1   **Title:**

2   Assessment of Metagenomic MinION and Illumina sequencing as an approach for the

3   recovery of whole genome sequences of chikungunya and dengue viruses directly from

4   clinical samples.

5

6   **Running title:**

7   Metagenomic sequencing of dengue & chikungunya

8

9   **Key Words:**

10   chikungunya dengue metagenomic nanopore

11

12

13   **Summary Line:**

14   Whole genomes of chikungunya and dengue viruses can be recovered directly from patient

15   blood samples in the majority of clinically relevant cases, without the need for enrichment or

16   viral isolation, on a portable sequencing device.

17

18   **Authors:**

19   Liana E. Kafetzopoulou[1,2], Kyriakos Efthymiadis[3], Kuiama Lewandowski[1], Ant Crook[1], Dan

20   Carter[1,2], Jane Osborne[4], Emma Aarons[4], Roger Hewson[1,2], Julian A. Hiscox[2,5], Miles W.

21   Carroll[1,2], Richard Vipond[1,2], Steven T. Pullan[1,2].

22

23   **Affiliations:**

24   1: Public Health England, National Infections Service, Porton Down, UK.

25   2: NIHR Health Protection Research Unit in Emerging and Zoonotic Infections, Liverpool,

26   UK.

27   3: Artificial Intelligence Laboratory, Vrije Universiteit Brussel, Brussels, BE

28   4: Rare and Important Pathogens Laboratory, Public Health England, Porton Down, UK

29   5: Institute of Infection and Global Health, University of Liverpool, UK.

30

31   **Corresponding author:**

32   Steven T. Pullan

33   Public Health England

34   National Infection Service

35   Porton Down, Salisbury,SP4 0JG

36   Tel: 01980 619678

37   Email: steven.pullan@phe.gov.uk

38

39

40   **Word count**

41   Abstract: 149

42   Main text: 3,497

43

44

45

46

47

48

49

50   **Abstract**

51

52   The recent global emergence and re-emergence of arboviruses has caused significant human

53   disease. Common vectors, symptoms and geographical distribution make differential

54   diagnosis both important and challenging. We performed metagenomic sequencing using

55   both the Illumina MiSeq and the portable Oxford Nanopore MinION to study the feasibility

56   of whole genome sequencing from clinical samples containing chikungunya or dengue virus,

57   two of the most important arboviruses. Direct metagenomic sequencing of nucleic acid

58   extracts from serum and plasma without viral enrichment allowed for virus and coinfection

59   identification, subtype determination and in the majority of cases elucidated complete or

60   near-complete genomes adequate for phylogenetic analysis. This work demonstrates that

61   metagenomic whole genome sequencing is feasible for over 90% and 80% of chikungunya

62   and dengue virus PCR-positive patient samples respectively. It confirms the feasibility of

63   field metagenomic sequencing for these and likely other RNA viruses, highlighting the

64   applicability of this approach to front-line public health.

65

66   **Introduction**

67

68   Arboviruses are predominantly RNA viruses that replicate in hematophagous (blood-

69   sucking) arthropod vectors such as ticks, mosquitoes and other biting flies to maintain their

70   transmission cycle [1]. These pathogens circulate through sylvatic or enzootic cycles in wild

71   animals causing disease in incidental or dead-end hosts, such as humans, in spill-over events

72   [2]. Human disease outbreaks caused by arboviruses have increased in prevalence over recent

73   decades as ecological factors, human activity and increased vector distribution range have led

74   to the emergence of previously unknown viruses and the re-emergence of viruses previously

75 not thought to be of public health significance [3]. This resurgence has been led by the spread

76 of mosquito-borne arboviruses such as yellow fever (YFV), chikungunya (CHIKV), dengue

77 (DENV), West Nile (WNV), and Zika (ZIKV) viruses across both hemispheres [4]. CHIKV

78 and DENV are of particular global health concern, as they have lost the need for enzootic

79 amplification and have consequently been responsible for extensive epidemics [2].

80   CHIKV is a single-stranded positive-sense RNA virus of the alphavirus genus which

81 causes the debilitating arthritic disease, chikungunya fever [5]. Originally restricted to

82 regions of Africa and Asia, it has recently spread globally and been designated a serious

83 emerging disease by the WHO [6]. Over the last 15 years, recurring outbreaks of CHIKV

84 appear to have been associated with an increased incidence of more acute presentation of the

85 disease, with an increase in morbidity and possibly mortality [7,8].

86   DENV is a single-stranded positive-sense RNA virus, member of the flavivirus genus

87 and the most prevalent human arboviral pathogen. DENV fever occurs following infection

88 with one of the four dengue serotypes (DENV1-4). A minority of cases develop severe

89 dengue, with acute hemorrhagic manifestations and multi-organ failure. Despite DENV cases

90 being underreported and misclassified, an increase of 143.1% in dengue fever cases has been

91 estimated between 2005 and 2015 [9]. Approximately 500,000 DENV infected patients

92 require hospitalisation annually [10].

93   Both CHIKV and DENV are predominantly transmitted to humans via *Aedes* species

94 mosquitoes, particularly *A. aegypti* and *A. albopictus* [11][12], and share clinical

95 presentations of high fever, arthralgia, myalgia, rash and headache. The circulation of

96 CHIKV, DENV (and other arboviruses) in the same geographical locations leads to

97 challenges in differential diagnosis, especially in endemic and epidemic regions in which

98 diagnosis is predominantly symptom-based [13]. Additionally, co-circulation and common

99 vector transmission increase the risk of coinfections. Arboviral co-infection reports have

100   increased in recent years, however, little is known about their clinical presentation and

101   consequences [14–17]. Metagenomic sequencing allows for identification of multiple

102   pathogens within a sample in a non-targeted and unbiased approach. Metagenomic RNA

103   sequencing enables target independent detection and discovery of RNA viruses which has

104   been used for causative agent identification in outbreaks, e.g. Lujo virus in South Africa [18],

105   Bundibugyo ebolavirus in Uganda [19] and novel virus discovery such as the Rhabdovirus

106   associated with hemorrhagic fever in central Africa [20]. In addition to pathogen

107   identification it also provides genomic information for downstream analysis including typing

108   and surveillance. Real-time genomic surveillance was facilitated on-site by the portable

109   Oxford nanopore MinION sequencer during the 2014-2016 EBOV epidemic in West Africa

110   and the ZIKV outbreak in the Americas [21–24] for epidemiological and transmission chain

111   investigations [25]. In both examples, an amplicon sequencing approach was utilised, which

112   has proven extremely useful for known pathogen outbreak samples with low viral titres, but

113   this approach does not allow detection of coinfections or unknown pathogens. Metagenomic

114   approaches on the MinION have in principle sequenced viruses and bacteria from clinical,

115   environmental and vector samples [26–29]. Metagenomic sequencing of CHIKV was also

116   demonstrated in principle on the MinION by Greninger *et al.* in 2015 reporting the detection

117   of CHIKV from a human blood sample [29]. Additionally, metagenomics identified CHIKV

118   coinfections within a ZIKV sample cohort [30], with the high proportion of CHIKV reads

119   identified making it a promising target for the approach.

120   The resurgence of CHIKV and DENV, and the prevalence of mixed arboviral

121   infections [31–33] create an issue that could ideally be resolved in a single, unbiased and

122   rapid virus detection. In this study we set out to test the feasibility of direct metagenomic

123   sequencing of DENV and CHIKV genomes from a cohort of clinical serum and plasma

124   samples with a representative range of viral loads seen at the time of diagnosis. The objective

125 was to assess the proportion of total nucleic acid present in these samples, that was viral in

126 origin, and to determine the limits of retrieval for whole genome sequencing using both the

127 laboratory based Illumina technology and the portable MinION platform.

128

129 **Methods**

130 **Sample Collection and Nucleic Acid Extraction**

131 26 routine diagnostic samples, 9 plasma and 17 serum, were obtained from the Rare and

132 Imported Pathogens Laboratory (RIPL), Public Health England, Porton Down. All had

133 previously tested positive by real-time reverse transcription-PCR (qRT-PCR) for

134 chikungunya or dengue virus. A range of representative clinical Ct values was selected. Total

135 nucleic acid was extracted from 140 μl of each using the QIAamp viral RNA kit (Qiagen)

136 replacing carrier RNA with linear polyacrylamide and eluting in 60 ul, followed by treatment

137 with TURBO DNase (Thermo Fisher Scientific) at 37°C for 30 min. RNA was purified and

138 concentrated to 8 μl using the RNA Clean & Concentrator™-5 kit (Zymo Research)

139

140 **Molecular confirmation and quantification**

141 Drosten *et al.* [34] and Edwards *et al. [35]* RT-PCR assays were used for confirmation of

142 DENV and CHIKV respectively. RNA oligomers were used as standards for genome copy

143 quantitation.

144

145 **Metagenomic Library Preparation**

146 cDNA was prepared using a Sequence Independent Single Primer Amplification (SISPA)

147 approach adapted from Greninger *et al.* [36]. Reverse transcription and second strand cDNA

148   synthesis were as described. cDNA amplification was performed using AccuTaq LA (Sigma),

149   in which 5μl of cDNA and 1 μl (100 pmol/μl)Primer B (5′-GTTTCCCACTGGAGGATA-3′)

150   were added to a 50 μl reaction, according to manufacturer's instructions. PCR conditions: 98

151   °C for 30s; 30 cycles of 94 °C for 15s, 50 °C for 20s, and 68 °C for 5min, followed by 68 °C

152   for 10min. Amplified cDNA was purified using a 1:1 ratio of AMPure XP beads (Beckman

153   Coulter, Brea, CA) and quantified using the Qubit High Sensitivity dsDNA kit (Thermo

154   Fisher).

155

156   **MinION Library Preparation and Sequencing**

157   MinION sequencing libraries were prepared using total amplified cDNA of each sample to a

158   maximum of 1 μg. Oxford Nanopore kits SQK-NSK007 or SQK-LSK208 (2D), SQK-

159   LSK308 (1D^2) and SQK-RBK001 (Rapid) were used and each sample was run individually

160   on the appropriate flow cell (FLO-MIN105, FLO-MIN106 or FLO-MIN107) and run on a

161   48hr script. Basecalling was performed using Metrichor (ONT) for SQK-NSK007 and SQK-

162   LSK208 or Albacore v1.2 for SQK-LSK308 and SQK-RBK001. Poretools [37] was used to

163   extract fastq files from Metrichor fast5 files.

164

165   **Illumina Library Preparation and Sequencing**

166   Nextera XT V2 kit (Illumina) sequencing libraries were prepared using 1.5 ng of amplified

167   cDNA as per manufacturer's instructions. Samples were multiplexed in batches of a

168   maximum of 16 samples per run and sequenced on a 2x150bp paired end Illumina MiSeq run,

169   by Genomics Services Development Unit, Public Health England.

170

171 **Data handling**

172 BWA MEM v0.7.15 [38] was used to align reads to the following references (Genbank ID):

173 DENV Serotype 1 (NC_001477.1), DENV Serotype 2 (NC_001474.2), DENV Serotype 3

174 (NC_001475.2), DENV Serotype 4 (NC_002640.1) and CHIKV (NC_004162.2) using -x

175 ont2d mode for nanopore and MEM defaults for Illumina reads. Samtools v1.4 was used to

176 compute percentage reads mapped and coverage depth. Bedtools v2.26.0 was used to

177 calculate genome coverage at 10x and 20x. Mapping consensus sequences for Illumina were

178 generated using in-house software QuasiBam [39] and for MinION using Racon [40] with

179 pileup function. Taxonomic classification was performed using Kraken (0.10.4-beta)[41] and

180 a locally built database populated with bacterial, viral, and archaeal genomes [42].

181 De novo assemblies were generated using Spades 3.8.2 [43] in combination with SSPACE

182 Standard v3.0 [44] for Illumina generated sequences and CANU v1.6 [44,45] for Nanopore

183 sequences (settings: corOutCoverage=1000, genomeSize=12000, minReadLength=300,

184 minOverlapLength=50).

185

186 **Results & Discussion**

187

188 **Metagenomic Illumina sequencing of CHIKV and DENV patient samples**

189 To assess the feasibility of direct metagenomic sequencing of CHIKV and DENV genomes

190 from patient serum and plasma we selected a set of positive samples collected during 2016 in

191 RIPL diagnostic laboratories, PHE Porton Down. For each virus, samples were chosen to

192 represent the range of viral titres seen during 2016, based on real-time qRT-PCR Ct value

193 (Figure 1). Ct values were confirmed with a second qRT-PCR and genome copy numbers

194 estimated. CHIKV samples selected ranged from Ct 14.72 to Ct 32.57, corresponding to 10^{10}

195   and 10^5 genome copies per ml of serum or plasma. DENV samples selected ranged from Ct

196   16.29 to Ct 31.29, corresponding to 10^9 and 10^5 estimated genome copies per ml (Table 1).

197   All samples were subject to DNase digestion prior to reverse-transcription and SISPA. In

198   order to measure the proportion of viral nucleic acid present relative to host/background and

199   assess genome coverage, all samples were sequenced using an Illumina MiSeq, generating

200   between 219,000 and 2,070,000 paired-end reads per sample (Table 1).

201

202   **Reference mapping of Illumina data**

203   Reads were mapped to appropriate reference sequences, the percentage of reads

204   mapping and genome coverage at a minimum depth of 20 reads (20x) are shown in Figure 2.

205   The proportion of total reads mapping to the respective viral reference was high for both

206   viruses. In some low Ct samples, over 90% of reads mapped to the viral reference and high

207   proportions were still observed at mid-Ct range. The lowest proportion of reads mapping to

208   the viral reference was 5.03% and 0.47% for CHIKV and DENV respectively (Table 1,

209   Figure 2). The majority of samples returned over 95% genome coverage at 20x and over 98%

210   genome coverage at 10x. Irrespective of lower mapping percentages in high Ct value

211   samples, genome coverage of 88.5% (20x) and 89.4% (10x) for CHIKV and 75.0% (20x) and

212   77.8% (10x) for DENV was observed. These results clearly show that, due to the significant

213   levels of viral nucleic acid present in CHIKV and DENV clinical samples, relatively modest

214   metagenomic sequencing is capable of elucidating significant portions of viral genome even

215   for samples with Ct values at the higher end of clinical range. A weak correlation was

216   observed between Ct value and proportion of viral reads, with a significant level of variation

217   between samples, likely due to patient-to-patient variability or sample handling during

218   collection, storage and testing. This affected the total level of non-viral host/background

219   nucleic acid present. For example, the two lowest viral titre CHIKV samples (13 & 14) have

220   similar Ct values (32.2 & 32.57) but varied significantly in the proportion of reads that were

221   viral in origin (5.03% & 21.72%). The 5.03% viral reads in CHIKV13 is the lowest for

222   CHIKV, yet still sufficient to generate 88.5% of the CHIKV genome at 20x depth from just

223   ~662,000 paired-end reads. This amount of genomic information is highly informative and

224   further sequencing would very likely increase coverage. Of all diagnostic samples tested in

225   2016 only 7 of 73 CHIKV samples had a Ct greater than 32.2 (including sample CHIKV14)

226   (Table 1), which suggests that for the majority (>90%) of CHIKV PCR positive samples viral

227   load is sufficient for genome sequencing directly from patient samples without further

228   enrichment beyond a simple DNAse digestion (Figure 1). For the DENV samples, the lowest

229   viral read proportion observed was 0.47% in DENV12, Ct 31.29. This generated 71.5%

230   coverage at 20x depth (increased to 77.8 at 10x depth) from just 687,000 paired end Illumina

231   reads and allowed for DENV serotype identification. Only 62 of 368 DENV cases in 2016

232   had a higher Ct, predicting that >80% of PCR positive DENV samples have a viral load

233   sufficient for genome sequencing (Figure 1).

234

235   **Metagenomic MinION sequencing of CHIKV and DENV patient samples**

236   The high yield of viral sequences from clinical CHIKV and DENV samples and

237   recent improvements in data output from the Oxford Nanopore MinION portable sequencing

238   device highlight the exciting prospect of metagenomic MinION viral whole-genome-

239   sequencing, even for lower viral titre samples. To assess this, four samples for each virus

240   from the Illumina sequenced samples were selected to evaluate the performance of the

241   MinION with the metagenomic protocol. Samples were selected with low, mid and high Ct

242   values; CHIKV 1, CHIKV 3, CHIKV 4 and CHIKV 9 (Ct: 14.72, 21.41, 22.74 and 30.08)

243   and DENV 1, DENV 2, DENV 6, DENV 11 (Ct: 16.29, 16.85, 21.22 and 28.69). Total cDNA

244   from the SISPA preparation was used as input for sequencing library preparation and

245   sequenced individually on a single flow cell. Details on total input amounts, sequencing kit

246   and flow cell information for each can be found in Table 2. Total 1D reads generated for each

247   run varied between 203,000 and 3,481,000 per sample and mean read length ranged from 564

248   to 886 bases (Table 2).

249

250   **Reference mapping of Nanopore data**

251   Extracted 1D reads were mapped to the reference genomes used in the Illumina data

252   analysis. Figure 3 shows percentages of reads mapping to viral reference which were

253   generally concordant with the Illumina data, although a slight decrease is observed across the

254   range of Ct values. In the MinION data the highest mapped read percentages observed were

255   85.12% and 72.14% for CHIKV 9 and DENV 2 respectively, compared to 95.23% and

256   92.56% in the Illumina data from the same samples. Whilst in high Ct samples the viral

257   proportion drops to 4.08% for CHIKV 9 and 2.90% for DENV 11, from 6.66% and 3.73% in

258   the Illumina data. Despite the decrease in proportion of mapped viral reads, comparable

259   genome coverage is observed at both 20x and 10x (Figure 3, Table 3) and is even increased

260   compared to Illumina data at lower viral titres, e.g 100% at 20x for CHIKV 9 compared to

261   95.98% in the Illumina data and 95.25% for the high Ct DENV 11 sample which generated

262   94.44% coverage from the Illumina data. Differences in precise proportions of viral reads

263   seen are likely due to inter-library variation and do not appear significant. Differences in

264   genome coverage achieved are due to both differences in total reads generated per sample

265   (not normalised between platforms) as well as differences in average individual read length.

266   Average read lengths in Nanopore data ranged from 564 to 886 bp (Table 2).

267   Figure 4 shows coverage depth of reads mapped across the relevant genome for each

268   sample sequenced by both Illumina and Nanopore. Read levels are not normalised thus actual

269   depth is a function of total reads sequenced, but it can clearly be observed that the overall

270   pattern of coverage seen is highly similar for both methods suggesting it is more dependent

271   upon the SISPA methodology than sequencing library preparation. Simple voting consensus

272   sequences generated from each alignment ranged from 99.5% to 99.9% agreement between

273   Illumina and Nanopore data.

274   Viral read proportions are generally in agreement with that seen for Illumina

275   sequencing thus it is expected that a similar proportion of qRT-PCR positive patient samples

276   would be suitable for direct metagenomic sequencing on the MinION. Of the samples tested

277   on the MinION, lowest titre samples CHIKV 9 and DENV 11 both generated near complete

278   genome coverage.

279

280   **Metagenomic data analysis and co-infection**

281   The data presented clearly demonstrate that direct metagenomic sequencing of serum and

282   plasma samples is a feasible approach for characterisation of CHIKV and DENV with typical

283   viral loads for imported cases sent to PHE's Rare and Imported Pathogens Laboratory. The

284   true power of a metagenomic approach is unbiased virus identification without prior

285   knowledge of the pathogen. To test the applicability of the metagenomic approach for the

286   data generated, we investigated the target virus detection in each sample by read taxonomic

287   classification using Kraken (Figure 5). The distribution of reads classified as CHIKV, DENV,

288   other viruses, bacteria, archaea/eukaryota and unclassified show a similar pattern for both

289   Illumina and MinION generated reads. The proportion of unclassified reads for each sample

290   increased with Ct value, as the proportion of human origin reads is higher in samples with

291   lower viral load and the human genome is not represented in the Kraken database. A

292   decrease in the percentage of CHIKV and DENV classified reads is observed for MinION

293   data compared to Illumina, due to the lower per read accuracy, but was sufficient to identify

294   the correct predominant virus for all samples.

295

296   Kraken analysis also allowed for the identification of a DENV co-infection in sample CHIKV

297   3, the consensus sequence of which (derived via reference mapping) was unique in the

298   sample set, eliminating cross-contamination from the DENV positive samples. For sample

299   CHIKV 3 (CHIKV Ct 21.41) Kraken classified 0.08% of Illumina reads and 0.15% of

300   MinION reads as DENV. Using reference mapping to validate the finding, 0.22% of Illumina

301   reads and 0.43% of MinION reads mapped to DENV-1 reference genome. Genome coverage

302   at 20x of 99.73% and 95.99% was achieved for the primary CHIKV and secondary DENV

303   co-infection respectively, with a single MinION flow cell. This demonstrates the clear benefit

304   of non-targeted approaches for the capacity of libraries generated to detect co-infections even

305   in the presence of another high-titre infection.

306

307   *De novo* **assembly**

308   As alternative reference-free approach to read classification *de novo* assembly of the data was

309   attempted using Canu [45] and contig identification by BLASTn. Table 3. lists the longest

310   viral contig length identified in each sample, ranging from 4.2 Kb (36% of reference genome

311   size) to 10.8 Kb (91%) for CHIKV and 4.7 Kb (44%) to 10.1 Kb (95%) for DENV.

312   Identification of the pathogen present without prior knowledge would have therefore been

313   possible for all samples.

314
315   **Latest MinION library kits.**

316   Initial experiments used analysis of 1D reads from MinION 2D library kit (SQK-

317   LSK208) to successfully identify and genome sequence CHIKV and DENV present in the co-

318   infected CHIKV 3 sample. In order to keep up with the rapidly evolving nanopore

319   chemistries and flowcell developments, we repeated the sequencing utilising the MinION $1D^2$

320   (SQK-LSK308) and Rapid (SQK-RBK001) kits, currently the most accurate and the fastest

321   library preparation kit available, respectively. The Rapid kit has a library preparation time of

322   just 10 minutes. From this study 74.5% of reads generated with the $1D^2$ kit mapped to

323   CHIKV and 0.37% to DENV, whilst from the Rapid kit the result was 66.26% and 0.29%

324   respectively. A reduction in viral proportion of total reads was observed compared to the 2D

325   kit, which may be due in part to the extended storage time of the original samples prior to

326   retesting. In the case of the $1D^2$ kit, the lower proportion was outweighed by a substantial

327   increase in total data generated per flow cell (5 M vs 1.8 M reads). For the Rapid kit, the

328   lower proportions should be considered in the light of the greatly simplified sample workflow

329   and turnaround-time. Coverage at 20x for CHIKV was above 99% for both kits, and for

330   DENV was 95.04% from the $1D^2$ and 81.09% from the Rapid kit (Table 4). Coverage depth

331   across the genome for both viruses can be found in Figure 6, presenting a similar pattern of

332   coverage independent of the kit used. *De novo* assembly of the data (Table 4) produced a

333   CHIKV contig of 10.7, 11.3 and 11.4 Kb for the 2D, $1D^2$ and Rapid libraries respectively and

334   the longest contigs generated for DENV were 7.5, 2.3 and 4.2 Kb. An advantage of the

335   MinION system is the production of data in real-time during the sequencing run. Analysis

336   showed that near-maximum coverage for both viruses was obtained within 30 minutes with

337   the 2D kit, 8 minutes with the $1D^2$ kit and 85 minutes with the Rapid kit (Supplementary

338   Figure 1).

339

340   **Conclusions**

341   Globalisation, environmental changes and other human interventions, have increased

342   human risk from emerging and re-emerging viral pathogens [46]. The improvement of

343   diagnostic and genomic sequencing methods, both in speed and detection, will provide an

344   important advantage for disease monitoring and control. We demonstrate that across the

345   clinically relevant range of viral loads an unexpectedly high proportion of reads generated are

346   viral in origin and therefore metagenomic sequencing provides an effective approach for the

347   analysis of CHIKV and DENV genomes directly from the majority of qRT-PCR positive

348   serum and plasma samples, without the need for culture or viral nucleic acid enrichment

349   beyond a simple DNA degradation step. This and recent improvements in the read numbers

350   generated by the MinION, make metagenomic whole genome sequencing feasible in the

351   field. Its use in such a scenario can complement the use of targeted sequencing, as although

352   lower throughput, it has significant additional power in the initial assessment of syndromic

353   outbreaks that lack a clear suspected agent. Once defined, higher throughput targeted

354   methods can be employed alongside. Additionally, the metagenomic approach can detect

355   unexpected causes of disease in the midst of other outbreaks, as well as co-infections. It is

356   ideally suited for the investigation of viral species with high levels of genetic diversity which

357   are difficult to assess using targeted methods, without constant reappraisal. Such issues

358   equally affect diagnostics in non-outbreak scenarios in which the range of RNA viruses to be

359   tested for is both wide and genetically diverse, such as second-line testing of undiagnosed

360   fevers in returning travellers, which therefore may also benefit from a metagenomic

361   approach.

362

363

364

365

366

367

368

369   **Figure Legends**

370

371   **Figure 1. Ct values distribution of RIPL CHIKV and DENV positive samples 2016.**

372   A total of 73 samples were positive for CHIKV, and 368 were positive for DENV. Median Ct

373   for CHIKV was 26.1, for DENV it was 26.8. The 14 CHIKV and 12 DENV samples selected

374   for this work are indicated by circles, and were selected to represent a diverse spread of

375   clinical viral loads.

376

377   **Figure 2. Proportion of viral reads and reference genome coverage per sample.**

378   For each sample, identified by Ct value, the percentage of total reads mapping to the

379   appropriate reference sequence, and therefore attributed as viral in origin, is plotted in the

380   upper panel. Lower panels display the percentage of the reference genome sequenced to a

381   minimum depth of 20-fold in the Illumina data generated. See Table 1 for total read numbers.

382

383   **Figure 3. Proportion of viral reads and reference genome coverage comparison.**

384   For each sample, identified by Ct value, the percentage of total reads mapping to the

385   appropriate reference sequence, and therefore attributed as viral in origin, is plotted in the

386   upper. Lower panels display the percentage of the reference genome sequenced to a minimum

387   depth of 20-fold in the data generated, in black or grey for the Illumina sequence data, in blue

388   or green for Nanopore data. See Tables 1 & 3 for total read numbers.

389

390   **Figure 4. Coverage depth across the CHIKV or DENV genome**

391   Read depth across the genome following reference alignment is shown for all eight samples

392   sequenced by both Nanopore and Illumina method. Illumina coverage is shown in gray,

393   nanopore in blue and green for CHIKV and DENV respectively. Total depth has not been

394   normalised, comparison is to show overall pattern of coverage is highly similar across the

395   methods.

396

397   **Figure 5. Kraken classification of reads from metagenomic data**

398   Kraken classification distribution comparison for Illumina (cross-hatched) and nanopore data

399   for each sample, for reads grouped as either CHIKV (blue), DENV (green), other viruses

400   (brown), archea/eukaryota (orange), bacteria (brown) or unclassified (grey).

401
402   **Figure 6. Comparison of genome coverage depth across the CHIKV or DENV genome**

403   **for the different sequencing library preparation methods.**

404   Read depth across both CHIKV and DENV genomes following reference alignment is shown

405   for co-infection sample CHIKV 3, sequenced using for different sequencing library

406   preparation/sequencing methods. Total depth has not been normalised, comparison is to show

407   overall pattern of coverage is highly similar across the methods.

408

409   **Supplementary Figure 1. Proportion of genome covered over the course of each**

410   **sequencing run**

411   The percentage of the CHIKV or DENV genome sequenced plotted over the course of the

412   nanopore sequencing run for each kit version tested.

413
414   **Acknowledgements**

420 **<u>References</u>**

421 <u>1.</u>   Papa A. Emerging arboviral human diseases in Southern Europe. J Med Virol. 2017;89:
422   1315–1322.

423 <u>2.</u>   Weaver SC, Reisen WK. Present and future arboviral threats. Antiviral Res. 2010;85:
424   328–345.

425 <u>3.</u>   Weaver SC, Reisen WK. Present and future arboviral threats. Antiviral Res. 2010;85:
426   328–345.

427 <u>4.</u>   Gould E, Pettersson J, Higgs S, Charrel R, de Lamballerie X. Emerging arboviruses:
428   Why today? One Health. 2017;4: 1–13.

429 <u>5.</u>   Thiberville S-D, Moyen N, Dupuis-Maguiraga L, Nougairede A, Gould EA, Roques P,
430   et al. Chikungunya fever: Epidemiology, clinical syndrome, pathogenesis and therapy.
431   Antiviral Res. 2013;99: 345–370.

432 <u>6.</u>   WHO | WHO publishes list of top emerging diseases likely to cause major epidemics.
433   World Health Organization; 2017; Available: http://www.who.int/medicines/ebola-
434   treatment/WHO-list-of-top-emerging-diseases/en/

435 <u>7.</u>   Burt FJ, Chen W, Miner JJ, Lenschow DJ, Merits A, Schnettler E, et al. Chikungunya
436   virus: an update on the biology and pathogenesis of this emerging pathogen. Lancet
437   Infect Dis. 2017;17: e107–e117.

438 <u>8.</u>   Mavalankar et al D. Increased Mortality Rate Associated with Chikungunya Epidemic,
439   Ahmedabad, India - Volume 14, Number 3—March 2008 - Emerging Infectious Disease
440   journal - CDC. doi:10.3201/eid1403.070720

441 <u>9.</u>   Vos T, Allen C, Arora M, Barber RM, Bhutta ZA, Brown A, et al. Global, regional, and
442   national incidence, prevalence, and years lived with disability for 310 diseases and
443   injuries, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015.
444   Lancet. 2016;388: 1545–1602.

445 <u>10.</u>   Patterson J, Sammon M, Garg M. Dengue, Zika and Chikungunya: Emerging
446   Arboviruses in the New World. West J Emerg Med. 2016;17: 671–679.

447 <u>11.</u>   Burt FJ, Rolph MS, Rulli NE, Mahalingam S, Heise MT. Chikungunya: a re-emerging
448   virus. Lancet. 2012;379: 662–671.

449 <u>12.</u>   Simmons CP, Farrar JJ, Nguyen van VC, Wills B. Dengue. N Engl J Med. 2012;366:
450   1423–1432.

451 <u>13.</u>   Furuya-Kanamori L, Liang S, Milinovich G, Soares Magalhaes RJ, Clements ACA,
452   Hu W, et al. Co-distribution and co-infection of chikungunya and dengue viruses. BMC
453   Infect Dis. 2016;16: 84.

454 <u>14.</u>   Perera-Lecoin M, Luplertlop N, Surasombatpattana P, Liégeois F, Hamel R,
455   Thongrungkiat S, et al. Dengue and Chikungunya Coinfection – The Emergence of an
456   Underestimated Threat. In: Rodriguez-Morales AJ, editor. Current Topics in

457   Chikungunya. InTech; 2016.

458   15.___Omarjee R, Prat CM, Flusin O, Boucau S, Tenebray B, Merle O, et al. Importance of
459   case definition to monitor ongoing outbreak of chikungunya virus on a background of
460   actively circulating dengue virus, St Martin, December 2013 to January 2014.
461   Eurosurveillance. European Centre for Disease Prevention and Control; 2014;19: 20753.

462   16.___Brito CAA, Azevedo F, Cordeiro MT, Marques ETA Jr, Franca RFO. Central and
463   peripheral nervous system involvement caused by Zika and chikungunya coinfection.
464   PLoS Negl Trop Dis. Public Library of Science; 2017;11: e0005583.

465   17.__Wilder-Smith A, Gubler DJ, Weaver SC, Monath TP, Heymann DL, Scott TW.
466   Epidemic arboviral diseases: priorities for research and public health. Lancet Infect Dis.
467   2017;17: e101–e106.

468   18.___Briese T, Paweska JT, McMullan LK, Hutchison SK, Street C, Palacios G, et al.
469   Genetic Detection and Characterization of Lujo Virus, a New Hemorrhagic Fever–
470   Associated Arenavirus from Southern Africa. PLoS Pathog. 2009;5: e1000455.

471   19.__Towner JS, Sealy TK, Khristova ML, Albariño CG, Conlan S, Reeder SA, et al.
472   Newly discovered ebola virus associated with hemorrhagic fever outbreak in Uganda.
473   PLoS Pathog. 2008;4: e1000212.

474   20.___Grard G, Fair JN, Lee D, Slikas E, Steffen I, Muyembe J-J, et al. A novel rhabdovirus
475   associated with acute hemorrhagic fever in central Africa. PLoS Pathog. 2012;8:
476   e1002924.

477   21.__Metsky HC, Matranga CB, Wohl S, Schaffner SF, Freije CA, Winnicki SM, et al.
478   Zika virus evolution and spread in the Americas. Nature. 2017;546: 411–415.

479   22.___Grubaugh ND, Ladner JT, Kraemer MUG, Dudas G, Tan AL, Gangavarapu K, et al.
480   Genomic epidemiology reveals multiple introductions of Zika virus into the United
481   States. Nature. 2017;546: 401–405.

482   23.__Faria NR, Quick J, Claro IM, Thézé J, de Jesus JG, Giovanetti M, et al. Establishment
483   and cryptic transmission of Zika virus in Brazil and the Americas. Nature. 2017;546:
484   406–410.

485   24.___Quick J, Grubaugh ND, Pullan ST, Claro IM, Smith AD, Gangavarapu K, et al.
486   Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus
487   genomes directly from clinical samples. Nat Protoc. 2017;12: 1261–1276.

488   25.__Quick J, Loman NJ, Duraffour S, Simpson JT, Severi E, Cowley L, et al. Real-time,
489   portable genome sequencing for Ebola surveillance. Nature. 2016;530: 228–232.

490   26.___Walter MC, Zwirglmaier K, Vette P, Holowachuk SA, Stoecker K, Genzel GH, et al.
491   MinION as part of a biomedical rapidly deployable laboratory. J Biotechnol. 2017;250:
492   16–22.

493   27.__Brown BL, Watson M, Minot SS, Rivera MC, Franklin RB. MinION^{TM} nanopore
494   sequencing of environmental metagenomes: a synthetic approach. Gigascience. 2017;6:
495   1–10.

496   28.___Batovska J, Lynch SE, Rodoni BC, Sawbridge TI, Cogan NO. Metagenomic
497   arbovirus detection using MinION nanopore sequencing. J Virol Methods. 2017;249:
498   79–84.

499   29. _ Greninger AL, Naccache SN, Federman S, Yu G, Mbala P, Bres V, et al. Rapid
500   metagenomic identification of viral pathogens in clinical samples by real-time nanopore
501   sequencing analysis. Genome Med. 2015;7: 99.

502   30.___Sardi SI, Somasekar S, Naccache SN, Bandeira AC, Tauro LB, Campos GS, et al.
503   Coinfections of Zika and Chikungunya Viruses in Bahia, Brazil, Identified by
504   Metagenomic Next-Generation Sequencing. J Clin Microbiol. 2016;54: 2348–2353.

505   31. _ Rodriguez-Morales AJ, Villamil-Gómez WE, Franco-Paredes C. The arboviral burden
506   of disease caused by co-circulation and co-infection of dengue, chikungunya and Zika in
507   the Americas. Travel Med Infect Dis. 2016;14: 177–179.

508   32.___Cardoso CW, Kikuti M, Prates APPB, Paploski IAD, Tauro LB, Silva MMO, et al.
509   Unrecognized Emergence of Chikungunya Virus during a Zika Virus Outbreak in
510   Salvador, Brazil. PLoS Negl Trop Dis. 2017;11: e0005334.

511   33.___Edwards T, del Carmen Castillo Signor L, Williams C, Donis E, Cuevas LE, Adams
512   ER. Co-infections with Chikungunya and Dengue Viruses, Guatemala, 2015. Emerg
513   Infect Dis. 2016;22: 2003–2005.

514   34.___Drosten C, Göttig S, Schilling S, Asper M, Panning M, Schmitz H, et al. Rapid
515   detection and quantification of RNA of Ebola and Marburg viruses, Lassa virus,
516   Crimean-Congo hemorrhagic fever virus, Rift Valley fever virus, dengue virus, and
517   yellow fever virus by real-time reverse transcription-PCR. J Clin Microbiol. 2002;40:
518   2323–2330.

519   35.___Edwards CJ, Welch SR, Chamberlain J, Hewson R, Tolley H, Cane PA, et al.
520   Molecular diagnosis and analysis of Chikungunya virus. J Clin Virol. 2007;39: 271–275.

521   36. _ Greninger AL, Naccache SN, Federman S, Yu G, Mbala P, Bres V, et al. Rapid
522   metagenomic identification of viral pathogens in clinical samples by real-time nanopore
523   sequencing analysis. Genome Med. 2015;7: 99.

524   37.___Loman NJ, Quinlan AR. Poretools: a toolkit for analyzing nanopore sequence data.
525   Bioinformatics. 2014;30: 3399–3401.

526   38. _ Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler
527   transform. Bioinformatics. 2009;25: 1754–1760.

528   39.___Penedos AR, Myers R, Hadef B, Aladin F, Brown KE. Assessment of the Utility of
529   Whole Genome Sequencing of Measles Virus in the Characterisation of Outbreaks.
530   PLoS One. Public Library of Science; 2015;10: e0143081.

531   40.___Vaser R, Sović I, Nagarajan N, Šikić M. Fast and accurate de novo genome assembly
532   from long uncorrected reads. Genome Res. 2017;27: 737–746.

533   41.___Wood DE, Salzberg SL. Kraken: ultrafast metagenomic sequence classification using
534   exact alignments. Genome Biol. 2014;15: R46.

535   42.   Lewandowski K, Bell A, Miles R, Carne S, Wooldridge D, Manso C, et al. The Effect
536   of Nucleic Acid Extraction Platforms and Sample Storage on the Integrity of Viral RNA
537   for Use in Whole Genome Sequencing. J Mol Diagn. 2017;19: 303–312.

538   43.   Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al.
539   SPAdes: a new genome assembly algorithm and its applications to single-cell
540   sequencing. J Comput Biol. 2012;19: 455–477.

541   44.   Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. Scaffolding pre-assembled
542   contigs using SSPACE. Bioinformatics. 2011;27: 578–579.

543   45.   Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu:
544   scalable and accurate long-read assembly via adaptive k-mer weighting and repeat
545   separation. Genome Res. 2017;27: 722–736.

546   46.   Lipkin WI, Firth C. Viral surveillance and discovery. Curr Opin Virol. 2013;3: 199–
547   204.

548

549

550

551

552

553

554

555

556

557 **Table 1. Summary of samples and Illumina mapping data.**

Sample	Ct Value	Estimated Copy number (/ml)	Sample Type	Total Reads (R1+R2)	% Reads Mapping	% 20x Coverage	% 10x Coverage	Reference Virus	Reference Size (nts)
CHIKV 1	14.72	2.12E+10	Plasma	1113560	78.32%	99.59%	99.72%	Chikungunya	11826
CHIKV 2	20.06	5.49E+08	Serum	1278624	98.48%	99.14%	99.47%	Chikungunya	11826
CHIKV 3	21.41	2.18E+08	Plasma	1391258	95.23%	98.86%	99.37%	Chikungunya	11826
CHIKV 4	22.74	8.76E+07	Plasma	888968	19.16%	97.08%	97.32%	Chikungunya	11826
CHIKV 5	23.77	4.33E+07	Plasma	1357606	97.13%	99.16%	99.58%	Chikungunya	11826
CHIKV 6	25.4	1.42E+07	Serum	3236848	34.88%	97.80%	98.40%	Chikungunya	11826
CHIKV 7	25.76	1.11E+07	Plasma	3748070	72.77%	99.04%	99.56%	Chikungunya	11826
CHIKV 8	28.17	2.13E+06	Plasma	1499952	28.41%	98.69%	99.00%	Chikungunya	11826
CHIKV 9	30.08	5.76E+05	Serum	1035026	6.66%	95.98%	98.22%	Chikungunya	11826
CHIKV 10	30.37	4.72E+05	Serum	1575222	16.84%	97.39%	98.01%	Chikungunya	11826
CHIKV 11	30.66	3.87E+05	Serum	1143054	13.52%	95.36%	96.96%	Chikungunya	11826
CHIKV 12	30.95	3.17E+05	Serum	1507380	10.93%	96.11%	96.52%	Chikungunya	11826
CHIKV 13	32.2	1.35E+05	Serum	1323920	5.03%	88.47%	89.38%	Chikungunya	11826
CHIKV 14	32.57	1.05E+05	Serum	1479404	21.72%	96.32%	96.93%	Chikungunya	11826
DENV 1	16.29	4.21E+09	Plasma	439292	93.44%	99.51%	99.58%	Dengue 1	10735
DENV 2	16.85	2.83E+09	Serum	513472	92.56%	99.40%	99.58%	Dengue 1	10735
DENV 3	17.67	1.58E+09	Plasma	738814	92.53%	99.58%	99.58%	Dengue 2	10723
DENV 4	18.20	1.09E+09	Serum	477368	93.97%	98.73%	99.12%	Dengue 2	10723
DENV 5	19.73	3.67E+08	Serum	915554	89.65%	99.14%	99.40%	Dengue 2	10723
DENV 6	21.22	3.61E+07	Serum	3587926	83.87%	99.68%	99.69%	Dengue 4	10649
DENV 7	23.76	2.11E+07	Serum	4146678	2.17%	86.99%	89.13%	Dengue 1	10735
DENV 8	24.8	1.01E+07	Serum	777264	69.23%	99.56%	99.58%	Dengue 3	10707
DENV 9	25.28	7.17E+06	Plasma	787728	26.97%	98.77%	98.81%	Dengue 2	10723
DENV 10	26.98	2.15E+06	Serum	596240	6.58%	93.47%	93.97%	Dengue 3	10707
DENV 11	28.69	6.39E+05	Serum	1034698	3.73%	94.44%	94.70%	Dengue 1	10735
DENV 12	31.29	1.01E+05	Serum	1374766	0.47%	71.46%	77.76%	Dengue 1	10735

558

559

560

561

562   **Table 2. Summary of samples and Nanopore sequencing.**

Sample	Ct Value	Input (ng)	Sequencing Kit	Flow cell	1D Total bp	1D Total Reads	1D Mean Read Length (nt)	1D Max Read Length (nt)
CHIKV 1	14.7	431.5	SQK-NSK007	MIN105	1.51E+08	267171	564	92712
CHIKV 3	21.4	928.8	SQK-LSK208	MIN106	1.63E+09	1891028	862	99031
CHIKV 4	22.7	113.4	SQK-NSK007	MIN105	1.74E+08	216493	805	125387
CHIKV 9	30.1	212.4	SQK-LSK208	MIN106	2.12E+09	3481358	608	121711
DENV 1	16.3	1626.0	SQK-NSK007	MIN105	2.42E+08	284622	851	115494
DENV 2	16.9	1626.0	SQK-NSK007	MIN105	1.55E+08	203700	760	52157
DENV 6	21.2	475.0	SQK-LSK208	MIN106	1.22E+09	1377721	886	118733
DENV 11	28.7	65.8	SQK-LSK208	MIN106	7.07E+08	1111566	636	119438

563

564   **Table 3. Summary of Nanopore mapping data.**

Sample	Ct Value	Total Reads	% Reads Mapping	% 20x Coverage	20x Genome Length (nt)	% 10x Coverage	Reference	Reference Size (nt)	Max *de Novo* Contig (nt)
CHIKV 1	14.7	267171	65.1%	98.57%	11658	99.2%	Chikungunya	11826	5263
CHIKV 3	21.4	1891028	85.1%	99.76%	11798	99.9%	Chikungunya	11826	10793
CHIKV 4	22.7	216493	11.6%	94.11%	11130	97.2%	Chikungunya	11826	4256
CHIKV 9	30.08	3481358	4.08%	100%	11826	100%	Chikungunya	11826	9860
DENV 1	16.3	284622	71.3%	99.9%	10719	99.9%	Dengue 1	10735	8281
DENV 2	16.9	203700	72.1%	99.6%	10692	99.6%	Dengue 1	10735	10157
DENV 6	21.2	1377721	71.1%	99.9%	10634	99.9%	Dengue 4	10649	7877
DENV 11	28.7	1111566	2.9%	95.3%	10226	96.3%	Dengue 1	10735	4699

565

566

567

568

569

570

571   **Table 4. Comparison of Nanopore mapping data across library kits**.

Platform	Kit Information	Virus Identified	Total Reads (nt)	% Reads Mapping	% 20x Coverage	% 10x Coverage	Reference	Reference Size (nt)	Max *de Novo* Contig (nt)
Illumina	Nextera XT	Chikungunya	1391258	95.23%	98.86%	99.37%	CHIKV	11826	7321
Illumina	Nextera XT	Dengue	1391258	0.22%	63.66%	77.82%	DENV1	10735	6613
MinION 2D	SQK-LSK208	Chikungunya	1891028	85.12%	99.73%	99.91%	CHIKV	11826	10793
MinION 2D	SQK-LSK208	Dengue	1891028	0.43%	95.99%	96.09%	DENV1	10735	7549
MinION 1D^2	SQK-LSK308	Chikungunya	5080906	74.50%	99.94%	100%	CHIKV	11826	11369
MinION 1D^2	SQK-LSK308	Dengue	5080906	0.37%	95.04%	96.42%	DENV1	10735	2199
MinION Rapid	SQK-RBK001	Chikungunya	611110	66.26%	99.66%	99.68%	CHIKV	11826	11473
MinION Rapid	SQK-RBK001	Dengue	611110	0.29%	81.09%	90.83%	DENV1	10735	4227

572

573

574

575

Figure 1

PCR Positive Cases - 2016

Figure 2

Figure 3

Chikungunya Virus Samples

Dengue Virus Sampes

Figure 4

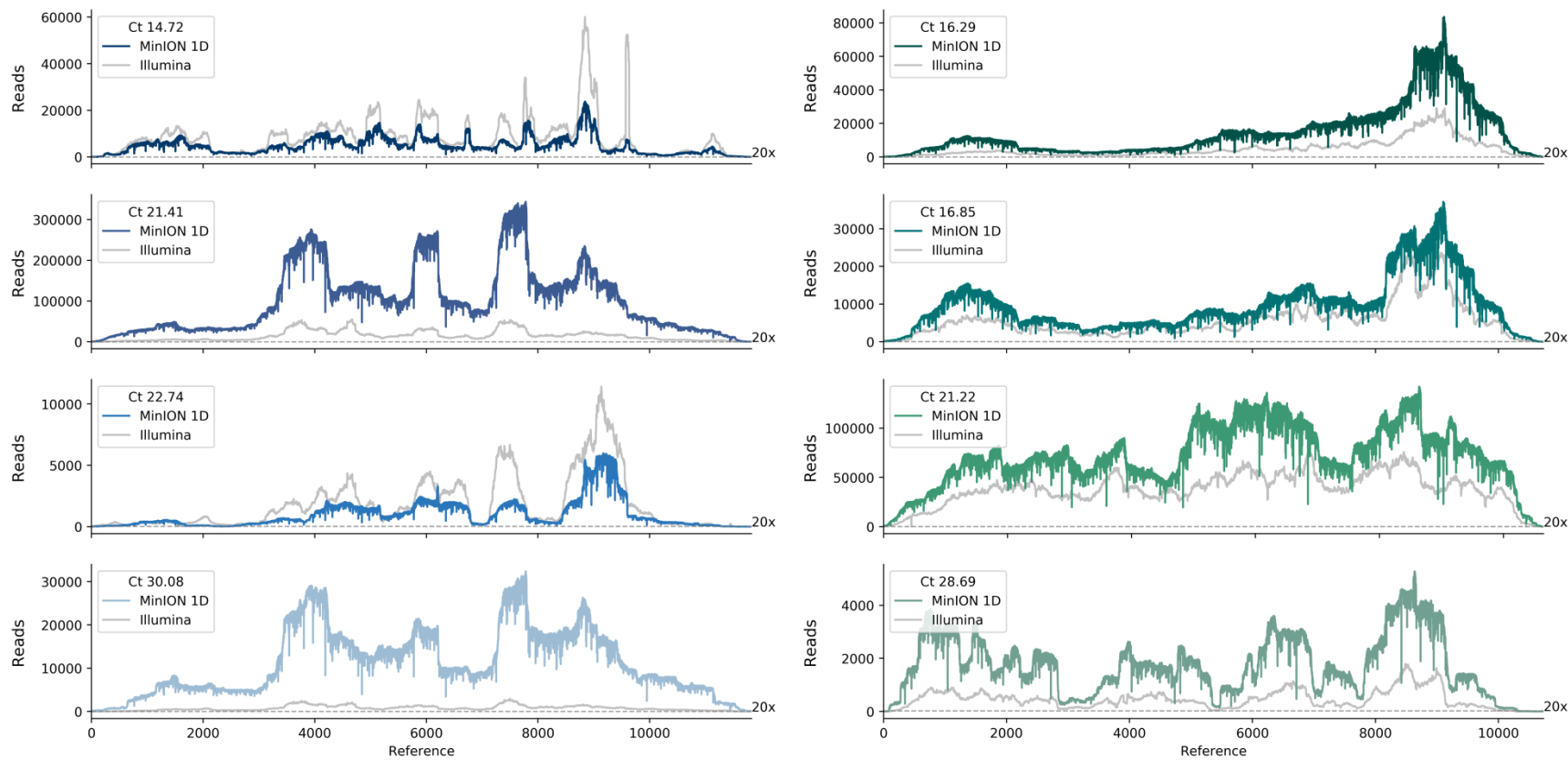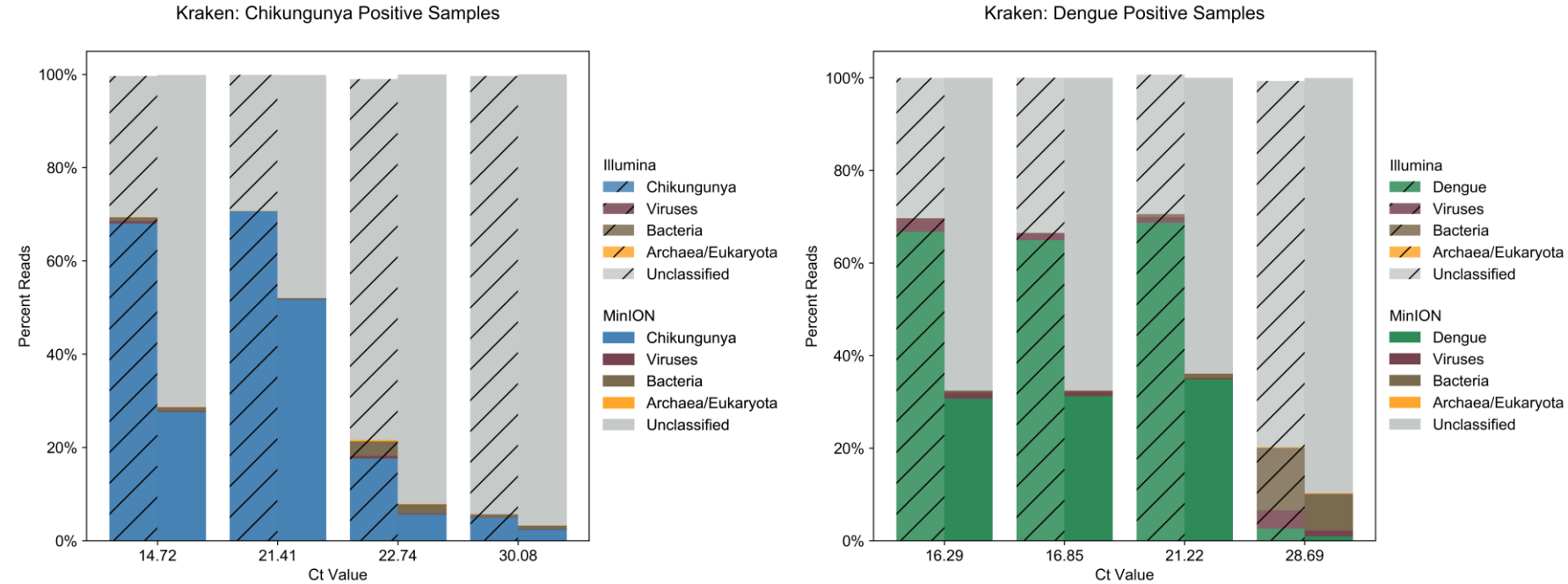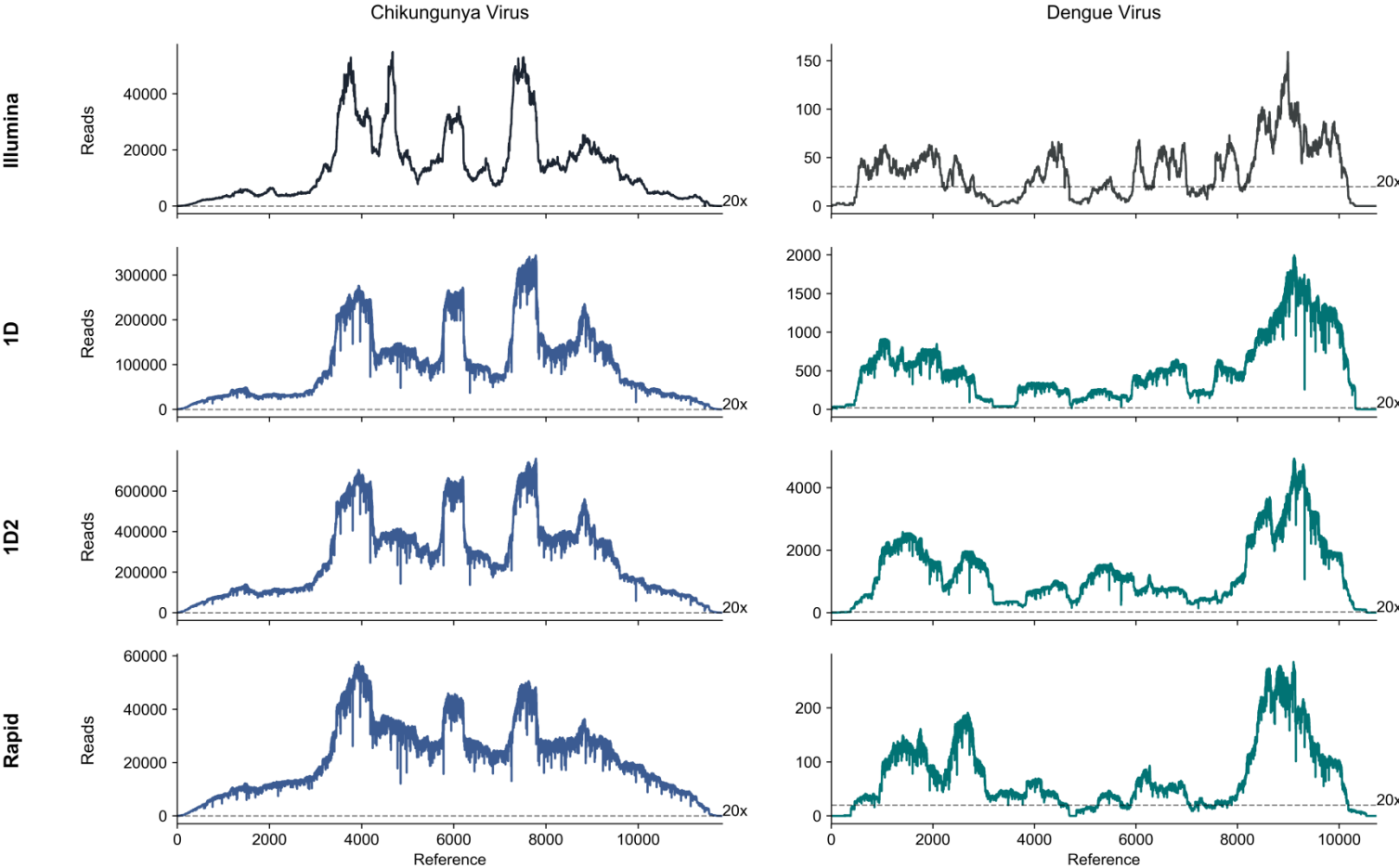# Figure 5

Kraken: Chikungunya Positive Samples

Kraken: Dengue Positive Samples
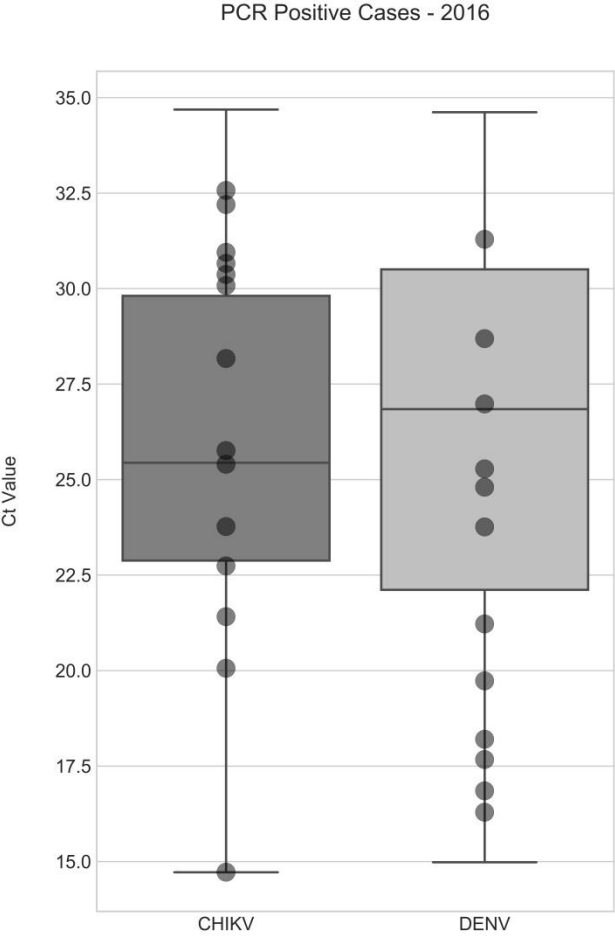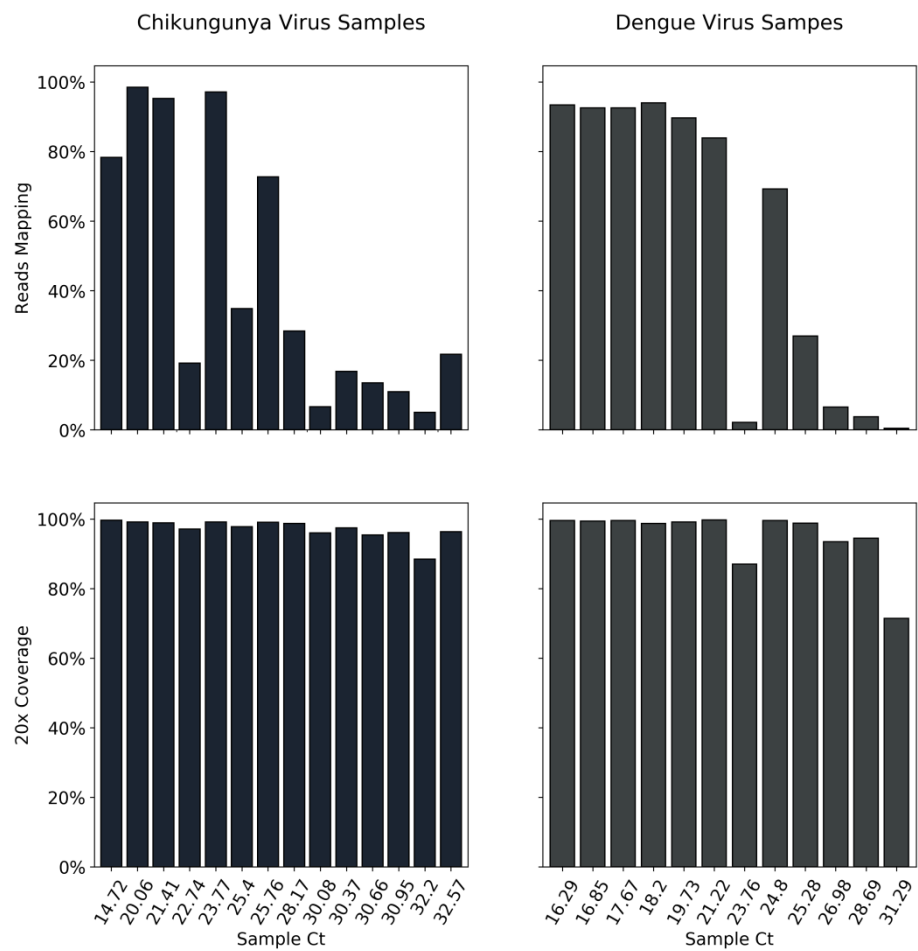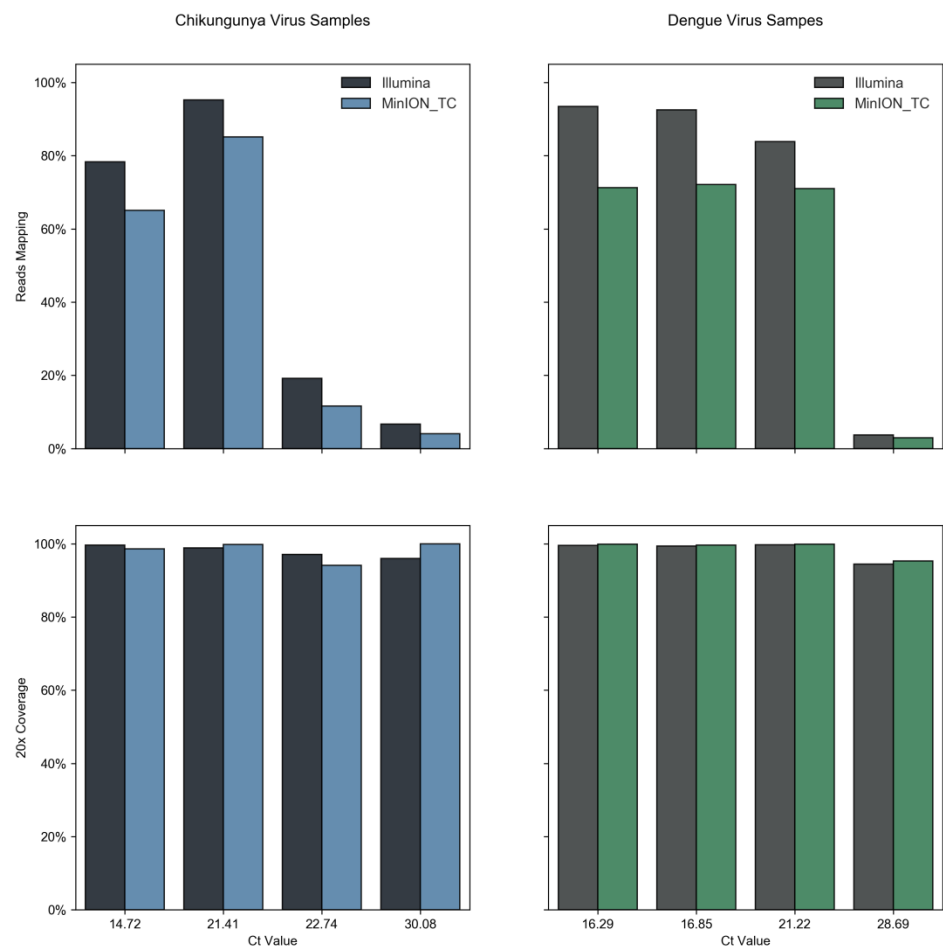
Figure 6

Supplementary
Figure 1