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 ACT 2601, 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 ACT 2601, Australia; **John P. Rathjen**† and **Benjamin Schwessinger**†, Research School of Biology, The Australian National University, Acton ACT 2601, 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 shot-gun

sequencing with a portable DNA sequencing device as a method for the detection of fungal pathogens from wheat (*Triticum aestivum*) in a standard molecular biology laboratory. 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 the bacterial genus *Pseudomonas* co-present with *Puccinia* and *Zymoseptoria* but not *Pyrenophora* infections. One limitation of the method is the over-representation of redundant wheat genome sequences from samples. This could be addressed by long-range amplicon-based sequencing approaches in future studies, which specifically target non-host organisms. Our work outlines a new approach for detection of a broad range of plant pathogens and associated microbes using a portable sequencer in a standard laboratory, providing the basis for future development of an on-site disease monitoring system.

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

45 whom identify diseases initially by eye in the field. While molecular methods of pathogen
46 detection also exist, these are generally only capable of identifying specific diseases with low
47 amounts of variation (De Boer and López 2011). For example, the enzyme-linked
48 immunosorbent assay (ELISA) have been applied for the detection of tree pathogen *Xylella*
49 *fastidiosa* (Sherald and Lei 1991). Recently, other optical sensor-based methods have been
50 developed for the detection, identification, and quantification of plant diseases (Mahlein 2015).
51 This method has been used to help smallholder farmers for the disease and pest control
52 (PlantVillage, https://plantvillage.psu.edu/). Conventional pathology, however, remains the
53 primary means of pathogen identification and harbors certain limitations: 1. It relies heavily on
54 the physical appearance of disease symptoms; 2. It has difficulty detecting pathogens that do
55 not infect aerial parts of the plant; 3. Co-infections may lead to conflicting visual symptoms; 4.
56 No method currently exists for the rapid identification of unknown pathogens during outbreaks
57 – a limitation epitomized by the recent, devastating incursion of South American races of wheat
58 blast fungus 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* (*EF1*α)
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
65 multilocus sequence typing (Maiden et al. 1998), can discriminate closely related species based
66 on single nucleotide polymorphisms (SNPs) within amplicon sequences. This has been used, for

67 example, 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
70 microbiome can impact plant function, for example, on altering flowering time (Panke-Buisse et
71 al. 2015) or promoting colonization of fungal pathogens such as *Botrytis cinerea* (Ritpitakphong
72 et al. 2016). Furthermore, the microbiome is likely to contain organisms that interact with
73 disease-causing agents and, as some of these organisms have anti-microbial activities against
74 plant pathogens, could potentially be harnessed for biocontrol strategies (Ellis 2017). Restrictive
75 read lengths generated by the Illumina sequencing platform limits sequence-based
76 identification of microbial species. Standard Illumina sequencing only allows for a maximum
77 amplicon length of 500 base pairs (bp) to be sequenced and, as such, only the genetic variation
78 present within 500 bp is available for taxonomic assignment of the organism in question
79 (Benítez-Páez and Sanz 2017). Use of PCR also biases the abundance of individual amplicons,
80 making estimation of population 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
83 (Benítez-Páez and Sanz 2017). Multiple single molecule, long-read sequencing technologies are
84 currently in use. Most predominant among these are single molecular real-time (SMRT)
85 sequencing (PacBio) and nanopore sequencing (MinION; Oxford Nanopore Technologies (ONT)).
86 PacBio SMRT sequencing relies on the detection of fluorescent nucleotides incorporated into a
87 single DNA molecule during its synthesis within a nanoscale observation chamber (Ardui et al.
88 2018). Nanopore sequencing enables direct sequencing of native DNA by measuring voltage

89 changes across an artificial membrane when a single DNA molecule passes through a nanopore
90 embedded in the membrane of a flowcell (Jain et al. 2016). Advantages of nanopore sequencing
91 include its relatively low cost, pocketsize form and the potential for real-time data analysis,
92 while current disadvantages include relatively low per read accuracy of around 90% (Leggett
93 and Clark 2017). Because the MinION is a mobile technology independent of large sequencing
94 facilities, it has been useful for on-site sample sequencing in extreme environments such as
95 microbial paleo mats in the Antarctic (Johnson et al. 2017). For rapid pathogen identification,
96 studies have shown a turnaround time of six hours for bacterial from clinical samples with
97 MinION (Charalampous et al. 2018), and from seconds to up to four hours for detecting cassava
98 virus in Africa (Boykin et al. 2018) – both unachievable with previous technologies. For
99 antibiotic resistance, Břinda et al. (2018) reported their identification of known resistant
100 bacterial strain within five minutes from real time MinION data. Moreover, the MinION
101 provides significantly longer sequencing reads (~10 kb) than other sequencing platforms such as
102 Illumina, allowing for more genetic information per read and thus more precise identification of
103 the organism from which the read is derived (Benítez-Páez, Portune, and Sanz 2016; Benítez-
104 Páez and Sanz 2017). Although recent studies have shown important advances in identification
105 of bacterial pathogens using nanopore sequencing from clinical samples (Mitsuhashi et al. 2017;
106 Schmidt et al. 2017), little has been done testing this new sequencing platform in terms of plant
107 pathogens detection and 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
113 sampled from the field and a fungicide treated control. By mapping sequencing reads to
114 databases, we successfully detected all fungal diseases present in the infected samples
115 including stripe rust (caused by *Puccinia striiformis* f. sp. *tritici*), septoria tritici blotch (caused by
116 *Zymoseptoria tritici*) and yellow leaf spot (caused by *Pyrenophora tritici repentis*), and found
117 one disease (septoria nodorum blotch, caused by *Parastagonospora nodorum*). We also
118 conducted preliminary characterisation of the microbiomes associated with those diseases,
119 identifying the bacterial genus *Pseudomonas* as co-present with *Puccinia* and *Zymoseptoria*
120 infections. Our results suggest that portable nanopore sequencing has a considerable potential
121 for adaptation to a broad range of crop disease diagnoses and environmental monitoring
122 applications under field conditions.
123

124 **MATERIALS AND METHODS**

125 **Sample collection**

126 Wheat cultivar Crusader (Advanta Seeds Pty. Ltd.) seeds were planted in the disease nursery
127 field on May 2016 at the Wagga Wagga Agricultural Institute (New South Wales (NSW),
128 35°02'24.3"S 147°19'09.5"E) with daily irrigation. Wheat straw infected the previous season
129 with septoria tritici blotch and yellow leaf spot was allowed to stand in the field to promote
130 development of pseudothecia. This infected material was used to manually inoculate plants in
131 the nursery field at the tillering growth stage (4-5 weeks post emergence) while stripe rust
132 infections occurred naturally at jointing stage (7-9 weeks post emergence). Healthy or diseased

133 plants were diagnosed by a plant pathologist from the NSW Department of Primary Industries
134 via symptomatology before samples were collected during September 2016. Five different
135 treatment groups were collected into five separate containers. These included four infection
136 types: three single infections (*Puccinia striiformis* f. sp. *tritici*, *Z. tritici* and *Pyrenophora tritici*
137 *repentis*) and one double infection (*Puccinia striiformis* f. sp. *tritici* + *Z. tritici*), and a fungicide-
138 treated control. For each treatment group, three separate entire wheat tillers were pooled in
139 one container and shipped to the Rathjen laboratory, Research School of Biology, The
140 Australian National University. Each biological replicate (n=4) consisted of two independent leaf
141 cuts (~100 mg) from each treatment group containing visible disease symptoms or not in case
142 of the control. The two leaf of each biological replicate were transferred into 2 mL Eppendorf
143 tubes containing one 5 mm metal bead. All tubes were labelled and stored at -80°C prior to
144 DNA extraction.

145 **DNA extraction**

146 DNA extractions were performed according to Hu (2016). Briefly, 100 mg leaf tissue was
147 homogenized and cells were lysed using cetyl trimethylammonium bromide (CTAB, Sigma-
148 Aldrich Cas. No: 57-09-0) buffer (added RNAse T1, Thermo Fisher, Cat#EN0541, 1000 units per
149 1750 µL), followed by a phenol:chloroform:isoamyl alcohol (25:24:1, Sigma-Aldrich, Cat#P2069)
150 extraction to remove protein and lipids. The DNA was precipitated with 700 µL isopropanol,
151 washed with 1 mL 70 % ethanol, dried for 5 minutes at room temperature and resuspended in
152 TE buffer containing 10 mM Tris and 1 mM EDTA at pH 8. DNA extractions were performed
153 independently for each biological replicate (n=4), with the operator blind to the identity of each
154 sample. After extraction, DNA was purified with one volume of Agencourt AMPure XP beads

155 (Beckman Coulter, Inc. Cat#BECLA63880) according to the manufacturer's protocol and stored
156 at 4°C. Quality and average size of genomic DNA was visualized by gel electrophoresis with a 1%
157 agarose gel for 1 hour at 100 volts. Representative gel images for replicate 3 and 4 are shown in
158 Supplementary Fig. S1. DNA was quantified by NanoDrop and Qubit (Life Technologies)
159 according to the manufacturer's protocol. DNA quality and purity was evaluated as described
160 previously (Schalamun et al. 2018).

161 **Library construction and sequencing**

162 For each biological replicate, one PCR-based barcoded 1D sequencing libraries were
163 constructed using 1D PCR barcoding workflow (SQK-LSK107 and EXP-PBC001, batch number
164 EP01.10.0005) as previously described in Hu and Schwessinger (2018) with the omission of the
165 DNA shearing step. Briefly, a dA-tailing reaction was performed on 700 ng genomic DNA from
166 each sample using the NEBNext Ultra II End-repair/dA-tailing kit (New England Biolabs, Catalog
167 #E7645). Repaired DNA was cleaned up using one volume of AMPure XP beads, washed on a
168 magnetic rack using 70% ethanol, and eluted with 31 µL nuclease-free water. Thirty µL eluate
169 was used for barcode adapter ligation via the NEB Blunt/TA Ligase Master Mix (New England
170 Biolabs, Catalog #M0367S) and Barcode Adapter Mix (ONT 1D PCR barcoding kit EXP-PBC001),
171 followed by another wash step on one volume of AMPure XP beads. Twenty ng of the adapter-
172 ligated template in 2 µL of nuclease free water was added to each barcoding PCR reaction with
173 50 µL LongAmp Taq 2X Master Mix (New England Biolabs, Catalog #M0287S), 46 µL water and 2
174 µL PCR Barcodes (ONT 1D PCR barcoding kit EXP-PBC001). The PCR was performed as follows:
175 denaturation at 95°C for 3 min, 15 cycles of 95°C/15 s; 62°C/15 s; 65°C/150 s, a final extension
176 step at 65°C for 5 min. Barcoded DNA was purified using one volume of AMPure XP beads and

177 pooled into 1 µg fractions in a total of 45 µL to which 5 µL of DNA CS control (ONT 1D ligation
178 sequencing kit SQK-LSK107) was added. A further dA-tailing reaction was performed as above
179 followed by purification on with one volume of AMPure XP beads. DNA was eluted in 31 µL
180 nuclease free water. Adapter ligation was carried out using 30 µL end-prepped DNA, 20 µL
181 Adapter Mix (ONT 1D ligation sequencing kit SQK-LSK107) and 50 µL NEB Blunt/TA Ligase
182 Master Mix. The adapter-ligated DNA library was purified with one volume of AMPure XP beads
183 using the Adapter Beads Binding buffer (ONT 1D ligation sequencing kit SQK-LSK107) for
184 washing and eluted in 15 µL of Elution Buffer (ONT 1D ligation sequencing kit SQK-LSK107).
185 Sequencing reactions were performed independently for each biological replicate on a MinION
186 flowcell (R9.4, ONT) connected to a MK1B device (ONT) operated by the MinKNOW software
187 (version 1.5.2 and version 1.6.11). Each flowcell was primed with 1 mL of priming buffer
188 comprising 480 µL Running Buffer Fuel Mix (RBF, ONT) and 520 µL nuclease free water. Twelve
189 µL of amplicon library was added to a loading mix including 35 µL RBF, 25.5 µL Library Loading
190 beads (ONT library loading bead kit EXP-LLB001, batch number EB01.10.0012) and 2.5 µL water
191 with a final volume of 75 µL and then added to the flowcell via the SpotON sample port. The
192 "NC_48Hr_sequencing_FLO-MIN106_SQK-LSK107" protocol was executed through MinKNOW
193 after loading the library. The sequencing run was restarted after 12 h and stopped after 48 h.

194 **Sequencing analysis**

195 Raw fast5 files were processed via the Albacore 2.0.2 software (ONT) for basecalling, barcode
196 de-multiplexing and quality filtering (Phred quality score of >7) following manufacturer's
197 recommendations. All reads that passed quality filtering for each barcode were treated in
198 parallel as follows: firstly, NanoLyse (De Coster et al. 2018) was used to remove lambda phage

199 control DNA sequences; secondly, barcode and adapter sequences were trimmed from the
200 ends of reads using Porechop (Wick, Judd, and Holt 2018). To identify middle adapter
201 sequences by Porechop, a 95% threshold was set, and reads with middle adapter sequences
202 were split into two sequences; thirdly, seqtk (https://github.com/lh3/seqtk) was used to
203 convert the processed fastq file into fasta format for further analysis. The two nucleotide BLAST
204 (blastn, Basic Local Alignment Search Tool, Altschul et al. 1990) searches were performed using
205 the fasta files for each barcode as queries: first, against a custom genome reference database
206 that contained the wheat reference genome (Clavijo et al. 2017), accessed 09/2017, and the
207 genome sequence of three fungal species used in our study: *Puccinia striiformis* f. sp. *tritici*:
208 Schwessinger et al. (2018), accessed 09/2017; *Z. tritici*: Goodwin et al. (2011), version 2,
209 accessed 09/2017; *Pyrenophora tritici-repentis*: Manning et al. (2013), unmasked assembly,
210 accessed 09/2017. The genome of a fourth species suggested to be absent from the studied
211 wheat growing region, *Parastagonospora nordorum*, as a negative control (Hane et al. (2007),
212 accessed 09/2017. Individual reads were assigned to have originated from a specific DNA
213 sequence based on their best blastn hit (e-value < 0.01) (Zhang et al. 2000). Reads that did not
214 hit the custom-made database were captured using the filterbyname.sh script from the bbmap
215 script package (https://sourceforge.net/projects/bbmap/). These reads were used as queries to
216 search against the National Center for Biotechnology Information (NCBI) nucleotide (nt)
217 database (ftp://ftp.ncbi.nlm.nih.gov/blast/db/, downloaded 09/2017) using blastn and the
218 same settings and species assignment criteria mentioned above. Two Python scripts
219 (QC_and_BLAST.py and creating_final_dataframe.py) that combine all steps from quality
220 control to BLAST outputs and construct dataframes containing statistical and information from

221 BLAST searches for each read were deposited at
222 https://github.com/Yiheng323/Pathogen_Detection_and_Microbiome_scripts to facilitate the
223 reproducibility of our results. All basecalled fastq data for each sequencing run were deposited
224 at NCBI Short Reads Archive (SRA) under accession number PRJNA493553. The sequencing
225 summary files and final dataframes were deposited as .tab files for each replicate at
226 https://doi.org/10.6084/m9.figshare.7138262. The python scripts used for generating Fig. 2 – 5
227 and Supplementary Fig. S2 – S7 were also deposited at
228 https://github.com/Yiheng323/Pathogen_Detection_and_Microbiome_scripts.
229

230 **RESULTS**

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

232 We designed a pipeline for identification of three known wheat pathogens from infected field
233 samples using MinION sequencing. We infected three experimental field plots as disease
234 nurseries with their corresponding pathogens (*Puccinia. striiformis* f. sp. *tritici* for stripe rust
235 disease, *Z. tritici* for septoria tritici blotch and *Pyrenophora. tritici-repentis* for yellow leaf spot)
236 commonly found in New South Wales, Australia, and conducted sampling during the 2016-17
237 growing season (September). The sequencing experiment included three steps (Fig. 1): 1.
238 Pathologists collected three plants from each disease nursery field, with visual confirmation of
239 disease symptoms from three single infections plus one double infection (*Puccinia striiformis* f.
240 sp. *tritici* + *Z. tritici*) and a negative control comprising the same wheat variety treated with
241 commercial fungicides at the 10- to 11-week growth stage (three months post germination).

242 Individual plants were pooled by treatment type and shipped to the laboratory on ice for
243 processing. 2. We extracted DNA from two independent leaf cuts of each treatment type
244 without knowing the identity of each treatment to eliminate potential bias from operators. 3.
245 We barcoded each sample to enable pooling of all five sample types into one sequencing run.
246 The five wheat leaf samples were barcoded as follows: leaves infected with one of the three
247 fungal pathogens (barcodes 1-3); leaves infected with two pathogens (barcode 4); and fungicide
248 treated control without disease symptoms (barcode 5). We performed steps 2 to 3 four times
249 independently as biological replicates, with Step 3 performed using a MinION 1D PCR barcoding
250 workflow (EXP-PBC001 and SQK-LSK107). The raw data of each nanopore sequencing run were
251 converted into fastq sequencing reads using Albacore software (ONT), with a quality (Q) score
252 for each base estimating the likelihood of the identified base being correct using the Phred
253 scale (Ewing et al. 1998; Ewing and Green 1998). Based on read mean Q scores, Albacore
254 grouped reads into pass (Q score > 7) or fail (Q score <= 7) based on manufacturer's
255 recommendations. We binned reads based on the barcode sequences identified by Albacore at
256 both ends of each sequencing read. If no barcode sequence could be identified, we named the
257 corresponding read as unclassified. Overall, ~92-95% of the total sequence length of all four
258 sequencing runs passed quality filtering (Supplementary Table 1). The total sequence length
259 from the four runs varied from 682 to 5100 megabase pairs (Mbp). The mean read length
260 ranged from 1380 to 3054 bp. Proportions of each barcode in each run were relatively even,
261 although 23.3-25.7% of the sequences were unclassified in all four runs (Fig. 2).

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

263 We performed blastn searches for all quality-filtered reads against a custom genome reference
264 database. We assigned the origin of each read based on its best blastn hit within each database
265 (e-value < 0.01) using a 'winner-takes-all' strategy which only considers the best match of each
266 sequencing read. The custom genome reference database contained the wheat genome and
267 four fungal genomes including *Puccinia striiformis* f. sp. *tritici*, *Z. tritici*, *Pyrenophora tritici-*
268 *repentis* and *Parastagonospora nodorum*. These represent the host plant genome, three fungal
269 species present in infected field samples, and one fungal species (*Parastagonospora nordorum*)
270 not expected to be present in the sampling area as a negative control. The majority of reads
271 were assigned to wheat sequences, comprising total sequence length of 90.1% in flowcell 1,
272 80.1% in flowcell 2, 92.7% in flowcell 3 and 91.4% in flowcell 4 (Fig. 3). We found that healthy
273 leaf samples (barcode 5) contained no or minor total length percentages of pathogen-related
274 sequences in all four replicates (< 0.01%, Fig. 4). This indicates that our analysis method is
275 robust and does not lead to high background signals in healthy wheat samples (barcode 5).
276 Significantly higher proportions of total sequencing length (0.5–5.7%) from infected samples
277 (barcode 1-4) were assigned as pathogen genome-derived. In all cases, the pathogen genome
278 identified by the highest total sequence length corresponded to the pathogen causing the
279 disease as identified by symptomatology (Fig. 4). This indicates that we are clearly able to
280 identify the disease-causing pathogen via our sequencing workflow in cases of single infections.
281 Infections with two pathogens (barcode 4) were identified less clearly. In barcode 4 from all
282 biological replicates, total sequence length percentages assigned to *Puccinia striiformis* f. sp.
283 *tritici* were between 0.1% and 0.4%; close to or lower than the proportion assigned to negative
284 control species *Parastagonospora nodorum.* The total relative sequence length identifying each

285 major pathogen species within each barcode were similar across all replicates, with around 0.6%
286 of barcode 1 from *Puccinia striiformis* f. sp. *tritici*, 1.9-5.0% of barcode 2 from *Z. tritici*, 2.1-5.7%
287 of barcode 3 from *Pyrenophora tritici-repentis*, and 0.1-0.4% of barcode 4 from *Puccinia*
288 *striiformis* f. sp. *tritici* and 0.6-1.2% from *Z. tritici*. Unexpectedly, we also identified
289 *Parastagonospora nodorum* in samples derived from different treatment groups in all four
290 biological replicates. To ascertain whether *Parastagonospora nodorum* was truly present, we
291 extracted all 915 reads that were assigned to have originated from the *Parastagonospora*
292 *nodorum* genome in all four biological replicates. We used these reads to perform an open-
293 ended blastn search against the entire NCBI non-redundant nucleotide (nt) database. We found
294 that 344 of 915 reads were assigned to genus *Parastagonospora*, including 339 reads identifing
295 *Parastagonospora nodorum* SN15 genome (Supplementary Fig. S2). A further 371 reads hit
296 related species, including 226 from the same *Pleosporales* order: *Leptosphaeria* spp. (84 reads),
297 *Alternaria* spp. (81), *Pyrenophora* spp. (10), *Bipolaris* spp. (9), *Shiraia* spp. (9) and other genus
298 (Supplementary Fig. S2). The remaining 200 reads fell above the given e-value threshold (0.01)
299 and failed to be assigned. The difference between both BLAST searches potentially stems from
300 the relatively high error rate of the MinION sequencer and the distinct conserved regions from
301 genomes of closely related species (Jain et al. 2018; Laver et al. 2015). We tested for the
302 presence of *Parastagonospora nodorum* in one of our four biological replicates using an
303 independent molecular biology approach. We performed a PCR screen in replicate 3 using the
304 *PnTox3* gene that is universally found in all Australian *Parastagonospora nodorum* isolates but
305 not in any other fungal species (Liu et al. 2009). We were able to amplify a specific band
306 corresponding to *PnTox3* (Supplementary Fig. S3) in the *Z. tritici* treatment group (barcode 2)

307 which displayed the most *Parastagonospora nodorum* assigned sequencing reads (Fig. 4). The
308 *Puccinia striiformis* f. sp. *tritici* treatment group (barcode 1) and the *Puccinia striiformis* f. sp.
309 *tritici* + *Z. tritici* treatment group (barcode 4) contained a lesser number of *Parastagonospora*
310 *nodorum*-assigned sequencing reads, and these appeared to be below the detection threshold
311 of PCR analysis. These results support the finding that *Parastagonospora nodorum* is present in
312 our samples from the Wagga Wagga wheat-growing region despite the disease not being
313 commonly found in the region nor the pathogen being purposefully inoculated in the trial.

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

315 To obtain an overview of the microbiome profile in and on wheat leaves infected by different
316 fungal pathogens, we attempted to identify other microbial species present on the leaves. We
317 searched the non-redundant NCBI nt database with reads that were not assigned to any of the
318 five reference genomes present in our initial custom genome reference database. We found
319 that the proportion of reads assigned to a genome in the restricted and the nt-NCBI databases
320 were high for flowcells 1, 3 and 4 (92.9, 94.3 and 93.8%, respectively), but only 84.6% for
321 flowcell 2 likely due to technical issues. Overall, from each flowcell, less than 1.4% of passed
322 reads identified microbiota other than the major pathogens in the custom genome reference
323 database. We pooled the reads from each barcode in all biological replicates, and calculated the
324 total sequence length percentage that identified a specific genus in each sample using NCBI
325 taxonomic identifiers (Fig. 5A). Genus *Pseudomonas* took up the highest proportion in all
326 diseased samples, ranging from 0.08% to more than 0.88%, but was nearly absent in healthy
327 wheat (barcode 5, 0.01%). Genus *Alternaria* is also found in all samples with a proportion
328 between 0.05% to 0.2% in diseased samples and 0.005% in healthy wheat. Uniquely in *Puccinia*

329 *striiformis* f. sp. *tritici* treatment group (barcode 1), bacterial *Erwinia* comprised 0.09% of total
330 sequence length as represented by 219 reads and the second highest proportion. In contrast,
331 *Erwinia* only represented 0.003% and 0.001% in *Z. tritici* treatment group (barcode 2) and
332 fungicide treatment group (barcode 5) while being absent in any other treatment group. In
333 treatment groups infected with *Z. tritici* (barcodes 2 and 4), hits on *Z. tritici* were still present
334 and may represent sequences in the six *Z. tritici* genomes in the NCBI nt database that are
335 absent from the *Z. tritici* genome we used to generated our initial custom database. At the
336 species level (Fig. 5B), total sequence length percentages identifying *Pseudomonas syringae*
337 were much more frequent in treatment groups infected with *Puccinia striiformis* f. sp. *tritici*
338 (barcode 1, 0.59%) and/or *Z. tritici* (barcode 2, 0.33%). In contrast, the treatment group
339 infected wtih *Pyrenophora tritici-repentis* (barcode 3) displayed a much lower relative total
340 sequence length of 0.02% assigned to *Pseudomonas syringae*. Relative sequence length
341 assigned to other *Pseudomonas* species such as *Pseudomonas poae* followed a similar trend.
342 This finding is consisted with previous reports were *Pseudomonas* spp. have been descripted to
343 be associated with fungal infections of wheat (Al-Sallami et al. 1997; Mehrabi et al. 2016). At
344 the upper taxonomic classifications, *Proteobacteria* and *Ascomycota* were the most abundant
345 bacterial and fungal phyla across all treatment groups (Supplementary Fig. S4) followed by
346 *Actinobacteria* and *Basidiomycota*. At the class level (Supplementary Fig. S5), either
347 *Gammaproteobacteria* or *Dothideomycetes* were the most abundant. Overall, the microbiome
348 structure of the upper taxonomy resembles that of the lower taxonomy (Supplementary Fig. S4,
349 Fig. S5, Fig. S6, Fig. S7), as most sequences were assigned to genus *Pseudomonas*.
350

351 **DISCUSSION**

352 We have developed a whole genome shot-gun sequencing (WGS) workflow which accurately
353 identified three fungal pathogens responsible for three major wheat diseases – stripe rust,
354 septoria tritici blotch and yellow spot – from field samples. We designed a two-step BLAST
355 process to identify major agents in diseased samples and characterise the associated
356 microbiota. We identified a number of *Pseudomonas* species present in stripe rust and septoria
357 tritici blotch infected wheat leaves which appear far less abundant in yellow spot or non-
358 diseased leaves.

359 WGS provides a wealth of information that could be used to accurately identify disease-causing
360 agents and begin to describe microbial community composition in a sample (Brown et al. 2017;
361 Thomas, Gilbert, and Meyer 2012). Without PCR amplification process to potentially introduce
362 bias, WGS data is obtained directly from sample extracts and illustrates features closest to
363 reality dependent only on sampling and DNA extraction strategies. Omitting PCR steps, WGS
364 requires less processing time and proves advantageous for on-site diagnostic purposes where
365 technical simplicity and fast processing time are prioritised. It has a significant drawback,
366 however, in that most sequences are derived from the host (wheat) genome, representing 90.1%
367 of total sequence length in flowcell 1, 80.1% in flowcell 2, 92.7% in flowcell 3 and 91.4% in
368 flowcell 4. This result is due to our DNA extraction method in which the whole leaf sample was
369 ground to retain all the microbial population in and on the leaf. For the purpose of microbial
370 species identification, host sequences are unwanted and can overwhelm community DNA
371 (Sharpton 2014). To enrich microbial sequences, one strategy is to apply intricate molecular
372 methods during DNA extraction such as gradient centrifugation and sonication (Bulgarelli et al.

373 2012; Lim et al. 2014). As these techniques require sophisticated instruments, this can be a
374 challenge to apply to environmental samples, especially in field conditions. Another strategy to
375 enrich microbiome sequences involves sequencing of microbial metabarcode markers. With the
376 improvement of sequencing length, species-level resolution using long read amplicon
377 sequencing has been achieved in few cases (Benítez-Páez and Sanz 2017; Martijn et al. 2017).
378 Long-read amplicon sequencing seems to be a promising method for species identification of
379 known microbes, as it requires less input DNA than WGS and metabarcode databases contain
380 information from comparatively more species than whole genome databases (Cruaud et al.
381 2017; Tessler et al. 2017; Ranjan et al. 2016). Amplicon sequencing is also possible in applied
382 cases given the rapid development of remote PCR such as the miniPCR system (Amplyus,
383 Cambridge, USA) and the Bento Lab platform (Bento Bioworks Ltd, London, United Kingdom).
384 Drawbacks of amplicon strategy include biases related to PCR amplification, which lead to
385 inaccuracy when calculating relative proportions of microbial species.
386 While many workflows are available for microbiome profiling using metagenomics data, there is
387 no standard method for analysing long read data. This makes validation of both methods and
388 results crucial (Quince et al. 2017), and mock microbial communities or samples containing
389 known microbial species have been used to validate methods as a proof-of-concept (Nicholls et
390 al. 2018). Verifying the diagnosis of a specific pathogen strain based on SNPs called from
391 MinION reads, Votintseva et al. (2017) first identified pathogens using clinical methods and
392 then generated sequencing data to map this data to the targeted species in real-time.
393 Sequencing reactions were stopped once sufficient SNPs were obtained to confidentially
394 identify the targeted pathogen strain. Similarly, our study used wheat samples inoculated

395 manually by known pathogens and confirmed visually prior to sequencing. The major pathogen
396 identifications from the first BLAST analysis against the custom genome reference database
397 validated the capacity of our workflow for identification of those pathogen species. To validate
398 the result, different analysis approaches should be applied on the same dataset. Brown et al.
399 (2017) compared the performance of four bioinformatics pipelines on identifying species within
400 different synthetic bacterial communities using MinION. Compared to the composition of a
401 mock community, even the pipeline that resulted in the most similar classification still contains
402 reads classified to species known not to be present. These false positive results are due largely
403 to sequencing errors, which illustrates the need for more stringent filtering of the raw data and
404 improvement of per read accuracy.
405 Nanopore metagenomics has great potential for quick diagnosis of suddenly emerging crop
406 disease and large-scale disease monitoring at centralized agricultural institutions. As shown
407 from other sequencing studies performed in a more extreme environment (Johnson et al. 2017;
408 Boykin et al. 2018), all processes including DNA extraction and purification, library preparation
409 and sequencing can be completed within a few hours. A tailored database can be created for
410 each crop species containing the genomes of pathogens known to cause diseases to minimise
411 processing time. For rapidly evolving pathogens such as wheat rusts, the SNP-based
412 classification method described by Votintseva et al. (2017) could be adapted to define rust
413 strains once enough SNPs are obtained. However, before running the sequencer, sampling
414 procedures and tissue processing strategies still need to be carefully curated. In our study, we
415 randomly sampled multiple leaves from different plants belonging to five treatment groups and
416 performed complete sequencing and analysis independently on all four biological replicate

417 sample sets. Consistency in barcode classification and BLAST analysis across all replicates
418 suggested reliability of the workflow. In centralized agricultural institutions, amplicon
419 sequencing can potentially play a key role and, to-date, there have been a few preliminary
420 successes (Benítez-Páez, Portune, and Sanz 2016; Benítez-Páez and Sanz 2017; Pomerantz et al.
421 2018). Given the multiplexing capacity of the 96 native barcoding kit (ONT), many samples can
422 be sequenced in parallel when larger scale sequencing platforms are employed (e.g. GridION,
423 ONT). To achieve this goal, long read databases, sequencing accuracy and access to large-scale
424 sample processing power need to be improved.
425 In conclusion, our workflow demonstrates the potential of this technology for plant pathogen
426 diagnosis, field applications and microbiome characterisation. A combination of on-site and
427 centralized sequencing approaches would, in future, revolutionize management of agricultural
428 biosecurity and reduce crop loss.
429

430 **ACKNOWLEDGEMENTS**

431 We thank Dr. Megan McDonald for providing photos of infected wheat leaves, and thank Mr.
432 Michael McCaig, Mr. Tony Goldthorpe for their help during field sampling. We also thank Prof.
433 Peter Solomon for sharing knowledge and PCR primers for *Parastagonospora nodorum Tox3*
434 gene.
435
436

437 **LITERATURE CITED**

438 Al-Sallami, F., Karov, S., Vassileva, P., Popova, R., and Vassilev, V. 1997. Pseudomonas syringae pv.
439 atrofaciens associated with fungal black point of wheat (Triticum aestivum) grain. In *Pseudomonas*
440 *Syringae Pathovars and Related Pathogens*, Springer, p. 505–508.
441 Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. 1990. Basic local alignment search
442 tool. J. Mol. Biol. 215:403–410.
443 Ardui, S., Ameur, A., Vermeesch, J. R., and Hestand, M. S. 2018. Single molecule real-time (SMRT)
444 sequencing comes of age: applications and utilities for medical diagnostics. Nucleic Acids Res. 46:2159–
445 2168.
446 Benítez-Páez, A., Portune, K. J., and Sanz, Y. 2016. Species-level resolution of 16S rRNA gene amplicons
447 sequenced through the MinION™ portable nanopore sequencer. GigaScience. 5:4.
448 Benítez-Páez, A., and Sanz, Y. 2017. Multi-locus and long amplicon sequencing approach to study
449 microbial diversity at species level using the MinION™ portable nanopore sequencer. GigaScience. 6:1–
450 12.
451 Bhattacharya, S. 2017. Deadly new wheat disease threatens Europe's crops. Nat. News. 542:145.
452 Boykin, L., Ghalab, A., Rossitto De Marchi, B., Savill, A., M. Wainaina, J., Kinene, T., et al. 2018. Real time
453 portable genome sequencing for global food security. F1000Research. 7:1101.
454 Břinda, K., Callendrello, A., Cowley, L., Charalampous, T., Lee, R. S., MacFadden, D. R., et al. 2018.
455 Lineage calling can identify antibiotic resistant clones within minutes. bioRxiv. :403204.
456 Brown, B. L., Watson, M., Minot, S. S., Rivera, M. C., and Franklin, R. B. 2017. MinION™ nanopore
457 sequencing of environmental metagenomes: a synthetic approach. GigaScience. 6:1–10.
458 Bulgarelli, D., Rott, M., Schlaeppi, K., Themaat, E. V. L. van, Ahmadinejad, N., Assenza, F., et al. 2012.
459 Revealing structure and assembly cues for *Arabidopsis* root-inhabiting bacterial microbiota. Nature.
460 488:91–95.
461 Callaway, E. 2016. Devastating wheat fungus appears in Asia for first time. Nat. News. 532:421.

462 Charalampous, T., Richardson, H., Kay, G. L., Baldan, R., Jeanes, C., Rae, D., et al. 2018. Rapid Diagnosis
463 of Lower Respiratory Infection using Nanopore-based Clinical Metagenomics. bioRxiv. :387548.
464 Clavijo, B. J., Venturini, L., Schudoma, C., Accinelli, G. G., Kaithakottil, G., Wright, J., et al. 2017. An
465 improved assembly and annotation of the allohexaploid wheat genome identifies complete families of
466 agronomic genes and provides genomic evidence for chromosomal translocations. Genome Res.
467 27:885–896.
468 Cruaud, P., Rasplus, J.-Y., Rodriguez, L. J., and Cruaud, A. 2017. High-throughput sequencing of multiple
469 amplicons for barcoding and integrative taxonomy. Sci. Rep. 7:41948.
470 De Boer, S. H., and López, M. M. 2011. New Grower-Friendly Methods for Plant Pathogen Monitoring.
471 Annu. Rev. Phytopathol. 50:197–218.
472 De Coster, W., D'Hert, S., Schultz, D. T., Cruts, M., Van Broeckhoven, C., and Berger, B. 2018. NanoPack:
473 visualizing and processing long-read sequencing data. Bioinformatics. 34:2666–2669.
474 Ellis, J. G. 2017. Can Plant Microbiome Studies Lead to Effective Biocontrol of Plant Diseases? Mol. Plant.
475 Microbe Interact. 30:190–193.
476 Ewing, B., and Green, P. 1998. Base-Calling of Automated Sequencer Traces Using Phred. II. Error
477 Probabilities. Genome Res. 8:186–194.
478 Ewing, B., Hillier, L., Wendl, M. C., and Green, P. 1998. Base-Calling of Automated Sequencer Traces
479 UsingPhred. I. Accuracy Assessment. Genome Res. 8:175–185.
480 Fisher, M. C., Henk, D. A., Briggs, C. J., Brownstein, J. S., Madoff, L. C., McCraw, S. L., et al. 2012.
481 Emerging fungal threats to animal, plant and ecosystem health. Nature. 484:186–194.
482 Goodwin, S. B., M'Barek, S. B., Dhillon, B., Wittenberg, A. H. J., Crane, C. F., Hane, J. K., et al. 2011.
483 Finished Genome of the Fungal Wheat Pathogen Mycosphaerella graminicola Reveals Dispensome
484 Structure, Chromosome Plasticity, and Stealth Pathogenesis. PLOS Genet. 7:e1002070.

485 Hane, J. K., Lowe, R. G. T., Solomon, P. S., Tan, K.-C., Schoch, C. L., Spatafora, J. W., et al. 2007.
486 Dothideomycete–Plant Interactions Illuminated by Genome Sequencing and EST Analysis of the Wheat
487 Pathogen Stagonospora nodorum. Plant Cell. 19:3347–3368.
488 Hu, Y. 2016. High quality DNA extraction from Fungi_small scale. protocols.io. Available at:
489 https://www.protocols.io/view/high-quality-dna-extraction-from-fungi-small-scale-exmbfk6 [Accessed
490 September 26, 2018].
491 Hu, Y., and Schwessinger, B. 2018. Amplicon sequencing using MinION optimized from 1D native
492 barcoding genomic DNA. protocols.io. Available at: https://www.protocols.io/view/amplicon-
493 sequencing-using-minion-optimized-from-1d-mhkc34w [Accessed September 27, 2018].
494 Jain, M., Koren, S., Miga, K. H., Quick, J., Rand, A. C., Sasani, T. A., et al. 2018. Nanopore sequencing and
495 assembly of a human genome with ultra-long reads. Nat. Biotechnol. 36:338–345.
496 Jain, M., Olsen, H. E., Paten, B., and Akeson, M. 2016. The Oxford Nanopore MinION: delivery of
497 nanopore sequencing to the genomics community. Genome Biol. 17:239.
498 Janda, J. M., and Abbott, S. L. 2007. 16S rRNA Gene Sequencing for Bacterial Identification in the
499 Diagnostic Laboratory: Pluses, Perils, and Pitfalls. J. Clin. Microbiol. 45:2761–2764.
500 Johnson, S. S., Zaikova, E., Goerlitz, D. S., Bai, Y., and Tighe, S. W. 2017. Real-Time DNA Sequencing in the
501 Antarctic Dry Valleys Using the Oxford Nanopore Sequencer. J. Biomol. Tech. JBT. 28:2–7.
502 Kennedy, K., Hall, M. W., Lynch, M. D. J., Moreno-Hagelsieb, G., and Neufeld, J. D. 2014. Evaluating Bias
503 of Illumina-Based Bacterial 16S rRNA Gene Profiles. Appl. Environ. Microbiol. 80:5717–5722.
504 Laver, T., Harrison, J., O'Neill, P. A., Moore, K., Farbos, A., Paszkiewicz, K., et al. 2015. Assessing the
505 performance of the Oxford Nanopore Technologies MinION. Biomol. Detect. Quantif. 3:1–8.
506 Leggett, R. M., and Clark, M. D. 2017. A world of opportunities with nanopore sequencing. J. Exp. Bot.
507 68:5419–5429.

508 Lim, Y. W., Haynes, M., Furlan, M., Robertson, C. E., Harris, J. K., and Rohwer, F. 2014. Purifying the
509 Impure: Sequencing Metagenomes and Metatranscriptomes from Complex Animal-associated Samples.
510 JoVE J. Vis. Exp. :e52117–e52117.
511 Lin, S.-Y., Chooi, Y.-H., and Solomon, P. S. 2018. The global regulator of pathogenesis PnCon7 positively
512 regulates Tox3 effector gene expression through direct interaction in the wheat pathogen
513 Parastagonospora nodorum. Mol. Microbiol. 109:78–90.
514 Liu, Z., Faris, J. D., Oliver, R. P., Tan, K.-C., Solomon, P. S., McDonald, M. C., et al. 2009. SnTox3 Acts in
515 Effector Triggered Susceptibility to Induce Disease on Wheat Carrying the Snn3 Gene. PLOS Pathog.
516 5:e1000581.
517 Mahlein, A.-K. 2015. Plant Disease Detection by Imaging Sensors – Parallels and Specific Demands for
518 Precision Agriculture and Plant Phenotyping. Plant Dis. 100:241–251.
519 Maiden, M. C. J., Bygraves, J. A., Feil, E., Morelli, G., Russell, J. E., Urwin, R., et al. 1998. Multilocus
520 sequence typing: A portable approach to the identification of clones within populations of
521 pathogenic microorganisms. Proc. Natl. Acad. Sci. U. S. A. 95:3140–3145.
522 Manning, V. A., Pandelova, I., Dhillon, B., Wilhelm, L. J., Goodwin, S. B., Berlin, A. M., et al. 2013.
523 Comparative Genomics of a Plant-Pathogenic Fungus, Pyrenophora tritici-repentis, Reveals
524 Transduplication and the Impact of Repeat Elements on Pathogenicity and Population Divergence. G3
525 Genes Genomes Genet. 3:41–63.
526 Martijn, J., Lind, A. E., Spiers, I., Juzokaite, L., Bunikis, I., Pettersson, O. V., et al. 2017. Amplicon
527 sequencing of the 16S-ITS-23S rRNA operon with long-read technology for improved phylogenetic
528 classification of uncultured prokaryotes. bioRxiv. :234690.
529 Mehrabi, Z., McMillan, V. E., Clark, I. M., Canning, G., Hammond-Kosack, K. E., Preston, G., et al. 2016.
530 *Pseudomonas* spp. diversity is negatively associated with suppression of the wheat take-all pathogen. Sci.
531 Rep. 6:29905.

532 Mitsuhashi, S., Kryukov, K., Nakagawa, S., Takeuchi, J. S., Shiraishi, Y., Asano, K., et al. 2017. A portable
533 system for rapid bacterial composition analysis using a nanopore-based sequencer and laptop computer.
534 Sci. Rep. 7:5657.
535 Nezhad, A. S. 2014. Future of portable devices for plant pathogen diagnosis. Lab. Chip. 14:2887–2904.
536 Nicholls, S. M., Quick, J. C., Tang, S., and Loman, N. J. 2018. Ultra-deep, long-read nanopore sequencing
537 of mock microbial community standards. bioRxiv. :487033.
538 Odds, F. C., and Jacobsen, M. D. 2008. Multilocus Sequence Typing of Pathogenic Candida Species.
539 Eukaryot. Cell. 7:1075–1084.
540 Panke-Buisse, K., Poole, A. C., Goodrich, J. K., Ley, R. E., and Kao-Kniffin, J. 2015. Selection on soil
541 microbiomes reveals reproducible impacts on plant function. ISME J. 9:980–989.
542 PlantVillage. Available at: https://plantvillage.psu.edu/ [Accessed January 4, 2019].
543 Pomerantz, A., Peñafiel, N., Arteaga, A., Bustamante, L., Pichardo, F., Coloma, L. A., et al. 2018. Real-time
544 DNA barcoding in a rainforest using nanopore sequencing: opportunities for rapid biodiversity
545 assessments and local capacity building. GigaScience. 7.
546 Pretorius, Z. A., Singh, R. P., Wagoire, W. W., and Payne, T. S. 2000. Detection of Virulence to Wheat
547 Stem Rust Resistance Gene Sr31 in Puccinia graminis. f. sp. tritici in Uganda. Plant Dis. 84:203–203.
548 Quince, C., Walker, A. W., Simpson, J. T., Loman, N. J., and Segata, N. 2017. Shotgun metagenomics,
549 from sampling to analysis. Nat. Biotechnol. 35:833–844.
550 Raja, H. A., Miller, A. N., Pearce, C. J., and Oberlies, N. H. 2017. Fungal Identification Using Molecular
551 Tools: A Primer for the Natural Products Research Community. J. Nat. Prod. 80:756–770.
552 Ranjan, R., Rani, A., Metwally, A., McGee, H. S., and Perkins, D. L. 2016. Analysis of the microbiome:
553 Advantages of whole genome shotgun versus 16S amplicon sequencing. Biochem. Biophys. Res.
554 Commun. 469:967–977.

555 Ritpitakphong, U., Falquet, L., Vimoltust, A., Berger, A., Métraux, J.-P., and L'Haridon, F. 2016. The
556 microbiome of the leaf surface of Arabidopsis protects against a fungal pathogen. New Phytol.
557 210:1033–1043.
558 Schalamun, M., Nagar, R., Kainer, D., Beavan, E., Eccles, D., Rathjen, J. P., et al. 2018. Harnessing the
559 MinION: An example of how to establish long-read sequencing in a laboratory using challenging plant
560 tissue from Eucalyptus pauciflora. Mol. Ecol. Resour.
561 Schmidt, K., Mwaigwisya, S., Crossman, L. C., Doumith, M., Munroe, D., Pires, C., et al. 2017.
562 Identification of bacterial pathogens and antimicrobial resistance directly from clinical urines by
563 nanopore-based metagenomic sequencing. J. Antimicrob. Chemother. 72:104–114.
564 Schwessinger, B., Sperschneider, J., Cuddy, W. S., Garnica, D. P., Miller, M. E., Taylor, J. M., et al. 2018. A
565 Near-Complete Haplotype-Phased Genome of the Dikaryotic Wheat Stripe Rust Fungus Puccinia
566 striiformis f. sp. tritici Reveals High Interhaplotype Diversity. mBio. 9.
567 Sharpton, T. J. 2014. An introduction to the analysis of shotgun metagenomic data. Front. Plant Sci.
568 5:209.
569 Sherald, J. L., and Lei, J. D. 1991. Evaluation of a rapid ELISA test kit for detection of Xylella fastidiosa in
570 landscape trees. Plant Dis. USA. Available at: http://agris.fao.org/agris-
571 search/search.do?recordID=US9115875 [Accessed January 4, 2019].
572 Tessler, M., Neumann, J. S., Afshinnekoo, E., Pineda, M., Hersch, R., Velho, L. F. M., et al. 2017. Large-
573 scale differences in microbial biodiversity discovery between 16S amplicon and shotgun sequencing. Sci.
574 Rep. 7:6589.
575 Thomas, T., Gilbert, J., and Meyer, F. 2012. Metagenomics - a guide from sampling to data analysis.
576 Microb. Inform. Exp. 2:3.

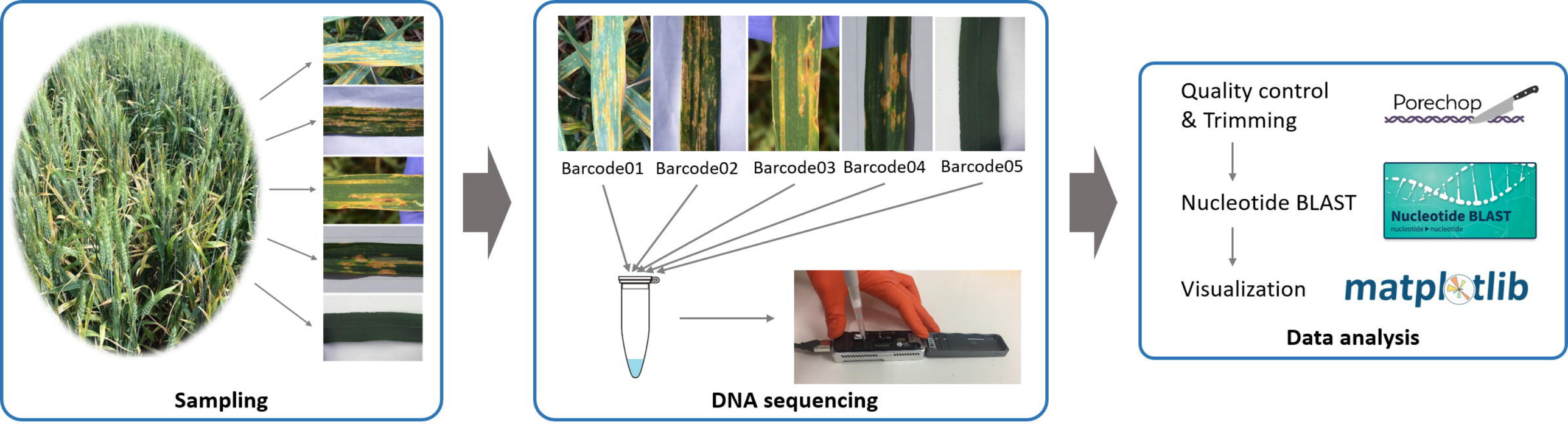
577 Votintseva, A. A., Bradley, P., Pankhurst, L., Elias, C. del O., Loose, M., Nilgiriwala, K., et al. 2017. Same-
578 Day Diagnostic and Surveillance Data for Tuberculosis via Whole-Genome Sequencing of Direct
579 Respiratory Samples. J. Clin. Microbiol. 55:1285–1298.
580 Wick, R. R., Judd, L. M., and Holt, K. E. 2018. Deepbinner: Demultiplexing barcoded Oxford Nanopore
581 reads with deep convolutional neural networks. PLOS Comput. Biol. 14:e1006583.
582 Zhang, Z., Schwartz, S., Wagner, L., and Miller, W. 2000. A greedy algorithm for aligning DNA sequences.
583 J. Comput. Biol. J. Comput. Mol. Cell Biol. 7:203–214.
584
585
586 Fig. 1. The Nanopore sequencing workflow for fungal pathogen identification. 1) Sampling:
587 plant pathologists collected wheat leaf samples from disease nursery fields, with visual
588 confirmation of disease symptoms for four infection types (three single fungal infections and
589 one double fungal infection). Fungicide-treated wheat leaves of the same variety and growth
590 stage were collected as a control treatment. 2) DNA sequencing: to minimise operator bias.
591 DNA was extracted from all five samples without knowledge of the causative agent(s) affecting
592 each sample. We labelled samples with DNA barcodes through a PCR reaction step before
593 sample pooling and 1D library preparations (SQK-LSK107); 3) Data analysis: We quality-
594 controlled sequences based on Phred quality scores and trimmed adapter sequence using
595 Porechop (Wick, Judd, and Holt 2018). We performed BLAST (Altschul et al., 1990) searches
596 using processed reads and identified the best hits. We summarized the total sequence length of
597 hits to each species using the python *matplotlib* module. Images used in step 3 are from the
598 official websites of each program.
599

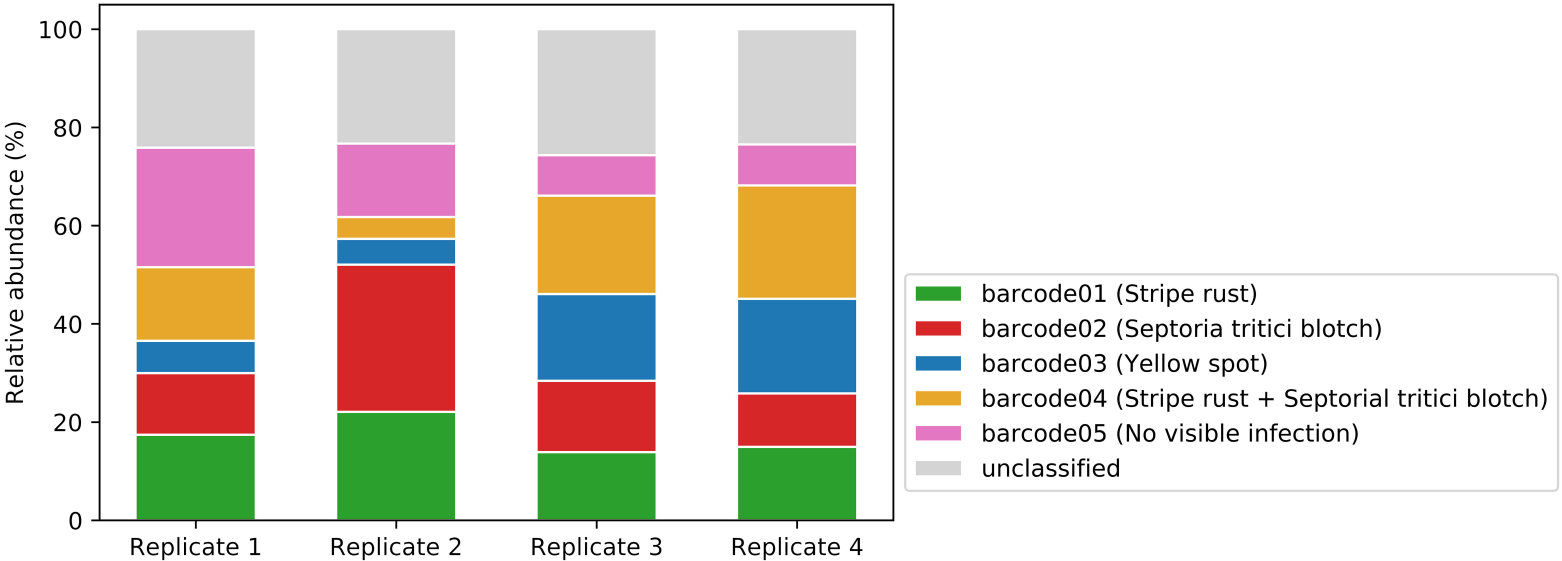
600 Fig. 2. Uneven proportions of barcode-labelled DNAs across all biological replicates. Columns
601 indicate four biological replicates from five different treatment types. The y-axis indicates
602 relative abundance of total sequence lengths labelled with one of five DNA barcodes within
603 each sample set. Each barcode represents one wheat sample. Reads without a detectable
604 barcode were labelled as 'unclassified'.
605
606 Fig. 3. Wheat is predominantly identified in the whole genome sequencing analysis using BLAST.
607 The graph shows the relative sequence length per species identified using a 'winner-takes-all'
608 approach when searching the custom genome reference database. Wheat (grey), fungal
609 pathogen genomes (blue), NCBI nucleotide (nt) database (red) and unmapped (light green) (see
610 results and methods for details). The numbers at the bottom of each bar show the total
611 sequence length (bp) of processed reads for each replicate. Replicates 1, 3, 4 showed similar
612 classification levels while replicate 2 contained a higher number of unclassified sequences.
613
614 Fig. 4. The total sequence length per pathogen genome per barcode correlated with the
615 phenotypic disease identification. The graphs show the relative sequence length assigned to
616 each pathogen genome using a 'winner-takes-all' approach in the BLAST analysis against the
617 custom genome reference database. The proportion of reads hitting the wheat host genome
618 are listed in the top-right corner of each graph. Unassigned reads are omitted from the
619 visualization.
620
621

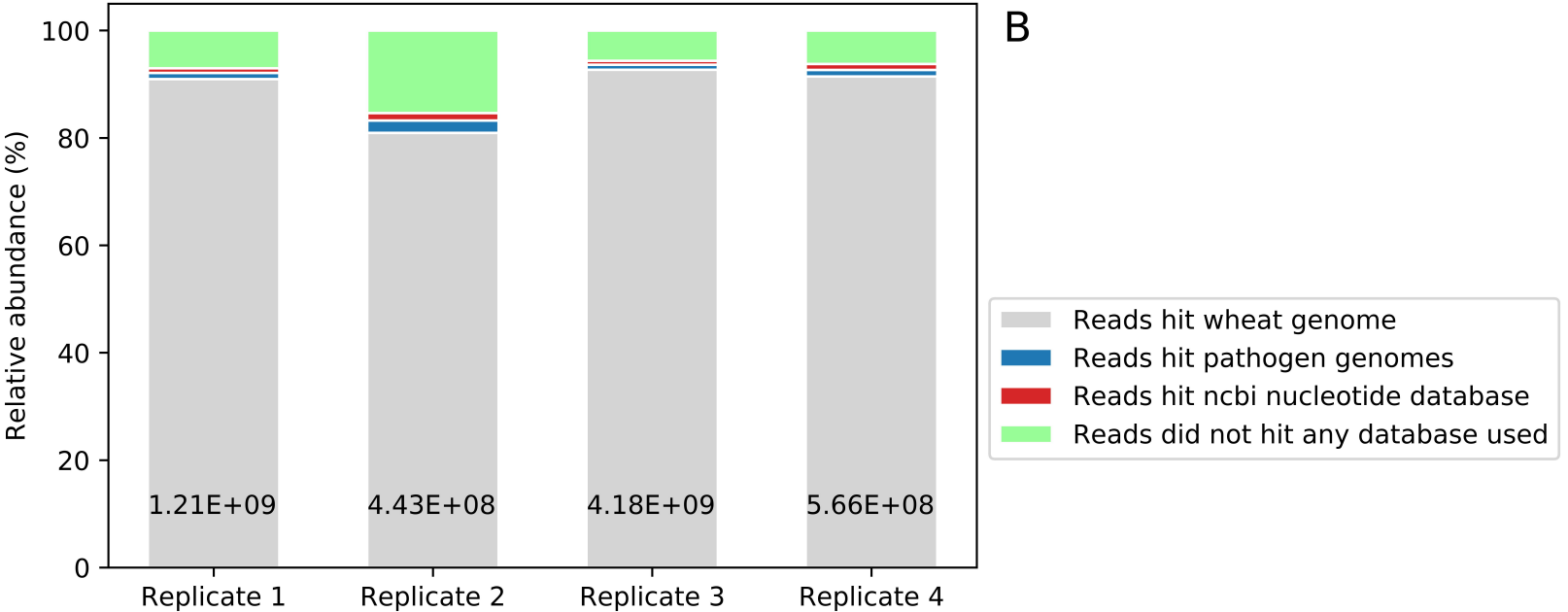
622 Fig. 5. Fungal infections of wheat increase the microbial population, especially for *Puccinia*
623 *striiformis* f. sp. *tritici* and *Z. tritici* infections. Percentage of sequences that did not hit the
624 custom genome reference database but hit the non-redundant NCBI nt database combined
625 from all replicates, and grouped by genus (A) or species (B). Different colours represented
626 different treatment types. Reads that were assigned to the restricted database were eliminated,
627 and the top ten highest classifications were displayed on the y-axis for each replicate. Genus
628 *Pseudomonas* represented the highest proportion across all barcodes containing *Puccinia*
629 *striiformis* f. sp. *tritici* or *Z. tritici*, with *Pseudomonas syringae* as the highest at the species level.
630
631 Supplementary Fig. S1. Example of DNA quality control by 1% agarose gel electrophoresis.
632 About 200 nanograms of DNA from replicates 3 and 4 prior to the adapter ligation step of the
633 sequencing library preparation were loaded onto the gel. The molecular weight standards are
634 HyperLadder 1 kb (Bioline, Catalog number BIO-33025).
635
636 Supplementary Fig. S2. BLAST analysis against the non-redundant NCBI nt database using the
637 reads assigned to *Parastagonospora nodorum* from BLAST against the restricted database. Arc
638 angle is proportional to the number of reads assigned to the indicated genus or order. The
639 unclassified (grey) arc angle refers to the proportion of reads assigned as unclassified by BLAST
640 analysis. Genus or orders assigned to two reads or less are grouped as 'others' under each
641 category.
642

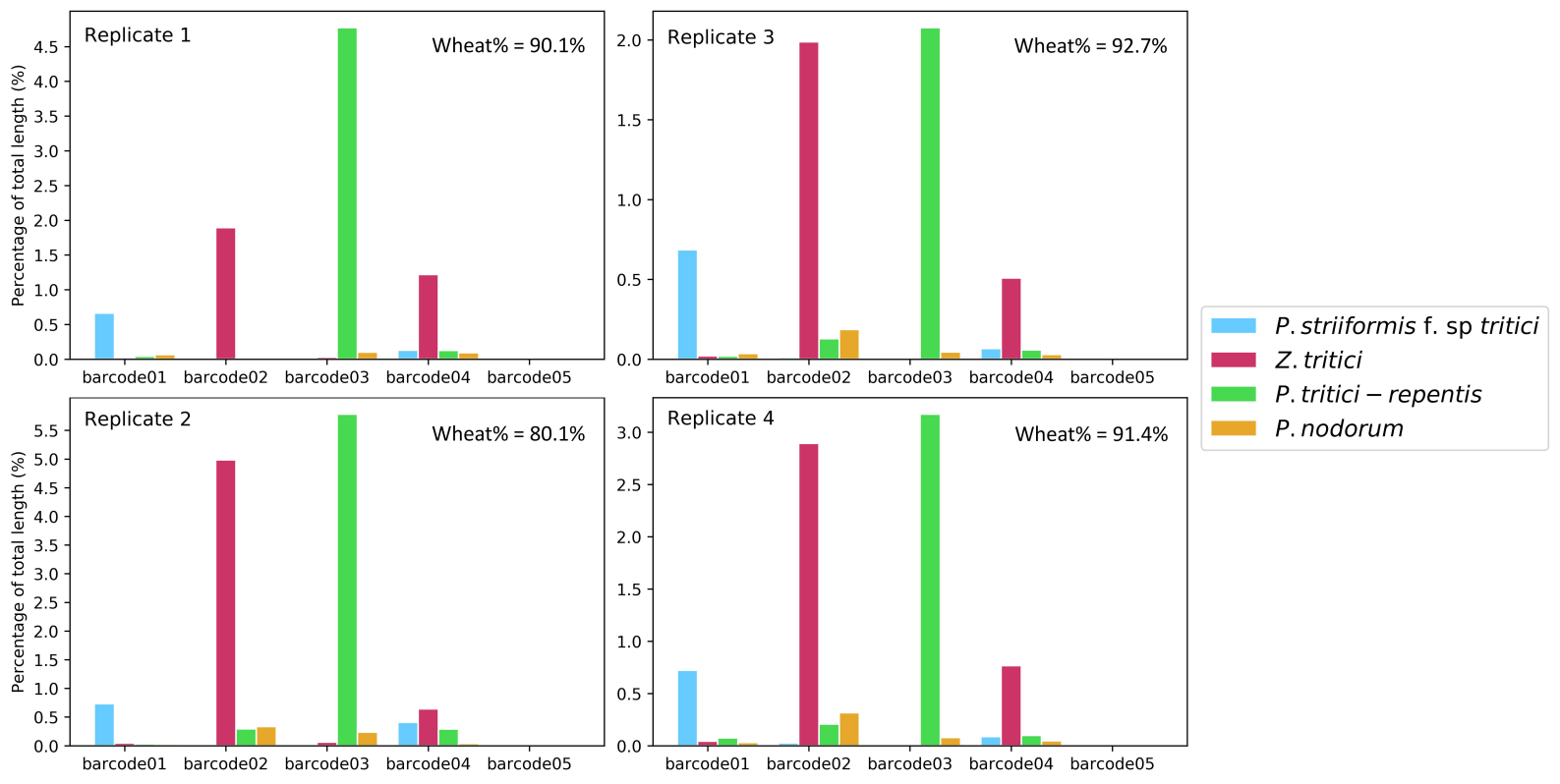
643 Supplementary Fig. S3. PCR amplification of the *Parastagonospora nodorum Tox3* gene from
644 the DNA sample of replicate 3. The wheat *Rubisco* gene was used as a control for the quality of
645 DNA in each sample. The positive control for *Tox3* amplification is pure genomic DNA from
646 *Parastagonospora nodorum* strain SN15, and the negative control is pure genomic DNA from *Z.*
647 *tritici*. The PCR reaction used NEB Q5 High-Fidelity 2X Master Mix (Catalog Number M0492S)
648 following the manufacturer's protocol and the amplification conditions were: 30 s at 98°C; 30
649 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
650 before holding at 4°C. The primer sequences for the wheat *Rubisco* gene are: Forward:
651 AGCAAGGTTGGCTTTGTCTT, Reverse: TTAACCTCCTCCACCTCGTT. The primer pair used for the
652 *Tox3* gene are: Forward: AATGTCGACCGTTTTGACC, Reverse: GGTTGCCGCAGTTGATATAA
653 obtained from Lin, Chooi, and Solomon (2018).
654
655 Supplementary Fig. S4. Percentage of sequences that did not hit the custom genome reference
656 database but hit the non-redundant NCBI nt database combined from all replicates, and
657 grouped by phylum. Reads that are assigned to the restricted database were eliminated, and
658 the top ten highest classifications were displayed for each replicate. Different colours
659 represented different barcodes.
660
661 Supplementary Fig. S5. Percentage of sequences that did not hit the custom genome reference
662 database but hit the non-redundant NCBI nt database combined from all replicates, and
663 grouped by class. Reads that are assigned to the restricted database were eliminated, and the

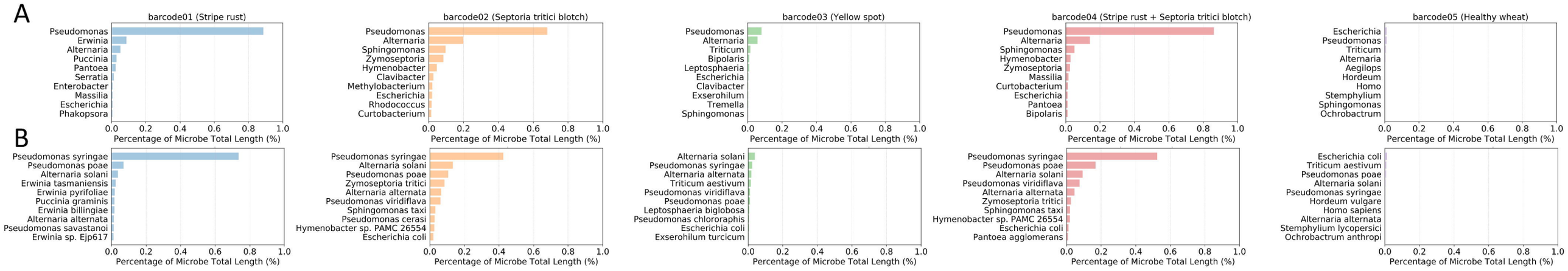
664 top ten highest classifications were displayed for each replicate. Different colours represented
665 different barcodes.
666
667 Supplementary Fig. S6. Percentage of sequences that did not hit the custom genome reference
668 database but hit the non-redundant NCBI nt database combined from all replicates, and
669 grouped by order. Reads that are assigned to the restricted database were eliminated, and the
670 top ten highest classifications were displayed for each replicate. Different colours represented
671 different barcodes.
672
673 Supplementary Fig. S7. Percentage of sequences that did not hit the custom genome reference
674 database but hit the non-redundant NCBI nt database combined from all replicates, and
675 grouped by family. Reads that are assigned to the restricted database were eliminated, and the
676 top ten highest classifications were displayed for each replicate. Different colours represented
677 different barcodes.


Sampling
Barcode01
Barcode02
Barcode03
Barcode04
Barcode05
DNA sequencing
Quality control
& Trimming
Porechop
Nucleotide BLAST
Nucleotide BLAST
nucleotide ▸ nucleotide
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


B
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 did not hit 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
0.0 0.2 0.4 0.6 0.8 1.0
Percentage of Microbe Total Length (%)