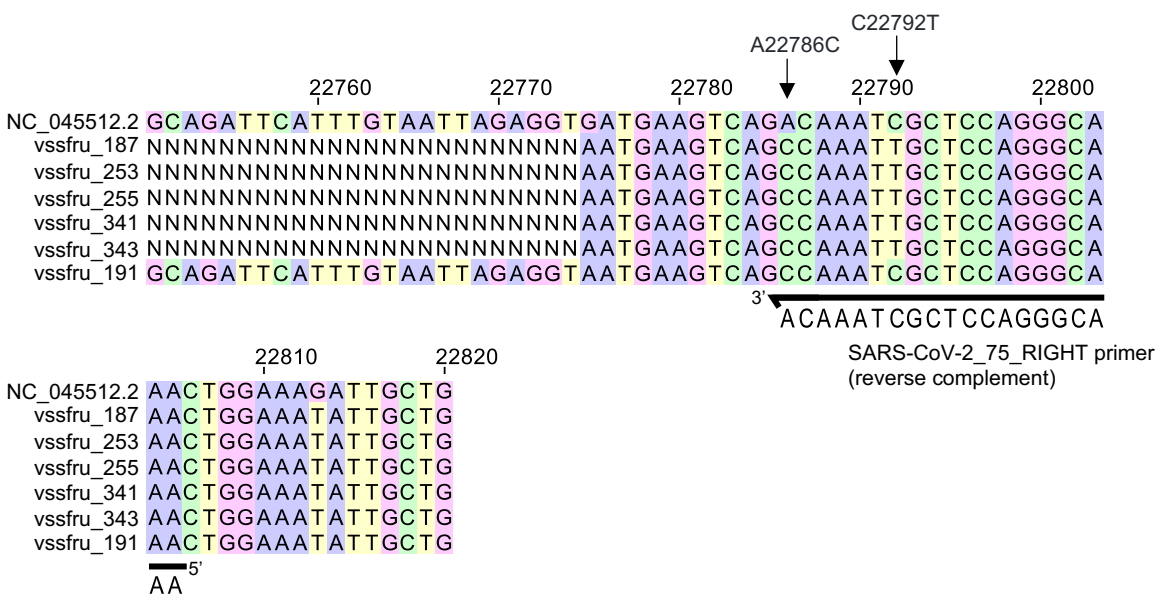
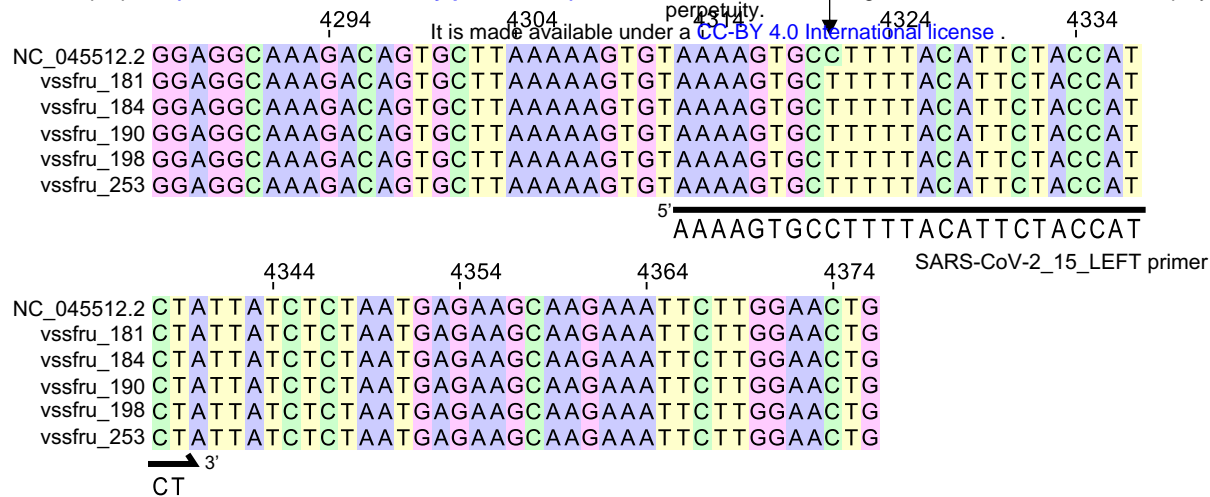


**Figure 3. Increased sequencing depth at BA.1 variant defining positions 22898, 26530 and 26577.** Analysis of 59 BA.1 samples with V4 and V4.1 primers. **A.** Sequencing depth of variant defining mutations at all 17 nucleotide positions as defined by `aln2type` for BA.1 with V4 (blue) and V4.1 (orange) primers. **B.** Number of samples with masked bases (N's) at variant defining nucleotide positions in BA.1 samples with V4 (blue) and V4.1 (orange) primers.



**Figure 4. Coverage profile of SARS-CoV-2 BA.2 using V4.1 primers.** Percentage of samples with >50x coverage at each nucleotide position for the assembly of 168 SARS-CoV-2 BA.2 samples with V4.1 primers (in green). The depth of each nucleotide position was generated using SAMtools on the primer trimmed alignment files.

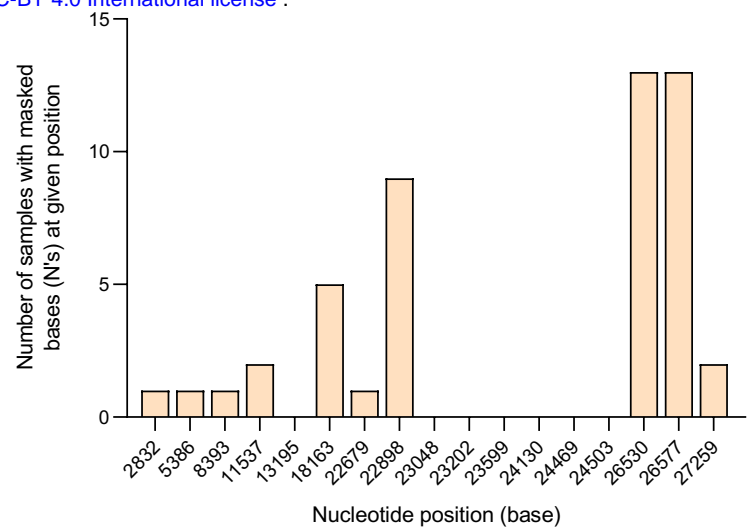
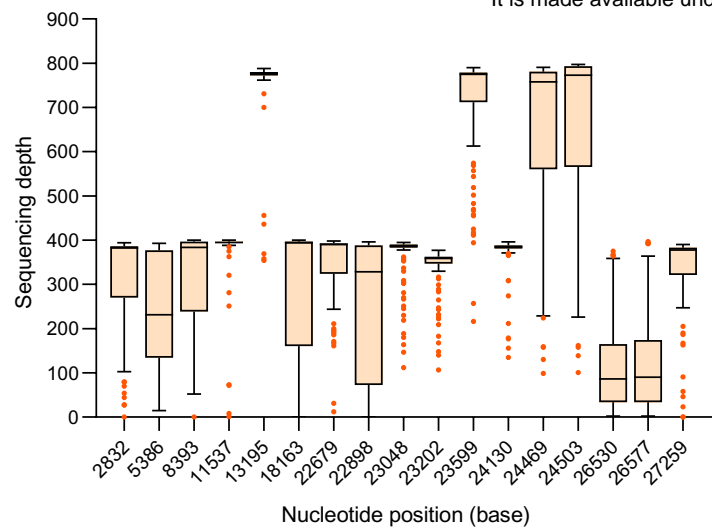
It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/).



**Figure 5. Identification of SNPs responsible for low coverage of amplicons 15 and 75 in BA.2 samples.** Multiple sequence alignment using MAFFT (v7.505) of NC\_045512.2 reference with selected BA.2 samples with **A.** reduced coverage of amplicon 15 or **B.** reduced coverage of amplicon 75. Alignments viewed using Jalview (v2.11.2.2).

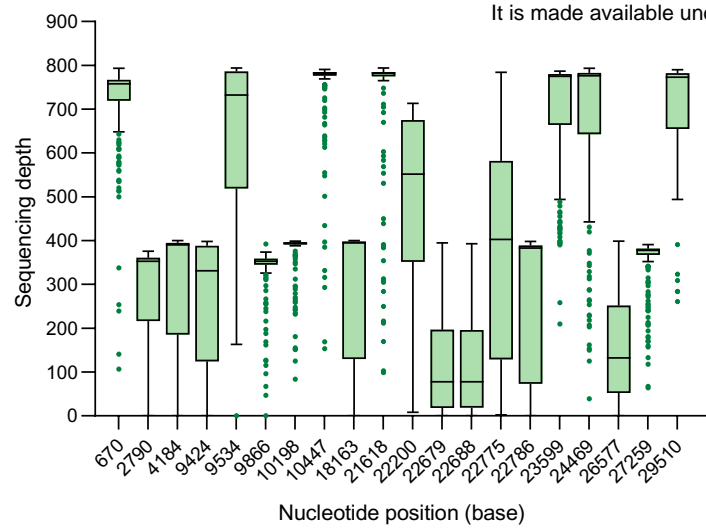


A

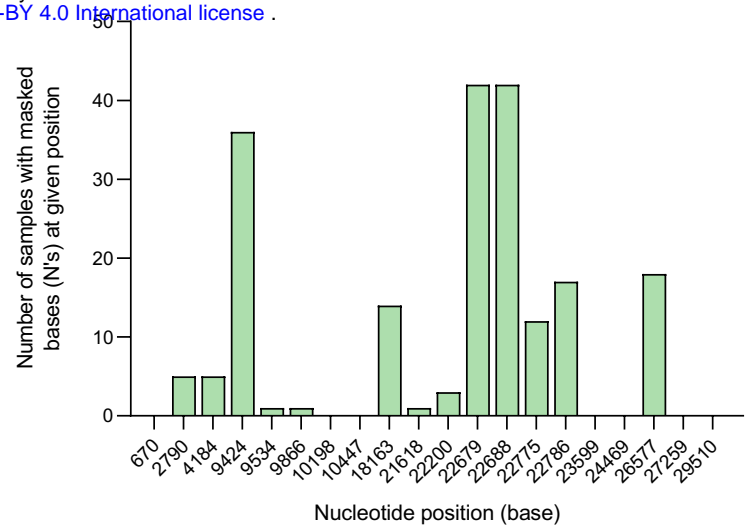


**Figure S1. Sequencing coverage and aln2type VOC-21NOV-01 defining mutations.** Analysis of 106 BA.1 samples with V4.1 primers. **A.** Sequencing depth of variant defining mutations at all 17 nucleotide positions as defined by aln2type for BA.1. **B.** Number of samples with masked bases (N's) at variant defining nucleotide positions in BA.1 samples.

A

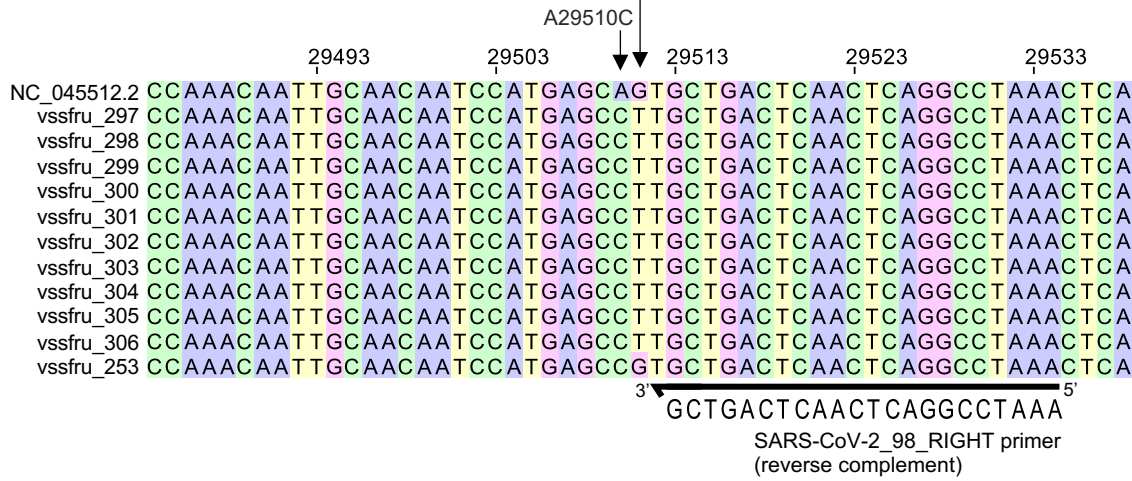


B



**Figure S2. Sequencing coverage and *aln2*type VUI-22JAN-01 defining mutations.** Analysis of 168 BA.2 samples. **A.** Sequencing depth of variant defining mutations at all 20 nucleotide positions as defined by *aln2*type for BA.2. **B.** Number of samples with masked bases (N's) at variant defining nucleotide positions in BA.2 samples.

It is made available under a [CC-BY 4.0 International license](#).



**Figure S3. A29510C aln2type mixed-call status due to downstream mutation present in subset of samples.**

Multiple sequence alignment using MAFFT (v7.505) of NC\_045512.2 reference with 10 BA.2 samples with mixed-call status for A29510C variant defining mutation and vssfru\_253 (A29510C detected). Alignment viewed using Jalview (v2.11.2.2).