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

Yiheng Hu and **Gamran S. Green**, Research School of Biology, The Australian National University, Acton ACT2601, Australia; **Andrew W. Milgate**, Wagga Wagga Agricultural Institute, New South Wales Department of Primary Industry, Wagga Wagga NSW 2650, Australia; **Eric A. Stone**, ANU-CSIRO Centre for Genomics, Metabolomics and Bioinformatics, Acton ACT2601, Australia; **John P. Rathjen**† and **Benjamin Schwessinger**†, Research School of Biology, The Australian National University, Acton ACT2601, Australia.

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

Funding: This project was supported by The Hermon Slade Foundation, The Australian National University and Australian Research Council.

Current address of G. S. Green: School of Biotechnology and Biomolecular Sciences, The University of New South Wales, Kensington, NSW2052, Australia.

ABSTRACT

Fungal diseases of plants are responsible for major losses in agriculture, highlighting the need for rapid and accurate identification of plant pathogens. Disease outcomes are often defined not only by the main pathogen but are influenced by diverse microbial communities known as the microbiome at sites of infection. Here we present the first use of whole genome sequencing

from a portable DNA sequencing device as a method for the detection of fungal pathogens from wheat (*Triticum aestivum*). The data revealed that our method is robust and applicable to the diagnosis of fungal diseases including wheat stripe rust (caused by *Puccinia striiformis* f. sp. *tritici*), septoria tritici blotch (caused by *Zymoseptoria tritici*) and yellow leaf spot (caused by *Pyrenophora tritici repentis*). We also identified a bacterial genus *Pseudomonas* co-present with *Puccinia* and *Zymoseptoria* infections but not *Pyrenophora* infections. One limitation of the method is the over-representation of redundant wheat genome sequences in the sample. This could be addressed by amplicon-based sequencing approaches in future studies. Our work outlines a new approach for the detection of a broad range of plant pathogens and associated microbes using a portable sequencer, providing the basis for future development of a system for on-site disease monitoring.

Plant diseases, especially those caused by fungal pathogens, jeopardise global crop biosecurity. Increased global trade, human migration, environmental changes and the accelerated emergence of virulence have been identified as causes for increasing prevalence of diseases in plants and animals (Fisher et al. 2012). To prevent and manage diseases, rapid detection and identification of their causal agents are crucial. This is particularly important for outbreak management during the incursion of virulent new isolates that overcome classical methods of control such as the use of resistant crop varieties (Bhattacharya 2017; Pretorius et al. 2000). Traditional methods for crop disease diagnosis rely largely on the expertise of pathologists who identify diseases initially by eye in the field. While molecular methods of pathogen detection

45 also exist, these are generally only capable of identifying a limited number of species (De Boer
46 and López 2012). Recently, other methods based on the use of optical sensors have been
47 developed for the detection, identification, and quantification of plant diseases (Mahlein 2015).
48 Conventional pathology, however, remains the primary means of pathogen identification and
49 harbors certain limitations: 1. It relies heavily on the physical appearance of disease symptoms;
50 2. It has difficulty detecting pathogens which do not infect aerial parts of the plant or prior to
51 the emergence of visible disease symptoms; 3. Co-infections may lead to conflicting visual
52 symptoms; 4. No method currently exists for the rapid identification of unknown pathogens
53 during outbreaks – a limitation epitomized by the recent, devastating incursion of South
54 American races of wheat blast fungus in Bangladesh (Callaway 2016).
55 DNA sequencing-based methods have been applied to address limitations to detection and
56 monitoring of environmental microbial communities (Nezhad 2014). Such assays rely largely on
57 amplification and sequencing of small conserved regions called amplicons from pathogen
58 genomes such as the internal transcribed spacer region (ITS), *Elongation Factor 1 alpha* (*EF1α*)
59 and *Beta-tubulin* genes for fungal species and 16S ribosomal RNA sequences for bacteria (Janda
60 and Abbott 2007; Raja et al. 2017). Closely-related species can be discriminated based on single
61 nucleotide polymorphisms (SNPs) within the amplicon sequences. This has been used for
62 example to separate risks and different epidemiological or regulatory outcomes posed by the
63 species in question (Chemidlin Prévost-Bouré et al. 2011; Thonar, Erb, and Jansa 2012).
64 Amplicon sequencing also allows characterisation of the microbial community - the microbiome
65 - within a sample. The microbiome can impact plant function for example on shifting flowering
66 time (Panke-Buisse et al. 2015) and promoting the colonization of fungal pathogen such as

67 *Botrytis cinerea* (Ritpitakphong et al. 2016). The microbiome is likely to contain organisms that
68 interact with disease-causing agents that could potentially be harnessed for biocontrol
69 strategies (Ellis 2017). However, restrictive read lengths generated by the Illumina sequencing
70 platform limits sequence-based identification of microbial species. Only 500 base pairs (bp) of
71 each DNA amplicon can be sequenced in each reaction. This means that only that genetic
72 variation present within the 500 bp is available for taxonomic assignment of the organism in
73 question (Benítez-Páez and Sanz 2017). Use of PCR also biases abundance of individual
74 amplicons, making estimation of population frequencies problematic (Kennedy et al. 2014).
75 Third-generation DNA sequencing platforms have developed rapidly during the last half-decade,
76 overcoming the read length issue inherent in second-generation sequencing methods and
77 facilitating the identification of microbial species (Benítez-Páez and Sanz 2017). Two competing
78 long-read sequencing technologies are currently in use. Single molecular real-time (SMRT)
79 sequencing (PacBio) enables detection of the incorporation of a fluorescent nucleotide into a
80 single molecule during DNA synthesis within a nanoscale observation chamber (Ardui et al.
81 2018). Nanopore sequencing (MinION; Oxford Nanopore Technologies (ONT)) enables direct
82 sequencing of native DNA by measuring the voltage change across an artificial membrane when
83 a single DNA molecule passes through a nanopore embedded in the membrane within a
84 flowcell (Jain et al. 2016). Advantages of nanopore sequencing include its relatively low cost,
85 pocket-size form, and real-time data processing. This means that it is a mobile technology
86 independent of large sequencing facilities, and as such has been useful for on-site sample
87 sequencing in extreme environments such as microbial paleo mats in the Antarctic (Johnson et
88 al. 2017), or providing rapid bacterial pathogen identification from clinical samples with a

89 turnaround time of six hours (Charalampous et al. 2018), both of which could not be achieved
90 with previous technologies. Most importantly, this technology provides dramatically longer
91 sequencing reads (~10 kb) than second-generation sequencing platforms such as Illumina.
92 Longer reads contain more genetic information, allowing more accurate identification of the
93 organism from which the sequence is derived (Benítez-Páez, Portune, and Sanz 2016; Benítez-
94 Páez and Sanz 2017). Although recent studies have shown important advances in identification
95 of bacterial pathogens using nanopore sequencing from clinical samples (Mitsuhashi et al. 2017;
96 Schmidt et al. 2017), little has been done in the detection of plant pathogens and microbiome
97 profiling of diseased plants using this approach.

98 Here, we tested nanopore sequencing for disease diagnosis in field-grown wheat and explored
99 its potential for microbiome profiling. We designed a whole-genome sequencing workflow for
100 the identification of major pathogen species from field samples, as well as the profiling of
101 microbes present in these samples to the species level. As a proof-of-concept, we applied our
102 sequencing workflow to wheat leaves infected with known fungal pathogens sampled from the
103 field, and a fungicide treated control. By mapping sequencing reads to databases, we
104 successfully detected all fungal diseases present in the infected samples including stripe rust
105 (caused by *Puccinia striiformis* f. sp. *tritici*), septoria tritici blotch (caused by *Zymoseptoria tritici*)
106 and yellow leaf spot (caused by *Pyrenophora tritici repentis*). We also conducted a preliminary
107 characterisation of the microbiomes associated with those diseases, and found bacterial genus
108 *Pseudomonas* co-present with stripe rust and septoria tritici blotch. Our results suggest this
109 technology has considerable potential for adaptation to a broad range of crop disease
110 diagnoses and environmental monitoring applications.

111

112 **MATERIALS AND METHODS**

113 **Sample collection**

114 Wheat cultivar Crusader (Advanta Seeds Pty. Ltd.) seeds were planted in the disease nursery
115 field on May 2016 at the Wagga Wagga Agricultural Institute (New South Wales (NSW),
116 35°02'24.3"S 147°19'09.5"E) with daily irrigation. Wheat straw infected the previous season
117 with septoria tritici blotch and yellow leaf spot separately then allowed to stand in the field to
118 promote development of pseudothecia were used to manually inoculate plants in the nursery
119 field at the tillering growth stage (4-5 weeks post emergence) while stripe rust infections
120 occurred naturally at jointing stage (7-9 weeks post emergence). Healthy or diseased plants
121 were diagnosed via symptomatology by a plant pathologist from the NSW Department of
122 Primary Industries before leaf samples were collected during September 2016. For sampling, an
123 entire tiller from each plant was removed and stored on ice. Samples from three separate
124 plants infected with the same disease were pooled and wrapped in aluminium foil before
125 shipping to the Rathjen laboratory, Research School of Biology, The Australian National
126 University. Two representative leaves from each symptomatology group containing visible
127 disease symptoms or fungicide treated control were cut into pieces roughly weighing 100 mg
128 and transferred 2 mL Eppendorf tubes containing one 5 mm metal bead. All tubes were labelled
129 and stored at -80°C prior to DNA extraction.

130 **DNA extraction**

131 The DNA extraction protocol is described in (Hu 2016). Briefly, 100 mg leaf tissue was
132 homogenized and cells were lysed using cetyl trimethylammonium bromide (CTAB, Sigma-

133 Aldrich) buffer (added RNAse T1, Thermo Fisher, 1000 units per 1750 µL), followed by a
134 phenol:chloroform:isoamyl alcohol (25:24:1, Sigma-Aldrich) extraction to remove protein and
135 lipids. The DNA was precipitated with 700 µL isopropanol, washed with 1 mL 70 % ethanol
136 before dried for 5 minutes at room temperature and resuspended in TE buffer containing 1mM
137 Tris and 1mM EDTA at pH8. DNA extractions were performed four times independently, with
138 the operator blind to the identity of each sample. After extraction, DNA was purified with one
139 volume of Agencourt AMPure XP Beads (Beckman Coulter, Inc.) using the manufacturer's
140 protocol and stored at 4°C. Quality and average size of genomic DNA was visualized by gel
141 electrophoresis using 1% agarose gel for 1 hour at 100 volts. DNA was quantified by NanoDrop
142 and Qubit (Life Technologies) following the manufacture's protocol.

143 **Library construction and sequencing**

144 Sequencing libraries were constructed as previously described (Hu and Schwessinger 2018) with
145 the omission of the DNA shearing step. Briefly, a dA-tailing reaction was performed on 700 ng
146 genomic DNA from each sample using the NEBNext Ultra II End-repair/dA-tailing kit (New
147 England Biolabs). Repaired DNA was washed on AMPure XP Beads on a magnetic rack using 70 %
148 ethanol, and eluted with 31 µL nuclease-free water. Thirty µL eluate was used for barcode
149 adapter ligation via the NEB Blunt/TA Ligase Master Mix (New England Biolabs) and the Barcode
150 Adapter Mix (ONT), followed by another wash step on AMPure XP Beads. Twenty ng of the
151 adapter-ligated template in 2 µL of nuclease free water was added to each barcoding PCR
152 reaction with 50 µL LongAmp Taq 2X Master Mix (New England Biolabs), 46 µL water and 2 µL
153 PCR Barcodes (ONT). The PCR was performed as follows: Denaturation at 95°C for 3 min,
154 followed by 15 cycles of 95°C/15 s; 62°C/15 s; 65°C/150 s, and a final extension step at 65°C for

155 5 min. Barcoded DNAs were purified using 1 volume of AMPure XP Beads then pooled into 1 µg
156 fractions in a total of 45 µL to which 5 µL of DNA CS control (ONT) was added. A further dA-
157 tailing reaction was performed as above followed by purification on AMPure XP Beads, with the
158 DNA eluted in 31 µL nuclease free water. Adapter ligation was carried out using 30 µL end-
159 prepped DNA, 20 µL Adapter Mix (ONT) and 50 µL NEB Blunt/TA Ligase Master Mix. The
160 adapter-ligated DNA library was then purified on AMPure XP Beads using the Adapter Beads
161 Binding buffer (ONT) for washing, and eluted in 15 µL of Elution Buffer (ONT).
162 Sequencing reactions were performed on a MinION Flowcell (R9.4, ONT) connected to a MK1B
163 device (ONT) operated by the MinKNOW software (version 1.5.2 and version 1.6.11). Each
164 Flowcell was primed with 1 mL of priming buffer which consists of 480 µL Running Buffer Fuel
165 Mix (RBF, ONT) and 520 µL nuclease free water. 12 µL of amplicon library was added to a
166 loading mix including 35 µL RBF, 25.5 µL Library Loading Beads (ONT) and 2.5 µL water for a
167 final volume of 75 µL added to the Flowcell via the SpotON sample port. The
168 "NC_48Hr_sequencing_FLO-MIN106_SQK-LSK107" protocol was executed by the MinKNOW
169 after loading the library. The sequencing run was restarted after 12 h and stopped after 48 h.
170 **Sequencing analysis**
171 Raw fast5 files were processed by the Albacore 2.0.2 software (ONT) for basecalling, barcode
172 de-multiplexing and quality filtering (Phred quality score of >7). All reads that passed quality
173 filtering for each barcode were treated in parallel as follows: firstly, NanoLyse (De Coster et al.
174 2018) was used to remove lambda phage control DNA sequences; secondly, barcode and
175 adapter sequences were trimmed from the ends of reads using Porechop
176 (https://github.com/rrwick/Porechop). To identify middle adapter sequence by Porechop, a 95%

177 threshold was set, and reads with middle adapter sequences were split into two sequences;
178 thirdly, seqtk (https://github.com/lh3/seqtk) was used to convert the processed fastq file into
179 fasta format for further analysis. Two BLAST (Basic Local Alignment Search Tool, Altschul et al.,
180 1990) searches were performed on the fasta files for each barcode: first, against a wheat
181 reference genome (Clavijo et al. 2017, accessed 09/2017) and a custom database containing the
182 genomes of the four fungal species responsible for the diseases studied here: *Puccinia*
183 *striiformis* f. sp. *tritici*: Schwessinger et al. (2018), accessed 09/2017; *Z. tritici*: Goodwin et al.
184 (2011), version 2, accessed 09/2017; *Pyrenophora tritici-repentis*: Manning et al. (2013),
185 unmasked assembly, accessed 09/2017. The genome of a further species suggested to be
186 absent from the studied wheat growing region, *Parastagonospora nordorum*, was included as a
187 negative control (Hane et al. 2007, accessed 09/2017). Fasta files representing raw sequence
188 reads were searched against the reference genomes using BLAST. The e-value cut-off for each
189 BLAST search was set to 0.01, and only the top hit was taken for further analysis. Reads that did
190 not hit the reference genomes were captured using the filterbyname.sh script from the bbmap
191 script package (https://sourceforge.net/projects/bbmap/). These reads were searched against
192 the National Center for Biotechnology Information (NCBI) nucleotide (nt) database
193 (ftp://ftp.ncbi.nlm.nih.gov/blast/db/) using BLAST with the same settings as previously. Those
194 reads providing hits were separated and retained again using filterbyname.sh script. Two
195 Python scripts (QC_and_BLAST.py and creating_final_dataframe.py) that combine all steps from
196 quality control to BLAST outputs and construct dataframes containing statistical and
197 information from BLAST searches for each read were deposited at
198 https://github.com/Yiheng323/Pathogen_Detection_and_Microbiome_scripts. All basecalled

199 fastq data for each sequencing run were deposited at NCBI Short Reads Archive (SRA) under the
200 accession number PRJNA493553. The sequencing summary files and final dataframes as .tab files
201 for each replicate were deposited at https://doi.org/10.6084/m9.figshare.7138262. The python
202 scripts used for generating Fig. 2 – 5, Supplementary Fig. S1 – S5 were also deposited at
203 https://github.com/Yiheng323/Pathogen_Detection_and_Microbiome_scripts.
204

205 **RESULTS**

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

207 We designed a pipeline for identification of three known wheat pathogens from infected field
208 samples using MinION sequencing. We infected three experimental field plots as disease
209 nurseries with their corresponding pathogens (*Puccinia. striiformis* f. sp. *tritici* for stripe rust
210 disease, *Z. tritici* for Septorial tritici blotch and *Pyrenophora. tritici-repentis* for Yellow leaf spot)
211 commonly found in New South Wales, Australia, and conducted sampling during the 2016-17
212 growing season (September). The sequencing experiment included three steps (Fig. 1): 1.
213 Pathologists collected leaf samples from each disease nursery field, with visual confirmation of
214 disease symptoms from the three single infections plus one double infection (*Puccinia*
215 *striiformis* f. sp. *tritici* and *Z. tritici*), and a negative control which was the same wheat variety
216 treated with commercial fungicides at the 10- to 11-week growth stage (three months post
217 germination), before shipping to the laboratory on ice for processing. 2. We extracted DNA
218 without knowing the identity of each sample to eliminate potential bias from operators. 3. We
219 barcoded each sample to enable pooling of all five samples to reduce sequencing costs. The five
220 wheat leaf samples were barcoded as follows: leaves infected with one of the three fungal

221 pathogens (barcodes 1-3); leaves infected with two pathogens (barcode 4); and no visible
222 infection (barcode 5). We performed step 2 and 3 four times as biological replicates, using a
223 MinION 1D PCR barcoding workflow (EXP-PBC001 and SQK-LSK107). The raw data were
224 converted into sequence reads using the software Albacore, with a quality (Q) score for each
225 base estimating the likelihood of the identified base being correct using the Phred scale (Ewing
226 et al., 1998; Ewing and Green, 1998). Based on read mean Q scores, we grouped reads into pass
227 (Q score > 7) or fail (Q score <= 7). We binned reads based on the barcode sequences identified
228 by Albacore at both ends of each sequence. If no barcode sequence could be identified, we
229 named the corresponding read as unclassified. Overall, ~92-95% of the total sequence length of
230 all four sequencing runs passed quality filtering (Supplementary Table 1). The total sequence
231 length from the four runs varied from 682 to 5100 megabase pairs (Mbp). The mean read
232 length ranged from 1380 to 3054 bp. Proportions of each barcode in each run were relatively
233 even, although 23.3-25.7% of the sequences were unclassified from all four runs (Fig. 2).

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

235 We performed Basic Local Alignment Search Tool (BLAST, Altschul et al., 1990) searches for all
236 quality-filtered reads against a manually constructed reference database. This contained the
237 wheat genome and four fungal genomes including *Puccinia striiformis* f. sp. *tritici*, *Z. tritici*,
238 *Pyrenophora tritici-repentis* and *Parastagonospora nodorum*. These represent the host plant
239 genome, three of the fungal species present in the field samples, and one fungal species
240 (*Parastagonospora nordorum*) not expected to be present in the sampling area included as a
241 negative control. We took the hit with the lowest expectation value (e-value) from BLAST

242 analysis of each read as it represented the most significant match. The majority of reads are
243 wheat sequences, comprising total sequence length of 90.1% in flowcell 1, 80.1% in flowcell 2,
244 92.7% in flowcell 3 and 91.4% in flowcell 4 (Fig. 3). We found that healthy leaf samples (barcode
245 5) contained no or minor total length percentages of pathogen-related sequences in all four
246 replicates (< 0.01%, Fig. 4). This indicates that our analysis method is robust and does not lead
247 to high background signals in healthy wheat samples (barcode 5). Significantly higher
248 proportions of total sequencing length (0.5–5.7%) from infected samples (barcode 1-4) were
249 assigned as derived from pathogen genomes. In all cases, the pathogen genome identified by
250 the highest total sequence length corresponded to the pathogen causing the disease as
251 identified by symptoms (Fig. 4). This indicates that we are clearly able to identify the disease
252 causing pathogen via our sequencing workflow in the case of single infections. Infections with
253 two pathogens (barcode 4) were identified less clearly. In barcode 4 from all replicates, the
254 total sequence lengths assigned to *Puccinia striiformis* f. sp. *tritici* were between 0.1% and 0.4%,
255 which were close to or lower than the proportion of total length assigned to the negative
256 control species *Parastagonospora nodorum.* The percentages of total sequence length by
257 barcode that identified pathogen genomes were similar across all replicates, with around 0.6%
258 of barcode 1 sequences from *Puccinia striiformis* f. sp. *tritici*, 1.9-5.0% of barcode 2 sequences
259 from *Z. tritici*, 2.1-5.7% of barcode 3 sequences from *Pyrenophora tritici-repentis*, and 0.1-0.4%
260 of barcode 4 sequences contained *Puccinia striiformis* f. sp. *tritici* and 0.6-1.2% from *Z. tritici*.
261 Unexpectedly, we also identified *Parastagonospora nodorum* from different samples in all four
262 technical sequencing replicates. To ascertain whether *Parastagonospora nodorum* was truly
263 present, we extracted all reads that hit the *Parastagonospora nodorum* genome by BLAST

264 analysis against the reference database. We used these reads to perform an open BLAST search
265 against the entire NCBI non-redundant (nt) nucleotide database which contains all publically
266 available genome sequences. We found that 344 of 915 reads were assigned to genus
267 *Parastagonospora*, including 339 reads identified *Parastagonospora nodorum* SN15 genome
268 (Supplementary Fig. S1). A further 371 reads hit related species, including 226 from the same
269 *Pleosporales* order: *Leptosphaeria* spp. (84 reads), *Alternaria* spp. (81), *Pyrenophora* spp. (10),
270 *Bipolaris* spp. (9), *Shiraia* spp. (9) and other genus (Supplementary Fig. S1). The remaining 200
271 reads fell under the given e-value threshold (0.01) and failed to be classified. The difference
272 between both BLAST searches potentially stems from the relatively high error rate of the
273 MinION sequencer (Jain et al., 2018; Laver et al., 2015). Nevertheless, this result supports the
274 finding that *Parastagonospora nodorum* is present in the sample from the Wagga Wagga wheat
275 growing region despite the disease not purposefully inoculated in the trial.

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

277 To obtain an overall picture of the microbiome profile in wheat leaves infected by different
278 fungal pathogens, we attempted to identify other microbial species present on the leaves. To
279 do this, we searched the non-redundant NCBI nucleotide database with reads that were not
280 assigned to any of the five reference genomes present in the initial restricted database. We
281 found that the proportion of reads assigned to a genome in the restricted and the nt-NCBI
282 databases were high for flowcells 1, 3 and 4 (92.9, 94.3 and 93.8%, respectively), but only 84.6%
283 for flowcell 2, likely due to technical issues. Overall, fewer than 1.4% of passed reads from each
284 flowcell identified microbiota other than the major pathogens in the restricted database in and
285 on the sampled wheat leaves.

286 We pooled the reads from each barcode in all four biological replicates, and calculated the
287 percentage of total length that identified a specific genus in each sample using NCBI taxonomic
288 identifiers (Fig. 5A). Genus *Pseudomonas* took up the highest proportion of reads in all diseased
289 samples, ranging from 0.08% to more than 0.88%, but was nearly absent in healthy wheat
290 (barcode 5, 0.01%). Genus *Alternaria* is also found in all samples with a proportion between
291 0.05% to 0.2% in diseased samples and 0.005% in healthy wheat. Uniquely in barcode 1,
292 bacterial *Erwinia* comprised 0.09% of the total read length as represented by 219 reads and the
293 second highest proportion. In contrast, *Erwinia* only represented 0.003% and 0.001% in barcode
294 2 and 5, and none in barcode 3 and 4, while *Alternaria* ranked second highest in other diseased
295 samples (barcode 2 - 4). In barcodes 2 and 4, hits on *Z. tritici* were still present and may
296 represent the six *Z. tritici* strains genomes in the NCBI nt database not included in the *Z. tritici*
297 genome we used in the first BLAST analysis. At the species level (Fig. 5B), reads identifying
298 *Pseudomonas syringae* were much more frequent in samples infected with *Puccinia striiformis* f.
299 sp. *tritici* (barcode 1, 0.59%) and/or *Z. tritici* (barcode 2, 0.33%). This was not seen in the sample
300 infected with *Pyrenophora tritici-repentis* (barcode 3), where only 0.02% reads matched
301 *Pseudomonas syringae*. Populations of other *Pseudomonas* species such as *Pseudomonas poae*
302 followed a similar trend to the relative *Pseudomonas syringae* accross different wheat samples,
303 presenting much more in barcode 1, 2, 4 than in barcode 3 and 5. This finding corresponds to
304 other reports in which *Pseudomonas* spp. were associated with fungal infections of wheat (Al-
305 Sallami et al. 1997; Mehrabi et al. 2016). Considering the upper taxonomic classifications,
306 Proteobacteria and Ascomycota were the most abundant bacterial and fungal Phylums across
307 all samples (Supplementary Fig. S2) followed by Actinobacteria and Basidiomycota. At the Class

308 level (Supplementary Fig. S3), either Gammaproteobacteria or Dothiodeomycetes were the
309 most abundant. Overall, the microbiome structure of the upper taxonomy resembles the lower
310 taxonomy (Supplementary Fig. S2-S5), since most sequences were assigned to genus
311 *Pseudomonas*.
312

313 **DISCUSSION**

314 We have developed a WGS workflow which accurately identified three fungal pathogens
315 responsible for three major wheat diseases - stripe rust, septoria tritici blotch and yellow spot -
316 from field samples. We designed a two-step BLAST process to identify major agents in diseased
317 samples and characterise the associated microbiota. We identified a number of *Pseudomonas*
318 species in stripe rust and septoria tritici blotch infected wheat samples but not in yellow spot or
319 non-diseased samples.
320 WGS provides a wealth of information that could be used to accurately identify disease-causing
321 agents, and to begin to describe the composition of the microbial community (Brown et al.
322 2017; Thomas, Gilbert, and Meyer 2012). Without any amplification process which potentially
323 includes biases, the WGS data is obtained directly from the sample extracts, illustrated the
324 most close-to-reality features dependent only by sampling strategies. WGS requires less
325 processing time, omitting the PCR step, providing advantageous for on-site diagnostic purposes
326 where technical simplicity and fast processing time are prioritised. However, it has a significant
327 drawback in that most sequences are derived from the host (wheat) genome, representatively
328 90.1% in flowcell 1, 80.1% in flowcell 2, 92.7% in flowcell 3 and 91.4% in flowcell 4. This is due
329 to our DNA extraction method in which the whole leaf sample was grinded to retain all the

330 microbial population in and on the leaf. For the purpose of microbial species identification,
331 host sequences are unwanted and can overwhelm the community DNA (Sharpton 2014). To
332 enrich microbial sequences, one strategy is to apply intricate molecular methods such as
333 gradient centrifugation and sonication during the extraction process (Bulgarelli et al. 2012; Lim
334 et al. 2014). This can be challenging when applying to environmental samples, especially in field
335 conditions, as these techniques require sophisticated instruments. Another strategy to enrich
336 microbiome sequences involves sequencing microbial marker gene amplicons. With the
337 improvement of sequencing length, species level resolution using long read amplicon
338 sequencing can be achieved in few cases (Benítez-Páez and Sanz 2017; Martijn et al. 2017).
339 Amplicon sequencing seems to be a promising method for species identification, as it requires
340 less input DNA than WGS and marker gene databases contain information from many more
341 species than whole genome databases (Cruaud et al. 2017; Tessler et al. 2017; Ranjan et al.
342 2016). Amplicon sequencing is also possible in applied cases given remote PCR platforms are
343 rapidly developing, such as the miniPCR system (Amplyus, Cambridge, USA) and the Bento Lab
344 platform (Bento Bioworks Ltd, London, United Kingdom). Drawbacks of amplicon strategy
345 include biases related to PCR amplification, which lead to inaccuracy when calculating the
346 relative proportions of microbial species.

347 While many workflows are available for microbiome profiling using metagenomics data, there is
348 no standard method for analysing long read data. This makes validation of both methods and
349 results crucial (Quince et al. 2017). To validate the method, mock microbial communities or
350 samples containing known microbial species were commonly used for proof-of-concept. For
351 example, to verify the diagnosis of a specific pathogen strain based on SNPs called from MinION

352 reads, Votintseva et al. (2017) first identified pathogens using clinical methods then generated
353 sequencing data and mapped the data to the targeted species in real-time. The reactions were
354 stopped once enough SNPs were obtained to confidentially identify the targeted pathogen
355 strain. Similarly, our study used wheat samples manually inoculated with infections and visually
356 confirmed known pathogens prior to sequencing, while the same identifications from the first
357 BLAST analysis validated the capacity of our workflow for identification of those pathogen
358 species. To validate the result, different analysis approaches should be applied on the same
359 dataset. Brown et al. (2017) compared the performance of four bioinformatics pipelines on
360 identifying species within different synthetic bacterial communities using MinION. Compared to
361 the composition of mock community, even the pipeline that resulted in the most similar
362 classification still contains reads classified to species known not to be present. These false
363 positive results are due largely to sequencing errors, which illustrates the need for more
364 stringent filtering of the raw data.

365 Nanopore metagenomics will have great potential for quick diagnosis of suddenly emerging
366 crop disease and large-scale disease monitoring at centralized agricultural institutions. As
367 shown from other sequencing studies performed in a more extreme environment (Johnson et al.
368 2017; Pomerantz et al. 2018), all processes including DNA extraction and purification, library
369 preparation and sequencing can be completed within a few hours. A tailored database can be
370 created for each crop species containing the genomes of pathogens known to cause diseases to
371 minimise processing time. For rapidly evolving pathogens such as wheat rusts, a SNP-based
372 classification method as described by Votintseva et al. (2017) can be adapted to define rust
373 strains once enough SNPs are obtained. However, before running the sequencer, sampling

374 procedures and tissue processing strategies still need to be carefully curated. These vary in crop
375 species and tissue types, so economic efficiency and statistical significance should be
376 considered during the sampling process (Knight et al. 2012). Our study randomly sampled
377 multiple leaves for each sample and extracted four times as technical replicates. Consistency in
378 barcode classification and BLAST analysis through all four replicates indicated reliability of the
379 workflow. In centralized agricultural institutions, amplicon sequencing can potentially play a key
380 role and, to-date, there have been a few preliminary successes (Benítez-Páez, Portune, and
381 Sanz 2016; Benítez-Páez and Sanz 2017; Pomerantz et al. 2018). Given the multiplexing capacity
382 of the 96 native barcoding kit (ONT), hundreds of samples can be sequenced in parallel when
383 larger scale sequencing platform were employed (e.g. GridION, ONT). To achieve this goal, long
384 read databases, sequencing accuracy and access to large-scale sample processing power need
385 to be improved.
386 In conclusion, our workflow demonstrates the potential of this technology for plant pathogen
387 diagnosis with the potential for field applications, with the additional possibility of microbiome
388 characterization. A combination of on-site and centralized sequencing approaches would, in
389 future, revolutionize management of agricultural biosecurity and prevent crop loss to meet the
390 increasing food demand of human beings.
391

392 **LITERATURE CITED**

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

396 Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. 1990. Basic local alignment
397 search tool. J. Mol. Biol. 215:403–410.
398 Ardui, S., Ameur, A., Vermeesch, J. R., and Hestand, M. S. 2018. Single molecule real-time
399 (SMRT) sequencing comes of age: applications and utilities for medical diagnostics. Nucleic
400 Acids Res. 46:2159–2168.
401 Benítez-Páez, A., Portune, K. J., and Sanz, Y. 2016. Species-level resolution of 16S rRNA gene
402 amplicons sequenced through the MinION™ portable nanopore sequencer. GigaScience. 5:4.
403 Benítez-Páez, A., and Sanz, Y. 2017. Multi-locus and long amplicon sequencing approach to
404 study microbial diversity at species level using the MinION™ portable nanopore sequencer.
405 GigaScience. 6:1–12.
406 Bhattacharya, S. 2017. Deadly new wheat disease threatens Europe's crops. Nat. News. 542:145.
407 Brown, B. L., Watson, M., Minot, S. S., Rivera, M. C., and Franklin, R. B. 2017. MinION™
408 nanopore sequencing of environmental metagenomes: a synthetic approach. GigaScience. 6:1–
409 10.
410 Bulgarelli, D., Rott, M., Schlaeppi, K., Themaat, E. V. L. van, Ahmadinejad, N., Assenza, F., et al.
411 2012. Revealing structure and assembly cues for *Arabidopsis* root-inhabiting bacterial
412 microbiota. Nature. 488:91–95.
413 Callaway, E. 2016. Devastating wheat fungus appears in Asia for first time. Nat. News. 532:421.
414 Charalampous, T., Richardson, H., Kay, G. L., Baldan, R., Jeanes, C., Rae, D., et al. 2018. Rapid
415 Diagnosis of Lower Respiratory Infection using Nanopore-based Clinical Metagenomics.
416 bioRxiv. :387548.

417 Chemidlin Prévost-Bouré, N., Christen, R., Dequiedt, S., Mougel, C., Lelièvre, M., Jolivet, C., et al.
418 2011. Validation and application of a PCR primer set to quantify fungal communities in the soil
419 environment by real-time quantitative PCR. PloS One. 6:e24166.
420 Clavijo, B. J., Venturini, L., Schudoma, C., Accinelli, G. G., Kaithakottil, G., Wright, J., et al. 2017.
421 An improved assembly and annotation of the allohexaploid wheat genome identifies complete
422 families of agronomic genes and provides genomic evidence for chromosomal translocations.
423 Genome Res. 27:885–896.
424 Cruaud, P., Rasplus, J.-Y., Rodriguez, L. J., and Cruaud, A. 2017. High-throughput sequencing of
425 multiple amplicons for barcoding and integrative taxonomy. Sci. Rep. 7:41948.
426 De Boer, S. H., and López, M. M. 2012. New grower-friendly methods for plant pathogen
427 monitoring. Annu. Rev. Phytopathol. 50:197–218.
428 De Coster, W., D'Hert, S., Schultz, D. T., Cruts, M., Van Broeckhoven, C., and Berger, B. 2018.
429 NanoPack: visualizing and processing long-read sequencing data. Bioinformatics. 34:2666–2669.
430 Ellis, J. G. 2017. Can Plant Microbiome Studies Lead to Effective Biocontrol of Plant Diseases?
431 Mol. Plant. Microbe Interact. 30:190–193.
432 Ewing, B., and Green, P. 1998. Base-Calling of Automated Sequencer Traces Using Phred. II.
433 Error Probabilities. Genome Res. 8:186–194.
434 Ewing, B., Hillier, L., Wendl, M. C., and Green, P. 1998. Base-Calling of Automated Sequencer
435 Traces UsingPhred. I. Accuracy Assessment. Genome Res. 8:175–185.
436 Fisher, M. C., Henk, D. A., Briggs, C. J., Brownstein, J. S., Madoff, L. C., McCraw, S. L., et al. 2012.
437 Emerging fungal threats to animal, plant and ecosystem health. Nature. 484:186–194.

438 Goodwin, S. B., M'Barek, S. B., Dhillon, B., Wittenberg, A. H. J., Crane, C. F., Hane, J. K., et al.
439 2011. Finished Genome of the Fungal Wheat Pathogen Mycosphaerella graminicola Reveals
440 Dispensome Structure, Chromosome Plasticity, and Stealth Pathogenesis. PLOS Genet.
441 7:e1002070.
442 Hane, J. K., Lowe, R. G. T., Solomon, P. S., Tan, K.-C., Schoch, C. L., Spatafora, J. W., et al. 2007.
443 Dothideomycete–Plant Interactions Illuminated by Genome Sequencing and EST Analysis of the
444 Wheat Pathogen Stagonospora nodorum. Plant Cell. 19:3347–3368.
445 Hu, Y. 2016. High quality DNA extraction from Fungi_small scale. protocols.io. Available at:
446 https://www.protocols.io/view/high-quality-dna-extraction-from-fungi-small-scale-exmbfk6
447 [Accessed September 26, 2018].
448 Hu, Y., and Schwessinger, B. 2018. Amplicon sequencing using MinION optimized from 1D
449 native barcoding genomic DNA. protocols.io. Available at:
450 https://www.protocols.io/view/amplicon-sequencing-using-minion-optimized-from-1d-
451 mhkc34w [Accessed September 27, 2018].
452 Jain, M., Koren, S., Miga, K. H., Quick, J., Rand, A. C., Sasani, T. A., et al. 2018. Nanopore
453 sequencing and assembly of a human genome with ultra-long reads. Nat. Biotechnol. 36:338–
454 345.
455 Jain, M., Olsen, H. E., Paten, B., and Akeson, M. 2016. The Oxford Nanopore MinION: delivery of
456 nanopore sequencing to the genomics community. Genome Biol. 17:239.
457 Janda, J. M., and Abbott, S. L. 2007. 16S rRNA Gene Sequencing for Bacterial Identification in
458 the Diagnostic Laboratory: Pluses, Perils, and Pitfalls. J. Clin. Microbiol. 45:2761–2764.

459 Johnson, S. S., Zaikova, E., Goerlitz, D. S., Bai, Y., and Tighe, S. W. 2017. Real-Time DNA
460 Sequencing in the Antarctic Dry Valleys Using the Oxford Nanopore Sequencer. J. Biomol. Tech.
461 JBT. 28:2–7.
462 Kennedy, K., Hall, M. W., Lynch, M. D. J., Moreno-Hagelsieb, G., and Neufeld, J. D. 2014.
463 Evaluating Bias of Illumina-Based Bacterial 16S rRNA Gene Profiles. Appl. Environ. Microbiol.
464 80:5717–5722.
465 Knight, R., Jansson, J., Field, D., Fierer, N., Desai, N., Fuhrman, J. A., et al. 2012. Unlocking the
466 potential of metagenomics through replicated experimental design. Nat. Biotechnol. 30:513–
467 520.
468 Laver, T., Harrison, J., O'Neill, P. A., Moore, K., Farbos, A., Paszkiewicz, K., et al. 2015. Assessing
469 the performance of the Oxford Nanopore Technologies MinION. Biomol. Detect. Quantif. 3:1–8.
470 Lim, Y. W., Haynes, M., Furlan, M., Robertson, C. E., Harris, J. K., and Rohwer, F. 2014. Purifying
471 the Impure: Sequencing Metagenomes and Metatranscriptomes from Complex Animal-
472 associated Samples. JoVE J. Vis. Exp. :e52117–e52117.
473 Mahlein, A.-K. 2015. Plant Disease Detection by Imaging Sensors – Parallels and Specific
474 Demands for Precision Agriculture and Plant Phenotyping. Plant Dis. 100:241–251.
475 Manning, V. A., Pandelova, I., Dhillon, B., Wilhelm, L. J., Goodwin, S. B., Berlin, A. M., et al. 2013.
476 Comparative Genomics of a Plant-Pathogenic Fungus, Pyrenophora tritici-repentis, Reveals
477 Transduplication and the Impact of Repeat Elements on Pathogenicity and Population
478 Divergence. G3 Genes Genomes Genet. 3:41–63.

479 Martijn, J., Lind, A. E., Spiers, I., Juzokaite, L., Bunikis, I., Pettersson, O. V., et al. 2017. Amplicon
480 sequencing of the 16S-ITS-23S rRNA operon with long-read technology for improved
481 phylogenetic classification of uncultured prokaryotes. bioRxiv. :234690.
482 Mehrabi, Z., McMillan, V. E., Clark, I. M., Canning, G., Hammond-Kosack, K. E., Preston, G., et al.
483 2016. *Pseudomonas* spp. diversity is negatively associated with suppression of the wheat take-
484 all pathogen. Sci. Rep. 6:29905.
485 Mitsuhashi, S., Kryukov, K., Nakagawa, S., Takeuchi, J. S., Shiraishi, Y., Asano, K., et al. 2017. A
486 portable system for rapid bacterial composition analysis using a nanopore-based sequencer and
487 laptop computer. Sci. Rep. 7:5657.
488 Nezhad, A. S. 2014. Future of portable devices for plant pathogen diagnosis. Lab. Chip.
489 14:2887–2904.
490 Panke-Buisse, K., Poole, A. C., Goodrich, J. K., Ley, R. E., and Kao-Kniffin, J. 2015. Selection on
491 soil microbiomes reveals reproducible impacts on plant function. ISME J. 9:980–989.
492 Pomerantz, A., Peñafiel, N., Arteaga, A., Bustamante, L., Pichardo, F., Coloma, L. A., et al. 2018.
493 Real-time DNA barcoding in a rainforest using nanopore sequencing: opportunities for rapid
494 biodiversity assessments and local capacity building. GigaScience. 7.
495 Pretorius, Z. A., Singh, R. P., Wagoire, W. W., and Payne, T. S. 2000. Detection of Virulence to
496 Wheat Stem Rust Resistance Gene Sr31 in Puccinia graminis. f. sp. tritici in Uganda. Plant Dis.
497 84:203–203.
498 Quince, C., Walker, A. W., Simpson, J. T., Loman, N. J., and Segata, N. 2017. Shotgun
499 metagenomics, from sampling to analysis. Nat. Biotechnol. 35:833–844.

500 Raja, H. A., Miller, A. N., Pearce, C. J., and Oberlies, N. H. 2017. Fungal Identification Using
501 Molecular Tools: A Primer for the Natural Products Research Community. J. Nat. Prod. 80:756–
502 770.
503 Ranjan, R., Rani, A., Metwally, A., McGee, H. S., and Perkins, D. L. 2016. Analysis of the
504 microbiome: Advantages of whole genome shotgun versus 16S amplicon sequencing. Biochem.
505 Biophys. Res. Commun. 469:967–977.
506 Ritpitakphong, U., Falquet, L., Vimoltust, A., Berger, A., Métraux, J.-P., and L'Haridon, F. 2016.
507 The microbiome of the leaf surface of Arabidopsis protects against a fungal pathogen. New
508 Phytol. 210:1033–1043.
509 Schmidt, K., Mwaigwisya, S., Crossman, L. C., Doumith, M., Munroe, D., Pires, C., et al. 2017.
510 Identification of bacterial pathogens and antimicrobial resistance directly from clinical urines by
511 nanopore-based metagenomic sequencing. J. Antimicrob. Chemother. 72:104–114.
512 Schwessinger, B., Sperschneider, J., Cuddy, W. S., Garnica, D. P., Miller, M. E., Taylor, J. M., et al.
513 2018. A Near-Complete Haplotype-Phased Genome of the Dikaryotic Wheat Stripe Rust Fungus
514 Puccinia striiformis f. sp. tritici Reveals High Interhaplotype Diversity. mBio. 9.
515 Sharpton, T. J. 2014. An introduction to the analysis of shotgun metagenomic data. Front. Plant
516 Sci. 5:209.
517 Tessler, M., Neumann, J. S., Afshinnekoo, E., Pineda, M., Hersch, R., Velho, L. F. M., et al. 2017.
518 Large-scale differences in microbial biodiversity discovery between 16S amplicon and shotgun
519 sequencing. Sci. Rep. 7:6589.
520 Thomas, T., Gilbert, J., and Meyer, F. 2012. Metagenomics - a guide from sampling to data
521 analysis. Microb. Inform. Exp. 2:3.

522 Thonar, C., Erb, A., and Jansa, J. 2012. Real-time PCR to quantify composition of arbuscular
523 mycorrhizal fungal communities--marker design, verification, calibration and field validation.
524 Mol. Ecol. Resour. 12:219–232.
525 Votintseva, A. A., Bradley, P., Pankhurst, L., Elias, C. del O., Loose, M., Nilgiriwala, K., et al. 2017.
526 Same-Day Diagnostic and Surveillance Data for Tuberculosis via Whole-Genome Sequencing of
527 Direct Respiratory Samples. J. Clin. Microbiol. 55:1285–1298.

528

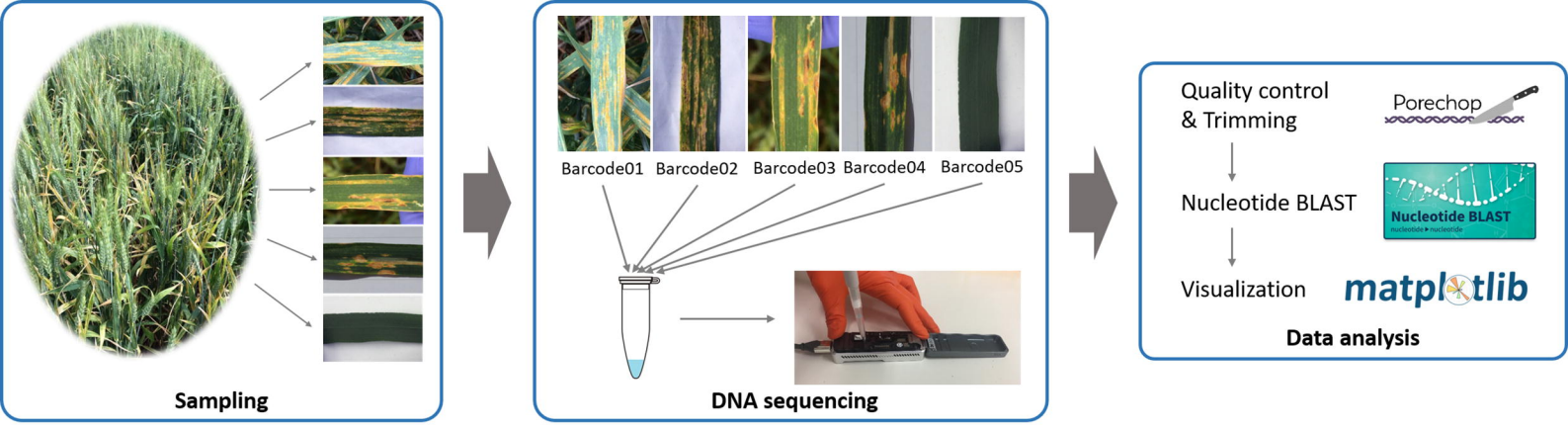
529 **Figures and figure legends**

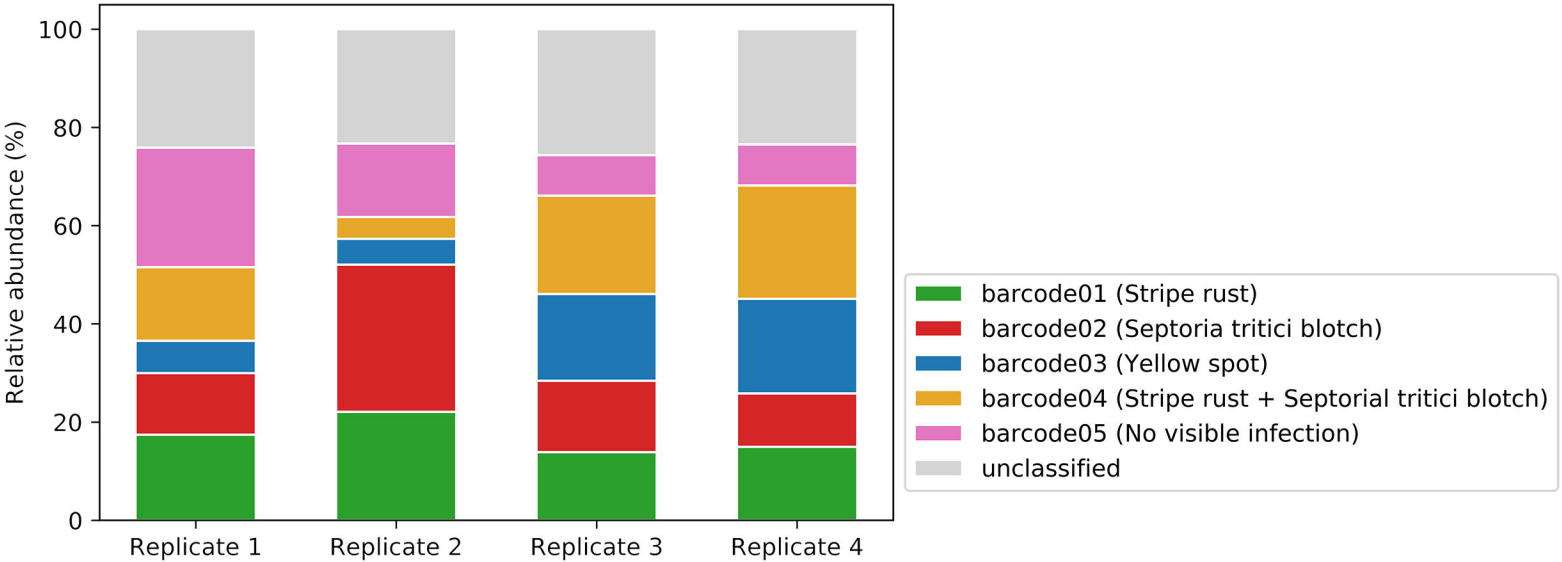
530 Fig. 1. The Nanopore sequencing workflow for fungal pathogen identification. 1) Sampling:
531 plant pathologists collected wheat leaf samples from disease nursery fields, with visual
532 confirmation of disease symptoms for four diseased sample (three single fungal infections and
533 one double fungal infection). A fungicide treated wheat sample from the same variety and
534 growth stage was collected as a negative control. 2) DNA sequencing: to minimise operator bias,
535 DNA was extracted from all five samples without knowledge of the causative agent(s) affecting
536 each sample. We labelled samples with DNA barcodes through a PCR reaction step before
537 sample pooling and 1D library preparations (SQK-LSK107); 3) Data analysis: We quality
538 controlled sequences based on Phred quality scores and trimmed adapter sequence using
539 Porechop (https://github.com/rrwick/Porechop). We performed BLAST (Altschul et al., 1990)
540 searches using processed reads and saved the information of the best hits. We summarized the
541 number of total sequence length from hits by each species using the python *matplotlib* module.
542 Images used in step 3 are from the official websites of each program.

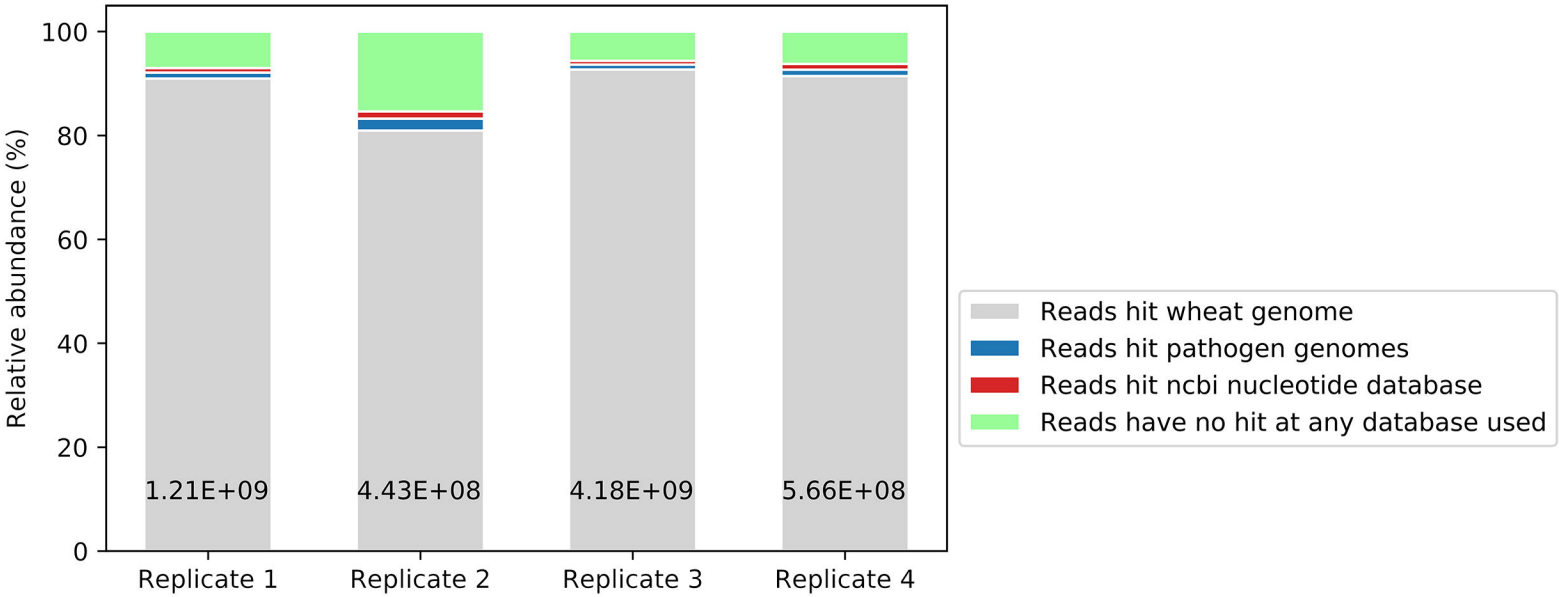
543

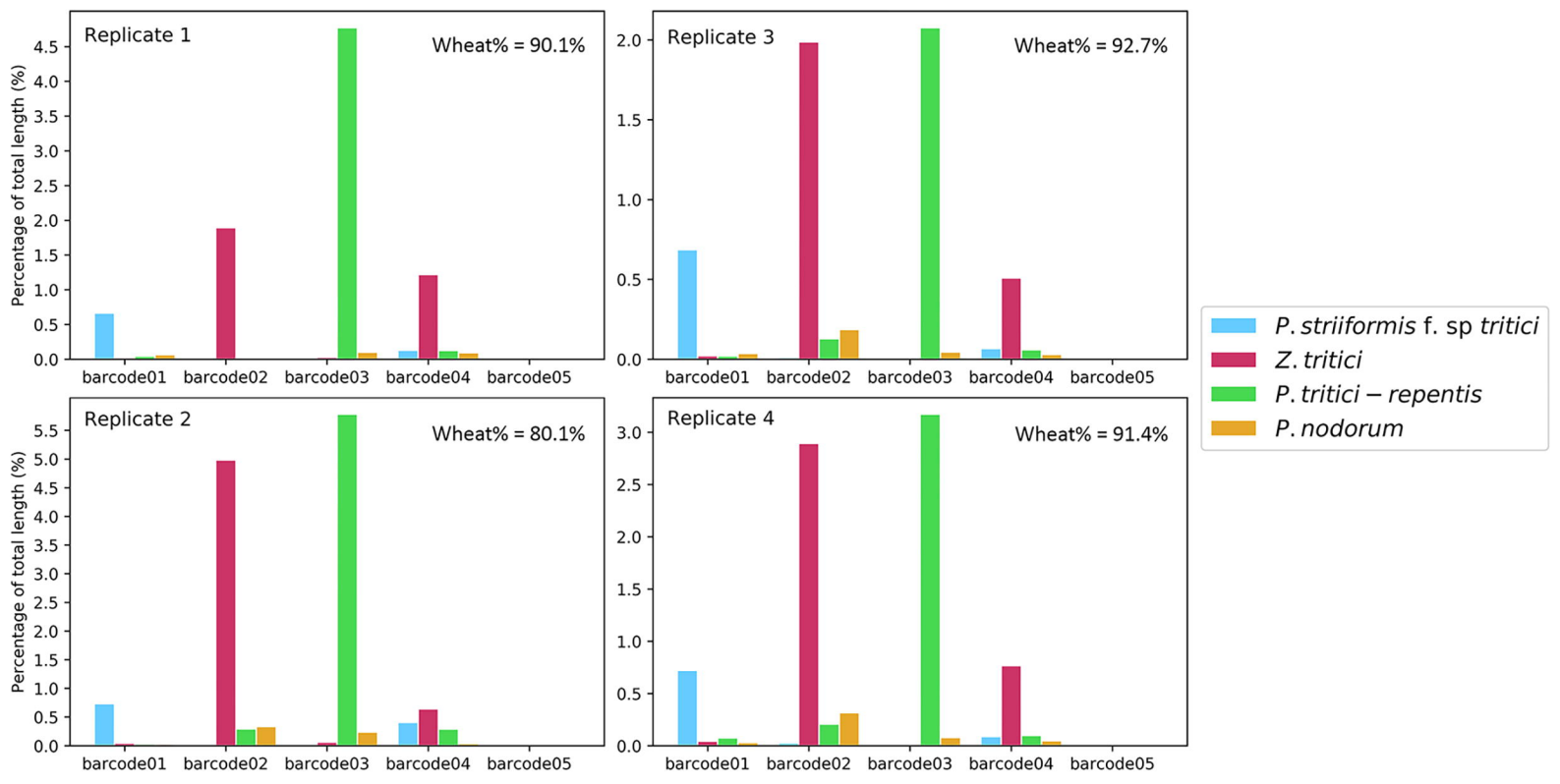
544 Fig. 2. Uneven proportions of barcode-labelled DNA across replicates. Columns indicate four
545 sets of wheat leaf samples collected as biological replicates. The y-axis indicates relative
546 abundance of total sequence lengths labelled with one of five DNA barcodes within each
547 sample set. Each barcode represents one wheat sample. Reads without a detectable barcode
548 were labelled as 'unclassified'.
549
550 Fig. 3. Wheat is the predominant identified species in the whole genome sequencing analysis
551 using BLAST. The graph shows the relative sequence length per species identified using a
552 'winner-takes-all' approach when searching a targeted BLAST database Wheat (grey), fungal
553 pathogen genomes (blue), NCBI nucleotide (nt) database (red) and unmapped (light green) (see
554 results and methods for details). The numbers at the bottom of each bar show the total
555 sequence length [bp] of processed reads of each replicate. Replicate 1, 3, 4 showed similar
556 classifications while replicate 2 contains more unclassified sequences.
557
558 Fig. 4. The total sequence length per pathogen genome per barcode correlate with the
559 phenotypic disease identification. The graphs show the relative sequence length assigned to
560 each pathogen genome using a 'winner-takes-all' approach in the BLAST analysis. The
561 proportion of reads hitting the wheat host genome are listed in the top-right corner of each
562 graph. Unassigned reads are omitted from the visualization.
563
564 Fig. 5. Fungal pathogen infections of wheat increase the microbial population especially for stripe rust
565 infection and Septoria tritici blotch infection. Percentage of sequences that did not hit the

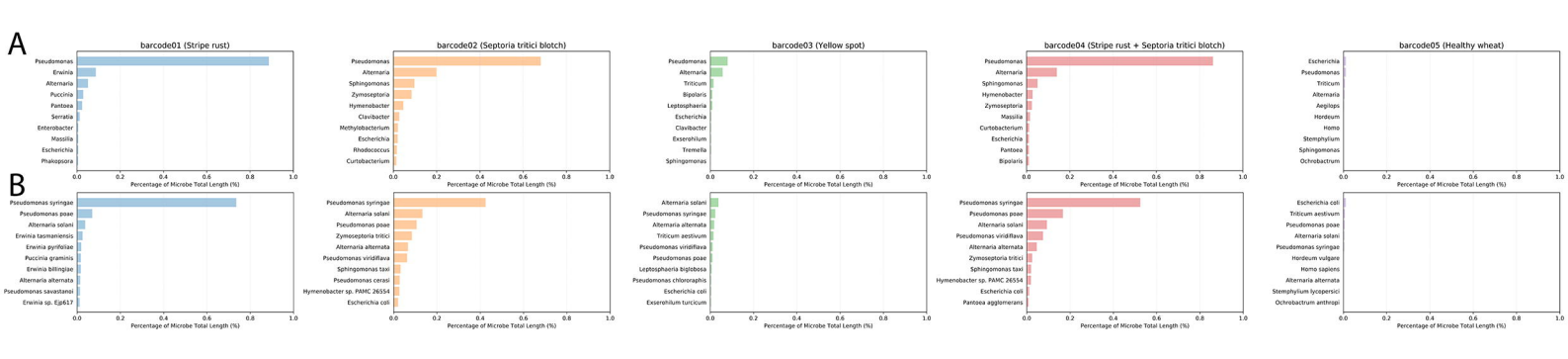
566 restricted database but hit the non-redundant NCBI nucleotide database combined from all
567 replicates, and grouped by genus (A) or species (B). Different colours represented different
568 barcodes. Reads that are assigned to the restricted database was eliminated, and the top ten
569 highest classifications were displayed for each replicate in the y-axis. Genus *Pseudomonas*
570 represented the highest proportion across all barcodes containing Stripe rust (*Puccinia*
571 *striiformis tritici*) or Septoria tritici blotch (*Z. tritici*), with *Pseudomonas syringae* as the highest
572 at the species level.
573


Sampling
Barcode01
Barcode02
Barcode03
Barcode04
Barcode05
DNA sequencing
Quality control
& Trimming
Porechop
Nucleotide BLAST
Nucleotide BLAST
Visualization
matplotlib
Data analysis


Relative abundance (%)
100
80
60
40
20
0
Replicate 1
Replicate 2
Replicate 3
Replicate 4
barcode01 (Stripe rust)
barcode02 (Septoria tritici blotch)
barcode03 (Yellow spot)
barcode04 (Stripe rust + Septorial tritici blotch)
barcode05 (No visible infection)
unclassified


Relative abundance (%)
100
80
60
40
20
0
1.21E+09
4.43E+08
4.18E+09
5.66E+08
Replicate 1
Replicate 2
Replicate 3
Replicate 4
Reads hit wheat genome
Reads hit pathogen genomes
Reads hit ncbi nucleotide database
Reads have no hit at any database used


Replicate 1
Wheat% = 90.1%
Replicate 2
Wheat% = 80.1%
Replicate 3
Wheat% = 92.7%
Replicate 4
Wheat% = 91.4%
Percentage of total length (%)
barcode01
barcode02
barcode03
barcode04
barcode05
P. striiformis f. sp tritici
Z. tritici
P. tritici – repentis
P. nodorum


A
barcode01 (Stripe rust)
Pseudomonas
Erwinia
Alternaria
Puccinia
Pantoea
Serratia
Enterobacter
Massilia
Escherichia
Phakopsora
barcode02 (Septoria tritici blotch)
Pseudomonas
Alternaria
Sphingomonas
Zymoseptoria
Hymenobacter
Clavibacter
Methylobacterium
Escherichia
Rhodococcus
Curtobacterium
barcode03 (Yellow spot)
Pseudomonas
Alternaria
Triticum
Bipolaris
Leptosphaeria
Escherichia
Clavibacter
Exserohilum
Tremella
Sphingomonas
barcode04 (Stripe rust + Septoria tritici blotch)
Pseudomonas
Alternaria
Sphingomonas
Hymenobacter
Zymoseptoria
Massilia
Curtobacterium
Escherichia
Pantoea
Bipolaris
barcode05 (Healthy wheat)
Escherichia
Pseudomonas
Triticum
Alternaria
Aegilops
Hordeum
Homo
Stemphylium
Sphingomonas
Ochrobactrum
0.0
0.2
0.4
0.6
0.8
1.0
Percentage of Microbe Total Length (%)
B
Pseudomonas syringae
Pseudomonas poae
Alternaria solani
Erwinia tasmaniensis
Erwinia pyrifoliae
Puccinia graminis
Erwinia billingiae
Alternaria alternata
Pseudomonas savastanoi
Erwinia sp. Ejp617
Pseudomonas syringae
Alternaria solani
Pseudomonas poae
Zymoseptoria tritici
Alternaria alternata
Pseudomonas viridiflava
Sphingomonas taxi
Pseudomonas cerasi
Hymenobacter sp. PAMC 26554
Escherichia coli
Alternaria solani
Pseudomonas syringae
Alternaria alternata
Triticum aestivum
Pseudomonas viridiflava
Pseudomonas poae
Leptosphaeria biglobosa
Pseudomonas chlororaphis
Escherichia coli
Exserohilum turcicum
Pseudomonas syringae
Pseudomonas poae
Alternaria solani
Pseudomonas viridiflava
Alternaria alternata
Zymoseptoria tritici
Sphingomonas taxi
Hymenobacter sp. PAMC 26554
Escherichia coli
Pantoea agglomerans
Escherichia coli
Triticum aestivum
Pseudomonas poae
Alternaria solani
Pseudomonas syringae
Hordeum vulgare
Homo sapiens
Alternaria alternata
Stemphylium lycopersici
Ochrobactrum anthropi