1 **Evaluation of NGS-based approaches for SARS-CoV-2 whole genome characterisation**

2 Caroline Charre[1,2,3], Christophe Ginevra[4,5], Marina Sabatier[1,2,6], Hadrien Regue[1,2], Grégory

3 Destras[1,2,6,7], Solenne Brun[1,7], Gwendolyne Burfin[1,7], Caroline Scholtes[1,2,3], Florence

4 Morfin[1,2,6,7], Martine Valette[1,7], Bruno Lina[1,2,6,7], Antonin Bal[1,2,6,7] and Laurence Josset[1,2,6,7]

5 [1]Laboratoire de Virologie, Institut des Agents Infectieux (IAI), Hospices Civils de Lyon,

6 Groupement Hospitalier Nord, Lyon cedex 4, France

7 [2]Université de lyon, Université Claude Bernard Lyon 1 (UCBL1), Lyon

8 [3]Centre de recherche en cancérologie (CRCL), INSERM U1052- 69008, Lyon

9 [4]CIRI, Centre International de Recherche en Infectiologie, (Team pathogenesis of

10 Legionella), Univ Lyon, Inserm, U1111, Université Claude Bernard Lyon 1, CNRS,

11 UMR5308, ENS de Lyon, 69007, Lyon, France.

12 [5]Hospices Civils de Lyon, Centre National de Référence des Légionelles, Lyon, France

13 [6]Université de Lyon, Virpath, CIRI, INSERM U1111, CNRS UMR5308, ENS Lyon,

14 Université Claude Bernard Lyon 1, Lyon, France

15 [7]Centre National de Référence France-Sud des Virus des Infections Respiratoires, Hospices

16 Civils de Lyon, Groupement Hospitalier Nord, Lyon cedex 4, France

17

18 **Abstract**

19

20 Since the beginning of the COVID-19 outbreak, SARS-CoV-2 whole-genome sequencing

21 (WGS) has been performed at unprecedented rate worldwide with the use of very diverse

22 Next Generation Sequencing (NGS) methods. Herein, we compare the performance of four

23 NGS-based approaches for SARS-CoV-2 WGS. Twenty four clinical respiratory samples

24 with a large scale of Ct values (from 10.7 to 33.9) were sequenced with four methods. Three

25 used Illumina sequencing: an in-house metagenomic NGS (mNGS) protocol and two newly

26 commercialized kits including a hybridization capture method developed by Illumina (DNA

27 Prep with Enrichment kit and Respiratory Virus Oligo Panel, RVOP) and an amplicon

28 sequencing method developed by Paragon Genomics (CleanPlex SARS-CoV-2 kit). We also

29 evaluated the widely used amplicon sequencing protocol developed by ARTIC Network and

30 combined with Oxford Nanopore Technologies (ONT) sequencing. All four methods yielded

31 near-complete genomes (>99%) for high viral loads samples, with mNGS and RVOP

32 producing the most complete genomes. For mid viral loads, 2/8 and 1/8 genomes were

33 incomplete (<99%) with mNGS and both CleanPlex and RVOP, respectively. For low viral

34 loads (Ct ≥25), amplicon-based enrichment methods were the most sensitive techniques

35 yielding complete genomes for 7/8 samples. All methods were highly concordant in terms of

36 identity in complete consensus sequence. Just one mismatch in two samples was observed in

37 CleanPlex *vs* the other methods, due to the dedicated bioinformatics pipeline setting a high

38 threshold to call SNP compared to reference sequence. Importantly, all methods correctly

39 identified a newly observed 34-nt deletion in ORF6 but required specific bioinformatic

40 validation for RVOP. Finally, as a major warning for targeted techniques, a default of

41 coverage in any given region of the genome should alert to a potential rearrangement or a

42 SNP in primer annealing or probe-hybridizing regions and would require regular updates of

43 the technique according to SARS-CoV-2 evolution.

44   **Introduction**

45   A novel human betacoronavirus, Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-

46   CoV-2), emerged in China in December 2019, rapidly spreading worldwide and resulting in

47   the coronavirus disease 2019 (COVID-19) pandemic (1). Whole-genome sequencing (WGS)

48   of SARS-CoV-2 has played a major role since the onset of the pandemic. Notably WGS has

49   contributed to design specific RT-PCRs (2), antiviral strategies (3), and vaccine candidates

50   (4). SARS-CoV-2 WGS also allowed to explore lineage transmission and might be useful to

51   assess the effectiveness of intervention measures (5–8). Furthermore SARS-CoV-2 genomic

52   surveillance allowed the characterization of some mutations for which the impact on virulence

53   and transmissibility needs to be confirmed (9–11). SARS-CoV-2 WGS is performed at an

54   unprecedented rate worldwide with the use of very diverse Next Generation Sequencing

55   (NGS) methods. Viral metagenomic NGS (mNGS) enabled early sequencing of the SARS-

56   CoV-2 genome (1) and is the method routinely used in the French National Reference Centre

57   for Respiratory Viruses (NRC, Lyon, France). However, this method lacks sensitivity and

58   cannot produce whole genomes for low viral load samples, except using a very high depth of

59   sequencing (10). Targeted methods such as amplicon- or capture-based enrichments are

60   promising candidates to overcome this issue. A formal comparative study of these different

61   approaches has not been performed so far. Herein, we aimed to assess performance of four

62   NGS-based approaches for SARS-CoV-2 WGS. Three used Illumina sequencing: an in-house

63   metagenomic NGS (mNGS) protocol and two newly commercialized kits including a

64   hybridization capture method developed by Illumina and an amplicon sequencing method

65   developed by Paragon Genomics. We also assessed a widely used amplicon sequencing

66   protocol developed by the ARTIC network and combined with Oxford Nanopore

67   Technologies (ONT) sequencing (7,12). The present evaluation covers experimental process

68   as well as the associated bioinformatic solutions recommended for each technique.

2

69 Importantly, this evaluation included two samples with a newly observed 34-nt deletion in

70 ORF6 [Queromes et al, in preparation] in order to assess the capacity to detect large

71 modification in SARS-CoV-2 genome.

72 **Methods**

73 **Sample selection**

74 A total of 24 clinical samples (nasopharyngeal swab) with a broad range of representative

75 SARS-CoV-2 cycle threshold (Ct) values (from 10.7 to 33.0) were selected for sequencing.

76 We defined 3 groups of samples: a low Ct values group (Ct<20), a medium Ct values group

77 (20≤Ct<25) and high Ct values group (Ct≥25). Each sample was tested with 2 real time

78 reverse-transcriptase polymerase chain reactions (RT-PCR) targeting distinct RdRp gene

79 regions (Institut Pasteur assay, Paris, France (13)). Total nucleic acid extraction was

80 performed with easymag platform (bioMérieux, Lyon, France). Samples used in this study

81 were collected as part of approved ongoing surveillance conducted by the NRC at the

82 Hospices Civils de Lyon. The investigations were carried out in accordance with the General

83 Data Protection Regulation (Regulation (EU) 2016/679 and Directive 95/46/EC) and the

84 French data protection law (Law 78–17 on 06/01/1978 and Décret 2019–536 on 29/05/2019).

85 **Sequencing approaches and technologies**

86 **Illumina sequencing**

87   **Viral metagenomics (mNGS)**

88 DNase treatment (Life Technologies, Carlsbad, CA, USA) was performed after nucleic acid

89 extraction in order to increase sensitivity of RNA virus detection and overcome human

90 contamination (14). Nucleic acids were randomly amplified using the WTA2 kit (WTA2,

91 Sigma-Aldrich, Darmstadt, Germany) and libraries were prepared using the Illumina Nextera

92 XT kit (Illumina, San Diego, CA, USA). This in-house approach is routinely used to perform

93 WGS SARS-CoV-2 genomic surveillance at the NRC (10,15) for samples with a Ct value

94 <20.

95   **Hybrid capture-based target enrichment**

96 Hybrid capture-based approach was performed using the Illumina DNA Prep with Enrichment

97 kit and Illumina Respiratory Virus Oligo Panel (RVOP) V1.0. Briefly, double stranded cDNA

98 synthesis was performed from DNase pretreated total nucleic acid extracts using the

99 NEBNext$^®$ Ultra™ II RNA First Strand Synthesis Module and NEBNext$^®$ Ultra™ II Non-

100 Directional RNA Second Strand kit (New England Biolabs, MA, USA). The enrichment was

101 based on a hybridization step with a respiratory viruses biotinylated adjacent oligoprobes

102 panel (16) recently expanded to include SARS-CoV-2.

103   **Amplicon-based target enrichment**

104 We implemented the CleanPlex SARS-CoV-2 panel (Paragon Genomics, Inc, Hayward, CA,

105 USA) protocol for target enrichment and library preparation (17). In short, from reverse-

106 transcribed RNA, multiplex PCR reactions were performed using 343 pairs of primers

107 separated into two pools covering the entire genome of SARS-CoV-2 ranging from 116bp to

108 196bp, with a median size of 149 bp. Illumina indexes were introduced by PCR.

109 For these three approaches the prepared libraries were sequenced on an Illumina NextSeq^{TM}

110 550 with mid-output 2x150 bp (mNGS and CleanPlex) or 2x75 bp (RVOP) flow cells.

111 **ONT sequencing using amplicon-based target enrichment**

112 We tested a multiplexed PCR amplicon approach implementing the ARTIC Network nCoV-

113 2019 sequencing protocol slightly modified by ONT (Oxford Nanopore Technologies,

114 Oxford, UK) for better performance. The multiplex PCR primers set v3 was used to span the

115 whole genome (12). Briefly, synthesized cDNA was used as template and tiling 400nt-

116 amplicons with 20 base pairs overlaps (not including primers) were generated using two pools

117 of primers for 35 cycles. Samples were multiplexed by using the native barcode kits from

4

118   ONT (EXP-NBD104 and EXP-NBD114). The library was prepared using the SQK-LSK109

119   kit and then sequenced on a FLO-MIN106 (R9.4.1) flow cell, multiplexing 24 samples per

120   run.

121   **Sequencing data analysis**

122   mNGS data were analysed with an in-house pipeline. Briefly, low quality and human reads

123   were filtered out (dehosting) and remaining reads were aligned to the SARS-CoV-2 reference

124   genome (isolate Wuhan-Hu-1, EPI_ISL_402125) using the BWA-MEM (v0.7.15-r1140)

125   algorithm. Consensus sequences were generated through a simple majority rule using custom

126   perl script. These sequences were used as the patients' own mapping reference for further

127   realignment of the reads. Final consensus sequence was called at 10x using no-clip alignment.

128   Visual inspection of read alignments was performed using Integrative Genomics Viewer

129   (IGV) (18) around regions with large drop in coverage to detect and define potential deletions.

130   For the CleanPlex SARS-CoV-2 amplicon approach, data analysis was performed according

131   to the supplier's recommendations with a pipeline developed by Sophia Genetics (V1) (17).

132   Sequencing data from the RVOP capture approach were parsed using the Illumina Dynamic

133   Read Analysis for GENomics (DRAGEN; v3.5.13; Illumina) Bio-IT Platform. Finally, ONT

134   sequencing data were analysed by implementing the recommended bioinformatics developed

135   by ARTIC and available online (v1.1.0) (19). Briefly, basecalling and demultiplexing were

136   performed using Guppy high accuracy models (v3.5.2). Local alignment was performed using

137   Minimap2 (v 2.17) and variant calling with nanopolish (v0.13.2). Five parameters were

138   specifically evaluated and compared: coverage metrics such as depth and breadth of coverage

139   at 10x, number of SNPs between methods after pairwise alignment of consensus sequences,

140   and detection of a particular 34nt-deletion detected with our in-house reference mNGS

141   method. The proportion (%) of genome coverage are presented as medians with interquartile

142   ranges [IQR] and compared using the non-parametric Friedman test.

143 **Results**

144 **Sensitivity assessment of four SARS-CoV-2 sequencing approaches**

145 Four widely used NGS methods were evaluated: an unbiased mNGS approach, and three

146 targeted approaches based on hybridization capture (RVOP) or amplicon sequencing

147 (CleanPlex and ARTIC-ONT). For high viral loads (Ct <20), the 4 sequencing approaches

148 yielded almost-complete genome (> 99 % covered at 10x) for all samples. A significant

149 difference in genome coverage distribution (p<0.001) was found; the highest median of

150 coverage was for RVOP and mNGS (Table 1). For mid Ct samples, one sample sequenced

151 with RVOP and two samples sequenced with mNGS had a genome coverage below 99%

152 (93.4% with RVOP, 72.6% and 56.5% with mNGS, respectively).

153 Among samples with low viral loads (Ct≥25Ct), significant discrepancies in genome coverage

154 were observed between methods (p=0.005); 8.8% median coverage in samples sequenced

155 with mNGS, 92.0% for RVOP, and highest median coverage were obtained with amplicon

156 sequencing (99.4%, CleanPlex; 99.6%, ARTIC-ONT).

157 While RVOP and mNGS enabled to generate the complete genome (maximum coverage was

158 100% for highest viral loads), amplicon-based target enrichment did not allow to cover

159 SARS-CoV-2 genome ends, as expected by considering the design of ARTIC-ONT and

160 CleanPlex protocols. The highest coverage was 99.6% and 99.7% for ARTIC-ONT and

161 CleanPlex, respectively.

162 Irrespective of Ct value, RVOP provided an even depth of coverage throughout the genome,

163 while uneven depth of coverage was obtained with mNGS and marked drops in sequencing

164 depth were observed for CleanPlex (Figure 1). We did not interpret the evenness of depth of

165 coverage for ARTIC-ONT as the dedicated pipeline limits the depth at 400x. These drops

166 corresponded to amplicons that were poorly amplified: from 5,159 to 5,199 (drop # 1), from

167 14,430 to 14,505 (drop # 2), from 19,337 to 19,399 (drop # 3), and from 22,641 to 22,715

168   (drop # 4; Figure 1). Regarding low and mid Ct samples, 1, 3, and 8 had a coverage <10x at

169   drops # 2, #1, and # 4, respectively. Importantly none of these regions involved crucial

170   positions used to define the Nextstrain major clades (20).

171

172   **Comparison of complete genome consensus sequence**

173   Almost-complete genomes generated with the 4 methods were further compared to define

174   accuracy of each approach (Figure 2). Of note, 9, 7, and 3 samples sequenced, respectively,

175   with mNGS, RVOP and CleanPlex were excluded of the analysis due to a coverage <99% at

176   10x, *vs* 1 sample for ARTIC-ONT protocol. For all other samples, all consensus sequences

177   were strictly identical, expect for 2 samples for which CleanPlex led to one mismatch

178   compared to the other methods (sample # 1, position 280086; sample # 20, position 11083).

179   By analysing VCF files, these 2 positions corresponded to SNPs (G280086T, G11083T) with

180   variant frequency <70% which is the threshold set for calling a SNP in Sophia Genetics

181   pipeline while other pipelines set this threshold at 50%.

182

183   **Performance of the methods for the detection of a large deletion in SARS-CoV-2 genome**

184   A 34 nt-deletion in the ORF6 at loci 27266 was detected in 2 samples (samples # 7 and # 8)

185   by all the evaluated approaches both in alignment files and consensus sequences. For RVOP,

186   it is noteworthy that DRAGEN output alignment files showed an absence of coverage in the

187   region but there were no spanning reads to visualise the breakpoint of the deletion. However,

188   DRAGEN output variant calling files clearly highlighted the 34 nt-deletion as well as the final

189   consensus sequence for both samples. Using Minimap2 as aligner, the deletion was clearly

190   identified with reads spanning the breakpoint of the deletion on the reference genome

191   (Supplementary figure 1).

192 **Discussion**

193 We present herein the evaluation of 4 representative NGS-based approaches for SARS-CoV-2

194 whole genome characterisation. Unbiased mNGS is the most appropriate method to identify

195 novel emerging pathogens such as SARS-CoV-2 (1,21) but did not produce all WGS for mid

196 and low Ct values samples in the present study. In contrast, the three targeted methods

197 evaluated herein exhibited higher sensitivity compared to mNGS by reaching a high

198 proportion of the genome covered. In particular, amplicon-based target enrichment using

199 CleanPlex or ARTIC-ONT was the most sensitive approach and allowed WGS of samples

200 with Ct up to 33.9. Although not reported herein, amplicon-based target enrichment can be

201 impacted by SNP or indels located within primer-annealing regions even if their tiling

202 amplicon designs (instead of adjacent amplicon designs) aims to reduce the impact of such

203 modifications. With SARS-CoV-2 evolving, primers for amplicon-based target enrichment

204 need to be constantly updated, as it has been done for ARTIC-ONT (v3) (22). Such updates

205 will be important for CleanPlex as we have observed coverage dropout issues in 4 regions of

206 the genome irrespective of Ct value, suggesting potential mismatches in primer-annealing

207 regions that may lead to variation in the optimal annealing temperature and thus to a decrease

208 in the amplification efficiency.

209 All 4 methods were highly concordant in terms of identity in complete consensus sequence.

210 Just one mismatch in two samples was observed in CleanPlex *vs* the other methods, due to the

211 dedicated pipeline setting very a high threshold to call SNP compared to reference genome.

212 Considering the very low evolution rate of this virus that can be explained by the editing

213 function of its polymerase, just one mutation can have high repercussions in terms of

214 evolutionary assessment.

215 Furthermore, a novel 34nt-deletion in the ORF6 previously observed with our in-house

216 mNGS method, was detected by the two evaluated tiling amplicon sequencing methods

8

217 (ARTIC and CleanPlex). However larger deletions spanning two primer annealing regions

218 may arise and would be difficult to define using amplicon-based target enrichment methods.

219 Missing regions with no coverage should be carefully investigated using other methods. The

220 34nt-deletion was also detected using RVOP despite an adjacent oligoprobes design. As the

221 oligoprobe set is larger than the primer panel used in tiling amplicons protocols, a more

222 comprehensive profiling of regions is obtained independently of genomic rearrangement.

223 Moreover, a major advantage of this capture protocol, is the enrichment of other respiratory

224 viruses that enables to detect coinfections (16).

225 The present study does have several limitations. First, the 4 methods were tested on a limited

226 number of European samples, all affiliated to the lineage B1, according to the Pangolin

227 classification (23); the present evaluation should be confirmed in larger studies including

228 other lineages in order to be more representative of the global ecology of this emerging virus.

229 As SARS-CoV-2 sequencing is an expanding market and other methods has been published

230 (24–26) or are probably under development, the present study is not exhaustive. However, the

231 approaches selected herein are representative of the methods mostly used during pandemics

232 and are easy to implement in a diagnostic laboratory. As mentioned above, dedicated

233 bioinformatic processes have been implemented as recommended by suppliers. Except for

234 RVOP for which we compared the aligner using in DRAGEN and minimap2, we chose not to

235 perform a comparison of the bioinformatic tools. In this context, further bioinformatic

236 evaluations are needed especially concerning parameters such as the variant frequency

237 threshold for calling SNP compared to reference sequence, as well as the minimum depth for

238 calling consensus. Additionally, *de novo* assembly should be evaluated as it may provide

239 better assessment of rearrangements such as large deletions, insertions and duplications that

240 can be missed by classical alignment analyses, but this method is time-consuming and may

241 not be appropriate for genomes with low coverage. More broadly, a repeatability test is

242   mandatory to comfort results obtained from the present evaluation. Finally, cost effectiveness

243   and turnaround time were not assessed in this study considering that many parameters such as

244   laboratory equipment and prices negotiated with suppliers may affect results.

245   Beyond these limitations, the data presented herein are useful for clinical and research teams

246   who want to implement SARS-CoV-2 sequencing and chose the most suitable protocol

247   according to the application. To summarise, mNGS remains the gold standard for samples

248   with high viral load to obtain a maximum of information without any bias. For low and mid

249   Ct values, RVOP leads to very high coverage as well, enabling genome end sequencing,

250   contrary to amplicon methods. For higher Ct values, amplicon-based enrichment are a very

251   interesting alternative, in particular ARTIC-ONT protocol that did not show any dropout

252   issues in this present evaluation. However, as a reminder, default of coverage in any given

253   region of the genome should alert to a potential rearrangement or an SNP in primer annealing

254   or probe-hybridizing regions and require regular updates according to SARS-CoV-2

255   evolution.

256

261   **Statement of data availability**

262   Data will be deposited on SRA database prior to publication.

265

266

267 **Legends**

268 **Table 1.** Proportion of coverage throughout the SARS-CoV-2 genome at 10x according to

269 evaluated methods. Cycles threshold (Ct) values were determined using the RdRp Institute

270 Pasteur RT-PCR assay (IP4). Using an R script, these proportions (%) were calculated from

271 depth files generated by bedtools from output aligned bam files of each specific-method

272 pipeline. The percentage of genome coverage are presented as medians with interquartile

273 ranges [IQR] and compared using the non-parametric Friedman test.

274

275 **Figure 1.** Plots of coverage according to evaluated methods and Cycle threshold (Ct) values

276 groups. Using an R script, these plots were constructed via ggplot2 from depth files generated

277 by bedtools from output aligned bam files of each specific-method pipeline. Of note, ARTIC-

278 ONT pipeline limits the depth at 400x, the real depth is not represented herein.

279

280 **Figure 2.** Mismatch count between consensus sequences generated by each method compared

281 2 by 2 for each sample. These matrices were generated only from consensus with determined

282 bases for more than 99% of the genome. If one sequence of the two had more than 1% of

283 undetermined bases (N), comparison was not assessed, grey tiles. Blue tiles correspond to

284 perfect identity and orange tile correspond to mismatches, the number of mismatches is

285 indicated inside the tile. Matrices were generated with an R script using Decipher (alignment),

286 ape (distance matrices) and ggplot2 (charts) libraries. Of note, undetermined bases and

287 deletions were not considered in the calculation of mismatches.

288

289 **Supplementary Figure 1.** Visualisation by Integrative Genomics Viewer (IGV) of the 34-nt

290 deletion in the ORF6 highlighted from capture sequencing data of one of the two samples

291 concerned (sample # 8). Plot of coverage from DRAGEN pipeline is represented in grey with

11

292   conventional numbering systems for SARS-CoV-2 reference (MN908947.3). Depth at

293   position 27079 is indicated into brackets. Details of output alignment files from DRAGEN did

294   not show any reads (grey) overlapping the deletion contrary to those obtained using

295   Minimap2 as aligner (black line within read).

References

1.   Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. N Engl J Med. 2020 Feb 20;382(8):727–33.

2.   Corman VM, Landt O, Kaiser M, Molenkamp R, Meijer A, Chu DK, et al. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. Eurosurveillance [Internet]. 2020 Jan 23 [cited 2020 May 8];25(3). Available from: https://www.eurosurveillance.org/content/10.2807/1560-7917.ES.2020.25.3.2000045

3.   Dai W, Zhang B, Su H, Li J, Zhao Y, Xie X, et al. Structure-based design of antiviral drug candidates targeting the SARS-CoV-2 main protease. Science. 2020 Apr 22 [cited 2020 May 24]; Available from: https://science.sciencemag.org/content/early/2020/04/21/science.abb4489

4.   Kames J, Holcomb DD, Kimchi O, DiCuccio M, Hamasaki-Katagiri N, Wang T, et al. Sequence analysis of SARS-CoV-2 genome reveals features important for vaccine design. bioRxiv. 2020 Mar 31;2020.03.30.016832.

5.   Grubaugh ND, Ladner JT, Lemey P, Pybus OG, Rambaut A, Holmes EC, et al. Tracking virus outbreaks in the twenty-first century. Nat Microbiol. 2019 Jan;4(1):10–9.

6.   Oude Munnink BB, Nieuwenhuijse DF, Stein M, O'Toole Á, Haverkate M, Mollers M, et al. Rapid SARS-CoV-2 whole genome sequencing for informed public health decision making in the Netherlands. Genomics; 2020 Apr [cited 2020 May 27]. Available from: http://biorxiv.org/lookup/doi/10.1101/2020.04.21.050633

7.   Consortium TC-19 GU (COG-U, Robson S, Scarlett G, Bourgeois YXC, Beckett AH. An integrated national scale SARS-CoV-2 genomic surveillance network. Lancet Microbe. 2020 Jun 2 [cited 2020 Jun 18]; Available from: https://researchportal.port.ac.uk/portal/en/publications/an-integrated-national-scale-sarscov2-genomic-surveillance-network(d537cd78-4d18-4935-90c3-2645ddf9b34f).html

8.   Preliminary analysis of SARS-CoV-2 importation & establishment of UK transmission lineages. Virological. 2020 [cited 2020 Jun 18]. Available from: https://virological.org/t/preliminary-analysis-of-sars-cov-2-importation-establishment-of-uk-transmission-lineages/507

9.   Batty EM, Kochakarn T, Panthan B, Kumpornsin K, Jiaranai P, Wangwiwatsin A, et al. Genomic surveillance of SARS-CoV-2 in Thailand reveals mixed imported populations, a local lineage expansion and a virus with truncated ORF7a. medRxiv. 2020 May 25;2020.05.22.20108498.

10.   Bal A, Destras G, Gaymard A, Bouscambert-Duchamp M, Valette M, Escuret V, et al. Molecular characterization of SARS-CoV-2 in the first COVID-19 cluster in France reveals an amino acid deletion in nsp2 (Asp268del). Clin Microbiol Infect. 2020 Mar 28 [cited 2020 Jun 18]; Available from: http://www.sciencedirect.com/science/article/pii/S1198743X20301658

11.   Holland LA, Kaelin EA, Maqsood R, Estifanos B, Wu LI, Varsani A, et al. An 81 nucleotide deletion in SARS-CoV-2 ORF7a identified from sentinel surveillance in Arizona (Jan-Mar 2020). J Virol. 2020 May 1 [cited 2020 Jun 18]; Available from: https://jvi-asm-org.proxy.insermbiblio.inist.fr/content/early/2020/04/30/JVI.00711-20

12.   https://artic.network/ncov-2019. (accessed on 24 March 2020).

13.   https://www.who.int/docs/default-source/coronaviruse/real-time-rt-pcr-assays-for-the-detectionof-sars-cov-2-institut-pasteur-paris.pdf?sfvrsn=3662fcb6_2 (accessed on 24 March 2020).

14.   Bal A, Pichon M, Picard C, Casalegno JS, Valette M, Schuffenecker I, et al. Quality control implementation for universal characterization of DNA and RNA viruses in clinical respiratory samples using single metagenomic next-generation sequencing workflow. BMC Infect Dis. 2018 Dec;18(1):1–10.

13

346 15.   Danesh G, Elie B, Michalakis Y, Sofonea MT, Bal A, Behillil S, et al. Early
347 phylodynamics analysis of the COVID-19 epidemics in France. medRxiv. 2020 Jun
348 4;2020.06.03.20119925.
349 16.   O'Flaherty BM, Li Y, Tao Y, Paden CR, Queen K, Zhang J, et al. Comprehensive
350 viral enrichment enables sensitive respiratory virus genomic identification and analysis by
351 next generation sequencing. Genome Res. 2018 Jun;28(6):869–77.
352 17.   https://www.paragongenomics.com/wp-content/uploads/2020/03/UG4001-01_-
353 CleanPlex-SARS-CoV-2-Panel-User-Guide.pdf. (accessed on 24 March 2020).
354 18.   Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al.
355 Integrative genomics viewer. Nat Biotechnol. 2011 Jan;29(1):24–6.
356 19.   https://artic.network/ncov-2019/ncov2019-bioinformatics-sop.html. (accessed on 24
357 March 2020).
358 20.   https://nextstrain.org/blog/2020-06-02-SARSCoV2-clade-naming. (accessed on 01
359 July 2020)
360 21.   Mokili JL, Rohwer F, Dutilh BE. Metagenomics and future perspectives in virus
361 discovery. Curr Opin Virol. 2012 Feb;2(1):63–77.
362 22.   https://artic.network/resources/ncov/ncov-amplicon-v3.pdf. (accessed on 24 March
363 2020).
364 23.   https://pangolin.cog-uk.io/. (accessed on 01 July 2020)
365 24.   Batty EM, Kochakarn T, Wangwiwatsin A, Joonlasak K, Huang AT, Panthan B, et al.
366 Comparing library preparation methods for SARS-CoV-2 multiplex amplicon sequencing on
367 the Illumina MiSeq platform. bioRxiv. 2020 Jun 16;2020.06.16.154286.
368 25.   Freed NE, Vlková M, Faisal MB, Silander OK. Rapid and Inexpensive Whole-
369 Genome Sequencing of SARS-CoV2 using 1200 bp Tiled Amplicons and Oxford Nanopore
370 Rapid Barcoding. bioRxiv. 2020 Jun 2;2020.05.28.122648.
371 26.   Paden CR, Tao Y, Queen K, Zhang J, Li Y, Uehara A, et al. Early Release - Rapid,
372 Sensitive, Full-Genome Sequencing of Severe Acute Respiratory Syndrome Coronavirus 2 -
373 Volume 26, Number 10—October 2020 - Emerging Infectious Diseases journal - CDC. [cited
374 2020 Jul 3]; Available from: https://wwwnc.cdc.gov/eid/article/26/10/20-1800_article
375