1 Journal: https://www.mdpi.com/journal/genes/special_issues/metagenomics_insitu
2
3
4 **Tree Lab: Portable genomics for early detection of plant viruses and pests in Sub-Saharan Africa**
5
6
7

8 Laura M. Boykin[1], Peter Sseruwagi[2], Titus Alicai[3], Elijah Ateka[4], Ibrahim Umar Mohammed[5], Jo-Ann L.
9 Stanton[6], Charles Kayuki[2], Deogratius Mark[2], Tarcisius Fute[2], Joel Erasto[2], Hilda Bachwenkizi[2],
10 Brenda Muga[4], Naomi Mumo[4], Jenniffer Mwangi[4], Phillip Abidrabo[3], Geoffrey Okao-Okuja[3], Geresemu
11 Omuut[3], Jacinta Akol[3], Hellen B. Apio[3], Francis Osingada[3], Monica A. Kehoe[7], David Eccles[8], Anders
12 Savill[1], Stephen Lamb[1], Tonny Kinene[1], Christopher B. Rawle[6], Abishek Muralidhar[9], Kirsty Mayall[9], Fred
13 Tairo[2], and Joseph Ndunguru[2]
14
15
16
17
18 [1] School of Molecular Sciences and Australian Research Council Centre of Excellence in Plant Energy
19 Biology, University of Western Australia, Crawley, Perth, WA 6009 Australia
20 [2] Mikocheni Agricultural Research Institute (MARI), P.O. Box 6226, Dar es Salaam, Tanzania
21 [3] National Crops Resources Research Institute (NaCRRI), P.O. Box 7084, Kampala, Uganda
22 [4] Jomo Kenyatta University of Agriculture and Technology (JKUAT), P.O. Box 62000 - 00200,
23 Nairobi, Kenya
24 [5] Kebbi State University of Science and Technology, Aliero, Nigeria
25 [6] Department of Anatomy, University of Otago, P.O. Box 56, Dunedin 9054, New Zealand
26 [7] Department of Primary Industries and Regional Development Diagnostic Laboratory Services, Plant
27 Pathology, South Perth WA 6151, Australia
28 [8] Malaghan Institute of Medical Research, PO Box 7060, Newtown, Wellington 6242, New Zealand
29 [9] MicroGEM Ltd, 9 Melody Ln, Ruakura, Hamilton 3216, New Zealand
30
31
32

Abstract

In this case study we successfully teamed the PDQeX DNA purification technology developed by MicroGEM, New Zealand, with the MinION and MinIT mobile sequencing devices developed by Oxford Nanopore Technologies to produce an effective point-of-need field diagnostic system. The PDQeX extracts DNA using a cocktail of thermophilic proteinases and cell wall degrading enzymes, thermo-responsive extractor cartridges and a temperature control unit. This single-step closed system delivers purified DNA with no cross contamination. The MinIT is a newly released data processing unit that converts MinION raw signal output into base called data locally in real time, removing the need for high specification computers and large file transfers from the field. All three devices are battery powered with an exceptionally small footprint that facilitates transport and set up.

To evaluate and validate capability of the system for unbiased pathogen identification by real-time sequencing in a farmer's field setting, we analysed samples collected from cassava plants grown by subsistence farmers in three sub-Sahara African countries (Tanzania, Uganda and Kenya). A range of viral pathogens, all with similar symptoms, greatly reduce yield or completely destroy cassava crops. 800 million people worldwide depend on cassava for food and yearly income, and viral diseases are a significant constraint on its production (https://cassavavirusactionproject.com). Early pathogen detection at a molecular level has great potential to rescue crops within a single growing season by providing results that inform decisions on disease management, use of appropriate virus resistant or replacement planting.

This case study presented conditions of working in-field with limited or no access to mains power, laboratory infrastructure, internet connectivity and highly variable ambient temperature. An additional challenge is that, generally, plant material contains inhibitors of downstream molecular processes

55   making effective DNA purification critical. We successfully undertook real-time on-farm genome

56   sequencing of samples collected from cassava plants on three farms, one in each country. Cassava

57   mosaic begomoviruses were detected by sequencing leaf, stem, tuber and insect samples. The entire

58   process, from arrival on farm to diagnosis including sample collection, processing and provisional

59   sequencing results was complete in under 4 hours. The need for accurate, rapid and on-site diagnosis

60   grows as globalized human activity accelerates. This technical breakthrough has applications that are

61   relevant to human and animal health, environmental management and conservation.

62

63   **Keywords:** cassava, cassava mosaic begomovirus, cassava mosaic disease, *Bemisia tabaci*, whitefly,

64   MinION, MinIT, PDQeX, Tanzania, Uganda, Kenya.

65

66   **Introduction**

67   Crop losses due to viral diseases and pests are major constraints on food security and income for millions

68   of households in sub-Saharan Africa (SSA). Such losses can be reduced if plant diseases and pests are

69   correctly diagnosed and identified early. To date, researchers have utilized conventional methods for

70   definitive identification of plant viruses and their vectors in SSA including PCR, qPCR, Next Generation

71   and Sanger sequencing, but these require laboratory infrastructure, are costly and time consuming, and

72   can delay time-sensitive corrective actions that could be taken. Direct rapid DNA/RNA sequencing of

73   infected material on-the-spot or near sample collection sites turns this conventional paradigm on its

74   head by taking the laboratory closer to farmers' fields. This reduces overall costs and gives crop

75   protection officers and farmers in rural communities' information that is critical for sustainable crop

76   production and management of pests and diseases, ensuring food and income security for millions of

77   Africans. Currently, provision of data on viruses which is essential for developing virus resistant varieties,

78   sharing virus-indexed germplasm between regions and deployment of virus-free certified planting

79   materials is hampered by the long time taken to receive results generated using the aforementioned

80   conventional diagnostic methods. Our innovation will simplify information flow and fast track the

81   deployments of virus resistant or tolerant cassava varieties directly to the farmers field. The emergence

82   of new tools for real-time diagnostics, such as the Oxford Nanopore MinION, has proved useful for the

83   early detection of Ebola (Quick et al. 2016) and Zika viruses (Faria et al. 2016, Quick et al. 2017). MinION

84   consensus sequence accuracy of 99% is sufficient to identify pathogen and strain type (Calus et al. 2018).

85   However, it can take months before results generated using other high throughput sequencing

86   approaches (e.g. Illumina, PacBio) are available, particularly when local scientists are reliant on third-

87   party service providers, who are often located in other countries. The delay in detecting or identifying

88   viruses impedes quick in-situ decision making necessary for early action, crop protection advice and

89   disease management strategies by farmers. This ultimately compounds the magnitude of crop losses and

90   food shortages suffered by farmers. We have decreased the time to precisely detect and identify

91   pathogens, vectors or pests, and increased resolution and reliability of results by utilizing the power of

92   low-cost portable DNA extraction, sequencing and data analysis devices, coupled with our innovative

93   data analysis pipelines. This real-time diagnosis in the field or located in regional laboratories quickly

94   provides high quality and reliable diagnostics data to help farmers, seed certification agencies, scientists,

95   crop protection and extension officers make timely and informed decisions. The immediate data

96   accessibility makes possible dissemination of results downstream to extension officers and farmers for

97   early disease control action via Information and Communication Technologies (ICT) applications. The

4

98   application of cutting-edge sequencing technology, genomics and bioinformatics for pest and disease

99   control has great potential to improve food security and agricultural development at large.

100

101   We propose using this technology to rapidly diagnose plant viruses and pests affecting smallholder

102   farmers' crops in SSA. Our case study has identified cassava DNA viruses on the farm allowing farmers,

103   researchers and development actors to take early, positive corrective action based on rapid diagnosis of

104   plants. This proof-of-concept shows that portable DNA sequencer technology has great potential to

105   reduce the risk of community crop failure. We have previously conducted pilot projects in Tanzania,

106   Uganda and Kenya testing symptomatic and asymptomatic cassava plants and already have shown that

107   sample collection to diagnosis and results delivered to the farmer or crop protection officer can be

108   completed within 48 hours (Boykin et al. 2018). This technology will put the power of genome

109   sequencing directly in the hands of agriculturalists and, in the work presented here, for the first time has

110   enabled pest and disease diagnosis within one day on-the-spot. This has significant implications for new

111   pest and disease outbreaks, monitoring of existing disease outbreaks and biosecurity monitoring at

112   borders between countries.

113

114   **Materials and Methods**:

115   **Tree Lab locations:** Three small scale family farms growing cassava were selected, one in each of the

116   following counties: Tanzania, Uganda and Kenya. In Tanzania (Kisamwene, Mara Region. GPS: 315N,

117   1°40'5"S, 33°55'55"E, 4380ft) on 1 August 2018, in Uganda (Wakiso. GPS: 255W 0°30'29"N, 32°37'19"E,

118   3730 ft) on 8 August 2018, and in Kenya (Kiambu. GPS: 87E 1°5'33" S, 37°19'33"E, 4570 ft) on 14 August

119   2018. A video of the Tanzanian Tree Lab is found here: https://vimeo.com/329068227.

120

121 **Sample selection:** All samples collected are documented in Table 1. For each sequencing run we

122 barcoded either 11, including DNA extractions from cassava leaves, stems and also whiteflies (*Bemisia*

123 *tabaci*) that were found feeding on cassava mosaic disease symptomatic cassava leaves.

124

125 **DNA extraction:** The PDQeX DNA extraction system from MicroGEM NZ Ltd (Stanton et al. 2019) was

126 used to prepare DNA from samples. Briefly, a Harris punch was used to collect four discs, 2 mm in

127 diameter, from each leaf, stem or root sample. Homogenization was performed by hand in 1x GREEN

128 plus buffer using a Dounce homogenizer made from sealing the end of a 1 ml pipette tip and a 1.5 ml

129 microfuge tube. Ninety microlitres of each homogenate and 10 µl of Enhancer (MicroGEM Ltd) was

130 added to a 200 µl tube containing a lyophilized 1x mix of the enzyme cocktail (Holmes et al. 2018);

131 (*phyto*GEM kit, MicroGEM Ltd, New Zealand). The reaction was re-suspended by gently flicking the 200

132 µl tube until all reagents were well mixed. All of the reaction mix was transferred to a PDQeX extraction

133 cartridge (Stanton et al. 2019) which was placed into the PDQeX1600 thermal incubation unit. PDQeX

134 extraction was performed by a series of heating steps. First, incubation at 52°C for five minutes to

135 promote cell lysis by activating cell wall degrading enzymes. Second, incubation at 75°C for five minutes

136 to activate thermophilic proteinases to degrade sample proteins and enzymes from the previous step.

137 Finally, heating to 95°C for 2 minutes to shrink the thermal responsive inner layer of the PDQeX

138 extraction cartridge forcing the digested sample through a burst valve and a cleanup column into a

139 collection tube (Stanton et al. 2019).

140

141   DNA was also extracted from whitefly. A single insect was fished from a pool of whiteflies in ethanol

142   collected from leaves using a Pooter. The whitefly was transferred by pipette, taking as little ethanol as

143   possible, to 98 µl of 1x BLUE buffer (MicroGEM Ltd) and pipetted up and down several times. The whole

144   mix was added to 1x lyophilized enzyme cocktail in a 200 µl tube (*prep*GEM, MicroGEM Ltd). Reagents

145   were re-suspended by gentle flicking and the contents transferred to a PDQeX extraction cartridge. The

146   cartridge was placed in the PDQeX1600 thermal unit and heated as follows: 35°C for five minutes; 52°C

147   for five minutes; 75°C for five minutes; 95°C for 2 minutes. DNA extraction took approximately 20

148   minutes in total and 7.5 µl of the collected elute was used directly for Rapid DNA library construction for

149   MinION Sequencing. The PDQeX1600 thermal unit was powered by a 12-volt Lithium Polymer battery.

150   The PDQeX1600 was operated using a purpose-made App from a Smart Phone that permitted run

151   initiation, temperature profile selection and editing, and monitoring of run progress.

152

153   **Library preparation and sequencing:** We utilized the Rapid Barcoding kit SQK-RBK004 with 9.4.1 flow

154   cells (Oxford Nanopore Technologies). The SQK-RBK004 protocols were performed as described by the

155   manufacturer (RBK_9054_v2_revB_23Jan2018). We completed the optional clean-up steps using

156   AMPure XP beads. The 30°C and 80°C steps were performed using the PDQeX1600 thermal incubation

157   unit. All libraries were loaded directly onto the MinION that was connected to a MinIT and live base

158   calling was enabled. For each Tree Lab the MinION and the MinIT were plugged into a 20000mAh laptop

159   powerbank (Comsol) set at 20V (Figure 1). The key to using a power bank for this purpose is to make

160   sure it not only has USB inputs but also has a DC port. It ran on average 4.5 hours set on 16.5V. When

161   the run stopped, we immediately plugged the devices into a second power bank and data generation

162   continued.

163

164   **Tree Lab data Analyses:** As the data was being basecalled on the MinIT we made a test folder on the

165   laptop called "treelab" and inside that folder we added a demultiplex folder, into which we then

166   transferred the first two .fastq files from the MinIT into. Demultiplexing was run with Porechop (Wick

167   2019),   preinstalled   on   the   laptop,   using   the   following   commands   >porechop   -i

168   /Users/lboykin/Desktop/treelab -b /Users/lboykin/Desktop/treelab/demultiplex.   A cassava mosaic

169   disease (CMD) reference data set was pre-curated and configured to work as a local database within

170   Geneious   vR11.1.2   (Kearse   et   al.   2012)

171   (https://figshare.com/articles/Nanopore_sequencing_of_cassava_from_Tanzania_Uganda_and_Kenya

172   /6667409). Twelve folders were created in Geneious, and the associated .fastq files from the

173   "demultiplex" folder were drag and dropped into the relevant folder created within Geneious. BLASTn

174   (Altschul et al. 1990) analysis was performed, ensuring the local CMD database was specified. The results

175   from the search against the CMD database were visualized in-situ within Geneious and discussed with

176   farmers and extension workers.

177

178   **Post Tree Lab data analyses:** Scripts from David Eccles' [Bioinformatics Scripts repository]

179   (https://doi.org/10.5281/zenodo.596663) were used to carry out subsequent read QC and analysis.

180   Sequenced read lengths were measured using [fastx-length.pl], and these lengths were used to generate

181   length-based QC plots using [length_plot.r].

182

183   **Assembly:** To determine whether any barcoded read sets could be assembled, an initial assembly

184   attempt was made on each subset using Canu v1.8, with a genome size of 400M, ignoring any warning

8

185 messages about coverage being too low. The large genome size ensured that no reads are discarded and

186 suppressing the coverage warning ensured that Canu would attempt an assembly with all the available

187 reads. Previous discussions with Canu developer Sergey Koren (pers. comm.) indicated that adjusting the

188 target genome size had no other effect on the contig assemblies that Canu produces.

189

190 **Blast**: We confirmed these in-field results by performing a post diagnostic blast of reads on the Nimbus

191 Cloud at the Pawsey Supercomputing Center with blast 2.2.31 against the full NCBI nucleotide database

192 to confirm results. For reference the specific database used was {$ blastcmd –db nt/nt –info} Database:

193 Nucleotide collection 49,266,009 sequences; 188,943,333,900 total bases Date: Aug 8, 2018 12:38 PM.

194 The data were processed into a blast archive using a blast script with the following parameters (Script

195 attached) {$blastn -query "$file" -db /mnt/nucdb/nt/nt -outfmt 11 -culling_limit 10 -out "out.$file.asn"

196 -num_threads 17 } then converted into XML (for loading into Geneious) and HTML for viewing.

197

198 **Blastn analysis – MEGAN:**

199 Blastn results produced from the Nimbus cloud analysis pipeline were also visualized using MEGAN

200 Community Edition version 6.12.6 (Huson et al. 2016) on the Zeus computing resource located at the

201 Pawsey Supercomputing Center.

202

203 **Blastn analysis - Kraken:**

204 We used kraken2 [https://github.com/DerrickWood/kraken2] to classify demultiplexed reads using the

205 Loman   Lab   "maxikraken2"   database

206   [https://lomanlab.github.io/mockcommunity/mc_databases.html#maxikraken2_1903_140gb-march-

207   2019-140gb], on the Zeus computing resource located at the Pawsey Supercomputing Center.

208

209   **Results**

210

211   <u>**Tree Lab**</u>

212   All essential equipment that were used in the Tree Labs in Kenya, Uganda and Tanzania are listed in Table

213   1 and shown in Figure 1. Summary statistics for our three Tree Labs are shown in Table 2 and Figure 2.

214   Table 2 summarizes DNA sequencing metrics from all three Tree Lab experiments. Each MinION run

215   contained 11 barcoded libraries representing 11 individual samples. All DNA samples, except the ACMV

216   and EACMV controls (Tanzania) were extracted using the PDQeX system with 11 samples prepared in lab

217   from exemplar material collected from scientific plots and 20 DNA samples extracted on farm. A total of

218   1,442,599 sequences were produced across all the experiments. Of these, barcodes could only be

219   resolved for 550,938 sequences using Porechop to demultiplex the samples. Mean sequence length

220   across all sequencing runs ranged from 355bp to 948bp with the longest read being 276,793 bp.

221

222   Raw reads of Cassava mosaic begomoviruses (CMBs) sequences were detected in 21 samples with the

223   longest CMB read reaching 2808 bp, close to the full genome size. A total of 18 leaf samples were

224   sequenced of which 15 were found to contain CMBs. Two of the 5 stem samples sequenced were found

225   to contain CMBs whereas neither of the two root samples sequenced presented CMB sequences. Six

226   single whiteflies were tested with 2 being positive for CMBs. All libraries, regardless of CMB content

227   produced DNA reads indicating that sequencing was successful for all samples. CMBs were detected in

10

228   plants with symptoms and there was a suggestion that the number of CMB sequences detected possibly

229   correlated with symptom severity scores, but more data will be required to prove this. Interestingly, a

230   known healthy plant taken from the scientific plot at JKUAT did not yield CMB sequences (Table 2).

231   Following assembly with Canu 8 of the 21 samples gave complete assembled virus genomes, however,

232   gave less than 10 fold coverage.

233

234   **Post Tree Lab data analyses**

235   We investigated whether there was any effect of sample type on read length. The cumulative density

236   curves (Figure 2) show the proportion of sequenced bases with length greater than a particular length

237   (with L10/L50/L90 highlighted). Additional length-based QC plots can be found in the supplemental

238   information (Supplemental File 1).

239

240   **MEGAN results**

241   The primary targets of this analysis were known cassava viruses, as well as the host, either cassava plant

242   (*Manihot esculenta*) or the whitefly (*Bemisia tabaci*) and its endosymbionts. Results are summarized in

243   Table 3, and in general the desired result of virus (EACMV or ACMV) and host DNA were recovered from

244   all symptomatic samples.

245

246   **Kraken2 results**

247   The analysis using Kraken2 had an approximately 50% classification success rate (IQR 45-52% unclassified

248   reads). This database is for human + microbial and viral sequence, so any eukaryote reads (e.g. from

249   cassava or whitefly) would probably be unclassified by Kraken2 or assigned to the human taxa. The

250   sample with the highest classification success was the ACMV positive control from Tanzania (mr_BC11,

251   5.6% unclassified), and the lowest classification success was the leaf tissue sample #2 from Kenya

252   (mm_BC02, 61% unclassified).

253

254   Results were aggregated into a table using Pavian (Supplemental File 2) to identify common elements of

255   each sample. Begomovirus reads were detected in 15 samples, with very high proportions of

256   Begomovirus (8.6%) in the ACMV Positive control from Tanzania (mr_BC11), and above-average

257   proportions (0.25%) in Kwatempale sample #5 from Sarah's Farm in Uganda (ut_BC05). ACMV and

258   EACMV were detected in 11 samples.

259

260   **Discussion**

261   This case study was designed to show the possibility to go from sample to diagnosis, in a regional setting,

262   on farm in three hours versus the normal 6 months with conventional methods. The results of this

263   research show that it is indeed possible, and that it is possible to use a range of battery powered devices

264   to achieve DNA extraction, long read sequencing and analysis all under a tree on the farm while the

265   farmers wait for results.

266

267   Access to next generation sequencing technology, or to services that offer access has been a major

268   barrier to their use in diagnostics for scientists, and particularly many agricultural scientists in SSA. The

269   advent of the Oxford Nanopore Technologies MinION has brought this technology to their door in recent

270   years, and with access to training through various institutions and especially the Oxford Nanopore

271   Technologies run "Pore Safari" there are more and more users in the region. Previous studies that have

12

272   used the technology for real time analysis of pathogen outbreaks, such as the Ebola and Zika studies

273   (Faria et al. 2016, Quick et al. 2016, Quick et al. 2017) have still relied heavily on the transport of bulky

274   laboratory equipment, or local acquisition of it to perform their work. Previous work by our team of

275   scientists (Boykin et al. 2018) showed that the turnaround time to result could be 48 hours, and now

276   with the addition of the PDQeX and MinIT to the system we have been able to reduce the time to 4 hours

277   and perform the entire process in the field and under a tree.

278

279   One of the major barriers to producing these outputs in the field has until now been the lack of a simple,

280   quick and effective methods to extract DNA from a sample without the need for laboratory equipment

281   requiring mains power and space, items such as benchtop centrifuges, fridges, freezers and temperature

282   sensitive extraction kits which can be bulky and rely on traditional laboratory infrastructure. The use of

283   the PDQeX in the system described in this case study was the real game changer: compact and able to

284   operate from a battery, it made nucleic acid extraction possible.

285

286   This study also highlighted where the next gains for in-field sequencing are to be made, as improvements

287   are required in rapid data analysis. The MinIT eliminated obstacles to base calling, by converting the raw

288   reads into .fast5 and .fastq reads in real time. This moves the data analysis bottleneck in the pipeline to

289   the Blast analysis. Blast is not fast analysis, and so for now we must rely on a pre-curated database of

290   known or expected pathogens and host genomes. This poses risks, in that new and emerging pathogens

291   or vectors could be missed in the first instance, and not seen until subsequent data analysis when the

292   scientist has returned to the lab or is within range of a good internet connection capable of uploading

293   large amounts of data to the cloud. In our case, we can predict what sorts of genomes should be in our

13

294   custom database, but for use in biosecurity, and at borders between countries a better solution is

295   required.

296

297   Read length distributions were generally quite similar for all samples. The Tanzania samples showed the

298   greatest difference in read length distribution (L50 range 400 - 2000 bp). The ACMV Positive control from

299   Tanzania (mr_BC11) showed a very pronounced spike of reads at around 4kb (presumably near full-

300   length ACMV sequence). Apart from that sample, there was no obvious association between read length

301   distribution and tissue type or variety. The Uganda samples had a moderate read length distribution

302   spread (L50 range 500-1000 bp), which split into two clusters of slightly shorter and longer reads (BC02,

303   BC04, BC05, BC06, BC07; BC01, BC03, BC08, BC12). These clusters did not appear to have any relationship

304   with tissue type or variety. The Kenya samples had a similarly moderate read length distribution spread

305   (L50 range 500-1000 bp), with no obvious clustering, or association of distribution with tissue type or

306   variety.

307

308   MinION Rapid libraries use transposase to fix sequencing adaptors to DNA fragments. The ratio of DNA

309   to transposase complex for the MinION Rapid kit has been optimized for 400 ng DNA and at lower

310   amounts DNA is susceptible to over fragmentation. This may account for DNA fragment length falling

311   around 900 bp, however, the control DNA also gave similar read length characteristics. Though there

312   was not enough data collected on farm for a thorough statistical analysis these results did show both

313   yield and integrity of the DNA extracted using the PDQeX was of sufficient quality for diagnostic

314   sequencing. We successfully retrieved enough data from each sample to establish whether the virus in

315   the plant was EACMV or ACMV. Assembly with Canu suggested that in this case, while there was enough

316　data to assemble whole genomes, the average coverage meant the quality was not sufficient for

317　downstream applications such as recombination detection and other evolutionary analyses. We

318　anticipate that as the DNA extraction methods improve, and in field library preparation becomes easier

319　this will be possible. An alternative would be to investigate the use of a panel-like targeted amplicon

320　approach or CRISPR/Cas9 enrichment, but again this removes the likelihood of detecting unknowns in

321　the samples, and could lead to samples giving negative results not being followed up, or the time to

322　result being blown out to days or weeks if they need to return to a laboratory to complete a different

323　type of library preparation.

324

325　Compared with other in-field diagnostic tools, this system involving the MinION is unique in its ability to

326　detect anything that might be present in the sample. Other in-field diagnostic tools, including serological

327　based dipsticks, LAMP-PCR, in field qPCR and AI driven applications on smart phones all have one single

328　thing in common – they require a prior knowledge of the suspected pathogen, coupled with targeted

329　design of antibodies, primers or training for known positives to function effectively. The only decision

330　required to run the MinION is whether to prepare a DNA or an RNA library.

331　**Executive summary:**
332
333　**Can we go from sample to answer on the farm?** Yes
334　**DNA extraction to library prep to sequencing?** Yes
335　**Can we detect virus in leaves off the grid at the farm?** Yes
336　**Can we detect virus in whiteflies off the grid on the farm?** Yes
337　**Can we detect virus in stems off the grid on the farm?** Yes
338　**Do we get enough to coverage of the viral genomes to generate polished genomes to track the**
339　**evolution of the viruses real-time?** No
340
341
342　**Video of Tree Lab:** https://vimeo.com/329068227
343
344

15

352
353
354
355
356

357
358   **Table 1.** Essential equipment for Tree Lab in Tanzania, Kenya and Uganda.
359
360   Sample collection plants
361   Envelopes
362   4 fine sharpies
363   Notebook
364   Scissors
365   Canvas sling bag to carry sample collection equipment
366   Gloves
367
368   Sample collection whiteflies
369   Pooter
370   70% ETOH
371   Eppendorf tubes
372   Transfer pipette to get the whitefly from the pooter to the Eppendorf tube
373
374   Lab under a tree/disease diagnostic camp
375   Blanket/floor covering
376   Hard flat surface- raised
377
378   Sample homogenization
379   1 microfuge tube/sample
380   1 ml tips with fine tip sealed to form pestle for crushing
381   Tip disposal/waste bags
382   1.5ml racks for holding
383   p10, p1000, p200 pipette (one set of pipettes)- tips for all 1 box
384
385   DNA extraction
386   2 MicroGEM kits (transported at ambient temperature)
387   PDQeX phytoGEM kit for plants
388   PDQex Universal kit for insects
389   Reagents lyophilized into 200μl tubes
390   Rack for 200μl microcentrifuge tubes
391   Nuclease free water
392   PDQeX thermal device
393   PDQeX Extraction tubes
394   Mobile phone being used as a remote controller- MicroGEM PDQeX app
395   Battery 2 12V batteries put together
396
397   Save the DNA for later quantification using fluorometry (e.g Qubit).
398
399   Library prep
400   Printed library prep protocol or iPad

17

401	Library kit
402	200µl tubes
403	1.5ml tubes
404	Ampure beads
405	Magnetic rack
406	Tube stands for the 0.2ml
407	Tris and sodium chloride
408	Nuclease free water
409	Syrafoam cooler and cold packs
410	Flow cell
411	MinIT
412	Laptop Power bank – one with a pin plug not just USB output- console
413	Laptop- with MinKnow installed
414	
415	
416	
417	
418	
419	
420	

Table 2. Summary statistics and locality information for the three Tree Labs in Tanzania, Kenya and Uganda. * indicates DNA extraction carried out using PDQeX in the laboratory before sequencing under the tree.

Sample	Barcode	Variety	Severity Score	Tissue type	Total reads	Max. Seq Length	CMV Blast hits	Max. CMV length	Min CMV length	% CMV reads	Canu Contigs
Tanzania											
1	1	Kilati - Local Variety	4	Leaf	61,317.0	26,107.0	125.0	2,808.0	196.0	0.204	0
3	3	Kilati - Local Variety	4	Leaf	21,468.0	16,098.0	14.0	815.0	288.0	0.065	0
4	4	Kasuxsali - Local Variety	3	Leaf	31,300.0	17,624.0	21.0	941.0	129.0	0.067	12
5	5	Mukombozi - Virus Resistant	1	Leaf	2,117.0	24,665.0	7.0	815.0	198.0	0.331	0
6	6	Mukombozi - Virus Resistant	1	Leaf	27,178.0	12,340.0	8.0	826.0	85.0	0.029	0
7	7	Mukombozi - Virus Resistant	1	Leaf	5,634.0	17,003.0	5.0	484.0	243.0	0.089	0
8	8	Whitefly close to Mukombozi - Virus Resistant	1	1x Whitefly, in EtOH	6,237.0	8,753.0	26.0	828.0	52.0	0.417	0
9	9	Whitefly close to Mukombozi - Virus Resistant	1	1x Whitefly, in EtOH	25,289.0	17,900.0	3.0	0.0	0.0	0	2
10	10	Whitefly close to Mukombozi - Virus Resistant	1	1x Whitefly, in EtOH	798.0	23,259.0	21.0	815.0	187.0	2.632	0
11*	11	ACMV - Positive Control DNA			10,966.0	28,541.0	4,311.0	1,598.0	31.0	39.312	0
12*	12	EACMV - Positive Control DNA			1,797.0	22,871.0	9.0	830.0	191.0	0.501	0
None	None	Porechop unable to match to Rapid Barcode			449,981.0	276,793.0	8,136.0	2,757.0	28.0	1.808	
Uganda											
1*	1	R39-B1-UG15F289P503	4	Leaf from stem 1	24,343.0	24,373.0	20.0	1,222.0	123.0	0.082	2
1.1*	2	R39-B1-UG15F289P503	4	Phloem, stem 1 - top	5,135.0	50,184.0	0.0	0.0	0.0	0	0
1.2*	3	R39-B1-UG15F289P503	4	Phloem, stem 1 - mid	4,010.0	9,768.0	0.0	0.0	0.0	0	0
1.3*	4	R39-B1-UG15F289P503	4	Phloem, stem 1 - bottom	10,314.0	48,265.0	1.0	402.0	402.0	0.010	0
5	5	Kwatempale from Sarah's Farm	4	Leaf	3,012.0	66,062.0	11.0	604.0	107.0	0.365	0
6	6	Kwatempale from Sarah's Farm	5	Leaf	5,074.0	5,243.0	1.0	235.0	235.0	0.020	0
7	7	Wild Plant from Naomi's Farm	3	Leaf	16,386.0	45,024.0	6.0	1,121.0	352.0	0.037	0
9	8	Sick branch from NAROCass1 from Naomi's Farm	3	Leaf	37,822.0	20,509.0	15.0	2,021.0	166.0	0.040	12
WF5	9	Whitefly from sample 5	4	1x Whitefly, no EtOH	347.0	41,599.0	0.0	0.0	0.0	0	0
WF7	10	Whitefly from sample 7	3	1x Whitefly, no EtOH	2,613.0	2,613.0	0.0	0.0	0.0	0	0
WF8	11	Whitefly from sample 8	1	1x Whitefly, no EtOH	705.0	15,596.0	0.0	0.0	0.0	0	0
None*	None	Porechop unable to match to Rapid Barcode			196,290.0	267,436.0	59.0	1,677.0	42.0	0.030	
Kenya											
1	1	Local	1	Leaf	43,049.0	39,851.0	1.0	123.0	123.0	0.002	1
2	2	Local	1	Leaf	15,890.0	66,540.0	0.0	0.0	0.0	0	0
4	4	Local	1	Leaf	38,291.0	53,843.0	3.0	890.0	251.0	0.008	1
5	5	Local	1	Leaf	9,213.0	44,230.0	0.0	0.0	0.0	0	0
L1*	6	Stem from CMV infected plant	4	Leaf	28,450.0	18,287.0	219.0	2,228.0	98.0	0.770	3
L2*	7	Stem from CMV infected plant	4	Leaf	17,320.0	42,191.0	76.0	2,127.0	78.0	0.439	0
S1*	8	Stem from CMV infected plant	4	Phloem, 22.5cm from tip	16,310.0	24,566.0	10.0	1,485.0	269.0	0.061	0
S2*	9	Stem from CMV infected plant	4	Phloem, 52.3cm from tip	5,346.0	54,245.0	0.0	0.0	0.0	0	0
R1*	10	Stem from CMV infected plant	4	Root 1, under outer bark	7,336.0	15,689.0	0.0	0.0	0.0	0	0
R2*	11	Stem from CMV infected plant	4	Root 2, under outer bark	21,576.0	24,853.0	0.0	0.0	0.0	0	0
H1*	12	Leaf from Healthy Plant	1	Leaf	44,295.0	33,307.0	0.0	0.0	0.0	0	3
None	None	Porechop unable to match to Rapid Barcode			245,390.0	265,898.0	75.0	2,404.0	28.0	0.031	
				TOTALS	**492,466.0**	**56,958.3**					

19

Table 3. Megan Results of Blastn from Nimbus cloud.

	Total reads	Reads classified	Manihot esculenta	EACMV	ACMV	TLCV	Begomo-associated DNA-III	Bemisia tabaci	Bemisia afer	Candidatus Portiera aleyrodidarum	Other
Tanzania											
1	61,317	23,481	15,862	63	-	1					
3	21,468	5,407	3,509	5							
4	31,309	10,561	6,462	2	5	1					
5	2,117	325	144		2			3		3	
6	27,178	2,127	76		1			521	67	802	
7	5,634	1,141	669		1						
8	6,237	1,449	88		5			912		15	
9	25,289	2,303	126	2				506	57	905	
10	789	166	66		3			11		2	
11	10,966	7,843	69	3	616					2	
12	1,797	356	171		2			13		7	
Uganda											
1	18,853	5,662	3,073		11		1				
1.1	5,135	876	402								
1.2	4,010	1,034	591								
1.3	10,314	1,933	864		1						
5	3,012	556	255	1	7						
6	5,074	758	39		1						
7	16,386	4,608	2,666		4						
9	37,822	12,768	7,268	10							
WF5	347	51	20					10			
WF7	243	48	21								
WF8	705	128	35					9		1	
Kenya											
1	43,049	10,283	9,947		1	1	1				
2	15,890	2,968	2,854								
4	38,291	8,648	8,836	1	1						
5	8,213	1,959	1,887								
L1	28,450	9,968	9,580	45	81						SLCV (1)
L2	17,320	4,367	2,050	29	4						
S1	16,310	3,718	1,933	5		1	1				
R1	7,336	2,303	1,016								
R2	21,578	6,810	2,000								
H1	44,295	15,537	8,389				1				

428 **Figure 1.** Tree Lab in Kenya. Essential equipment is listed in Table 1.
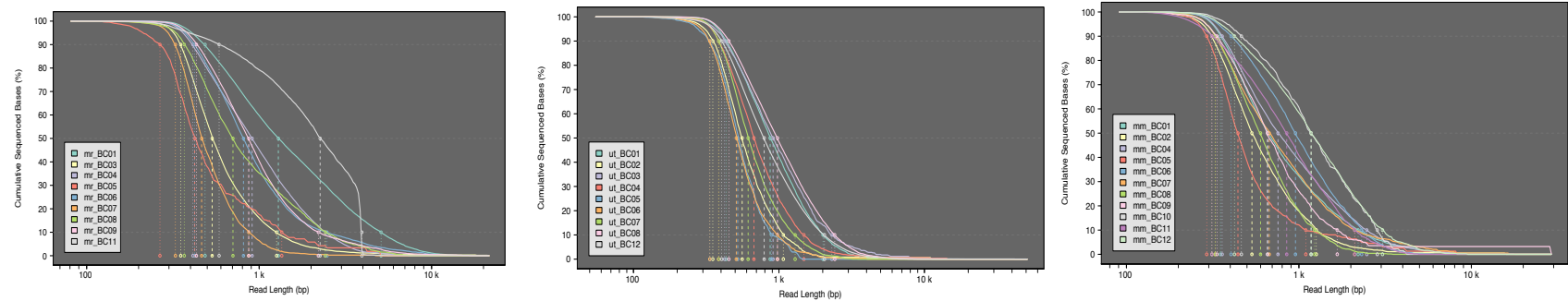429
430

431
432

433 **Figure 2.** Cumulative density curves showing the proportion of sequenced bases with length greater than a particular length (with
434 L10/L50/L90 highlighted).
435
436

437
438
439
440
441
442

References:

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. J Mol Biol **215**:403-410.

Boykin, L. M., A. Ghalab, B. R. De Marchi, A. Savill, J. M. Wainaina, T. Kinene, S. Lamb, M. Rodrigues, M. Kehoe, J. Ndunguru, F. Tairo, P. Sseruwagi, C. Kayuki, D. Mark, J. Erasto, H. Bachwenkizi, T. Alicai, G. Okao-Okuja, P. Abidrabo, E. Ogwok, J. F. Osingada, J. Akono, E. Ateka, B. Muga, and Kiarie. 2018. Real time portable genome sequencing for global food security. F1000 **7**:1101.

Calus, S. T., U. Z. Ijaz, and A. J. Pinto. 2018. NanoAmpli-Seq: a workflow for amplicon sequencing for mixed microbial communities on the nanopore sequencing platform. Gigascience **7**.

Faria, N. R., E. C. Sabino, M. R. Nunes, L. C. Alcantara, N. J. Loman, and O. G. Pybus. 2016. Mobile real-time surveillance of Zika virus in Brazil. Genome Med **8**:97.

Holmes, A. S., M. G. Roman, and S. Hughes-Stamm. 2018. In-field collection and preservation of decomposing human tissues to facilitate rapid purification of STR typing. . Forensic Sci. Int. Genet. **36**:124-129.

Huson, D. H., S. Beier, I. Flade, A. Gorska, M. El-Hadidi, S. Mitra, H. J. Ruscheweyh, and R. Tappu. 2016. MEGAN Community Edition - Interactive Exploration and Analysis of Large-Scale Microbiome Sequencing Data. PLoS Comput Biol **12**:e1004957.

Quick, J., N. D. Grubaugh, S. T. Pullan, I. M. Claro, A. D. Smith, K. Gangavarapu, G. Oliveira, R. Robles-Sikisaka, T. F. Rogers, N. A. Beutler, D. R. Burton, L. L. Lewis-Ximenez, J. G. de Jesus, M. Giovanetti, S. C. Hill, A. Black, T. Bedford, M. W. Carroll, M. Nunes, L. C. Alcantara, Jr., E. C. Sabino, S. A. Baylis, N. R. Faria, M. Loose, J. T. Simpson, O. G. Pybus, K. G. Andersen, and N. J. Loman. 2017. Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. Nat Protoc **12**:1261-1276.

Quick, J., N. J. Loman, S. Duraffour, J. T. Simpson, E. Severi, L. Cowley, J. A. Bore, R. Koundouno, G. Dudas, A. Mikhail, N. Ouedraogo, B. Afrough, A. Bah, J. H. Baum, B. Becker-Ziaja, J. P. Boettcher, M. Cabeza-Cabrerizo, A. Camino-Sanchez, L. L. Carter, J. Doerrbecker, T. Enkirch, I. G. G. Dorival, N. Hetzelt, J. Hinzmann, T. Holm, L. E. Kafetzopoulou, M. Koropogui, A. Kosgey, E. Kuisma, C. H. Logue, A. Mazzarelli, S. Meisel, M. Mertens, J. Michel, D. Ngabo, K. Nitzsche, E. Pallash, L. V. Patrono, J. Portmann, J. G. Repits, N. Y. Rickett, A. Sachse, K. Singethan, I. Vitoriano, R. L. Yemanaberhan, E. G. Zekeng, R. Trina, A. Bello, A. A. Sall, O. Faye, O. Faye, N. Magassouba, C. V. Williams, V. Amburgey, L. Winona, E. Davis, J. Gerlach, F. Washington, V. Monteil, M. Jourdain, M. Bererd, A. Camara, H. Somlare, A. Camara, M. Gerard, G. Bado, B. Baillet, D. Delaune, K. Y. Nebie, A. Diarra, Y. Savane, R. B. Pallawo, G. J. Gutierrez, N. Milhano, I. Roger, C. J. Williams, F. Yattara, K. Lewandowski, J. Taylor, P. Rachwal, D. Turner, G. Pollakis, J. A. Hiscox, D. A. Matthews, M. K. O'Shea, A. M. Johnston, D. Wilson, E. Hutley, E. Smit, A. Di Caro, R. Woelfel, K. Stoecker, E. Fleischmann, M. Gabriel, S. A. Weller, L. Koivogui, B. Diallo, S. Keita, A. Rambaut, P. Formenty, S. Gunther, and M. W. Carroll. 2016. Real-time, portable genome sequencing for Ebola surveillance. Nature **530**:228-232.

Stanton, J. L., A. Muralidhar, C. J. Rand, and D. J. Saul. 2019. Rapid extraction of DNA suitable for NGS workflows from bacterial cultures using the PDQeX. . BioTechniques **66**:1-7.