1 **Pathogen Detection and Microbiome Analysis of Infected Wheat Using a Portable DNA**

2 **Sequencer**

3 **Yiheng Hu** and **Gamran S. Green**, Research School of Biology, The Australian National University,

4 Acton ACT 2601, Australia; **Andrew W. Milgate**, Wagga Wagga Agricultural Institute, New South

5 Wales Department of Primary Industry, Wagga Wagga NSW 2650, Australia; **Eric A. Stone**, ANU-

6 CSIRO Centre for Genomics, Metabolomics and Bioinformatics, Acton ACT 2601, Australia; **John**

7 **P. Rathjen**† and **Benjamin Schwessinger**†, Research School of Biology, The Australian National

8 University, Acton ACT 2601, Australia.

9 †Corresponding authors: B. Schwessinger; E-mail: benjamin.schwessinger@anu.edu.au, and J. P.

10 Rathjen; Email: john.rathjen@anu.edu.au

11

14

15 Current address of G. S. Green: School of Biotechnology and Biomolecular Sciences, The

16 University of New South Wales, Kensington, NSW2052, Australia.

17

18 **ABSTRACT**

19 Fungal diseases of plants are responsible for major losses in agriculture, highlighting the need for

20 rapid and accurate identification of plant pathogens. Disease outcomes are often defined not

21 only by the main pathogen but are influenced by diverse microbial communities known as the

22 microbiome at sites of infection. Here we present the first use of whole genome shot-gun

23   sequencing with a portable DNA sequencing device as a method for the detection of fungal

24   pathogens from wheat (*Triticum aestivum*) in a standard molecular biology laboratory. The data

25   revealed that our method is robust and applicable to the diagnosis of fungal diseases including

26   wheat stripe rust (caused by *Puccinia striiformis* f. sp. *tritici*), septoria tritici blotch (caused by

27   *Zymoseptoria tritici*) and yellow leaf spot (caused by *Pyrenophora tritici repentis*). We also

28   identified the bacterial genus *Pseudomonas* co-present with *Puccinia* and *Zymoseptoria* but not

29   *Pyrenophora* infections. One limitation of the method is the over-representation of redundant

30   wheat genome sequences from samples. This could be addressed by long-range amplicon-based

31   sequencing approaches in future studies, which specifically target non-host organisms. Our work

32   outlines a new approach for detection of a broad range of plant pathogens and associated

33   microbes using a portable sequencer in a standard laboratory, providing the basis for future

34   development of an on-site disease monitoring system.

35

36

37   Plant diseases, especially those caused by fungal pathogens, jeopardise global crop biosecurity.

38   Increased global trade, human migration, environmental changes and the accelerated

39   emergence of virulence have been identified as causes for increasing prevalence of diseases in

40   plants and animals (Fisher et al. 2012). To prevent and manage diseases, rapid detection and

41   identification of their causal agents are crucial. This is particularly important for outbreak

42   management during the incursion of virulent new isolates that overcome classical methods of

43   control such as the use of resistant crop varieties (Bhattacharya 2017; Pretorius et al. 2000).

44   Traditional methods for crop disease diagnosis rely largely on the expertise of pathologists whom

45   identify diseases initially by eye in the field. While molecular methods of pathogen detection also

46   exist, these are generally only capable of identifying specific diseases with low amounts of

47   variation (De Boer and López 2011). For example, the enzyme-linked immunosorbent assay

48   (ELISA) have been applied for the detection of tree pathogen *Xylella fastidiosa* (Sherald and Lei

49   1991). Recently, other optical sensor-based methods have been developed for the detection,

50   identification, and quantification of plant diseases (Mahlein 2015). This method has been used

51   to help smallholder farmers for the disease and pest control (PlantVillage,

52   https://plantvillage.psu.edu/). Conventional pathology, however, remains the primary means of

53   pathogen identification and harbors certain limitations: 1. It relies heavily on the physical

54   appearance of disease symptoms; 2. It has difficulty detecting pathogens that do not infect aerial

55   parts of the plant; 3. Co-infections may lead to conflicting visual symptoms; 4. No method

56   currently exists for the rapid identification of unknown pathogens during outbreaks – a limitation

57   epitomized by the recent, devastating incursion of South American races of wheat blast fungus

58   in Bangladesh (Callaway 2016).

59   DNA sequencing-based methods have been applied to address limitations in detection and

60   monitoring of environmental microbial communities (Nezhad 2014). Such assays rely largely on

61   amplification and sequencing of small conserved regions called meta barcodes from pathogen

62   genomes, including the internal transcribed spacer region or *Elongation Factor 1 alpha* (*EF1α*)

63   and *Beta-tubulin* genes for fungal species and 16S ribosomal RNA sequences for bacteria (Janda

64   and Abbott 2007; Raja et al. 2017). Combining multiple conserved gene loci, known as multilocus

65   sequence typing (Maiden et al. 1998), can discriminate closely related species based on single

66   nucleotide polymorphisms (SNPs) within amplicon sequences. This has been used, for example,

67   to distinguish different closely related species of *Candida* that pose distinct risks and

68   epidemiological outcomes (Odds and Jacobsen 2008). Amplicon sequencing also allows for

69   characterisation of the microbial community - the microbiome - within a sample. The microbiome

70   can impact plant function, for example, on altering flowering time (Panke-Buisse et al. 2015) or

71   promoting colonization of fungal pathogens such as *Botrytis cinerea* (Ritpitakphong et al. 2016).

72   Furthermore, the microbiome is likely to contain organisms that interact with disease-causing

73   agents and, as some of these organisms have anti-microbial activities against plant pathogens,

74   could potentially be harnessed for biocontrol strategies (Ellis 2017). Restrictive read lengths

75   generated by the Illumina sequencing platform limits sequence-based identification of microbial

76   species. Standard Illumina sequencing only allows for a maximum amplicon length of 500 base

77   pairs (bp) to be sequenced and, as such, only the genetic variation present within 500 bp is

78   available for taxonomic assignment of the organism in question (Benítez-Páez and Sanz 2017).

79   Use of PCR also biases the abundance of individual amplicons, making estimation of population

80   frequencies problematic (Kennedy et al. 2014).

81   Next-generation DNA sequencing platforms have developed rapidly during the last half-decade,

82   overcoming the read length issue and facilitating the identification of microbial species (Benítez-

83   Páez and Sanz 2017). Multiple single molecule, long-read sequencing technologies are currently

84   in use. Most predominant among these are single molecular real-time (SMRT) sequencing (PacBio)

85   and nanopore sequencing (MinION; Oxford Nanopore Technologies (ONT)). PacBio SMRT

86   sequencing relies on the detection of fluorescent nucleotides incorporated into a single DNA

87   molecule during its synthesis within a nanoscale observation chamber (Ardui et al. 2018).

88   Nanopore sequencing enables direct sequencing of native DNA by measuring voltage changes

89   across an artificial membrane when a single DNA molecule passes through a nanopore embedded

90   in the membrane of a flowcell (Jain et al. 2016). Advantages of nanopore sequencing include its

91   relatively low cost, pocketsize form and the potential for real-time data analysis, while current

92   disadvantages include relatively low per read accuracy of around 90% (Leggett and Clark 2017).

93   Because the MinION is a mobile technology independent of large sequencing facilities, it has been

94   useful for on-site sample sequencing in extreme environments such as microbial paleo mats in

95   the Antarctic (Johnson et al. 2017). For rapid pathogen identification, studies have shown a

96   turnaround time of six hours for bacterial from clinical samples with MinION (Charalampous et

97   al. 2018), and from seconds to up to four hours for detecting cassava virus in Africa (Boykin et al.

98   2018) – both unachievable with previous technologies. For antibiotic resistance, Břinda et al.

99   (2018) reported their identification of known resistant bacterial strain within five minutes from

100 real time MinION data. Moreover, the MinION provides significantly longer sequencing reads

101 (~10 kb) than other sequencing platforms such as Illumina, allowing for more genetic information

102 per read and thus more precise identification of the organism from which the read is derived

103 (Benítez-Páez, Portune, and Sanz 2016; Benítez-Páez and Sanz 2017). Although recent studies

104 have shown important advances in identification of bacterial pathogens using nanopore

105 sequencing from clinical samples (Mitsuhashi et al. 2017; Schmidt et al. 2017), little has been

106 done testing this new sequencing platform in terms of plant pathogens detection and

107 microbiome profiling of diseased plants.

108 Here, we tested nanopore sequencing for disease diagnosis in field-grown wheat plants and

109 explored its potential for microbiome profiling. We designed a whole-genome shot-gun

110 sequencing workflow for identification of major pathogen species from field samples and

111 profiling of microbes present in these samples to the species level. As a proof-of-concept, we

112 applied our sequencing workflow to wheat leaves infected with known fungal pathogens sampled

113 from the field and a fungicide treated control. By mapping sequencing reads to databases, we

114 successfully detected all fungal diseases present in the infected samples including stripe rust

115 (caused by *Puccinia striiformis* f. sp. *tritici*), septoria tritici blotch (caused by *Zymoseptoria tritici*)

116 and yellow leaf spot (caused by *Pyrenophora tritici repentis*), and found one disease (septoria

117 nodorum blotch, caused by *Parastagonospora nodorum*). We also conducted preliminary

118 characterisation of the microbiomes associated with those diseases, identifying the bacterial

119 genus *Pseudomonas* as co-present with *Puccinia* and *Zymoseptoria* infections. Our results suggest

120 that portable nanopore sequencing has a considerable potential for adaptation to a broad range

121 of crop disease diagnoses and environmental monitoring applications under field conditions.

122

123 **MATERIALS AND METHODS**

124 **Sample collection**

125 Wheat cultivar Crusader (Advanta Seeds Pty. Ltd.) seeds were planted in the disease nursery field

126 on May 2016 at the Wagga Wagga Agricultural Institute (New South Wales (NSW), 35°02'24.3"S

127 147°19'09.5"E) with daily irrigation. Wheat straw infected the previous season with septoria

128 tritici blotch and yellow leaf spot was allowed to stand in the field to promote development of

129 pseudothecia. This infected material was used to manually inoculate plants in the nursery field

130 at the tillering growth stage (4-5 weeks post emergence) while stripe rust infections occurred

131 naturally at jointing stage (7-9 weeks post emergence). Healthy or diseased plants were

132 diagnosed by a plant pathologist from the NSW Department of Primary Industries via

133 symptomatology before samples were collected during September 2016. Five different

134 treatment groups were collected into five separate containers. These included four infection

135 types: three single infections (*Puccinia striiformis* f. sp. *tritici*, *Z. tritici* and *Pyrenophora tritici*

136 *repentis*) and one double infection (*Puccinia striiformis* f. sp. *tritici + Z. tritici*), and a fungicide-

137 treated control. For each treatment group, three separate entire wheat tillers were pooled in one

138 container and shipped to the Rathjen laboratory, Research School of Biology, The Australian

139 National University. Each biological replicate (n=4) consisted of two independent leaf cuts (~100

140 mg) from each treatment group containing visible disease symptoms or not in case of the control.

141 The two leaf of each biological replicate were transferred into 2 mL Eppendorf tubes containing

142 one 5 mm metal bead. All tubes were labelled and stored at -80°C prior to DNA extraction.

143 **DNA extraction**

144 DNA extractions were performed according to Hu (2016). Briefly, 100 mg leaf tissue was

145 homogenized and cells were lysed using cetyl trimethylammonium bromide (CTAB, Sigma-Aldrich

146 Cas. No: 57-09-0) buffer (added RNAse T1, Thermo Fisher, Cat#EN0541, 1000 units per 1750 μL),

147 followed by a phenol:chloroform:isoamyl alcohol (25:24:1, Sigma-Aldrich, Cat#P2069) extraction

148 to remove protein and lipids. The DNA was precipitated with 700 μL isopropanol, washed with 1

149 mL 70 % ethanol, dried for 5 minutes at room temperature and resuspended in 50 μL TE buffer

150 containing 10 mM Tris and 1 mM EDTA at pH 8. DNA extractions were performed independently

151 for each biological replicate (n=4), with the operator blind to the identity of each sample. After

152 extraction, DNA was purified with one volume of Agencourt AMPure XP beads (Beckman Coulter,

153 Inc. Cat#BECLA63880) according to the manufacturer's protocol and stored at 4°C. Quality and

154 average size of genomic DNA was visualized by gel electrophoresis with a 1% agarose gel for 1

155   hour at 100 volts. Representative gel images for replicate 3 and 4 are shown in Supplementary

156   Fig. S1. DNA was quantified by NanoDrop and Qubit (Life Technologies) according to the

157   manufacturer's protocol. DNA quality and purity was evaluated as described previously

158   (Schalamun et al. 2018).

159   **Library construction and sequencing**

160   For each biological replicate, one PCR-based barcoded 1D sequencing libraries were constructed

161   using 1D PCR barcoding workflow (SQK-LSK107 and EXP-PBC001, batch number EP01.10.0005) as

162   previously described in Hu and Schwessinger (2018) with the omission of the DNA shearing step.

163   Briefly, a dA-tailing reaction was performed on 700 ng genomic DNA from each sample using the

164   NEBNext Ultra II End-repair/dA-tailing kit (New England Biolabs, Catalog #E7645). Repaired DNA

165   was cleaned up using one volume of AMPure XP beads, washed on a magnetic rack using 70%

166   ethanol, and eluted with 31 µL nuclease-free water. Thirty µL eluate was used for barcode

167   adapter ligation via the NEB Blunt/TA Ligase Master Mix (New England Biolabs, Catalog #M0367S)

168   and Barcode Adapter Mix (ONT 1D PCR barcoding kit EXP-PBC001), followed by another wash

169   step on one volume of AMPure XP beads. Twenty ng of the adapter-ligated template in 2 µL of

170   nuclease free water was added to each barcoding PCR reaction with 50 µL LongAmp Taq 2X

171   Master Mix (New England Biolabs, Catalog #M0287S), 46 µL water and 2 µL PCR Barcodes (ONT

172   1D PCR barcoding kit EXP-PBC001). The PCR was performed as follows: denaturation at 95°C for

173   3 min, 15 cycles of 95°C/15 s; 62°C/15 s; 65°C/150 s, a final extension step at 65°C for 5 min.

174   Barcoded DNA was purified using one volume of AMPure XP beads and pooled into 1 µg fractions

175   in a total of 45 µL to which 5 µL of DNA CS control (ONT 1D ligation sequencing kit SQK-LSK107)

176   was added. A further dA-tailing reaction was performed as above followed by purification on with

177   one volume of AMPure XP beads. DNA was eluted in 31 µL nuclease free water. Adapter ligation

178   was carried out using 30 µL end-prepped DNA, 20 µL Adapter Mix (ONT 1D ligation sequencing

179   kit SQK-LSK107) and 50 µL NEB Blunt/TA Ligase Master Mix. The adapter-ligated DNA library was

180   purified with one volume of AMPure XP beads using the Adapter Beads Binding buffer (ONT 1D

181   ligation sequencing kit SQK-LSK107) for washing and eluted in 15 µL of Elution Buffer (ONT 1D

182   ligation sequencing kit SQK-LSK107).

183   Sequencing reactions were performed independently for each biological replicate on a MinION

184   flowcell (R9.4, ONT) connected to a MK1B device (ONT) operated by the MinKNOW software

185   (version 1.5.2 and version 1.6.11). Each flowcell was primed with 1 mL of priming buffer

186   comprising 480 µL Running Buffer Fuel Mix (RBF, ONT) and 520 µL nuclease free water. Twelve

187   µL of amplicon library was added to a loading mix including 35 µL RBF, 25.5 µL Library Loading

188   beads (ONT library loading bead kit EXP-LLB001, batch number EB01.10.0012) and 2.5 µL water

189   with a final volume of 75 µL and then added to the flowcell via the SpotON sample port. The

190   "NC_48Hr_sequencing_FLO-MIN106_SQK-LSK107" protocol was executed through MinKNOW

191   after loading the library. The sequencing run was restarted after 12 h and stopped after 48 h.

192   **Sequencing analysis**

193   Raw fast5 files were processed via the Albacore 2.0.2 software (ONT) for basecalling, barcode de-

194   multiplexing and quality filtering (Phred quality score of >7) following manufacturer's

195   recommendations. All reads that passed quality filtering for each barcode were treated in parallel

196   as follows: firstly, NanoLyse (De Coster et al. 2018) was used to remove lambda phage control

197   DNA sequences; secondly, barcode and adapter sequences were trimmed from the ends of reads

198   using Porechop (Wick, Judd, and Holt 2018). To identify middle adapter sequences by Porechop,

199   a 95% threshold was set, and reads with middle adapter sequences were split into two sequences;

200   thirdly, seqtk (https://github.com/lh3/seqtk) was used to convert the processed fastq file into

201   fasta format for further analysis. The two nucleotide BLAST (blastn, Basic Local Alignment Search

202   Tool, Altschul et al. 1990) searches were performed using the fasta files for each barcode as

203   queries: first, against a custom genome reference database that contained the wheat reference

204   genome (Clavijo et al. 2017), accessed 09/2017, and the genome sequence of three fungal species

205   used in our study: *Puccinia striiformis* f. sp. *tritici*: Schwessinger et al. (2018), accessed 09/2017;

206   *Z. tritici*:   Goodwin et al. (2011), version 2, accessed 09/2017; *Pyrenophora tritici-repentis*:

207   Manning et al. (2013), unmasked assembly, accessed 09/2017. The genome of a fourth species

208   suggested to be absent from the studied wheat growing region, *Parastagonospora nordorum*, as

209   a negative control (Hane et al. (2007), accessed 09/2017. Individual reads were assigned to have

210   originated from a specific DNA sequence based on their best blastn hit (e-value < 0.01) (Zhang et

211   al. 2000). Reads that did not hit the custom-made database were captured using the

212   filterbyname.sh   script   from   the   bbmap   script   package

213   (https://sourceforge.net/projects/bbmap/). These reads were used as queries to search against

214   the National Center for Biotechnology Information (NCBI) nucleotide (nt) database

215   (ftp://ftp.ncbi.nlm.nih.gov/blast/db/, downloaded 09/2017) using blastn and the same settings

216   and species assignment criteria mentioned above. Two Python scripts (QC_and_BLAST.py and

217   creating_final_dataframe.py) that combine all steps from quality control to BLAST outputs and

218   construct dataframes containing statistical and information from BLAST searches for each read

219   were   deposited   at

220   https://github.com/Yiheng323/Pathogen_Detection_and_Microbiome_scripts to facilitate the

221   reproducibility of our results. All basecalled fastq data for each sequencing run were deposited

222   at NCBI Short Reads Archive (SRA) under accession number PRJNA493553. The sequencing

223   summary files and final dataframes were deposited as .tab files for each replicate at

224   https://doi.org/10.6084/m9.figshare.7138262. The python scripts used for generating Fig. 2 – 5

225   and   Supplementary   Fig.   S2   –   S7   were   also   deposited   at

226   https://github.com/Yiheng323/Pathogen_Detection_and_Microbiome_scripts.

227

228   **RESULTS**

229   **Sequencing showed variations in yield and consistency in barcode classification**

230   We designed a pipeline for identification of three known wheat pathogens from infected field

231   samples using MinION sequencing. We infected three experimental field plots as disease

232   nurseries with their corresponding pathogens (*Puccinia. striiformis* f. sp. *tritici* for stripe rust

233   disease, *Z. tritici* for septoria tritici blotch and *Pyrenophora. tritici-repentis* for yellow leaf spot)

234   commonly found in New South Wales, Australia, and conducted sampling during the 2016-17

235   growing season (September). The sequencing experiment included three steps (Fig. 1): 1.

236   Pathologists collected three plants from each disease nursery field, with visual confirmation of

237   disease symptoms from three single infections plus one double infection (*Puccinia striiformis* f.

238   sp. *tritici* + *Z. tritici*) and a negative control comprising the same wheat variety treated with

239   commercial fungicides at the 10- to 11-week growth stage (three months post germination).

240   Individual plants were pooled by treatment type and shipped to the laboratory on ice for

241   processing. 2. We extracted DNA from two independent leaf cuts of each treatment type without

242   knowing the identity of each treatment to eliminate potential bias from operators. 3. We

243   barcoded each sample to enable pooling of all five sample types into one sequencing run. The

244   five wheat leaf samples were barcoded as follows: leaves infected with one of the three fungal

245   pathogens (barcodes 1-3); leaves infected with two pathogens (barcode 4); and fungicide treated

246   control without disease symptoms (barcode 5). We performed steps 2 to 3 four times

247   independently as biological replicates, with Step 3 performed using a MinION 1D PCR barcoding

248   workflow (EXP-PBC001 and SQK-LSK107). The raw data of each nanopore sequencing run were

249   converted into fastq sequencing reads using Albacore software (ONT), with a quality (Q) score

250   for each base estimating the likelihood of the identified base being correct using the Phred scale

251   (Ewing et al. 1998; Ewing and Green 1998). Based on read mean Q scores, Albacore grouped

252   reads into pass (Q score > 7) or fail (Q score <= 7) based on manufacturer's recommendations.

253   We binned reads based on the barcode sequences identified by Albacore at both ends of each

254   sequencing read. If no barcode sequence could be identified, we named the corresponding read

255   as unclassified. Overall, ~92-95% of the total sequence length of all four sequencing runs passed

256   quality filtering (Supplementary Table 1). The total sequence length from the four runs varied

257   from 682 to 5100 megabase pairs (Mbp). The mean read length ranged from 1380 to 3054 bp.

258   Proportions of each barcode in each run were relatively even, although 23.3-25.7% of the

259   sequences were unclassified in all four runs (Fig. 2).

260   **Sequence-based identification correctly identifies the fungal species causing wheat diseases**

261   We performed blastn searches for all quality-filtered reads against a custom genome reference

262   database. We assigned the origin of each read based on its best blastn hit within each database

263   (e-value < 0.01) using a 'winner-takes-all' strategy which only considers the best match of each

264   sequencing read. The custom genome reference database contained the wheat genome and four

265   fungal genomes including *Puccinia striiformis* f. sp. *tritici*, *Z. tritici*, *Pyrenophora tritici-repentis*

266   and *Parastagonospora nodorum*. These represent the host plant genome, three fungal species

267   present in infected field samples, and one fungal species (*Parastagonospora nordorum*) not

268   expected to be present in the sampling area as a negative control. The majority of reads were

269   assigned to wheat sequences, comprising total sequence length of 90.1% in flowcell 1, 80.1% in

270   flowcell 2, 92.7% in flowcell 3 and 91.4% in flowcell 4 (Fig. 3). We found that healthy leaf samples

271   (barcode 5) contained no or minor total length percentages of pathogen-related sequences in all

272   four replicates (< 0.01%, Fig. 4). This indicates that our analysis method is robust and does not

273   lead to high background signals in healthy wheat samples (barcode 5). Significantly higher

274   proportions of total sequencing length (0.5–5.7%) from infected samples (barcode 1-4) were

275   assigned as pathogen genome-derived. In all cases, the pathogen genome identified by the

276   highest total sequence length corresponded to the pathogen causing the disease as identified by

277   symptomatology (Fig. 4). This indicates that we are clearly able to identify the disease-causing

278   pathogen via our sequencing workflow in cases of single infections. Infections with two

279   pathogens (barcode 4) were identified less clearly. In barcode 4 from all biological replicates,

280   total sequence length percentages assigned to *Puccinia striiformis* f. sp. *tritici* were between 0.1%

281   and 0.4%; close to or lower than the proportion assigned to negative control species

282   *Parastagonospora nodorum.* The total relative sequence length identifying each major pathogen

283   species within each barcode were similar across all replicates, with around 0.6% of barcode 1

284   from *Puccinia striiformis* f. sp. *tritici*, 1.9-5.0% of barcode 2 from *Z. tritici*, 2.1-5.7% of barcode 3

285   from *Pyrenophora tritici-repentis*, and 0.1-0.4% of barcode 4 from *Puccinia striiformis* f. sp. *tritici*

286   and 0.6-1.2% from *Z. tritici*. Unexpectedly, we also identified *Parastagonospora nodorum* in

287   samples derived from different treatment groups in all four biological replicates. To ascertain

288   whether *Parastagonospora nodorum* was truly present, we extracted all 915 reads that were

289   assigned to have originated from the *Parastagonospora nodorum* genome in all four biological

290   replicates. We used these reads to perform an open-ended blastn search against the entire NCBI

291   non-redundant nucleotide (nt) database. We found that 344 of 915 reads were assigned to genus

292   *Parastagonospora*, including 339 reads identifing *Parastagonospora nodorum* SN15 genome

293   (Supplementary Fig. S2). A further 371 reads hit related species, including 226 from the same

294   *Pleosporales* order: *Leptosphaeria* spp. (84 reads), *Alternaria* spp. (81), *Pyrenophora* spp. (10),

295   *Bipolaris* spp. (9), *Shiraia* spp. (9) and other genus (Supplementary Fig. S2). The remaining 200

296   reads fell above the given e-value threshold (0.01) and failed to be assigned. The difference

297   between both BLAST searches potentially stems from the relatively high error rate of the MinION

298   sequencer and the distinct conserved regions from genomes of closely related species (Jain et al.

299   2018; Laver et al. 2015). We tested for the presence of *Parastagonospora nodorum* in one of our

300   four biological replicates using an independent molecular biology approach. We performed a PCR

301   screen in replicate 3 using the *PnTox3* gene that is universally found in all Australian

302   *Parastagonospora nodorum* isolates but not in any other fungal species (Liu et al. 2009). We were

303   able to amplify a specific band corresponding to *PnTox3* (Supplementary Fig. S3) in the *Z. tritici*

304   treatment group (barcode 2) which displayed the most *Parastagonospora nodorum* assigned

305   sequencing reads (Fig. 4). The *Puccinia striiformis* f. sp. *tritici* treatment group (barcode 1) and

306   the *Puccinia striiformis* f. sp. *tritici* + *Z. tritici* treatment group (barcode 4) contained a lesser

307   number of *Parastagonospora nodorum*-assigned sequencing reads, and these appeared to be

308   below the detection threshold of PCR analysis. These results support the finding that

309   *Parastagonospora nodorum* is present in our samples from the Wagga Wagga wheat-growing

310   region despite the disease not being commonly found in the region nor the pathogen being

311   purposefully inoculated in the trial.

312   **Bacterial species associated with fungal disease development**

313   To obtain an overview of the microbiome profile in and on wheat leaves infected by different

314   fungal pathogens, we attempted to identify other microbial species present on the leaves. We

315   searched the non-redundant NCBI nt database with reads that were not assigned to any of the

316   five reference genomes present in our initial custom genome reference database. We found that

317   the proportion of reads assigned to a genome in the restricted and the nt-NCBI databases were

318   high for flowcells 1, 3 and 4 (92.9, 94.3 and 93.8%, respectively), but only 84.6% for flowcell 2

319   likely due to technical issues. Overall, from each flowcell, less than 1.4% of passed reads identified

320   microbiota other than the major pathogens in the custom genome reference database. We

321   pooled the reads from each barcode in all biological replicates, and calculated the total sequence

322   length percentage that identified a specific genus in each sample using NCBI taxonomic identifiers

323   (Fig. 5A). Genus *Pseudomonas* took up the highest proportion in all diseased samples, ranging

324   from 0.08% to more than 0.88%, but was nearly absent in healthy wheat (barcode 5, 0.01%).

325   Genus *Alternaria* is also found in all samples with a proportion between 0.05% to 0.2% in diseased

326   samples and 0.005% in healthy wheat. Uniquely in *Puccinia striiformis* f. sp. *tritici* treatment

327   group (barcode 1), bacterial *Erwinia* comprised 0.09% of total sequence length as represented by

328   219 reads and the second highest proportion. In contrast, *Erwinia* only represented 0.003% and

329   0.001% in *Z. tritici* treatment group (barcode 2) and fungicide treatment group (barcode 5) while

330   being absent in any other treatment group. In treatment groups infected with *Z. tritici* (barcodes

331   2 and 4), hits on *Z. tritici* were still present and may represent sequences in the six *Z. tritici*

332   genomes in the NCBI nt database that are absent from the *Z. tritici* genome we used to generated

333   our initial custom database. At the species level (Fig. 5B), total sequence length percentages

334   identifying *Pseudomonas syringae* were much more frequent in treatment groups infected with

335   *Puccinia striiformis* f. sp. *tritici* (barcode 1, 0.59%) and/or *Z. tritici* (barcode 2, 0.33%). In contrast,

336   the treatment group infected wtih *Pyrenophora tritici-repentis* (barcode 3) displayed a much

337   lower relative total sequence length of 0.02% assigned to *Pseudomonas syringae*. Relative

338   sequence length assigned to other *Pseudomonas* species such as *Pseudomonas poae* followed a

339   similar trend. This finding is consisted with previous reports were *Pseudomonas* spp. have been

340   descripted to be associated with fungal infections of wheat (Al-Sallami et al. 1997; Mehrabi et al.

341   2016). At the upper taxonomic classifications, *Proteobacteria* and *Ascomycota* were the most

342   abundant bacterial and fungal phyla across all treatment groups (Supplementary Fig. S4) followed

343   by *Actinobacteria* and *Basidiomycota*. At the class level (Supplementary Fig. S5), either

344   *Gammaproteobacteria* or *Dothideomycetes* were the most abundant. Overall, the microbiome

345   structure of the upper taxonomy resembles that of the lower taxonomy (Supplementary Fig. S4,

346   Fig. S5, Fig. S6, Fig. S7), as most sequences were assigned to genus *Pseudomonas*.

347

348   **DISCUSSION**

349   We have developed a whole genome shot-gun sequencing (WGS) workflow which accurately

350   identified three fungal pathogens responsible for three major wheat diseases – stripe rust,

351   septoria tritici blotch and yellow spot – from field samples. We designed a two-step BLAST

352   process to identify major agents in diseased samples and characterise the associated microbiota.

353   We identified a number of *Pseudomonas* species present in stripe rust and septoria tritici blotch

354   infected wheat leaves which appear far less abundant in yellow spot or non-diseased leaves.

355   WGS provides a wealth of information that could be used to accurately identify disease-causing

356   agents and begin to describe microbial community composition in a sample (Brown et al. 2017;

357   Thomas, Gilbert, and Meyer 2012). Without PCR amplification process to potentially introduce

358   bias, WGS data is obtained directly from sample extracts and illustrates features closest to reality

359   dependent only on sampling and DNA extraction strategies. Omitting PCR steps, WGS requires

360   less processing time and proves advantageous for on-site diagnostic purposes where technical

361   simplicity and fast processing time are prioritised. It has a significant drawback, however, in that

362   most sequences are derived from the host (wheat) genome, representing 90.1% of total

363   sequence length in flowcell 1, 80.1% in flowcell 2, 92.7% in flowcell 3 and 91.4% in flowcell 4. This

364   result is due to our DNA extraction method in which the whole leaf sample was ground to retain

365   all the microbial population in and on the leaf. For the purpose of microbial species identification,

366   host sequences are unwanted and can overwhelm community DNA (Sharpton 2014). To enrich

367   microbial sequences, one strategy is to apply intricate molecular methods during DNA extraction

368   such as gradient centrifugation and sonication (Bulgarelli et al. 2012; Lim et al. 2014). As these

369   techniques require sophisticated instruments, this can be a challenge to apply to environmental

370   samples, especially in field conditions. Another strategy to enrich microbiome sequences involves

371   sequencing of microbial metabarcode markers. With the improvement of sequencing length,

372   species-level resolution using long read amplicon sequencing has been achieved in few cases

373 (Benítez-Páez and Sanz 2017; Martijn et al. 2017). Long-read amplicon sequencing seems to be a

374 promising method for species identification of known microbes, as it requires less input DNA than

375 WGS and metabarcode databases contain information from comparatively more species than

376 whole genome databases (Cruaud et al. 2017; Tessler et al. 2017; Ranjan et al. 2016). Amplicon

377 sequencing is also possible in applied cases given the rapid development of remote PCR such as

378 the miniPCR system (Amplyus, Cambridge, USA) and the Bento Lab platform (Bento Bioworks Ltd,

379 London, United Kingdom). Drawbacks of amplicon strategy include biases related to PCR

380 amplification, which lead to inaccuracy when calculating relative proportions of microbial species.

381 While many workflows are available for microbiome profiling using metagenomics data, there is

382 no standard method for analysing long read data. This makes validation of both methods and

383 results crucial (Quince et al. 2017), and mock microbial communities or samples containing

384 known microbial species have been used to validate methods as a proof-of-concept (Nicholls et

385 al. 2018). Verifying the diagnosis of a specific pathogen strain based on SNPs called from MinION

386 reads, Votintseva et al. (2017) first identified pathogens using clinical methods and then

387 generated sequencing data to map this data to the targeted species in real-time. Sequencing

388 reactions were stopped once sufficient SNPs were obtained to confidentially identify the targeted

389 pathogen strain. Similarly, our study used wheat samples inoculated manually by known

390 pathogens and confirmed visually prior to sequencing. The major pathogen identifications from

391 the first BLAST analysis against the custom genome reference database validated the capacity of

392 our workflow for identification of those pathogen species. To validate the usage of a narrow first

393 initial database consisting of suspected pathogens and host, we also performed an open-ended

394 BLAST analysis against the whole nt database with all sequencing reads of replicate 4

395   (Supplementary Table S2). The overall results of this unrestricted database search are consistent

396   with our targeted analysis approach, both in terms of pathogen identification and microbiome

397   analysis (Fig. 4, Fig. 5 and Supplementary Table S2). Yet, we observed that some reads were

398   assigned to closely related plant genera such as *Aegilops* or *Hordeum* instead of wheat (*Triticum*).

399   This is likely caused by the presence of highly similar DNA sequences in these related species and

400   the relatively high error rate of nanopore sequencing. To further explore and validate these

401   results different analysis approaches should be performed carefully on the same dataset and

402   compared in detail. Brown et al. (2017) compared the performance of four bioinformatics

403   pipelines on identifying species within different synthetic bacterial communities using MinION.

404   Compared to the composition of a mock community, even the pipeline that resulted in the most

405   similar classification still contains reads classified to species known not to be present. These false

406   positive results are due largely to sequencing errors, which illustrates the need for more stringent

407   filtering of the raw data and improvement of per read accuracy.

408   Nanopore metagenomics has great potential for quick diagnosis of suddenly emerging crop

409   disease and large-scale disease monitoring at centralized agricultural institutions. As shown from

410   other sequencing studies performed in a more extreme environment (Johnson et al. 2017; Boykin

411   et al. 2018), all processes including DNA extraction and purification, library preparation and

412   sequencing can be completed within a few hours. A tailored database can be created for each

413   crop species containing the genomes of pathogens known to cause diseases to minimise

414   processing time. For rapidly evolving pathogens such as wheat rusts, the SNP-based classification

415   method described by Votintseva et al. (2017) could be adapted to define rust strains once enough

416   SNPs are obtained. However, before running the sequencer, sampling procedures and tissue

417 processing strategies still need to be carefully curated. In our study, we randomly sampled

418 multiple leaves from different plants belonging to five treatment groups and performed

419 complete sequencing and analysis independently on all four biological replicate sample sets.

420 Consistency in barcode classification and BLAST analysis across all replicates suggested reliability

421 of the workflow. In centralized agricultural institutions, amplicon sequencing can potentially play

422 a key role and, to-date, there have been a few preliminary successes (Benítez-Páez, Portune, and

423 Sanz 2016; Benítez-Páez and Sanz 2017; Pomerantz et al. 2018). Given the multiplexing capacity

424 of the 96 native barcoding kit (ONT), many samples can be sequenced in parallel when larger

425 scale sequencing platforms are employed (e.g. GridION, ONT). To achieve this goal, long read

426 databases, sequencing accuracy and access to large-scale sample processing power need to be

427 improved.

428 In conclusion, our workflow demonstrates the potential of this technology for plant pathogen

429 diagnosis, field applications and microbiome characterisation. A combination of on-site and

430 centralized sequencing approaches would, in future, revolutionize management of agricultural

431 biosecurity and reduce crop loss.

432

439

440

LITERATURE CITED

Al-Sallami, F., Karov, S., Vassileva, P., Popova, R., and Vassilev, V. 1997. Pseudomonas syringae pv. atrofaciens associated with fungal black point of wheat (Triticum aestivum) grain. In *Pseudomonas Syringae Pathovars and Related Pathogens*, Springer, p. 505–508.

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

Ardui, S., Ameur, A., Vermeesch, J. R., and Hestand, M. S. 2018. Single molecule real-time (SMRT) sequencing comes of age: applications and utilities for medical diagnostics. Nucleic Acids Res. 46:2159–2168.

Benítez-Páez, A., Portune, K. J., and Sanz, Y. 2016. Species-level resolution of 16S rRNA gene amplicons sequenced through the MinION^{TM} portable nanopore sequencer. GigaScience. 5:4.

Benítez-Páez, A., and Sanz, Y. 2017. Multi-locus and long amplicon sequencing approach to study microbial diversity at species level using the MinION^{TM} portable nanopore sequencer. GigaScience. 6:1–12.

Bhattacharya, S. 2017. Deadly new wheat disease threatens Europe's crops. Nat. News. 542:145.

Boykin, L., Ghalab, A., Rossitto De Marchi, B., Savill, A., M. Wainaina, J., Kinene, T., et al. 2018. Real time portable genome sequencing for global food security. F1000Research. 7:1101.

Břinda, K., Callendrello, A., Cowley, L., Charalampous, T., Lee, R. S., MacFadden, D. R., et al. 2018. Lineage calling can identify antibiotic resistant clones within minutes. bioRxiv. :403204.

Brown, B. L., Watson, M., Minot, S. S., Rivera, M. C., and Franklin, R. B. 2017. MinION^{TM} nanopore sequencing of environmental metagenomes: a synthetic approach. GigaScience. 6:1–10.

461   Bulgarelli, D., Rott, M., Schlaeppi, K., Themaat, E. V. L. van, Ahmadinejad, N., Assenza, F., et al. 2012.

462   Revealing structure and assembly cues for *Arabidopsis* root-inhabiting bacterial microbiota. Nature.

463   488:91–95.

464   Callaway, E. 2016. Devastating wheat fungus appears in Asia for first time. Nat. News. 532:421.

465   Charalampous, T., Richardson, H., Kay, G. L., Baldan, R., Jeanes, C., Rae, D., et al. 2018. Rapid Diagnosis of

466   Lower Respiratory Infection using Nanopore-based Clinical Metagenomics. bioRxiv. :387548.

467   Clavijo, B. J., Venturini, L., Schudoma, C., Accinelli, G. G., Kaithakottil, G., Wright, J., et al. 2017. An

468   improved assembly and annotation of the allohexaploid wheat genome identifies complete families of

469   agronomic genes and provides genomic evidence for chromosomal translocations. Genome Res. 27:885–

470   896.

471   Cruaud, P., Rasplus, J.-Y., Rodriguez, L. J., and Cruaud, A. 2017. High-throughput sequencing of multiple

472   amplicons for barcoding and integrative taxonomy. Sci. Rep. 7:41948.

473   De Boer, S. H., and López, M. M. 2011. New Grower-Friendly Methods for Plant Pathogen Monitoring.

474   Annu. Rev. Phytopathol. 50:197–218.

475   De Coster, W., D'Hert, S., Schultz, D. T., Cruts, M., Van Broeckhoven, C., and Berger, B. 2018. NanoPack:

476   visualizing and processing long-read sequencing data. Bioinformatics. 34:2666–2669.

477   Ellis, J. G. 2017. Can Plant Microbiome Studies Lead to Effective Biocontrol of Plant Diseases? Mol. Plant.

478   Microbe Interact. 30:190–193.

479   Ewing, B., and Green, P. 1998. Base-Calling of Automated Sequencer Traces Using Phred. II. Error

480   Probabilities. Genome Res. 8:186–194.

481   Ewing, B., Hillier, L., Wendl, M. C., and Green, P. 1998. Base-Calling of Automated Sequencer Traces

482   UsingPhred. I. Accuracy Assessment. Genome Res. 8:175–185.

483   Fisher, M. C., Henk, D. A., Briggs, C. J., Brownstein, J. S., Madoff, L. C., McCraw, S. L., et al. 2012. Emerging

484   fungal threats to animal, plant and ecosystem health. Nature. 484:186–194.

485   Goodwin, S. B., M'Barek, S. B., Dhillon, B., Wittenberg, A. H. J., Crane, C. F., Hane, J. K., et al. 2011. Finished

486   Genome of the Fungal Wheat Pathogen Mycosphaerella graminicola Reveals Dispensome Structure,

487   Chromosome Plasticity, and Stealth Pathogenesis. PLOS Genet. 7:e1002070.

488   Hane, J. K., Lowe, R. G. T., Solomon, P. S., Tan, K.-C., Schoch, C. L., Spatafora, J. W., et al. 2007.

489   Dothideomycete–Plant Interactions Illuminated by Genome Sequencing and EST Analysis of the Wheat

490   Pathogen Stagonospora nodorum. Plant Cell. 19:3347–3368.

491   Hu, Y. 2016. High quality DNA extraction from Fungi_small scale. protocols.io. Available at:

492   https://www.protocols.io/view/high-quality-dna-extraction-from-fungi-small-scale-exmbfk6   [Accessed

493   September 26, 2018].

494   Hu, Y., and Schwessinger, B. 2018. Amplicon sequencing using MinION optimized from 1D native

495   barcoding genomic DNA. protocols.io. Available at: https://www.protocols.io/view/amplicon-sequencing-

496   using-minion-optimized-from-1d-mhkc34w [Accessed September 27, 2018].

497   Jain, M., Koren, S., Miga, K. H., Quick, J., Rand, A. C., Sasani, T. A., et al. 2018. Nanopore sequencing and

498   assembly of a human genome with ultra-long reads. Nat. Biotechnol. 36:338–345.

499   Jain, M., Olsen, H. E., Paten, B., and Akeson, M. 2016. The Oxford Nanopore MinION: delivery of nanopore

500   sequencing to the genomics community. Genome Biol. 17:239.

501   Janda, J. M., and Abbott, S. L. 2007. 16S rRNA Gene Sequencing for Bacterial Identification in the

502   Diagnostic Laboratory: Pluses, Perils, and Pitfalls. J. Clin. Microbiol. 45:2761–2764.

503   Johnson, S. S., Zaikova, E., Goerlitz, D. S., Bai, Y., and Tighe, S. W. 2017. Real-Time DNA Sequencing in the

504   Antarctic Dry Valleys Using the Oxford Nanopore Sequencer. J. Biomol. Tech. JBT. 28:2–7.

505   Kennedy, K., Hall, M. W., Lynch, M. D. J., Moreno-Hagelsieb, G., and Neufeld, J. D. 2014. Evaluating Bias

506   of Illumina-Based Bacterial 16S rRNA Gene Profiles. Appl. Environ. Microbiol. 80:5717–5722.

507   Laver, T., Harrison, J., O'Neill, P. A., Moore, K., Farbos, A., Paszkiewicz, K., et al. 2015. Assessing the

508   performance of the Oxford Nanopore Technologies MinION. Biomol. Detect. Quantif. 3:1–8.

509   Leggett, R. M., and Clark, M. D. 2017. A world of opportunities with nanopore sequencing. J. Exp. Bot.

510   68:5419–5429.

511   Lim, Y. W., Haynes, M., Furlan, M., Robertson, C. E., Harris, J. K., and Rohwer, F. 2014. Purifying the Impure:

512   Sequencing Metagenomes and Metatranscriptomes from Complex Animal-associated Samples. JoVE J. Vis.

513   Exp. :e52117–e52117.

514   Lin, S.-Y., Chooi, Y.-H., and Solomon, P. S. 2018. The global regulator of pathogenesis PnCon7 positively

515   regulates Tox3 effector gene expression through direct interaction in the wheat pathogen

516   Parastagonospora nodorum. Mol. Microbiol. 109:78–90.

517   Liu, Z., Faris, J. D., Oliver, R. P., Tan, K.-C., Solomon, P. S., McDonald, M. C., et al. 2009. SnTox3 Acts in

518   Effector Triggered Susceptibility to Induce Disease on Wheat Carrying the Snn3 Gene. PLOS Pathog.

519   5:e1000581.

520   Mahlein, A.-K. 2015. Plant Disease Detection by Imaging Sensors – Parallels and Specific Demands for

521   Precision Agriculture and Plant Phenotyping. Plant Dis. 100:241–251.

522   Maiden, M. C. J., Bygraves, J. A., Feil, E., Morelli, G., Russell, J. E., Urwin, R., et al. 1998. Multilocus

523   sequence typing: A portable approach to the identification of clones within populations of

524   pathogenic microorganisms. Proc. Natl. Acad. Sci. U. S. A. 95:3140–3145.

525   Manning, V. A., Pandelova, I., Dhillon, B., Wilhelm, L. J., Goodwin, S. B., Berlin, A. M., et al. 2013.

526   Comparative Genomics of a Plant-Pathogenic Fungus, Pyrenophora tritici-repentis, Reveals

527   Transduplication and the Impact of Repeat Elements on Pathogenicity and Population Divergence. G3

528   Genes Genomes Genet. 3:41–63.

529   Martijn, J., Lind, A. E., Spiers, I., Juzokaite, L., Bunikis, I., Pettersson, O. V., et al. 2017. Amplicon sequencing

530   of the 16S-ITS-23S rRNA operon with long-read technology for improved phylogenetic classification of

531   uncultured prokaryotes. bioRxiv. :234690.

532   Mehrabi, Z., McMillan, V. E., Clark, I. M., Canning, G., Hammond-Kosack, K. E., Preston, G., et al. 2016.

533   *Pseudomonas* spp. diversity is negatively associated with suppression of the wheat take-all pathogen. Sci.

534   Rep. 6:29905.

535   Mitsuhashi, S., Kryukov, K., Nakagawa, S., Takeuchi, J. S., Shiraishi, Y., Asano, K., et al. 2017. A portable

536   system for rapid bacterial composition analysis using a nanopore-based sequencer and laptop computer.

537   Sci. Rep. 7:5657.

538   Nezhad, A. S. 2014. Future of portable devices for plant pathogen diagnosis. Lab. Chip. 14:2887–2904.

539   Nicholls, S. M., Quick, J. C., Tang, S., and Loman, N. J. 2018. Ultra-deep, long-read nanopore sequencing

540   of mock microbial community standards. bioRxiv. :487033.

541   Odds, F. C., and Jacobsen, M. D. 2008. Multilocus Sequence Typing of Pathogenic Candida Species.

542   Eukaryot. Cell. 7:1075–1084.

543   Panke-Buisse, K., Poole, A. C., Goodrich, J. K., Ley, R. E., and Kao-Kniffin, J. 2015. Selection on soil

544   microbiomes reveals reproducible impacts on plant function. ISME J. 9:980–989.

545   PlantVillage. Available at: https://plantvillage.psu.edu/ [Accessed January 4, 2019].

546   Pomerantz, A., Peñafiel, N., Arteaga, A., Bustamante, L., Pichardo, F., Coloma, L. A., et al. 2018. Real-time

547   DNA barcoding in a rainforest using nanopore sequencing: opportunities for rapid biodiversity

548   assessments and local capacity building. GigaScience. 7.

549   Pretorius, Z. A., Singh, R. P., Wagoire, W. W., and Payne, T. S. 2000. Detection of Virulence to Wheat Stem

550   Rust Resistance Gene Sr31 in Puccinia graminis. f. sp. tritici in Uganda. Plant Dis. 84:203–203.

551   Quince, C., Walker, A. W., Simpson, J. T., Loman, N. J., and Segata, N. 2017. Shotgun metagenomics, from

552   sampling to analysis. Nat. Biotechnol. 35:833–844.

553   Raja, H. A., Miller, A. N., Pearce, C. J., and Oberlies, N. H. 2017. Fungal Identification Using Molecular Tools:

554   A Primer for the Natural Products Research Community. J. Nat. Prod. 80:756–770.

555 Ranjan, R., Rani, A., Metwally, A., McGee, H. S., and Perkins, D. L. 2016. Analysis of the microbiome:

556 Advantages of whole genome shotgun versus 16S amplicon sequencing. Biochem. Biophys. Res. Commun.

557 469:967–977.

558 Ritpitakphong, U., Falquet, L., Vimoltust, A., Berger, A., Métraux, J.-P., and L'Haridon, F. 2016. The

559 microbiome of the leaf surface of Arabidopsis protects against a fungal pathogen. New Phytol. 210:1033–

560 1043.

561 Schalamun, M., Nagar, R., Kainer, D., Beavan, E., Eccles, D., Rathjen, J. P., et al. 2018. Harnessing the

562 MinION: An example of how to establish long-read sequencing in a laboratory using challenging plant

563 tissue from Eucalyptus pauciflora. Mol. Ecol. Resour.

564 Schmidt, K., Mwaigwisya, S., Crossman, L. C., Doumith, M., Munroe, D., Pires, C., et al. 2017. Identification

565 of bacterial pathogens and antimicrobial resistance directly from clinical urines by nanopore-based

566 metagenomic sequencing. J. Antimicrob. Chemother. 72:104–114.

567 Schwessinger, B., Sperschneider, J., Cuddy, W. S., Garnica, D. P., Miller, M. E., Taylor, J. M., et al. 2018. A

568 Near-Complete Haplotype-Phased Genome of the Dikaryotic Wheat Stripe Rust Fungus Puccinia

569 striiformis f. sp. tritici Reveals High Interhaplotype Diversity. mBio. 9.

570 Sharpton, T. J. 2014. An introduction to the analysis of shotgun metagenomic data. Front. Plant Sci. 5:209.

571 Sherald, J. L., and Lei, J. D. 1991. Evaluation of a rapid ELISA test kit for detection of Xylella fastidiosa in

572 landscape trees. Plant Dis. USA. Available at: http://agris.fao.org/agris-

573 search/search.do?recordID=US9115875 [Accessed January 4, 2019].

574 Tessler, M., Neumann, J. S., Afshinnekoo, E., Pineda, M., Hersch, R., Velho, L. F. M., et al. 2017. Large-scale

575 differences in microbial biodiversity discovery between 16S amplicon and shotgun sequencing. Sci. Rep.

576 7:6589.

577 Thomas, T., Gilbert, J., and Meyer, F. 2012. Metagenomics - a guide from sampling to data analysis. Microb.

578 Inform. Exp. 2:3.

579   Votintseva, A. A., Bradley, P., Pankhurst, L., Elias, C. del O., Loose, M., Nilgiriwala, K., et al. 2017. Same-

580   Day Diagnostic and Surveillance Data for Tuberculosis via Whole-Genome Sequencing of Direct

581   Respiratory Samples. J. Clin. Microbiol. 55:1285–1298.

582   Wick, R. R., Judd, L. M., and Holt, K. E. 2018. Deepbinner: Demultiplexing barcoded Oxford Nanopore reads

583   with deep convolutional neural networks. PLOS Comput. Biol. 14:e1006583.

584   Zhang, Z., Schwartz, S., Wagner, L., and Miller, W. 2000. A greedy algorithm for aligning DNA sequences.

585   J. Comput. Biol. J. Comput. Mol. Cell Biol. 7:203–214.

586
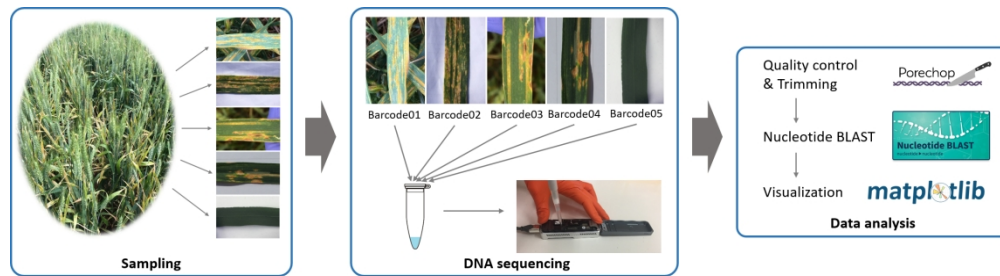
Fig. 1. The Nanopore sequencing workflow for fungal pathogen identification. 1) Sampling: plant pathologists collected wheat leaf samples from disease nursery fields, with visual confirmation of disease symptoms for four infection types (three single fungal infections and one double fungal infection). Fungicide-treated wheat leaves of the same variety and growth stage were collected as a control treatment. 2) DNA sequencing: to minimise operator bias. DNA was extracted from all five samples without knowledge of the causative agent(s) affecting each sample. We labelled samples with DNA barcodes through a PCR reaction step before sample pooling and 1D library preparations (SQK-LSK107); 3) Data analysis: We quality-controlled sequences based on Phred quality scores and trimmed adapter sequence using Porechop (Wick, Judd, and Holt 2018). We performed BLAST (Altschul et al., 1990) searches using processed reads and identified the best hits. We summarized the total sequence length of hits to each species using the python *matplotlib* module. Images used in step 3 are from the official websites of each program.

404x111mm (150 x 150 DPI)
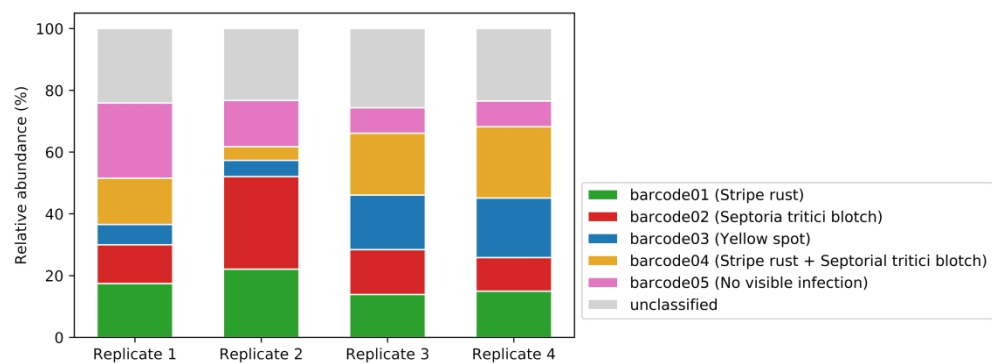
Fig. 2. Uneven proportions of barcode-labelled DNAs across all biological replicates. Columns indicate four biological replicates from five different treatment types. The y-axis indicates relative abundance of total sequence lengths labelled with one of five DNA barcodes within each sample set. Each barcode represents one wheat sample. Reads without a detectable barcode were labelled as 'unclassified'.
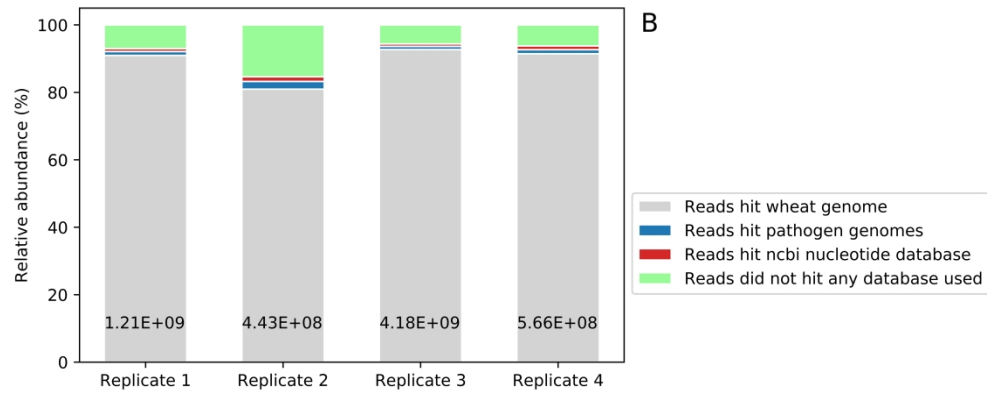
209x78mm (600 x 600 DPI)

Fig. 3. Wheat is predominantly identified in the whole genome sequencing analysis using BLAST. The graph shows the relative sequence length per species identified using a 'winner-takes-all' approach when searching the custom genome reference database. Wheat (grey), fungal pathogen genomes (blue), NCBI nucleotide (nt) database (red) and unmapped (light green) (see results and methods for details). The numbers at the bottom of each bar show the total sequence length (bp) of processed reads for each replicate. Replicates 1, 3, 4 showed similar classification levels while replicate 2 contained a higher number of unclassified sequences.

216x87mm (600 x 600 DPI)
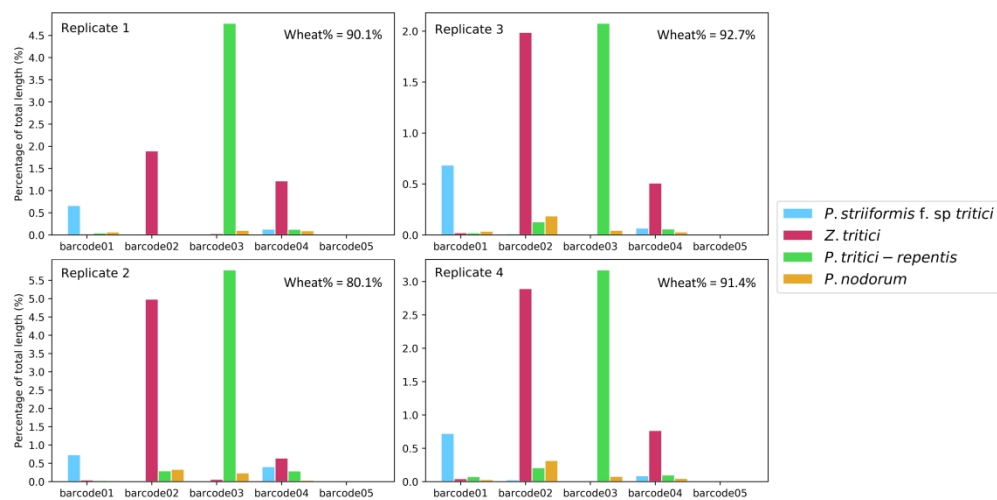
Fig. 4. The total sequence length per pathogen genome per barcode correlated with the phenotypic disease identification. The graphs show the relative sequence length assigned to each pathogen genome using a 'winner-takes-all' approach in the BLAST analysis against the custom genome reference database. The proportion of reads hitting the wheat host genome are listed in the top-right corner of each graph. Unassigned reads are omitted from the visualization.
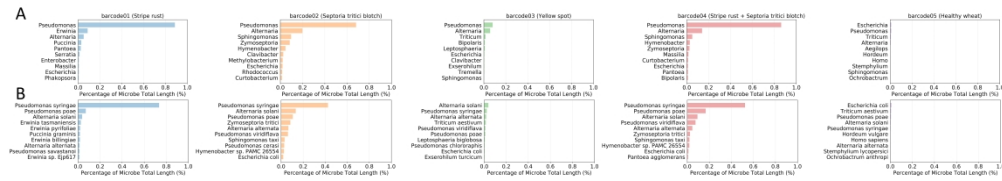
209x105mm (600 x 600 DPI)

Fig. 5. Fungal infections of wheat increase the microbial population, especially for *Puccinia striiformis* f. sp. *tritici* and *Z. tritici* infections. Percentage of sequences that did not hit the custom genome reference database but hit the non-redundant NCBI nt database combined from all replicates, and grouped by genus (A) or species (B). Different colours represented different treatment types. Reads that were assigned to the restricted database were eliminated, and the top ten highest classifications were displayed on the y-axis for each replicate. Genus *Pseudomonas* represented the highest proportion across all barcodes containing *Puccinia striiformis* f. sp. *tritici* or *Z. tritici*, with *Pseudomonas syringae* as the highest at the species level.

2994x575mm (72 x 72 DPI)

Supplementary Table S1. Basic statistical comparison of each sequencing replicate. The barcode column indicates DNA barcodes used in the library preparation for each replicate.

Replicate	No. of Q[a] > 7 reads/ No. of total reads (bp[b], %)	Q > 7 total read length/ Total read length (bp, %)	Total read length after Porechop (bp)	Mean read length from Q > 7 reads (bp)	Barcodes	Disease confirmed by pathologist
1	847,469/983,698 (86.2%)	1,587,046,441/1,707,799,548 (92.9%)	1,206,433,874	1872.7	barcode02	stripe rust
					barcode03	yellow leaf spot
					barcode04	None
					barcode05	stripe rust and septoria tritici blotch
					barcode06	septoria tritici blotch
2	568,686/659,039 (86.3%)	785,073,330/837,189,225 (93.8%)	442,667,426	1380.5	barcode07	stripe rust
					barcode08	septoria tritici blotch
					barcode09	None
					barcode10	stripe rust and septoria tritici blotch
					barcode11	yellow leaf spot
3	1,932,416/2,123,611 (91.0%)	5,051,730,065/5,304,946,873 (95.2%)	4,175,446,769	2614.2	barcode01	stripe rust
					barcode02	septoria tritici blotch
					barcode03	None
					barcode04	stripe rust and septoria tritici blotch
					barcode05	yellow leaf spot
4	223,192/242,106 (92.2%)	681,656,642/736,410,201 (92.6%)	565,771,763	3054.1	barcode01	stripe rust
					barcode02	septoria tritici blotch
					barcode03	None
					barcode04	stripe rust and septoria tritici blotch
					barcode05	yellow leaf spot

[a] Q: Phred quality

[b] bp: basepair

Supplementary Table S2. An example BLAST analysis against the non-redundant NCBI nt database using the total biological replicate 4 reads. This table presented the top 17 genera classified by BLAST in each treatment group (barcode). The yellow colored boxes represent the major pathogens as suggested by the pathologist in each sample, while the green colored boxes represented the bacterial genus *Pseudomonas* that we suggested is associated with different fungal pathogens.

| Barcode01 (18,909 reads)		Barcode02 (13,227 reads)		Barcode03 (23,489 reads)		Barcode04 (27,474 reads)		Barcode05 (11,256 reads)	
Blast hit Genera	# of reads	Blast hit Genera	# of reads	Blast hit Genera	# of reads	Blast hit Genera	# of reads	Blast hit Genera	# of reads
Triticum	17343	*Triticum*	11826	*Triticum*	21658	*Triticum*	25352	*Triticum*	10537
Aegilops	420	*Aegilops*	310	*Pyrenophora*	520	*Aegilops*	585	*Aegilops*	297
Pseudomonas	254	*Zymoseptoria*	143	*Aegilops*	467	*Pseudomonas*	178	*Hordeum*	33
Erwinia	32	*Pseudomonas*	89	*Hordeum*	50	*Zymoseptoria*	78	*Secale*	21
Hordeum	32	*Alternaria*	44	*Secale*	49	*Hordeum*	51	*Oryza*	7
Secale	24	*Hordeum*	37	*Alternaria*	39	*Alternaria*	40	*Escherichia*	3
Puccinia	18	*Parastagonospora*	23	*Parastagonospora*	7	*Secale*	23	*Alternaria*	2
Alternaria	15	*Hymenobacter*	12	*Oryza*	5	*Hymenobacter*	15	*Stipa*	2
Pantoea	11	*Sphingomonas*	10	*Bipolaris*	4	*Oryza*	13	*Brachypodium*	2
Brachypodium	7	*Escherichia*	9	*Brachypodium*	4	*Brachypodium*	10	*Lolium*	1
Escherichia	5	*Secale*	9	*Leptosphaeria*	4	*Curtobacterium*	7	*Streptomyces*	1
Oryza	5	*Oryza*	5	Unclassified	3	*Sphingomonas*	7	*Sorghum*	1
Serratia	3	*Pyrenophora*	4	*Setosphaeria*	2	*Escherichia*	5	*Sasa*	1
Massilia	3	*Rahnella*	4	*Pseudomonas*	2	*Parastagonospora*	5	*Saccharum*	1
Unclassified	3	*Leptosphaeria*	4	*Verticillium*	2	*Puccinia*	5	*Piptatherum*	1
Taeniatherum	2	*Thinopyrum*	2	*Amblyopyrum*	2	*Pantoea*	5	*Nannospalax*	1
Stipa	2	*Brachypodium*	2	*Lolium*	1	*Pyrenophora*	4	*Agrostis*	1
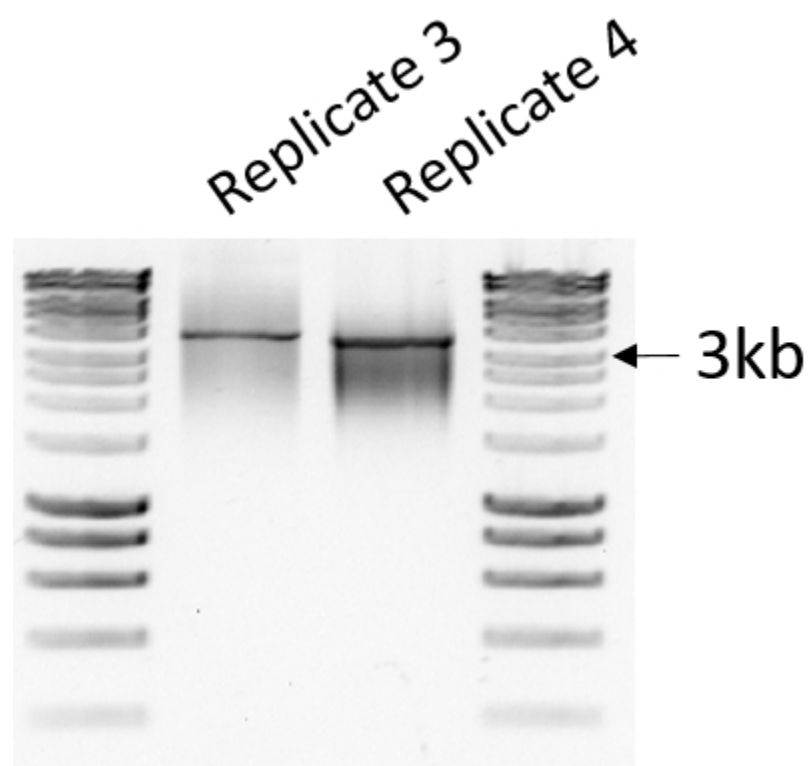
Supplementary Fig. S1. Example of DNA quality control by 1% agarose gel electrophoresis. About 200 nanograms of DNA from replicates 3 and 4 prior to the adapter ligation step of the sequencing library preparation were loaded onto the gel. The molecular weight standards are HyperLadder 1 kb (Bioline, Catalog number BIO-33025).

71x69mm (150 x 150 DPI)
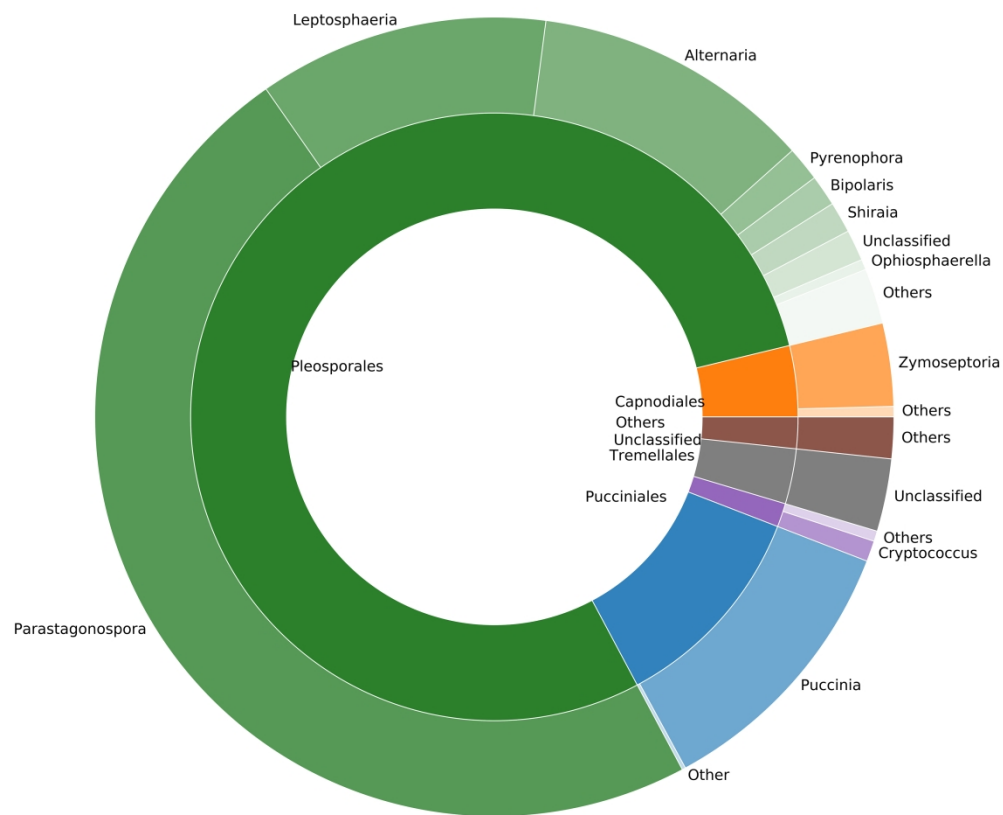
Supplementary Fig. S2. BLAST analysis against the non-redundant NCBI nt database using the reads assigned to *Parastagonospora nodorum* from BLAST against the restricted database. Arc angle is proportional to the number of reads assigned to the indicated genus or order. The unclassified (grey) arc angle refers to the proportion of reads assigned as unclassified by BLAST analysis. Genus or orders assigned to two reads or less are grouped as 'others' under each category.

1312x1087mm (72 x 72 DPI)
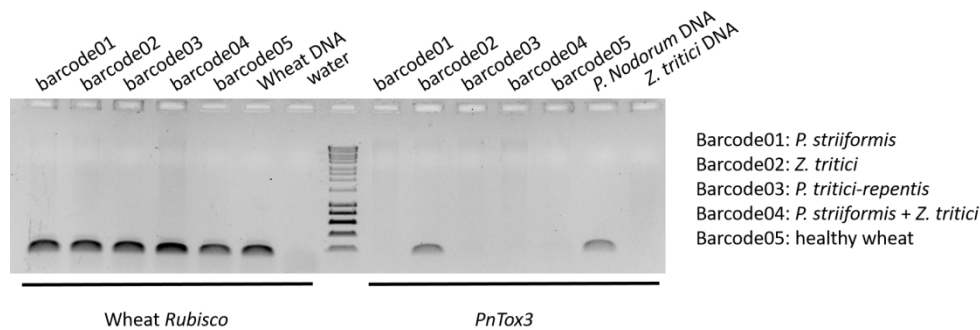
Supplementary Fig. S3. PCR amplification of the *Parastagonospora nodorum Tox3* gene from the DNA sample of replicate 3. The wheat *Rubisco* gene was used as a control for the quality of DNA in each sample. The positive control for *Tox3* amplification is pure genomic DNA from *Parastagonospora nodorum* strain SN15, and the negative control is pure genomic DNA from *Z. tritici*. The PCR reaction used NEB Q5 High-Fidelity 2X Master Mix (Catalog Number M0492S) following the manufacturer's protocol and the amplification conditions were: 30 s at 98°C; 30 cycles of 10 s at 98°C, 25 s at 56°C and 10 s at 72°C; then a final extension for 2 min at 72°C before holding at 4°C. The primer sequences for the wheat *Rubisco* gene are: Forward: AGCAAGGTTGGCTTTGTCTT, Reverse: TTAACCTCCTCCACCTCGTT. The primer pair used for the *Tox3* gene are: Forward: AATGTCGACCGTTTTGACC, Reverse: GGTTGCCGCAGTTGATATAA obtained from Lin, Chooi, and Solomon (2018).
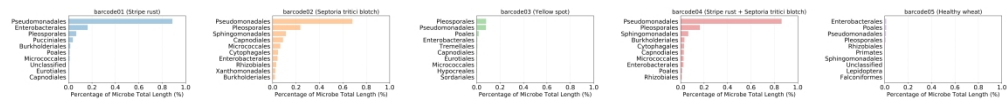
309x112mm (150 x 150 DPI)
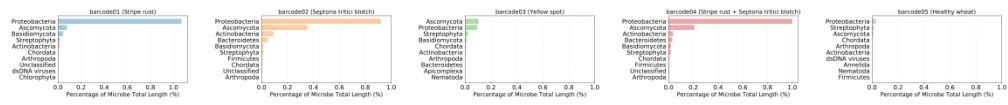
Supplementary Fig. S4. Percentage of sequences that did not hit the custom genome reference database but hit the non-redundant NCBI nt database combined from all replicates, and grouped by phylum. Reads that are assigned to the restricted database were eliminated, and the top ten highest classifications were displayed for each replicate. Different colours represented different barcodes.
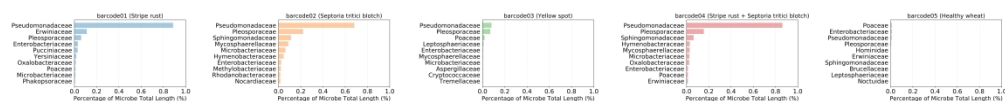
4398x408mm (72 x 72 DPI)
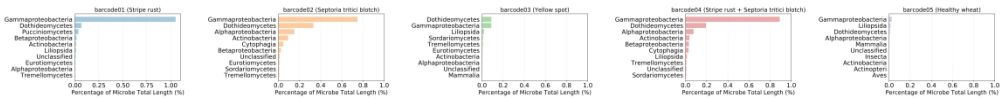
Supplementary Fig. S5. Percentage of sequences that did not hit the custom genome reference database but hit the non-redundant NCBI nt database combined from all replicates, and grouped by class. Reads that are assigned to the restricted database were eliminated, and the top ten highest classifications were displayed for each replicate. Different colours represented different barcodes.

4398x408mm (72 x 72 DPI)

Supplementary Fig. S6. Percentage of sequences that did not hit the custom genome reference database but hit the non-redundant NCBI nt database combined from all replicates, and grouped by order. Reads that are assigned to the restricted database were eliminated, and the top ten highest classifications were displayed for each replicate. Different colours represented different barcodes.

4399x408mm (72 x 72 DPI)

Supplementary Fig. S7. Percentage of sequences that did not hit the custom genome reference database but hit the non-redundant NCBI nt database combined from all replicates, and grouped by family. Reads that are assigned to the restricted database were eliminated, and the top ten highest classifications were displayed for each replicate. Different colours represented different barcodes.

4399x408mm (72 x 72 DPI)