

## Tools and Resources

### 1 **CropQuant: An automated and scalable field phenotyping platform for crop** 2 **monitoring and trait measurements to facilitate breeding and digital agriculture**

3  
4 Ji Zhou<sup>1,2,\*</sup>, Daniel Reynolds<sup>1</sup>, Thomas Le Cornu<sup>1,3</sup>, Danny Websdale<sup>1,3</sup>, Simon  
5 Orford<sup>2</sup>, Clare Lister<sup>2</sup>, Oscar Gonzalez-Navarro<sup>1,2</sup>, Stephen Laycock<sup>3</sup>, Graham  
6 Finlayson<sup>3</sup>, Tim Stitt<sup>1</sup>, Matthew D. Clark<sup>1,3</sup>, Michael W. Bevan<sup>2</sup>, Simon Griffiths<sup>2,\*</sup>

7  
8 <sup>1</sup>Earlham Institute, Norwich Research Park, Norwich UK

9 <sup>2</sup>John Innes Centre, Norwich Research Park, Norwich UK

10 <sup>3</sup>University of East Anglia, Norwich Research Park, Norwich UK

11 \*Correspondence should be addressed to: [ji.zhou@earlham.ac.uk](mailto:ji.zhou@earlham.ac.uk) &

12 [Simon.Griffiths@jic.ac.uk](mailto:Simon.Griffiths@jic.ac.uk)

### 13 14 **Abstract**

15 Automated phenotyping technologies are capable of providing continuous and precise  
16 measurements of traits that are key to today's crop research, breeding and agronomic  
17 practices. In addition to monitoring developmental changes, high-frequency and  
18 high-precision phenotypic analysis can enable both accurate delineation of the  
19 genotype-to-phenotype pathway and the identification of genetic variation influencing  
20 environmental adaptation and yield potential. Here, we present an automated and  
21 scalable field phenotyping platform called CropQuant, designed for easy and cost-  
22 effective deployment in different environments. To manage infield experiments and  
23 crop-climate data collection, we have also developed a web-based control system  
24 called CropMonitor to provide a unified graphical user interface (GUI) to enable real-  
25 time interactions between users and their experiments. Furthermore, we established a  
26 high-throughput trait analysis pipeline for phenotypic analyses so that lightweight  
27 machine-learning modelling can be executed on CropQuant workstations to study the  
28 dynamic interactions between genotypes (G), phenotypes (P), and environmental  
29 factors (E). We have used these technologies since 2015 and reported results  
30 generated in 2015 and 2016 field experiments, including developmental profiles of  
31 five wheat genotypes, performance-related traits analyses, and new biological insights  
32 emerged from the application of the CropQuant platform.  
33

## Tools and Resources

### 34 Introduction

35 The great American wheat breeder and agricultural innovator Orville Vogel once  
36 stated, the plant we are looking for is in our plots, but we have to be there when it is.  
37 In order to select varieties with greater yield potential and enhanced environmental  
38 adaptation, agricultural practitioners, including breeders, farmers and crop researchers,  
39 have been optimising trait combination since the beginning of agriculture<sup>1,2</sup>. Four  
40 decades after temporary success in ensuring global food security<sup>3</sup>, we are now facing  
41 an even bigger challenge to feed generations to come<sup>4</sup>. Due to a narrowing range of  
42 available genetic diversity of modern crop germplasm collections<sup>5</sup> and increased  
43 fluctuations in growing conditions<sup>6</sup>, there is increasing emphasis placed on exploiting  
44 new sources of genetic variation to enhance environmental adaptation and sustainable  
45 yield in crop landraces and wild relatives<sup>7</sup>. To identify and assess these types of traits,  
46 multiple regular measures of crop growth and development are required to quantify  
47 subtle and dynamic phenotypes from many plots in different growing environments,  
48 demanding new screening technologies to integrate field environmental datasets with  
49 multi-scale phenotypic analyses to understand genotype-by-environment interactions  
50 (GxE) and associate them to genetic variation<sup>8-10</sup>.

51  
52 In contrast to current field phenotyping methodologies, which are still involving  
53 laborious manual scoring and relatively subjective selection, modern genetic and  
54 genomics techniques are being rapidly deployed in breeding and crop research to  
55 identify and utilise traits such as improved stress tolerance and disease resistance<sup>11</sup>.  
56 For example, quantitative trait locus (QTL) analysis and genome-wide association  
57 studies (GWAS) are used to identify loci<sup>12</sup>, whole genome sequencing used to reveal  
58 gene content and genetic variation<sup>13</sup>, and marker-assisted selection (MAS) and  
59 genomic selection used to breed new lines with favourable alleles<sup>14,15</sup>. Therefore, field  
60 phenotyping and the integration of environmental data with meaningful phenotypic  
61 analyses are a key bottleneck that limits the potential of recent advances in crop  
62 genetic and genomic technologies<sup>16,17</sup>.

63  
64 Remote sensing platforms<sup>18</sup> and open image-based analytics software libraries<sup>19</sup>  
65 start to enable researchers, breeders and agronomists to develop new approaches to  
66 understand and improve crop performance. For instance, unmanned aerial vehicles  
67 (UAVs) and light aircraft are being used to study crop performance and field  
68 variability<sup>20,21</sup>. Satellite imaging<sup>22-24</sup> and ground-based portable devices<sup>25,26</sup> have been  
69 applied to take snapshots of crop growth to estimate yield-related traits using canopy  
70 photosynthesis rate and normalised difference vegetation indices (NDVI). Field-based  
71 agricultural vehicles have been developed to capture physiological and developmental  
72 traits during the growing season<sup>18,27</sup>. Finally, large imaging platforms equipped with  
73 3D laser scanners (e.g. near-infrared laser lines and Light Detection and Ranging,  
74 LiDAR) and multi- or hyper-spectral sensors are applied to automate crop monitoring  
75 of a fixed number of pots or plots, either in the field<sup>28,29</sup> or in greenhouses<sup>12,30</sup>.  
76 Although these advances are making important contributions to the research domain,  
77 there are limitations and challenges associated with their usage such as high costs,  
78 restricted mobility and scalability, limited frequency of screening, and inadequate  
79 software tools for phenotypic analyses<sup>31,32</sup>. In particular, while satellite imagery and  
80 UAVs are capable of screening tens of thousands of plots at multiple locations, their  
81 applications are subject to civil aviation rules, low spatial resolution and bad weather  
82 conditions such as heavy rainfall, strong wind and cloud coverage. Ground-based

## Tools and Resources

83 portable devices and vehicles have shown greater mobility and high-resolution field  
84 data between multiple sites; however, they require experienced specialists to operate,  
85 limiting their applications to infrequent phenotyping. Still, stationary large imaging  
86 platforms are providing key data on dynamic crop growth and GxE interactions; but  
87 their scale of operation is restricted and they are relatively expensive for less well-  
88 funded research laboratories to access. Furthermore, they mostly rely on proprietary  
89 analytics software for data management and trait analysis, requiring ongoing licensing  
90 maintenance to use software products and extra fees if tailored functions are needed.  
91 For these reasons, it is challenging for researchers and breeding communities to adopt  
92 new phenotyping approaches due to their expenses, lack of suitable software, limited  
93 scope of operation and maintenance costs<sup>33</sup>. The ability to facilitate crop improvement  
94 programmes at multiple scales and locations is still limited.

95  
96 To enable the next-generation breeding and associated crop research<sup>34</sup>, affordable  
97 field phenotyping technologies need to be developed. New methods should exploit  
98 up-to-date remote sensing technologies together with state-of-the-art computer vision  
99 and software solutions, to equip researchers with diverse tools for multi-scale field  
100 phenotyping needs. The work described here aims to address these challenges by the  
101 development of an automated field phenotyping platform called **CropQuant**, which  
102 integrates cost-effective hardware with open source software capable of complex  
103 analytic solutions. We demonstrate applications of CropQuant (CQ) through multiple  
104 growth phenotypes measurements based on defined wheat genotypes over two  
105 growing seasons. Dynamic predictive growth models were also created to forecast the  
106 performance of wheat genotypes under varied environment conditions.

### 108 **Materials and Methods**

#### 109 *An open and low-cost hardware design*

110 In order to carry out automated phenotyping, we have deployed many low-cost crop  
111 monitoring CQ workstations (*terminals*) operating jointly via a preinstalled or self-  
112 operating infield wireless network. **Figure 1** shows the system architecture of the CQ  
113 platform. A terminal is centred around a simple single-board computer (**Fig. 1a**),  
114 running a customised Python-based analytic software package on the Linux *Debian*  
115 operating system (OS), integrating *Pi* or USB camera sensors (e.g. red-green-blue  
116 (RGB) or no infrared filter, NoIR), climate sensors (ambient and soil-based), sensor  
117 circuit boards, with wired or wireless data communications (**Figs 1b&c**). The design  
118 of the CQ platform is driven by the concept of Internet of Things (IoT)<sup>35,36</sup> as well as  
119 how to utilise hardware and software resources that are widely available, so that crop  
120 phenotyping solution can be scalable and affordable for the communities.

121  
122 We have used these technologies in field experiments since 2015. **Figure 2** shows  
123 an experimental scale of how CQ workstations were deployed for onsite and offsite  
124 wheat assessment experiments in 2015 and 2016. To keep costs low so that the  
125 technology can be adopted by the communities easily, we have developed different  
126 CQ versions (**Figs 2a-d**). For example, an all-in-one CQ (**Fig. 2a**) uses a *Raspberry Pi*  
127 2 single-board computer to control internal hardware (**Figs 2b**), including (1) a *Pi*  
128 camera sensor for time-lapse crop photography, (2) a tailored circuit board (**Fig. 2c**)  
129 to integrate climate sensors for collecting environmental data (i.e. soil temperature  
130 and moisture, ambient temperature and humidity, and light levels), (3) a USB WiFi  
131 dongle (or a radio transmitter) for data transfer and remote interactions, and (4) a USB

## Tools and Resources

132 flash drive for local data storage. **Supplementary Figure 1** demonstrates how an all-  
133 in-one CQ was used in field experiments.

134

135 Besides the relatively costly all-in-one version (\$230-\$240 to build), other more  
136 specialised versions (**Fig. 2d**) were much cheaper to produce. For example, a business  
137 card size CQ (version 4, \$80-90) uses a *Pi Zero* computer and is tailored for crop  
138 photography. A version 2 CQ (**Fig. 2b**, \$170-180) can mount different sensor groups  
139 (ambient or soil-based) for assessing agronomic characterisation and crop adaptation.  
140 With hardware modularity in mind, we have tested a range of single-board computers  
141 (e.g. the *Raspberry Pi* series, Intel's Galileo and Edison) for performing simple infield  
142 image analysis as well as integrating modular components. Although we finally chose  
143 *Pi* computers due to its performance-to-price ratio and extensive community support,  
144 the platform can operate on other single-board computers, if *Pi* computers are not  
145 available. For the peripheral hardware, we used off-the-shelf weatherproof containers  
146 (IP67 rating) together with micro-USB and Ethernet couplers (IP66 rating) to ensure  
147 environmental endurance and outdoor deployment. **Supplementary Figure 2** shows  
148 the hardware design of a version 2 CQ. A full hardware list and a construction manual  
149 are included in **Supplementary Note 1**.

150

### 151 *Offsite self-operating mode*

152 **Figure 2e** demonstrates an offsite field experiment using the CQ platform in 2015,  
153 where 10 field workstations were deployed to monitor canopy development (**Movie 1**)  
154 and crop growth (**Movie 2**) on one-metre wheat plots. CQ devices were powered by  
155 lead acid batteries with trickle charging from solar panels. To operate the device with  
156 minimal energy requirements, we have implemented a headless access mode to carry  
157 out wired data transfer. Besides the programmed imaging task, the system was only  
158 wakened if an Ethernet connection (i.e. local area network, LAN) was established.  
159 Offsite CQs were self-operating and used to perform image-based phenotyping. The  
160 infield imaging script running on CQs can be seen in **Supplementary Note 2**.

161

### 162 *Onsite networking mode*

163 For onsite experiments (**Fig. 2f**), CQs were powered by 5V/2A power supplies and  
164 connected to a field WiFi network. There were 14 networked CQ terminals (21 at  
165 peak time, with two dedicated for tiller abortion studies) jointly operating in 2016,  
166 monitoring 12 six-metre wheat plots to study performance-related traits and yield  
167 production. Although not thoroughly tested, for pre-installed WiFi network, the scale  
168 of the CQ platform can be increased by adding more standard routers to allow more  
169 CQ connections. We have added new functions to networked CQs including wireless  
170 control, programmable imaging, and on-board quality control (Online Methods). For  
171 instance, end-users can access CQ workstations remotely for real-time inspection,  
172 either using a portable device (e.g. a tablet or a smartphone) in the field or an onsite  
173 office computer (**Fig. 2g** and **Supplementary Fig. 3**). They can check the field in  
174 different regions to: (1) review historical crop images, (2) initiate new experiments,  
175 and (3) transfer crop-climate data to external computing storage.

176

177 These monitoring activities are administered by our web-based control system,  
178 **CropMonitor** (**Figs 2h-j** and **Supplementary Fig. 4**), where the status of each CQ  
179 terminal is updated constantly with information such as online or offline status,  
180 operational mode (e.g. green for operating, amber for idle, and red for operation error  
181 or ending tasks), representative daily images, micro-environment readings, and the



## Tools and Resources

182 usage of computing resources (i.e. CPU and memory). Furthermore, the CropMonitor  
183 system can support a range of tasks. For example, when deploying CQs in the field,  
184 CropMonitor can activate live streaming between a CQ terminal and a smart device  
185 (e.g. a smartphone or a tablet) to enable the calibration and installation of CQ devices  
186 (**Supplementary Fig. 5**). During the experiment, the CropMonitor system can  
187 establish a mesh network (based on all-in-one CQs) to support data communications  
188 between the field and external servers (Online Methods and **Supplementary Note 3**).  
189 Users can reposition CQ terminals at any time to change or initiate new monitoring  
190 tasks. Notably, the IoT-style setting can improve the mobility of the CQ platform. For  
191 example, **Supplementary Figure 6** shows a speed breeding experiment<sup>37</sup> monitored  
192 by CQs over a 75-day period in 2017, which was accomplished by moving CQs to  
193 indoor, i.e. a growth chamber (the speed breeding condition) and a glasshouse (the  
194 control condition).

195

196 To increase the scalability of the phenotyping platform, we recently developed a  
197 mesh network system to connect infield terminals with or without any pre-installed  
198 network infrastructure. **Supplementary Figure 7** illustrates the network topology,  
199 where all-in-one CQs are operating jointly as backbone *routers* (i.e. cluster servers),  
200 running both dynamic host configuration protocol (DHCP) and virtual network  
201 computing (VNC) servers for networking. Cheaper CQs (versions 2-4) are connected  
202 to the *routers* as terminals (i.e. *end devices*). Depending on the number of *routers* and  
203 the coverage of WiFi dongles (25-30 metres) or radio transmitters (several hundred  
204 metres), the field mesh network can be expanded or downsized flexibly. We dedicate  
205 an all-in-one CQ as a *coordinator* to control data communications between terminals  
206 and external networks. Meanwhile, computing tasks such as image selection, quality  
207 control and initial data annotation are distributed to terminals to reduce computational  
208 burden of the in-depth trait analysis (**Supplementary Note 4**).

209

210 Additionally, to verify the outputs of low-cost remote sensors integrated in CQs, we  
211 have utilised the meteorological outputs of a commercial weather station (**Figs 2k&l**),  
212 including temperature, rainfall, photosynthetically active solar radiation, wind speed  
213 and relative humidity. Phenotypic and climate datasets were saved on an onsite high-  
214 performance computing (HPC) cluster (SGI UV2000 system with Intel Xeon cores)  
215 for durable data storage (**Fig. 2m**). Since the application of the CQ technology, we  
216 have successfully accomplished three tasks essential for the next-generation field  
217 phenotyping<sup>17</sup>: (1) *continuous monitoring* via time-lapse crop photography, (2) *infield*  
218 *evaluation* through networked terminals and the CropMonitor system, and (3) *efficient*  
219 *data transfer* using distributed computing and wireless data communications through  
220 an infield network (**Supplementary Fig. 8**).

221

### 222 *The high-throughput analysis pipeline*

223 In order to enable accurate delineation of the genotype-to-phenotype pathway and  
224 identify genetic variation influencing environmental adaptation and yield potential,  
225 we chose a high-frequency (two-three times per hour) and high-precision (2592x1944  
226 pixels per image) phenotyping approach to monitor the morphological change of  
227 crops. After the crop photography phase, we exploited open image processing and  
228 machine learning libraries such as OpenCV<sup>38</sup>, Scikit-learn<sup>39</sup> and Scikit-image<sup>40</sup> and  
229 developed an automated analysis pipeline to extract biologically relevant outputs.

230

## Tools and Resources

231 We designed the analysis pipeline to be executable on either a workstation PC or an  
232 HPC cluster. Firstly, to arrange the collected image series, we have developed a  
233 selection algorithm to choose representative images based on their size, clarity<sup>41</sup>,  
234 imaging dates and genotypes (**Fig. 3, Step 1**). Only high-quality images were retained  
235 for trait analysis (Online Methods and **Supplementary Note 5**). All datasets were  
236 archived in a central repository such as HPC clusters for future reference. Then, we  
237 developed a referencing algorithm to define the location of a monitored plot over time  
238 (**Fig. 3, Step 2**). In real-world agricultural and breeding situations for which CQs are  
239 deployed, strong wind, heavy rainfall, irrigation and chemical spraying can lead to  
240 modest camera movements, causing cross-reference problems when comparing trait  
241 analyses for a plot over time (**Movie 3**). To resolve this issue, we have designed the  
242 referencing algorithm to identify the initial plot location so that each image in the  
243 series can be transferred to the same position for comparison. For example, the  
244 algorithm detects 2D coordinates of white reference canes installed in the plot and  
245 dark markers on a ranging pole for height scales using colour- and shape-based  
246 feature selection<sup>39</sup>. Then, it classifies pixels into five groups to represent the canopy  
247 space and background objects such as wheel tracks, sky, and the reference canes using  
248 simple k-means<sup>42</sup> and spectral clustering<sup>43</sup> algorithms. Finally, a pseudo 3D reference  
249 system is established to record important coordinates of the plot region, the canopy  
250 space, and height markers, together with converting measurements from pixels to  
251 metric units such as centimetres (Online Methods and **Supplementary Note 6**).  
252

253 Following Step 2, we integrated the initial reference location into a performance-  
254 related trait analysis algorithm. For a given image series, the algorithm applies an  
255 adaptive intensity and gamma equalisation method<sup>44</sup> to minimise colour distortion  
256 caused by varied field illumination. Then, it tracks geometric differences<sup>38</sup> between  
257 the monitored plot and the reference location. If the plot location has changed in a  
258 given image, a geometric transformation method<sup>45</sup> will be applied to reposition the  
259 image, removing areas outside the plot region and may or may not generate a black  
260 bar to the top of the image (**Fig. 3, Step 3**). Within the plot region, the algorithm  
261 detects the visible part of the ranging pole (**Movie 4**) as well as the canopy space for  
262 height measurement. For instance, to measure the height of the canopy, an entropy-  
263 based texture analysis is used to determine whether the canopy region is changing  
264 between two consecutive images using gray-level co-occurrence matrices (GLCM)<sup>46</sup>.  
265 If positional changes (moving up or down, depending on growth stages) are identified,  
266 the canopy height is recorded and corner-featured points<sup>47</sup> are detected (**Fig. 3, Step 4**),  
267 generating many red pseudo points casting in the canopy region for measuring canopy  
268 height (**Movie 5** and **Supplementary Note 7**). These pseudo points can also be used  
269 to represent the tips of erect leaves at stem elongation or jointing (the Zadoks scale<sup>48</sup>,  
270 growth stages, GS 32-39), reflective surfaces of curving leaves or crop heads between  
271 booting and anthesis (GS 41-69), and corner points on spikelets during senescence  
272 (GS 71-95). Using the trait analysis algorithm, we have computed the dynamic height  
273 changes to present growth patterns for different wheat genotypes (**Fig. 3, Step 5.1**).  
274

### 275 *Developmental related trait measurements*

276 In addition to the canopy height, we also developed functions to calculate other traits.  
277 For example, vegetative greenness is calculated based on the normalised greenness  
278 value (0-255) for a given plot over time. The output was used to assess the change of  
279 green biomass and vegetation period. We used this trait to evaluate a Stay-Green  
280 mutant (prolonged green leaf area duration with delayed leaf senescence; **Fig. 3, Step**

## Tools and Resources

281 5.2). The main orientation of a plot (0-180°) is also quantified based on edge detection  
282 methods<sup>49</sup>, representing the alignment of stems to estimate the change of stem rigidity  
283 (**Fig. 3, Step 5.3**). Using this trait, we have identified lines with higher lodging risk  
284 either during ripening or when interacting with heavy rainfall or strong wind (Online  
285 Methods and **Supplementary Table 1**).  
286

### 287 Results

#### 288 *Use case 1 – Monitoring five wheat NILs*

289 The diverse environments for which wheat has been adapted to grow provide  
290 opportunities for us to explore the dynamic interactions between genetic diversity and  
291 phenotypic traits under varied environmental conditions<sup>50</sup>. To test the CQ platform,  
292 we chose wheat near-isogenic lines (NILs, Online Methods) to examine a number of  
293 key performance-related phenotypes in the same *Paragon* (a UK spring wheat variety)  
294 genetic background<sup>51</sup>. **Figure 4** demonstrates five dynamic developmental profiles  
295 generated from the experiment between May and August 2015, a 95-day period. The  
296 experiment was conducted in plots in a field which is 2.1 miles away from Norwich  
297 Research Park UK (see the plot layout in **Supplementary Table 2**) and all five NILs  
298 were monitored twice per hour. The genotypes were: (1) Late-DTEM (days to ear  
299 emergence<sup>48</sup>, the number of days between sowing and ear emergence; *late* means  
300 GS55 is delayed), with *Ppd-1* loss of function (lof); (2) Early-DTEM (GS55 is moved  
301 forward), with *Ppd-D1a* photoperiod insensitivity; (3) Short stems, *Rht-D1b* semi  
302 dwarfing; (4) Stay-Green, a stay green mutant; and (5) *Paragon* wild type (WT).  
303

304 To compare the performance of the NILs, we used *Paragon* WT as the reference  
305 line and highlighted six key growth stages, from stem elongation or jointing (GS 32-  
306 39) to ripening (GS 91-95). The thermal time (degree-day, °Cd, using a 0°C base<sup>52</sup>)  
307 was also used as a heuristic tool<sup>53</sup> to normalise the crop growth. The five growth  
308 curves (1258-2297 °Cd) approximately followed a sigmoid curve. At the beginning of  
309 the experiment, *Ppd-D1a* NIL (Early-DTEM, coloured amber) was already at the end  
310 of the jointing stage (GS37-39) and hence was the first to reach a maximum height;  
311 whereas *Ppd-1* lof (Late-DTEM, coloured blue) was the last to increase in height. By  
312 cross-referencing developmental profiles based on six growth stages, we noticed that  
313 although *Ppd-D1a* and *Rht-D1b* (Short-Stem, coloured red) had similar maximum  
314 heights (83.4cm and 80.6cm), the latter displayed a relatively steady rate of increase  
315 in stature. *Ppd-1* lof's growth was the most delayed line, resulting in an extended  
316 period of vegetation, stem extension, and overall time to ear emergence. As this  
317 genotype has received the most thermal time units, it was the tallest line in the field  
318 experiment. Although all NILs experienced some degree of height reduction due to a  
319 significant storm on 24<sup>th</sup> July 2015, *Paragon* WT (coloured purple) presented a much  
320 lower lodging risk, as it maintained its height afterwards. To verify the phenotypic  
321 observation, we scored heading dates and canopy heights manually on the same plots  
322 and obtained a Pearson correlation coefficient of 0.986 (**Supplementary Table 3**).  
323

324 We summarised different temperatures and accumulated degree-days (ADDs) in  
325 both 2015 and 2016 growing seasons (**Fig. 5a**). As the average temperature in 2015 is  
326 much lower than in 2016, we used a fixed ADD period (1250-2300 °Cd) to segment  
327 crop growth under different climates. Within the same ADD period, **Figure 5b** shows  
328 dissimilar growth curves of *Paragon* WT (**Supplementary Table 4**). The 2015 curve  
329 was much steeper during stem elongation (GS32-59, 1250-1750 °Cd), possibly

## Tools and Resources

330 reflecting the cold spring. Flowering half complete (GS65) was reached on 24<sup>th</sup> June  
331 2015 (an early drilling late maturity mode). While the 2016 curve (values are the  
332 means of two biological replicates) had a steadier and extended development due to a  
333 warm spring. GS65 was reached on 12<sup>th</sup> June 2016, 12 days ahead of 2015 (a late  
334 drilling early maturity mode). Using the CQ platform, not only have we collected  
335 high-frequency crop climate datasets, we also could identify dynamic developmental  
336 variances for genotypes under different climate patterns.

337

338 *Use case 2 – New biological insights into growth patterns*

339 High-frequency and high-resolution deep phenotyping has already been employed in  
340 human disease research to reveal the underling mechanisms of individual's disease<sup>54</sup>.  
341 In plant research, the similar approach is being adopted to characterise phenotypes of  
342 plant responses to environmental challenges for field experiments<sup>55</sup>. While applying  
343 CQs in wheat assessment experiments, we have explored new biological insights into  
344 dynamic growth patterns using the large phenotypic data captured in the field.

345

346 **Figure 6** presents some preliminary results of how we utilised the high-frequency  
347 phenotypic data to extract underlying growth patterns for *Paragon* genotypes. Initially,  
348 we calculated daily relative growth rates (RGR, comparing with the previous day) for  
349 the five NILs monitored in 2015 and compared them with two *Paragon* WT measured  
350 in 2016. All RGR data were aligned by the associated growth stages for comparison,  
351 showing that all lines were active from jointing to flowering and became inactive after  
352 grain-filling (**Fig. 6a**). After that, to study the change of RGR during the growth  
353 stages (GS32-69), we explored the frequency and the degree of the RGR data. To be  
354 precise, we converted the data series from its original time domain (with equal daily  
355 readings) to the frequency domain using a fast Fourier transform algorithm (FFT)<sup>56</sup>.  
356 After the conversion, we separated the frequencies (x-axis, cycles per day, i.e. the  
357 frequency of growth) and the magnitude spectrum (y-axis, normalised amplitudes, i.e.  
358 the degree of growth) and generated underlying growth patterns of all the monitored  
359 lines (**Supplementary Note 8**). Noticeably, for *Paragon* WT, although temperatures  
360 and developmental profiles were significantly different between 2015 and 2016, the  
361 underlying growth patterns for *Paragon* WT in both years were very similar (**Fig. 6b**).  
362 We identified two distinct growth peaks: (1) around 15 days (15.3 days and 15.6 days  
363 respectively) and (2) seven-eight days (8.4 days and 7.8 days), indicating that  
364 *Paragon* WT is likely to control its underlying growth pattern based on the number of  
365 elapsed days instead of other factors such as temperatures.

366

367 For Late-DTEM and Early-DTEM NILs whose genetic backgrounds only differ by  
368 carrying alleles such as *Ppd-D1a* and *Ppd-1* lof, their growth patterns also contain  
369 two peaks (**Fig. 6c**): (1) similar to *Paragon* WT, seven-eight days (7.1 days and 7.7  
370 days) and (2) 23.0 days for Early-DTEM and 15.3 days for Later-DTEM. For other  
371 *Paragon* NILs (e.g. Stay-Green mutant and Short-Stem), although the patterns were  
372 slightly different, we found that at least one growth peak was close to the region of  
373 seven-eight days (**Supplementary Note 8** and **Supplementary Fig. 9**). To verify the  
374 FFT approach, we created a hypothetical *Paragon* growth data by combining all  
375 *Paragon* NILs across two years as a technical replicate. **Figure 6d** shows that no clear  
376 growth peaks can be detected from the hypothetical datasets. Additionally, we have  
377 applied the FFT approach to converted the RGR series with equal degree-day  
378 readings; similarly, no clear growth peaks can be identified (**Supplementary Note 9**).

379



## Tools and Resources

380 As all the tested *Paragon* NILs show a growth peak at seven-eight days and only  
381 Late-DTEM had a growth peak at 23 days, this might provide some insights into the  
382 mechanism of *Ppd-D1a*. The cyclical 23-day peak in growth over the common 15-day  
383 peak might reflect a changed output of the circadian clock of which *Ppd1* (PRR7 in  
384 *Arabidopsis*<sup>57</sup>) is accelerating the development in a cyclical manner. We are currently  
385 conducting a number of experiments, from gene expression to cell biology to advance  
386 our understanding of this discovery generated by the CQ platform.

387

### 388 *Use case 3 – GxE predictive modelling*

389 Crop modelling is used in breeding and crop research for integrating complex external  
390 and internal variables to understand GxE interactions and genetic systems. Many  
391 existing models use genotypes (G) and environmental factors (E) as input parameters  
392 to predict phenotypes (P) as the output of the models<sup>58–62</sup>. Similarly, we established a  
393 light-weight GxE model to predict crop growth using continuous crop-climate data  
394 collected by the CQ platform. Also, we used the computational power of CQ's single-  
395 board computers to explore how to operate the GxE model on a daily basis together  
396 with infield phenotyping tasks.

397

398 The key input components (environmental factors, growth stages and growth traits)  
399 of the model and how it was utilised for predicting growth in fluctuating growing  
400 conditions are summarised in **Figure 7**. Firstly, we selected environmental factors that  
401 were strongly correlated with the performance-related traits such as RGR and height  
402 at four key growth stages (from jointing to flowering) using *Pearson* correlation  
403 (**Supplementary Table 5**). This approach has identified five out of the 12 factors  
404 ( $p < 0.01$ ), including degree-day, solar radiation, rainfall, temperature, and daily light  
405 duration. Two heat maps (**Fig. 7a**) were produced to present the selected factors at the  
406 four stages (Online Methods and **Supplementary Note 10**). After that, we built a  
407 stage-based predictive model using training datasets of growth stages in 2015 and  
408 2016 (Online Methods and **Supplementary Note 11**). We employed support vector  
409 machines (SVM)<sup>63</sup>, a popular machine learning algorithm for classification, with  
410 radial basis function kernels to classify growth stages. **Figure 7b** illustrates the  
411 classified growth stages (coloured blue) benchmarked against the stages scored by  
412 expert crop physiologists (coloured red). **Supplementary Figure 10** illustrates the  
413 performance of the model when classifying the timing and duration for other wheat  
414 genotypes. We found that SVMs trained on two-season *Paragon* WT data had the  
415 highest scores using the benchmarking approach. Although the model modestly  
416 mistimes in booting (GS41-49) and heading (GS51-59) due to their short duration, we  
417 are adding new training data acquired in 2016 and 2017 to improve the model.

418

419 On the basis of the identified environmental factors and growth stage modelling, we  
420 explored a set of linear regression models to establish a global predictive model to  
421 forecast the continuous growth curve of *Paragon* genotypes, an approach that can be  
422 used to help farmers and breeders to optimise crop growth and genotype selection in  
423 the future. **Figure 7c** shows how the growth predictive model performs on the  
424 hypothetical *Paragon* growth data (a technical replicate with mean squared error:  
425 46.593, correlation: 0.999). The environmental factors and corresponding daily RGR  
426 data are grouped together for each stage and a linear regression model is then  
427 constructed to fit growth rate within the forecast growth stages, together with an  
428 ordinary least squares method to determine model coefficients. A relative growth rate  
429 estimate equation  $y_t^s = \beta^s x_t + c^s$  is used, where  $x_t$  is the environmental data at time



## Tools and Resources

430 point  $t$ ,  $\beta^s$  is the model coefficients (weight vectors), and  $c^s$  are constant offsets (i.e.  
431 intercepts for each growth stage). The super-script  $s$  denotes different growth stages  
432 (Online Methods and **Supplementary Note 12**).

433

434 Growth estimates at each stage are concatenated to form a single vector, based on  
435 which height values (in centimetres) and stage-based growth rates are calculated.  
436 Using the predictive model, we produced growth estimates for four wheat NILs  
437 (**Supplementary Fig. 11**). In this way, we compared how well the model performed  
438 with respect to the trait analyses recordings obtained from the CQ platform. To link  
439 the predictive model with crop agronomy, we also calculated the average standard  
440 deviation (SD) of the predicted crop height values, so a real-time warning message  
441 can be triggered on the CropMonitor control system. For example, if crop is growing  
442 outside the safe region (the bounds of its estimated height region,  $\pm 1SD$ , **Fig. 7d**),  
443 warning messages will be generated to inform the users that the growth is either too  
444 quick or too slow.

445

### 446 **Discussion**

447 With the development of modern high-throughput and low-cost genotyping platforms,  
448 the current bottleneck in breeding and crop research lies in phenotyping. Here, we  
449 describe CropQuant, an automated and scalable field phenotyping system which we  
450 believe can enable researchers with a toolkit that can fulfil multi-scale and diverse  
451 phenotyping needs for a broader plant research and breeding communities. To deliver  
452 the technology, we used the IoT in agriculture ethos to combine networked sensors,  
453 distributed computing hardware, computer vision, image analysis, machine learning  
454 and modelling to provide high-frequency field phenotyping together with GxE growth  
455 predictions, in near real time and in a manner which closely match human scoring.

456

457 To enable the CQ system to monitor and compare developmental changes between  
458 wheat genotypes using phenotypic data collected under different climate patterns, we  
459 carried out an open hardware R&D strategy so that CQ devices could be built and  
460 reproduced by other research groups in a relatively easy and cost-effective way. We  
461 utilised widely available *Pi* single-board computers and off-the-shelf climate sensors  
462 to facilitate the hardware design. For example, a CQ device can be equipped with a  
463 range of imaging sensors (e.g. RGB or NoIR *Pi* cameras, or USB and IP cameras) for  
464 varied experiments, including using *Pi* cameras to perform side-view and top-view  
465 imaging for crop growth and canopy development studies, setting up USB endoscope  
466 cameras below the canopy to study tiller abortion, and connecting IP cameras for field  
467 level monitoring. Environment sensors were grouped by functions, i.e. ambient and  
468 soil-based. So, dissimilar sensor groups could be selected for different experiments.  
469 To increase the capability and usefulness of the low-cost CQ device, we provided the  
470 hardware construction manual and the circuit board design (**Supplementary Fig. 12**)  
471 so that imaging and remote sensing functions could be integrated as well as expanded.  
472 We believe that, following the current hardware design, crop-climate data collection  
473 in the field can be standardised, data evaluation and communication can be carried in  
474 the field, and terminal nodes can be scalable on the CQ platform. Notably, due to the  
475 limited manufacturing ability and R&D funding, the maximum indoor and outdoor  
476 terminal nodes operating on the CQ platform at the same time was 21, although in  
477 theory the system can operate at least 255 nodes simultaneously.

478

## Tools and Resources

479 From the software development perspective, we created the CropMonitor control  
480 system to provide users with a unified web-based platform which can be used to  
481 connect phenotyping hardware with ongoing experiments in an integrated design. Not  
482 only does it allow different users to monitor experiments in real time, but it can  
483 incorporate different solutions in one shared web place to support experiments at  
484 different phases, i.e. from device deployment to the completion of the experiment.  
485 With crop-climate data collected in a standard manner, we developed a number of  
486 open-source trait analytic algorithms to measure multiple performance-related traits to  
487 identify genetic variation under different climate patterns. The software solutions  
488 have been evaluated by noisy images caused by complex field environment. Still, it  
489 can reliably execute the trait analysis tasks. To verify the results generated by the  
490 analysis pipeline, we have scored the performance phenotypes manually on the same  
491 plots over two growing seasons and obtained a strong correlation. Furthermore, we  
492 established dynamic predictive growth models to forecast the performance of wheat  
493 genotypes under varied growing conditions, which could be valuable for agronomic  
494 practices. Although the results are promising, it is noticeable that more training  
495 datasets are required to improve these models, ideally from varied growing conditions.  
496 As the software solutions were implemented on open image analysis, computer vision  
497 and machine learning libraries, they can be easily adopted and expanded for other  
498 experiments by the communities. To support computational users to understand our  
499 work, we have provided detailed comments in our source code.

500

501 From a biological perspective, the use of key performance traits generated by the  
502 CQ platform can be an excellent tool for screening early establishment, vegetation  
503 period, flowering, growth patterns and lodging. For example, vegetative greenness is  
504 a useful marker to quantify senescence; utilising the side-view movie, we can closely  
505 monitor the process of wheat aging, from the lower stem to the canopy region, a new  
506 approach to determine physiological maturity which is important for researching grain  
507 development and ripening. Also, continuously monitored greenness can be used in  
508 plant pathogen interaction to analyse the activity of pathogens on the leaf surface, as  
509 broad yellowish symptoms can be observed from susceptible plants (e.g. rust in  
510 wheat). Moreover, crop-climate data acquired by the CQ platform can also assist us to  
511 carry out novel biological discoveries. For instance, we are using deep-learning neural  
512 network architectures to train a convolutional neural network classifier (CNN) to  
513 quantify yield component traits such as spike per unit area and spike/spikelet number  
514 (**Supplementary Fig. 13**).

515

516 The CQ platform, in combination with networked remote sensors, the web-based  
517 control system, computational analytic solutions, and machine-learning based growth  
518 modelling, has enabled a cost-effective and scalable field-scale phenotyping of wheat  
519 germplasm. Multiple performance-related measurements were quantified in near real  
520 time and related to growing conditions. This technology has the potential for multiple  
521 applications in breeding and crop production, for example, to optimise the timing of  
522 fertiliser applications, irrigation, and predicate harvest dates for maximising yields in  
523 different agronomic scenarios. In crop breeding systems, regular field monitoring  
524 using the CQ platform identified multiple growth and developmental variables that  
525 provided statistically significant phenotypic analysis. These can increase the accuracy  
526 of breeding values, particularly for environmental response factors. With more field  
527 experimental data collected from different environments feeding into the system, the  
528 GxE predictive model and analytics software pipeline can be continuously improved.

## Tools and Resources

529 In particular, as the field of machine learning has progressed enormously in the last  
530 few years, our ability to model complex nonlinear functions and extract high-level  
531 phenotypic features is also growing. For example, we are applying deep learning (i.e.  
532 CNN and recurrent neural networks, RNNs) to learn and extract features from  
533 multidimensional imaging data (including visible and invisible spectrums) that are  
534 exceptionally difficult to accomplish through traditional image analysis approaches.  
535 Hence, we are consistently exploring deep learning to provide more accurate crop  
536 growth and development scores as well as yield-related trait quantifications, offering  
537 considerable value to the communities. Our future plan for the CQ platform is to  
538 improve the hardware to enhance mobility and modularity, and work with a broader  
539 plant research communities to jointly increase the software package for capability and  
540 applicability in different growing conditions. So, we could finally deliver real-time  
541 infield analysis and integrate field-based phenotyping, UAVs, and satellite into a  
542 multi-level and multi-dimensional crop analytic system.  
543

### 544 **Conclusion**

545 We believe that the CropQuant technologies described here may have a significant  
546 impact on future crop research, breeding activities, and agronomic practices. The  
547 reasons are: (1) the low-cost and widely available hardware centred by single-board  
548 computers is capable of enabling tasks such as continuous crop monitoring, infield  
549 evaluation and efficient data transfer, which are essential for the next-generation field  
550 phenotyping; (2) automated trait analysis algorithms integrated in the CQ platform are  
551 open-source and expandable software solutions, which are easily accessible and based  
552 on community driven numeric and scientific libraries; (3) use cases presented in the  
553 paper explain how to apply the CQ platform to study dynamic interactions between  
554 genotypes, phenotypes, and environmental factors, which is capable of producing new  
555 biological insights of growth patterns through phenotypic analyses. Moreover, our  
556 work endeavours to address the affordability and scalability issue for the research  
557 communities, which are independent from specific commercial hardware platforms  
558 and proprietary or specialised software applications, allowing the utilisation of the CQ  
559 platform to accomplish data annotation, performance-related phenotypic analysis, and  
560 cross-referencing results freely by the academic communities. Our work confirms  
561 previously reported results in the literature and produces novel approaches to enhance  
562 the reproducibility of indoor and outdoor crop growth and development experiments.  
563 Our case studies of wheat NILs are not limited. Natural variation, mineral or nutrient  
564 stress and other crop species could also be monitored using the platform.  
565

### 566 **Methods**

567 Methods and any associated references are available in the online version of the paper.  
568 Note: Supplementary information is available in the online version of the paper.  
569

### 570 **Acknowledgements**

571 We thank members of the Zhou laboratory and the Griffiths laboratory for fruitful  
572 discussions. Prof C. Martin and Prof R. Morris at John Innes Centre for reading and  
573 improving the manuscript. We thank S. Cossey and Prof Caccamo for supporting the  
574 initiation of the project, M. Hewitt for helping to engineer the first prototype, C.  
575 Mumford and her team for field trials, NBI infrastructure teams for infield network  
576 testing, and research leaders at EI, JIC, TSL and UEA for constructive discussions.

## Tools and Resources

577 This work was strategically funded by the BBSRC, Institute Strategic Programme  
578 Grants (BB/J004669/1) and Core Strategic Programme Grant (BB/CSP17270/1) at the  
579 Earlham Institute, GRO (BB/J004588/1) to M.W.B. at JIC. J.Z. was partially funded  
580 by BBSRC's Designing Future Wheat Cross-institute Strategic Programme Grants  
581 (BB/P016855/1) to Prof G. Moore. The biological research and related software and  
582 hardware R&D led by J.Z. was partially supported by BBSRC's FoF award (GP105-  
583 JZ1-B), an NRP Translational Fund (GP069-JZ1-Q), and the UK government's  
584 Eastern Agri-Tech Initiative funding (GP080-JZ1-M).

585

### 586 **Author contributions**

587 J.Z, D.R., S.L., T.L.C., M.D.C., M.W.B. and S.G. designed research; J.Z., D.R., O.G.,  
588 C.L. and S.O. performed the research; J.Z., D.R., T.L.C., D.W., and G.F. conducted  
589 hardware design and the development of analytics software packages; J.Z., D.R.,  
590 T.L.C., D.W., O.G., C.L., S.L., M.W.B. and S.G. contributed to analyse data; and J.Z.,  
591 D.R., T.L.C., D.W., M.D.C., M.W.B and S.G. wrote the paper. All authors have read  
592 and approved the final manuscript.

593

### 594 **Competing financial interests**

595 The authors declare no competing financial interests.

596

### 597 **Open Access**

598 The source code is distributed under the terms of the Creative Commons Attribution 4.0  
599 International License (<http://creativecommons.org/licenses/by/4.0/>), permitting unrestricted  
600 use, distribution, and reproduction in any medium, provided you give appropriate credit to the  
601 original authors and the source, provide a link to the Creative Commons license, and indicate  
602 if changes were made. Unless otherwise stated The Creative Commons Public Domain  
603 Dedication waiver applies to the data and results made available in this paper.

604

### 605 **Source code**

606 Source code is freely available for academic usage, which can be downloaded at  
607 [https://drive.google.com/drive/folders/0B17ZL8AzLo8wNFJUVS11OFkzb3M?usp=s](https://drive.google.com/drive/folders/0B17ZL8AzLo8wNFJUVS11OFkzb3M?usp=s_haring)  
608 [haring](https://drive.google.com/drive/folders/0B17ZL8AzLo8wNFJUVS11OFkzb3M?usp=s_haring) (an online Github repository is being prepared and will be updated in bioRxiv  
609 as soon as possible)

610

### 611 References

- 612 1. Sauer, C. O. *Agricultural origins and dispersal*. (American Geographical Society, 1952).
- 613 2. Donald, C. M. The breeding of crop ideotypes. *Euphytica* **17**, 385–403 (1968).
- 614 3. Borlaug, N. E. Wheat breeding and its impact on world food supply. in *3rd International*  
615 *Wheat Genetics Symposium* 1–36 (Australian Academy of Science, 1968).
- 616 4. Pingali, P. Green Revolution: Impacts, Limits, and the path ahead. *Proc. Natl. Acad. Sci.* **109**,  
617 12302–12308 (2012).
- 618 5. Tester, M. & Langridge, P. Breeding Technologies to Increase Crop Production in a Changing  
619 World. *Science (80-. )*. **327**, 818–822 (2010).
- 620 6. Howden, S. M. *et al.* Adapting agriculture to climate change. *Proc. Natl. Acad. Sci.* **104**,  
621 19691–19696 (2007).
- 622 7. McCouch, S. R. *et al.* Feeding the Future. *Nature* **499**, 23–24 (2013).
- 623 8. Malosetti, M., Ribaut, J. M., Vargas, M., Crossa, J. & Van Eeuwijk, F. A. A multi-trait multi-  
624 environment QTL mixed model with an application to drought and nitrogen stress trials in  
625 maize (*Zea mays* L.). *Euphytica* **161**, 241–257 (2008).
- 626 9. Houle, D., Govindaraju, D. R. & Omholt, S. Phenomics: the next challenge. *Nat. Rev. Genet.*  
627 **11**, 855–866 (2010).
- 628 10. Hawkesford, M. J. & Lorence, A. Plant phenotyping : increasing throughput and precision at  
629 multiple scales. *Funct. Plant Biol.* (2017). doi:10.1071/FP
- 630 11. Barabaschi, D. *et al.* Next generation breeding. *Plant Sci.* **242**, 3–13 (2015).
- 631 12. Yang, W. *et al.* Combining high-throughput phenotyping and genome-wide association studies  
632 to reveal natural genetic variation in rice. *Nat. Commun.* **5**, 1–9 (2014).
- 633 13. Brenchley, R. *et al.* Analysis of the bread wheat genome using whole-genome shotgun  
634 sequencing. *Nature* **491**, 705–10 (2012).
- 635 14. Araus, J. L. & Cairns, J. E. Field high-throughput phenotyping: the new crop breeding frontier.  
636 *Trends Plant Sci.* **19**, 52–61 (2014).
- 637 15. Bevan, M. W. *et al.* Genomic innovation for crop improvement. *Nature* **543**, 346–354 (2017).
- 638 16. Furbank, R. T. & Tester, M. Phenomics--technologies to relieve the phenotyping bottleneck.  
639 *Trends Plant Sci.* **16**, 635–44 (2011).
- 640 17. Cobb, J. N., DeClerck, G., Greenberg, A., Clark, R. & McCouch, S. Next-generation  
641 phenotyping: Requirements and strategies for enhancing our understanding of genotype-  
642 phenotype relationships and its relevance to crop improvement. *Theor. Appl. Genet.* **126**, 867–  
643 887 (2013).
- 644 18. Deery, D., Jimenez-Berni, J., Jones, H., Sirault, X. & Furbank, R. Proximal Remote Sensing  
645 Buggies and Potential Applications for Field-Based Phenotyping. *Agronomy* **4**, 349–379 (2014).
- 646 19. Eliceiri, K. *et al.* *Biological imaging software tools*. *Nature methods* **9**, (2012).
- 647 20. Sankaran, S. *et al.* Low-altitude, high-resolution aerial imaging systems for row and field crop  
648 phenotyping: A review. *Eur. J. Agron.* **70**, 112–123 (2015).
- 649 21. Duan, T. *et al.* Comparison of ground cover estimates from experiment plots in cotton,  
650 sorghum and sugarcane based on images and ortho-mosaics captured by UAV. *Funct. Plant*  
651 *Biol.* (2016). doi:10.1071/FP16123
- 652 22. Yang, C., Everitt, J. H., Du, Q., Luo, B. & Chanussot, J. Using high-resolution airborne and  
653 satellite imagery to assess crop growth and yield variability for precision agriculture. *Proc.*  
654 *IEEE* **101**, 582–592 (2013).
- 655 23. Zaman-Allah, M. *et al.* Unmanned aerial platform-based multi-spectral imaging for field  
656 phenotyping of maize. *Plant Methods* **11**, 35 (2015).
- 657 24. Lobell, D. B. The use of satellite data for crop yield gap analysis. *F. Crop. Res.* **143**, 56–64  
658 (2013).
- 659 25. Pask, A., Pietragalla, J. & Mullan, D. *Physiological Breeding II: A Field Guide to Wheat*  
660 *Phenotyping*. CIMMYT (CIMMYT, 2012). doi:10.1017/CBO9781107415324.004
- 661 26. Barker, J. *et al.* Development of a field-based high-throughput mobile phenotyping platform.  
662 *Comput. Electron. Agric.* **122**, 74–85 (2016).
- 663 27. Andrade-Sanchez, P., Gore, M. & Heun, J. Development and evaluation of a field-based high-  
664 throughput phenotyping platform. *Funct. Plant Biol.* **41**, 68–79 (2014).
- 665 28. Vadez, V. *et al.* LeasyScan: A novel concept combining 3D imaging and lysimetry for high-  
666 throughput phenotyping of traits controlling plant water budget. *J. Exp. Bot.* **66**, 5581–5593  
667 (2015).
- 668 29. Virlet, N., Sabermanesh, P., Sadeghitehran, K. & Hawkesford, M. J. Field Scanalyzer: An  
669 automated robotic field phenotyping platform for detailed crop monitoring. *Funct. Plant Biol.*



## Tools and Resources

- 670            **44**, 143–153 (2016).
- 671    30.    Chen, D. *et al.* Dissecting the Phenotypic Components of Crop Plant Growth and Drought  
672            Responses Based on High-Throughput Image Analysis. *Plant Cell Online* **26**, 4636–4655  
673            (2014).
- 674    31.    White, J. W. *et al.* Field-based phenomics for plant genetics research. *F. Crop. Res.* **133**, 101–  
675            112 (2012).
- 676    32.    Kelly, D. *et al.* An opinion on imaging challenges in phenotyping field crops. *Mach. Vis. Appl.*  
677            **27**, 681–694 (2016).
- 678    33.    Ribaut, J.-M., de Vicente, M. C. & Delannay, X. Molecular breeding in developing countries:  
679            challenges and perspectives. *Curr. Opin. Plant Biol.* **13**, 213–218 (2010).
- 680    34.    Moose, S. P. & Mumm, R. H. Molecular plant breeding as the foundation for 21st century crop  
681            improvement. *Plant Physiol.* **147**, 969–977 (2008).
- 682    35.    Gubbi, J., Buyya, R., Marusic, S. & Palaniswami, M. Internet of Things (IoT): A Vision,  
683            Architectural Elements, and Future Directions. *Futur. Gener. Comput. Syst.* **29**, 1645–1660  
684            (2013).
- 685    36.    the UK Government Chief Scientific Adviser. *The Internet of Things: making the most of the*  
686            *Second Digital Revolution.* (2014). doi:GS/14/1230
- 687    37.    Watson, A. *et al.* Speed breeding: a powerful tool to accelerate crop research and breeding. 1–  
688            17 (2017). doi:dx.doi.org/10.1101/161182
- 689    38.    Howse, J. *OpenCV Computer Vision with Python.* (Packt Publishing Ltd., 2013).
- 690    39.    Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **12**, 2825–  
691            2830 (2011).
- 692    40.    van der Walt, S. *et al.* Scikit-image: image processing in Python. *PeerJ* **2**, 1–18 (2014).
- 693    41.    Szeliski, R. *Computer Vision : Algorithms and Applications.* (Springer Science & Business  
694            Media, 2010). doi:10.1007/978-1-84882-935-0
- 695    42.    Aloise, D., Deshpande, A., Hansen, P. & Papat, P. NP-hardness of Euclidean sum-of-squares  
696            clustering. *Mach. Learn.* **75**, 245–248 (2009).
- 697    43.    Shi, J. & Malik, J. Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Mach.*  
698            *Intell.* **22**, 888–905 (2000).
- 699    44.    Bassiou, N. & Kotropoulos, C. Color image histogram equalization by absolute discounting  
700            back-off. *Comput. Vis. Image Underst.* **107**, 108–122 (2007).
- 701    45.    Sezgin, M. & Sankur, B. Survey over image thresholding techniques and quantitative  
702            performance evaluation. *J. Electron. Imaging* **13**, 146–165 (2004).
- 703    46.    Haralick, R., Shanmugan, K. & Dinstein, I. Textural features for image classification. *IEEE*  
704            *Transactions on Systems, Man and Cybernetics* **3**, 610–621 (1973).
- 705    47.    Harris, C. & Stephens, M. A Combined Corner and Edge Detector. *Proceedings Alvey Vis. Conf.*  
706            *1988* 147–151 (1988). doi:10.5244/C.2.23
- 707    48.    Zadocks, J. C., CHANG, T. T. & KONZAK, C. F. A decimal code for the growth stages of  
708            cereals. *Weed Res.* **14**, 415–421 (1974).
- 709    49.    Kimmel, R. & Bruckstein, A. M. Regularized Laplacian zero crossings as optimal edge  
710            integrators. *Int. J. Comput. Vis.* **53**, 225–243 (2003).
- 711    50.    Semenov, M. A. & Doblas-Reyes, F. J. Utility of dynamical seasonal forecasts in predicting  
712            crop yield. *Clim. Res.* **34**, 71–81 (2007).
- 713    51.    Griffiths, S. *et al.* Meta-QTL analysis of the genetic control of ear emergence in elite European  
714            winter wheat germplasm. *Theor. Appl. Genet.* **119**, 383–95 (2009).
- 715    52.    McMaster, G. S. & Wilhelm, W. W. Growing degree-days: One equation, two interpretations.  
716            *Agric. For. Meteorol.* **87**, 291–300 (1997).
- 717    53.    Lancashire, P. D. *et al.* A uniform decimal code for growth stages of crops and weeds. *Ann.*  
718            *Appl. Biol.* **119**, 561–601 (1991).
- 719    54.    Delude, C. M. Deep phenotyping: The details of disease. *Nature* **527**, S14–S15 (2015).
- 720    55.    Fiorani, F. & Schurr, U. Future scenarios for plant phenotyping. *Annu. Rev. Plant Biol.* **64**,  
721            267–91 (2013).
- 722    56.    Heideman, M., Johnson, D. & Burrus, C. Gauss and the history of the fast fourier transform.  
723            *IEEE ASSP Mag.* **1**, 14–21 (1984).
- 724    57.    Nakamichi, N., Kita, M., Ito, S., Yamashino, T. & Mizuno, T. PSEUDO-RESPONSE  
725            REGULATORS, PRR9, PRR7 and PRR5, Together play essential roles close to the circadian  
726            clock of *Arabidopsis thaliana*. *Plant Cell Physiol.* **46**, 686–698 (2005).
- 727    58.    Chenu, K., Dehifard, R. & Chapman, S. C. Large-scale characterization of drought pattern:  
728            A continent-wide modelling approach applied to the Australian wheatbelt - spatial and  
729            temporal trends. *New Phytol.* **198**, 801–820 (2013).

## Tools and Resources

- 730 59. Tardieu, F. Why work and discuss the basic principles of plant modelling 50 years after the  
731 first plant models? *J. Exp. Bot.* **61**, 2039–2041 (2010).
- 732 60. Turc, O., Bouteill, M., Fuad-Hassan, A., Welcker, C. & Tardieu, F. The growth of vegetative  
733 and reproductive structures (leaves and silks) respond similarly to hydraulic cues in maize.  
734 *New Phytol.* **212**, 377–388 (2016).
- 735 61. Dawson, I. K. *et al.* Barley: a translational model for adaptation to climate change. *New Phytol.*  
736 **206**, 913–931 (2015).
- 737 62. Palosuo, T. *et al.* Simulation of winter wheat yield and its variability in different climates of  
738 Europe: A comparison of eight crop growth models. *Eur. J. Agron.* **35**, 103–114 (2011).
- 739 63. Hsu, C.-W. & Lin, C.-J. A comparison of methods for multiclass support vector machines.  
740 *IEEE Trans. Neural Networks* **13**, 415–425 (2002).
- 741 64. Shaw, L. M., Turner, A. S., Herry, L., Griffiths, S. & Laurie, D. A. Mutant alleles of  
742 Photoperiod-1 in Wheat (*Triticum aestivum* L.) that confer a late flowering phenotype in long  
743 days. *PLoS One* **8**, (2013).
- 744 65. Shaw, L. M., Turner, A. S. & Laurie, D. A. The impact of photoperiod insensitive Ppd-1a  
745 mutations on the photoperiod pathway across the three genomes of hexaploid wheat (*Triticum*  
746 *aestivum*). *Plant J.* **71**, 71–84 (2012).
- 747 66. R Development Core Team. *R: A Language and Environment for Statistical Computing*. **1**, (R  
748 Foundation for Statistical Computing, 2008).
- 749 67. Shannon, C. E. Editorial note on ‘Communication in the presence of noise’. *Proc. IEEE* **72**,  
750 1713–1713 (1984).
- 751 68. Kohavi, R. A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model  
752 Selection. *Int. Jt. Conf. Artif. Intell.* **14**, 1137–1143 (1995).
- 753

## Tools and Resources

### 754 **Online Methods**

755

756 **Five wheat NILs** used in the field trial represent a range of genetic variation all with  
757 the genetic background of the UK elite spring wheat ‘*Paragon*’. The development of  
758 the Late-DTEM: Par (Norstar + Gamma 319c) 3c-11, *Ppd-1* loss of function (lof)  
759 lines is described previously<sup>64</sup>. The development of the Early-DTEM NILs: Par  
760 (GS100 2A+CS2B+Son64 2D)-T10 B10 -3b16 and *Ppd-D1a* photoperiod insensitive  
761 has also been published<sup>65</sup>. The novel line Stay-Green is line 2316b selected on the  
762 basis of stay green phenotype from a population of 7000 *Paragon* EMS mutants  
763 carried through single seed descent up to M6 developed under the Wheat Genetic  
764 Improvement Network of the UK Department of Food and Rural Affairs (Defra). The  
765 semi-dwarf NILs (short) were produced by marker assisted backcrossing (to BC6)  
766 using *Rht-B1* and *Rht-D1* KASP markers (LGC). which is available online from  
767 <http://www.cerealsdb.uk.net/cerealgenomics/CerealsDB>. The sources of *Rht-D1b* and  
768 *Rht-B1b* were the UK winter wheat varieties ‘Alchemy’ and ‘Robigus’ respectively.  
769 The five wheat lines were sown in single 1 m<sup>2</sup> plots in autumn 2014 at Church Farm,  
770 Norfolk UK, and grown according to standard agronomic practice. The manual score  
771 of days to ear emergence (DTEM) was done when 50% of the plot showed 50%  
772 emergence of the ear from the flag leaf. The manual measurement of plant height was  
773 done from the ear tip to ground level.

774

775 **The CropQuant hardware** contains many components, of which the centre one is a  
776 *Raspberry Pi 2* or *Pi 3* single-board computer (we have also used Intel® Edison in a  
777 different version of CropQuant workstation). Based on a mobile ARM processor, the  
778 *Raspberry Pi* computer features on-board external connections in the form of USB  
779 and Ethernet to allow expansion using additional peripherals as well as an array of  
780 digital GPIO (general purpose input and output) pins to interface with. The crop  
781 growth image acquisition was performed using a 5MP RGB or NoIR (No InfraRed  
782 filter) camera module connected via a CSI (Camera Serial Interface) port on the *Pi*  
783 mother board. Digital temperature and humidity sensors are connected via  
784 manufacturer supplied circuits to the GPIO pins of the *Pi* for interactive control. The  
785 sensors themselves are mounted separate from the circuits, externally on the  
786 CropQuant’s housing, wired through the base of the device and sheltered by a smaller,  
787 open housing unit. The external mounting allows for accurate sensing of ambient air  
788 conditions while sheltering the electronics from direct water damage. The CropQuant  
789 terminal is housed within a weatherproof (IP66 rated) plastic container, sealed around  
790 all openings allowing operation in the field. Physical connection to the system for  
791 data transfer via USB or Ethernet and power (12/5V DC) is facilitated by water-  
792 resistant couplers designed to be sealed against the rain and air moisture.

793

794 **The CropQuant software package** runs on Linux-based operating system *Debian*. It  
795 contains two servers, NetATalk and VNC sever, to facilitate infield data transfer and  
796 remote systems control, which allows users to connect to each CQ terminal through a  
797 wireless (using a tablet or a smartphone) or a wired connection (using a laptop). To  
798 enable real-time systems interactions, a GUI-based imaging program has been  
799 developed and added into the software package to control the RGB or NoIR camera  
800 module for time-lapse crop monitoring. The program can automatically detect the IP  
801 address of a given CropQuant terminal so that the terminal can be associated with its  
802 specific experiment ID of the field trial. After that, the program requests users to

## Tools and Resources

803 specify information such as genotype, biological replicates and imaging duration via a  
804 GUI dialog box, where users can initiate the image acquisition. The program can  
805 automatically adjust white balance, exposure mode and shutter speed in relation to  
806 variable infield lighting conditions using the *picamera* package\*, a Python interface to  
807 the *Raspberry Pi* camera hardware. Both image resolution and imaging frequency  
808 (three times per hour in our field trials) can be changed if users want to modify their  
809 experimental settings. The program also conducts the initial quality control and data  
810 backup after each image is captured.

811

812 Besides the image acquisition, the software package contains a variety of functions  
813 such as performing simple workstation and network diagnostics and synchronising  
814 with the central server twice within an hour to upload sensor data and CropQuant  
815 hardware information (see CropMonitor). Representative daily images are routinely  
816 selected and transferred to the central server during the night, which provides a daily  
817 snapshot of the monitored crops. Image data backups held on the SD card of the  
818 device are routinely synchronised with the server to provide an external backup as  
819 needed, with verification of multiple separate backups being performed by the process  
820 before it removes the archived data to free storage space. Relying on the Linux  
821 *crontab* scheduling system, we can monitor the performance of the software package  
822 and resume it automatically in cases of software interruption or power disruption. The  
823 SD card image running on the current version of CropQuant can be downloaded via  
824 <https://drive.google.com/drive/folders/0B17ZL8AzLo8wNFJUVS11OFkzb3M?usp=s>  
825 [haring](#). Source code is freely available for academic usage, which was arranged into  
826 source trees and saved in both local and central repositories. We are also preparing an  
827 online Github repository for the CropQuant project.

828

829 \*<https://picamera.readthedocs.io/en/release-1.13/>

830

831 **CropMonitor** is an IoT-style control system developed to oversee the whole CQ  
832 platform. It is operated through an onsite central server, logging updates received  
833 from individual clients, i.e. CQ terminals. A Python application on each workstation  
834 is running at regular intervals, scheduled by the native *Cron* Linux command line  
835 utility. The application queries the terminal to determine workstation status  
836 information such as uptime, network addresses and storage usage. Sensor data and  
837 more variable system data such as CPU temperature and the usage of processor and  
838 memory are sampled at a higher frequency and a median average of the readings is  
839 recorded during the half-hourly query. Once the application has collected all  
840 necessary data it is encoded into a JSON data object and transmitted over HTTP to the  
841 central server which stores the data in an SQL database running on a HPC cluster.  
842 CropQuant status is displayed and automatically updated using a web-based interface,  
843 determining whether each node is online by the time of the most recent update. The  
844 web interface provides information, including the location of each CropQuant  
845 terminal in the field (a field map needs to be uploaded to the central server), graphs of  
846 collected terminal and sensor data, and facilitates device configuration, SSH and VNC  
847 linking to all active nodes. Nodes within the CropMonitor system are categorised into  
848 groups and projects as defined by the user, allowing the organisation of workstations  
849 and restriction of access to stored data. CropMonitor provides a centralised real-time  
850 monitoring system to administer the network of infield workstations and collate  
851 collected data for visualisation, batch processing and annotation.

852

## Tools and Resources

853 **The image selection algorithm** is designed to perform speedy assessment of large  
854 image datasets captured in field trials by comparing images to a number of fixed  
855 criteria. The Python-based algorithm can be executed either on a normal computer or  
856 a HPC cluster. All images which meet the analysis standards will be collated. Over  
857 200 GB data have been generated by ten offsite CropQuant terminals in the 2015  
858 season during a 95-day period, with 50GB data were actually analysed after the  
859 selection procedure. In turn, an image is measured based on its brightness, sharpness,  
860 and shadow percentage, allowing all images which perform above a set of thresholds  
861 to be retained for further traits analysis. To determine the brightness of an image, the  
862 median value of pixel intensity is taken by transforming the image into HSV colour  
863 space. If the median intensity value is lower than a set threshold, the image is culled  
864 and not used from this point forward. The image clarity is determined by applying a  
865 Sobel edge detection<sup>41</sup> to the image. The detectable edges are calculated and then  
866 correlated with sharpness and exposure range of the image. The result of the clarity  
867 detection is also compared to a set threshold, which will disqualify images if they are  
868 out of focus or unclear with ill-defined edges. The final image test is of the percentage  
869 shadow within the visible area. Dark pixels found in an image with an illumination  
870 value of below 20% are either too dark for feature extraction or containing too much  
871 shadow in monitored plots. Once all rules have been passed, selected images are  
872 included in a result folder with a CSV file recording image metadata for further high-  
873 throughput image analysis.

874

875 **The plot detection algorithm** detects initial reference positions of monitored plots.  
876 The algorithm identifies the coordinates of white reference canes (the plot region) and  
877 dark height markers on a ranging pole, using an ensemble of colour-based feature  
878 selection on the basis of HSV (hue, saturation and value) and *Lab* non-linear colour  
879 space. It also classifies pixels into different groups, including sky, soil between plots,  
880 crop canopy, shadow, and plot regions using simple unsupervised machine-learning  
881 techniques such as k-means and spectral clustering. After detecting initial reference  
882 objects in the image, the algorithm establishes a pseudo 3D reference system that  
883 records the 2D coordinates of the plot area, the canopy region, and height markers  
884 through a range of feature selection approaches. The pixel-metric conversion is also  
885 computed based on height markers on the ranging pole.

886

887 **The CropMeasurer algorithm** employs an adaptive intensity and dynamic gamma  
888 equalisation to adjust colour and contrast to minimise colour distortion caused by  
889 diverse infield lighting. The algorithm tracks geometric differences between the plot  
890 on a given image and the initial position. If different, a geometric transformation  
891 method will be applied to recalibrate the image, which removes areas outside the plot  
892 area and could generate different sizes of black bars to the top of the given image.  
893 Within a plot, CropMeasurer tracks the crop height by detecting the visible part of the  
894 ranging pole and defines the canopy region through a combined adaptive thresholding  
895 and local Otsu threshold methods. Finally, the algorithm applies Harris and Shi-  
896 Tomasi corner detection methods<sup>47</sup> to locate corner-featured points within the canopy  
897 region. Red pseudo points are generated to represent the tips of erect leaves, reflective  
898 surfaces of curving leaves, heads and the corner points on ears. The main orientation  
899 of a given plot is quantified based on an optimised Canny edge detection method<sup>49</sup>,  
900 which computes the alignment of crop stems.

901



## Tools and Resources

902 **Data interpolation and analysis** have been used to handle minor data loss during the  
903 field experiments. Four days' data gap (at the end of May 2015) has been recorded on  
904 a number of offsite CropQuant workstations, which was caused by SD card crash due  
905 to short-term battery failure. We used cubic spline interpolation method<sup>66</sup> to fill the  
906 small gap in the phenotypic datasets.

907

908 **RGR and FFT conversion** the data is recorded on a daily basis, the maximum  
909 frequency component visible is every two days (0.5 cycles-per-day) due to the  
910 Nyquist-Shannon sampling theorem<sup>67</sup>. The cycles-per-day can be viewed in the same  
911 manner as Hertz, which is known as cycles-per-second and indicates a measure of  
912 frequency. We represent the frequency in cycles-per-day as the data were recorded at  
913 daily intervals.

914

915 **The growth stage predictive model** is the basis of the GxExP model. The model is  
916 produced to explore how to predict growth stages for different wheat genotypes in  
917 relation to real-time performance traits and environment data. It employs support  
918 vector machines (SVM), a popular machine learning technique for classification, with  
919 radial basis function kernels to classify growth stages. The performance of the model  
920 is tested by *Paragon* WT (G1) growth data from 2015 and 2016. For *Paragon* WT  
921 (2015), the model is trained with all other 2015 genotype data, whereas for the 2016  
922 *Paragon* WT datasets, the model is trained with all 2015 data, all models utilise K-  
923 Fold<sup>2</sup> cross-validation for prediction. To simulate the real-world situation for the stage  
924 prediction, we did *not* allow the model to obtain knowledge of the previous stages.  
925 Hence, the model mainly modestly mistimes booting (GS41-49) and heading (GS51-  
926 59) due to the short duration of both stages.

927

928 The prediction in comparison with the manually recorded growth stages suggests a  
929 successful prediction of the timing and duration across all growth stages for both 2015  
930 and 2016 datasets, except for the short transition period during booting (GS41-49),  
931 where the duration of booting is two days short. Due to the limited data points for  
932 booting across all genotypes used for training, the model cannot differentiate booting  
933 from heading sufficiently. For this matter, we are planning to add training datasets  
934 from other varieties such as Watkins and Chinese Spring wheat in other field trials.  
935 The stage prediction is trained by *Paragon* WT growth data from both 2015 and 2016:  
936 (1) to predict the 2015 *Paragon* WT, the model is trained with all other 2015 NIL  
937 growth data; (2) whereas the 2016 *Paragon* WT was based on all 2015 growth data.  
938 Through this approach, the model can rectify itself using previous years' training data.  
939 After the training phase, the model utilises K-Fold<sup>68</sup> cross-validation for the growth  
940 stage prediction.

941

942 **The GxE interaction model** explores the interactions between the recorded crop  
943 growth of five wheat genotypes and a number of environmental factors. Correlations  
944 are performed for each environmental factor grouped over three days with the  
945 recorded growth data. The reason to group environmental factors into nested three-  
946 day periods is to remove outliers and smooth the input data. The correlations are  
947 determined for the first four growth stage for five genotypes. The analysis is  
948 performed on the grouped data as particular stages (e.g. booting and heading) contain  
949 few recorded growth data due to the short duration of both stages were present during  
950 the growth. To determine the interactions between relative growth rates (RGR) and  
951 environmental factors, we used the formula  $(e^{RGR})^{-1}$  to convert negative correlation

## Tools and Resources

952 values to positive counterparts, as the RGR series is a decreasing sequence in relation  
953 to the increasing nature of growth stages.

954

955 Based on significant environmental factors, linear regression models<sup>45</sup> have been  
956 explored and a single linear regression model is selected to estimate RGR of five  
957 genotypes in relation to given infield environment conditions. Environmental factors  
958 with insignificant correlations (where  $p > 0.01$ , with respect to the height over the  
959 entire time-series) are removed from the analysis as they provide little predictive  
960 power. Ordinary least squares are used to derive the model coefficients. The RGR  
961 data is normalised to present percentage changes in height between two consecutive  
962 days. To predict the canopy height for a given genotype, environment data at each  
963 growth stage is input to the global model. To derive the height of the plant over time,  
964 successive application  $h_t = h_{t-1}(1 + y_t)$  is applied, where  $h_t$  is taken from the  
965 above equation,  $h_{t-1}$  is the height of the plant at the previous time-point, and  $h_0$  is  
966 equal to the initial height.

967

968 The performance of the model is verified by estimating the growth of all five NILs,  
969 including the overall paragon growth data (GT). The estimation is displayed with  
970 respect to the true canopy height datasets. The mean squared error recorded for G2  
971 (*genotype two*, Late-DTEM), G3 (*genotype three*, Early-DTEM) and G4 (*genotype*  
972 *four*, Stay-Green) shows that the estimated height is close to the true growth curves.  
973 However, the error is much larger for G1 (*genotype one*, Paragon WT) and G5  
974 (*genotype five* Short). This is due to the majority of crop growth happens during the  
975 early stages (GS32-GS59), estimation deviation during these initial stages could affect  
976 the overall height results. As the global predictive model might not be sensitive  
977 towards specific genotypes, we are still seeking a better approach to incorporate all  
978 genotypes with a similar genetic background into the prediction. The stage predictions  
979 are used in the linear regression growth model that could be run on a single-board  
980 computer such as a *Pi* computer to give accurate quantifications. We have chosen to  
981 establish the predictive growth model based on data produced from the CQ platform  
982 and hence did not perform cross-validation tests to offer more rigorous evidence of  
983 how well the model will generalise to new data. Given larger datasets containing  
984 more biological replicates, conducting cross-validation produces more reliable growth  
985 models with increased precision. Warning messages will be triggered via the CQ  
986 platform, if the crop growth rate has deviated from the bounds of its estimated growth  
987 region ( $\pm 1SD$ ).

988

989 **High definition movies** referred in this manuscript can be freely downloaded at  
990 <https://drive.google.com/drive/folders/0B17ZL8AzLo8wNFJUVS11OFkzb3M?usp=s>  
991 [haring](#) (source code is freely available for academic usage and we are also preparing  
992 an online Github repository for the CropQuant project).

993

994 **Code availability.** We used a *Jupyter* Notebook (i.e. the iPython Notebook) to  
995 present and explain algorithms and software solutions associated with the CQ project.  
996 They are freely available for academic use. Software packages running on CQs, high-  
997 throughput trait analysis algorithms, and GxE modelling can be downloaded via  
998 GoogleDocs for academic usage and an online Github repository is being prepared.

999 <https://drive.google.com/drive/folders/0B17ZL8AzLo8wNFJUVS11OFkzb3M?usp=s>

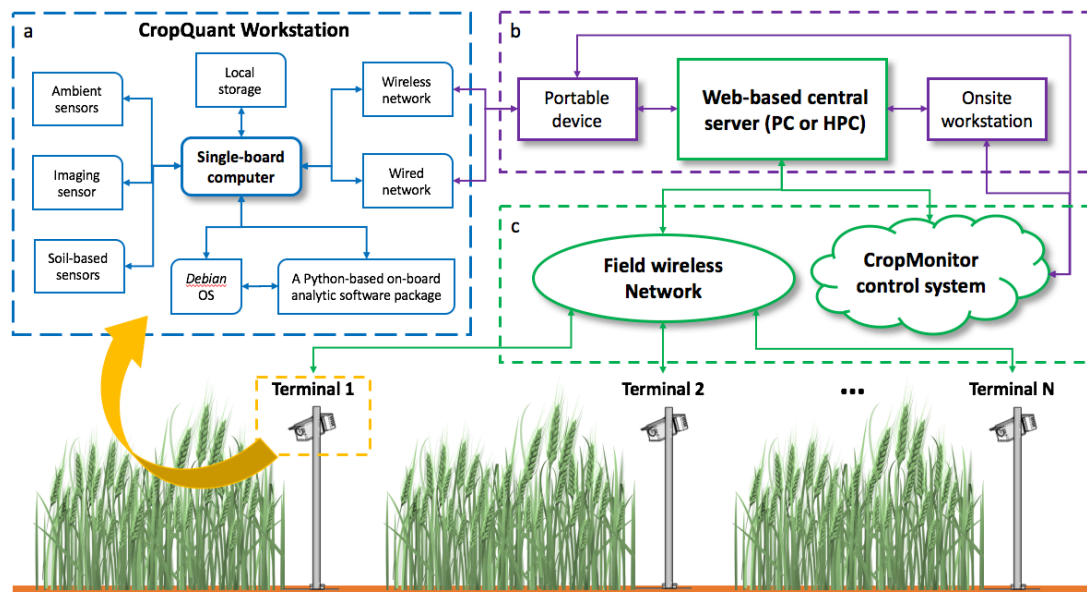
1000 [haring](#)

1001

## Tools and Resources

1002 **Figure 1**

1003



1004

1005

**Figure 1. A high-level system architecture of the CropQuant platform.**

1006

(a) The hardware and software design for a CQ workstation, including a single-board computer, climate sensors, a tailored circuit board to integrate sensors, an imaging sensor, local USB storage, wired and wireless network components, *Debian* operating system, and custom-made Python software package for crop images and climate data collection. (b) Data communications between CQs and either a portable device in the field or an onsite PC workstation. (c) The network setting that integrates CQ terminals, infield wireless network, and the CropMonitor control system.

1007

1008

1009

1010

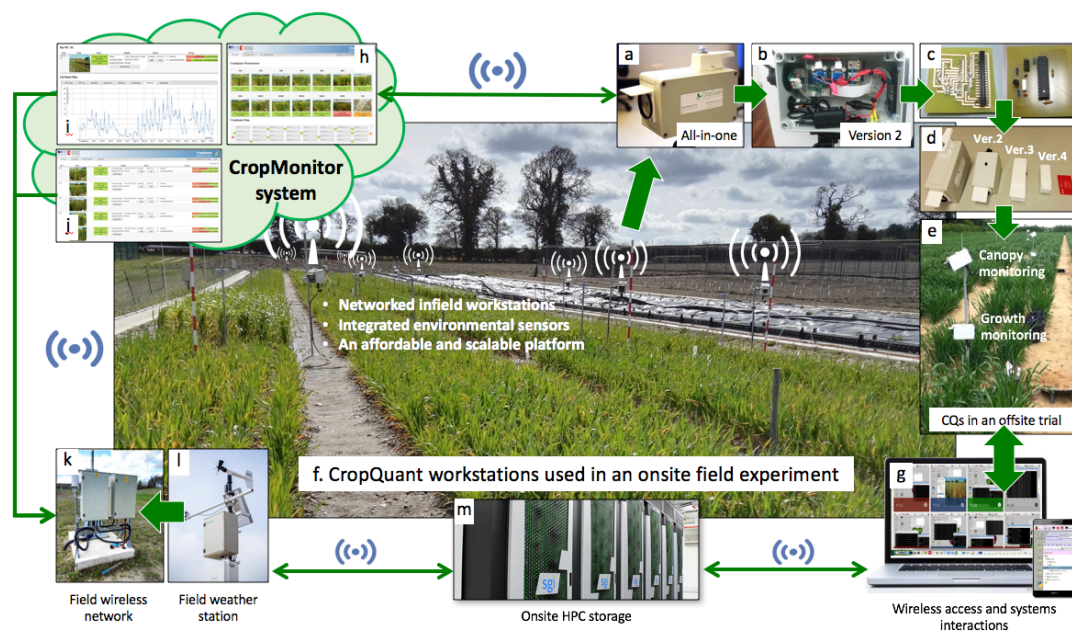
1011

1012

1013

## Tools and Resources

1014 **Figure 2**  
1015

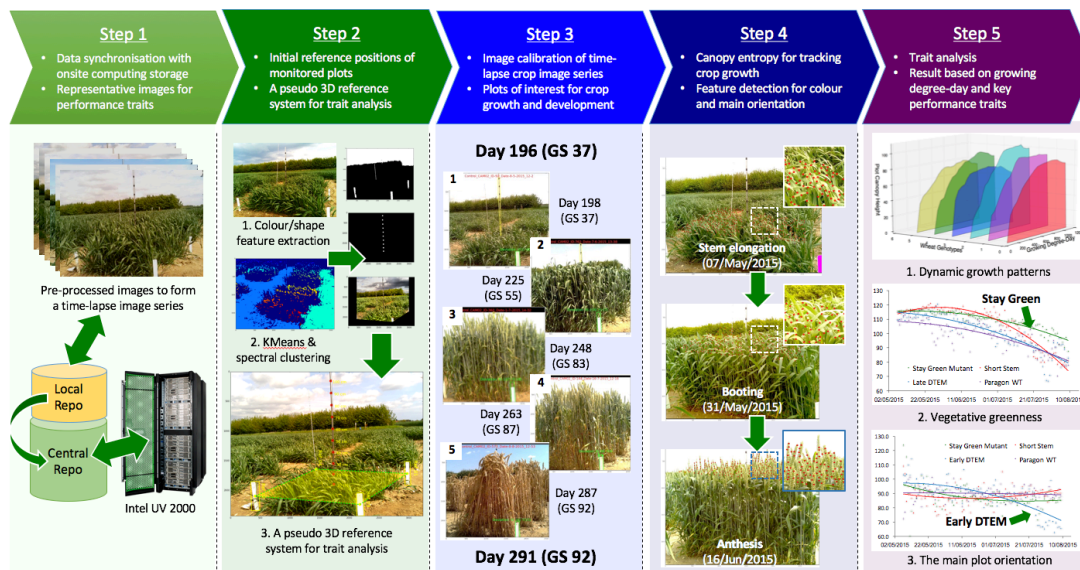


1016  
1017  
1018  
1019  
1020  
1021  
1022  
1023  
1024  
1025  
1026  
1027  
1028  
1029  
1030

**Figure 2. The CropQuant phenotyping platform used in field experiments.** (a) An all-in-one CQ workstation. (b) The internal hardware design of a version 2 CQ workstation. (c) A custom-made circuit board that integrates low-cost climate sensors. (d) Different versions of CQ built for dissimilar monitoring tasks. (e) CQs used in offsite field experiments, powered by batteries and solar panels. (f) The CQ platform established for onsite field experiments, powered by 5V/2A power supplies. (g) Real-time crop inspection using either a portal device in the field or a PC in an office. (h-j) the CropMonitor control system administers CQ terminals and records information such as online or offline status, operational mode, daily representative crop image, micro-environment for the plot region, and computational resource for each connected CQ terminal. (k) An infield WIFI system installed for field trials. (l) A commercial field weather station. (m) HPC clusters used for durable data storage and trait analysis.



1031 *Figure 3*  
1032



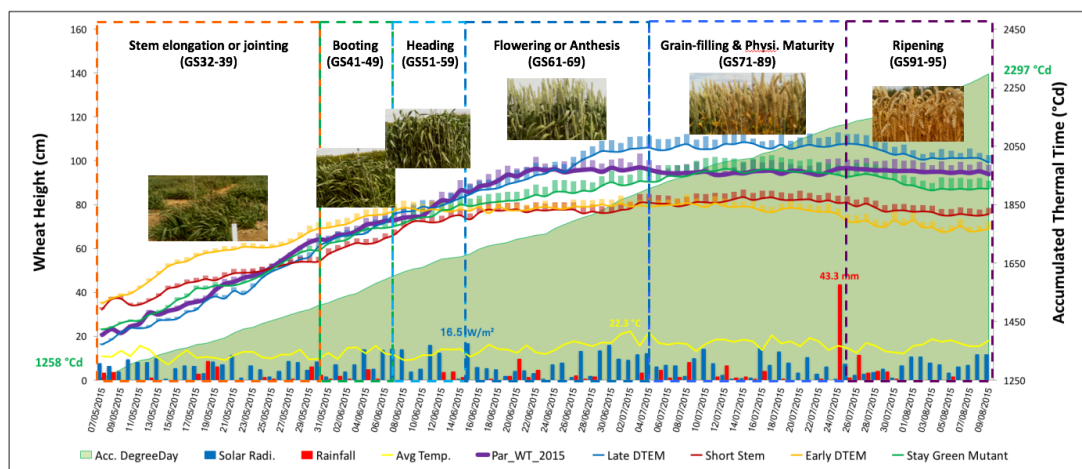
1033 **Figure 3. The high-throughput analysis pipeline established for batch processing**  
1034 **and measuring performance-related traits.**

1035 (1) *Step 1*, high-quality crop images were selected by the selection algorithm and  
1036 stored in both local and central repositories. (2) *Step 2*, initial reference positions of  
1037 monitored plots were detected by the referencing algorithm, which also calculated the  
1038 pixel-metric conversion rate. (3) *Steps 3 and 4*, the performance-related trait analysis  
1039 algorithm was designed to track plots of interest and conduct trait analyses to quantify  
1040 the canopy region and performance-related traits. (4) *Step 5*, traits such as dynamic  
1041 crop growth patterns in relation to thermal time (degree-day), vegetative greenness (0-  
1042 255), and the main plot orientation (0°-180°) were quantified and illustrated.  
1043  
1044



## Tools and Resources

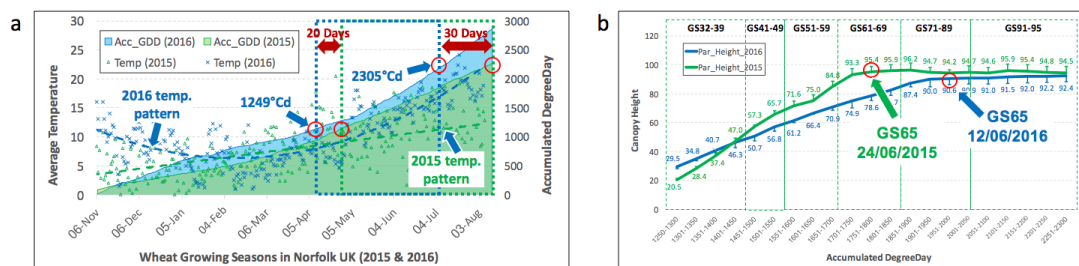
1045 **Figure 4**  
1046



1047  
1048 **Figure 4. The performance of five wheat NILs monitored by the CQ platform to**  
1049 **identify dynamic developmental profiles.**  
1050 Five wheat NILs (Late-DTEM, Early-DTEM; Short, Stay-Green, and *Paragon* WT)  
1051 and their dynamic performance in relation to environmental factors such as solar  
1052 radiation, rainfall, and temperature, during the 95-day monitoring period. Six growth  
1053 stages of *Paragon* WT were used as reference. Accumulated thermal time in degree-  
1054 day units was computed for comparison.  
1055

## Tools and Resources

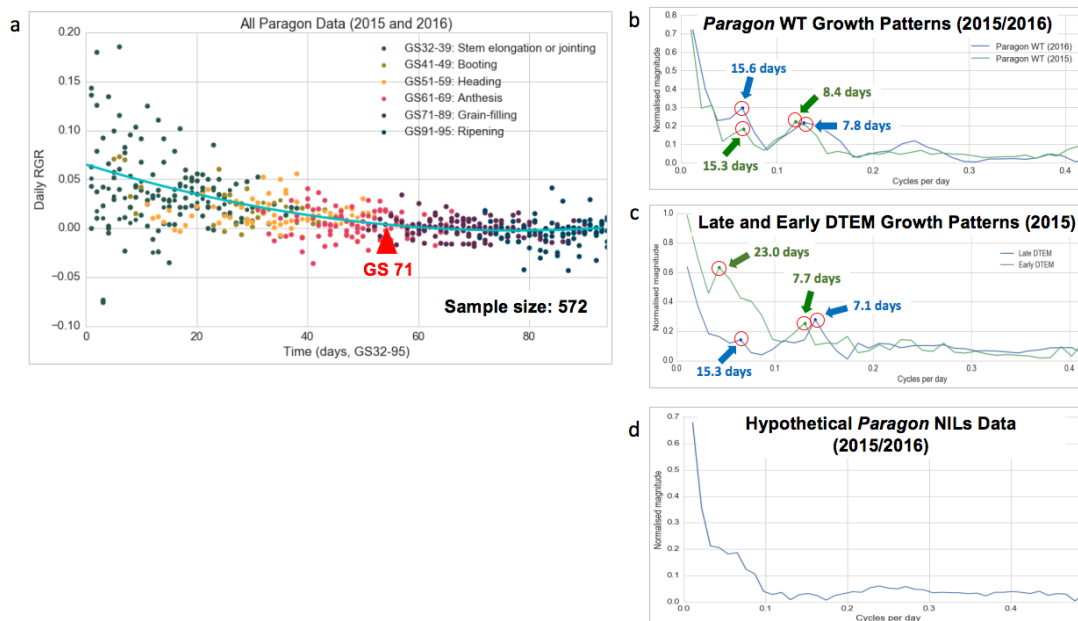
1056 **Figure 5**  
1057



1058  
1059 **Figure 5. Recognising subtle and dynamic developmental variances for wheat**  
1060 **genotypes under different climate patterns using the CQ platform.**  
1061 (a) Different temperatures and accumulated degree-day patterns recorded in 2015 and  
1062 2016. (b) Aligning and comparing growth curves of *Paragon* WT in 2015 and 2016  
1063 within similar growth stages and the degree-day period (1250-2300 °Cd).  
1064

## Tools and Resources

1065 **Figure 6**  
1066

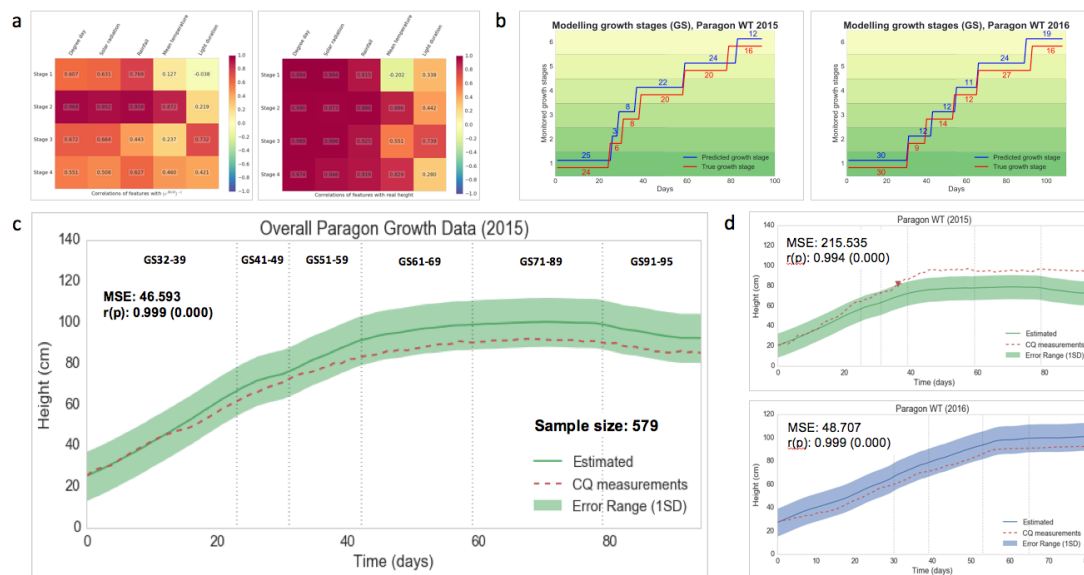


1067  
1068 **Figure 6. Extracting underlying growth patterns from continuous phenotypic**  
1069 **data using a fast Fourier transform (FFT).**

1070 (a) RGR (growth % of the previous day) is used to present the daily growth rate of all  
1071 the NILs at different growth stages. (b) The underlying growth patterns for *Paragon*  
1072 WT in 2015 and 2016 after the FFT conversion. (c) The underlying growth patterns  
1073 for Late-DTEM and Early-DTEM, after the FFT conversion. (d) Using a hypothetical  
1074 *Paragon* NILs growth data (merging all the NILs) to study the growth pattern.  
1075

## Tools and Resources

1076 **Figure 7**  
1077



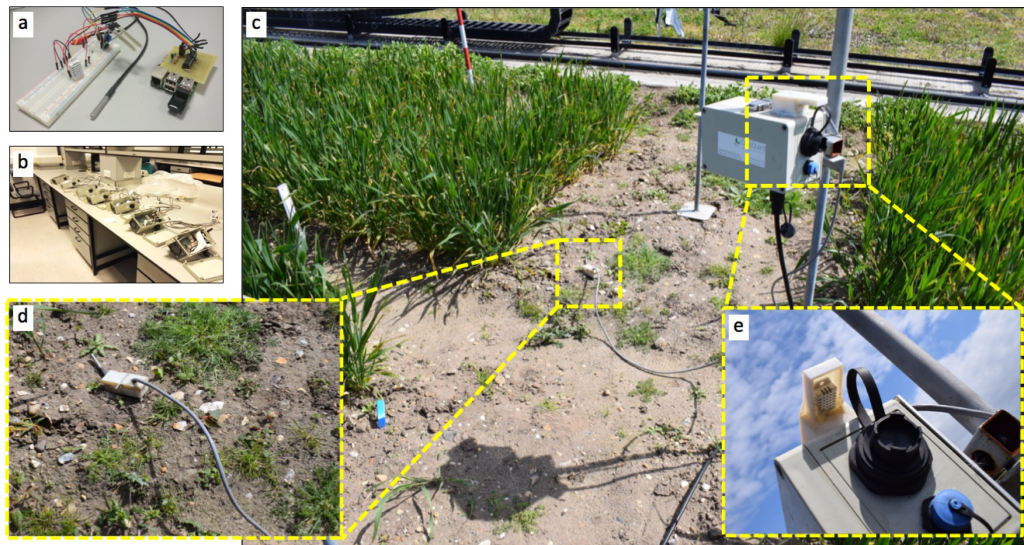
1078  
1079 **Figure 7. Establishing a GxE predictive model to forecast the performance of**  
1080 **wheat genotypes under different climate patterns.**  
1081 (a) Applying a correlation model to identify highly correlated environmental factors,  
1082 including thermal time, solar radiation, rainfall and growth stage duration using  
1083 *Pearson* correlation ( $p < 0.01$ ). (b) A growth stage-based predictive model applied to  
1084 estimate the key growth stages of *Paragon* WT in 2015 and 2016 compared with  
1085 manual scoring. (c) A global growth model comparing real growth curve measured by  
1086 CQ (red dotted line) with estimated growth curve (green line). (d) Warning messages  
1087 triggering mechanism to alert users if crop growth is outside the safe bounds ( $\pm 1SD$ )  
1088 of the estimated growth region.  
1089



## Tools and Resources

### 1090 *Supplementary figure 1*

1091



1092

1093 **Supplementary figure 1. An all-in-one CQ device deployed in field experiments.**

1094 (a) Low-cost remote sensors (e.g. light levels, ambient temperature and humidity, soil

1095 temperature and moisture) integrated by a tailored circuit board and then connected to

1096 a *Raspberry Pi* computer via GPIO (general purpose input/output) pins. (b) A number

1097 of all-in-one CQs being tested for establishing a mesh network in the Zhou laboratory.

1098 (c) An all-in-one CQ device deployed in a wheat field experiment in 2017. (d) A soil-

1099 based sensor installed to collect soil temperature and moisture for a six-metre wheat

1100 plot. (e) Light levels and ambient temperature and humidity sensors mounted on the

1101 top of the all-in-one CQ, together with an Ethernet coupler (black) and a micro-USB

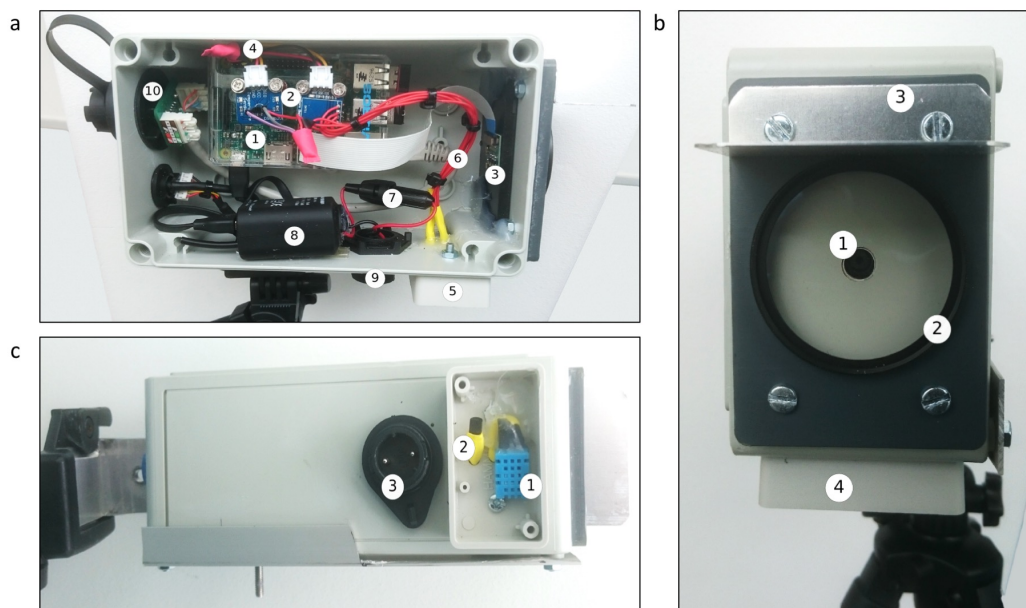
1102 coupler (blue).

1103

## Tools and Resources

### 1104 *Supplementary figure 2*

1105



1106

1107

### 1108 **Supplementary figure 2. The hardware design of a version 2 CropQuant.**

1109 (a) *Side view*: (1) a *Raspberry Pi 2*, (2) digital sensor circuits, (3) a *Pi* camera module,  
1110 (4) a sensor GPIO connection, (5) an external sensor housing, (6) a digital sensor  
1111 connection, (7) an inline power fuse, (8) a voltage converter, (9) an external power  
1112 connection, (10) an external Ethernet connection coupler. (b) *Front view*: (1) a  
1113 camera module, (2) an external camera UV lens, (3) a camera sunlight shield, (4) an  
1114 external sensor housing. (c) *Base view*: (1) a digital humidity sensor, (2) a digital  
1115 temperature sensor, and (3) an external power connection.

1116

## Tools and Resources

### 1117 *Supplementary figure 3*

1118



1119

1120

### **Supplementary figure 3. Real-time infield crop monitoring via portable devices.**

1121 (a) A crop scientist using an Android tablet to connect to infield CQ terminals to

1122 examine the performance of wheat growth. (b&c) After connecting to the VNC server

1123 running on a CQ terminal, different experiments can be inspected and managed by

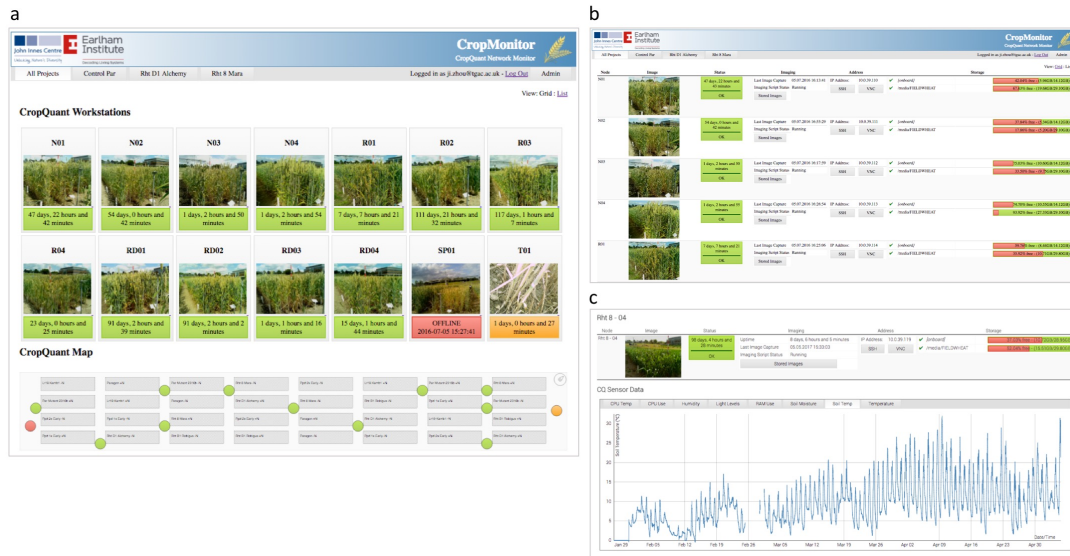
1124 crop scientists via a VNC viewer on the portable device, e.g. smartphones or tablets.

1125



## Tools and Resources

1126 *Supplementary figure 4*  
1127



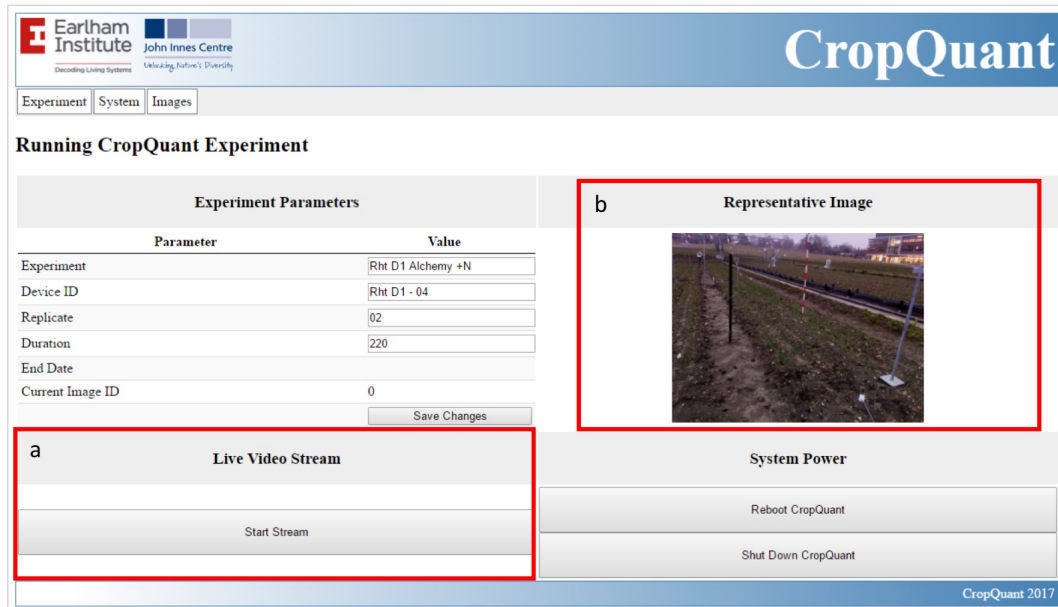
1128  
1129 **Supplementary figure 4. The CropMonitor control system.**  
1130 **(a)** Grid view of the CropMonitor system, which provides a regular overview of all  
1131 CQ terminals in the field, including online (green) or offline (red) status, operational  
1132 modes (amber means the imaging is either finished or halted), and the duration of  
1133 crop monitoring. An experimental layout of monitored plots is also provided, showing  
1134 the location of all CQ terminals and their operational modes. **(b)** List view of the  
1135 system showing CQ's online duration, network addresses, computing storage, and  
1136 SSH/VNC tools to access CQ terminals directly from the CropMonitor control system.  
1137 **(c)** Individual view of the system which illustrates an individual CQ workstation,  
1138 containing climate sensor data and systems information during the monitoring period.  
1139



## Tools and Resources

1140 *Supplementary figure 5*

1141



1142

1143 **Supplementary figure 5. A real-time stream function activated for deploying CQ**

1144 **workstations in the field.**

1145 **(a-b)** A live stream function showing the location of a CQ terminal in the field as well

1146 as assisting device deployment and systems calibration via the CropMonitor system

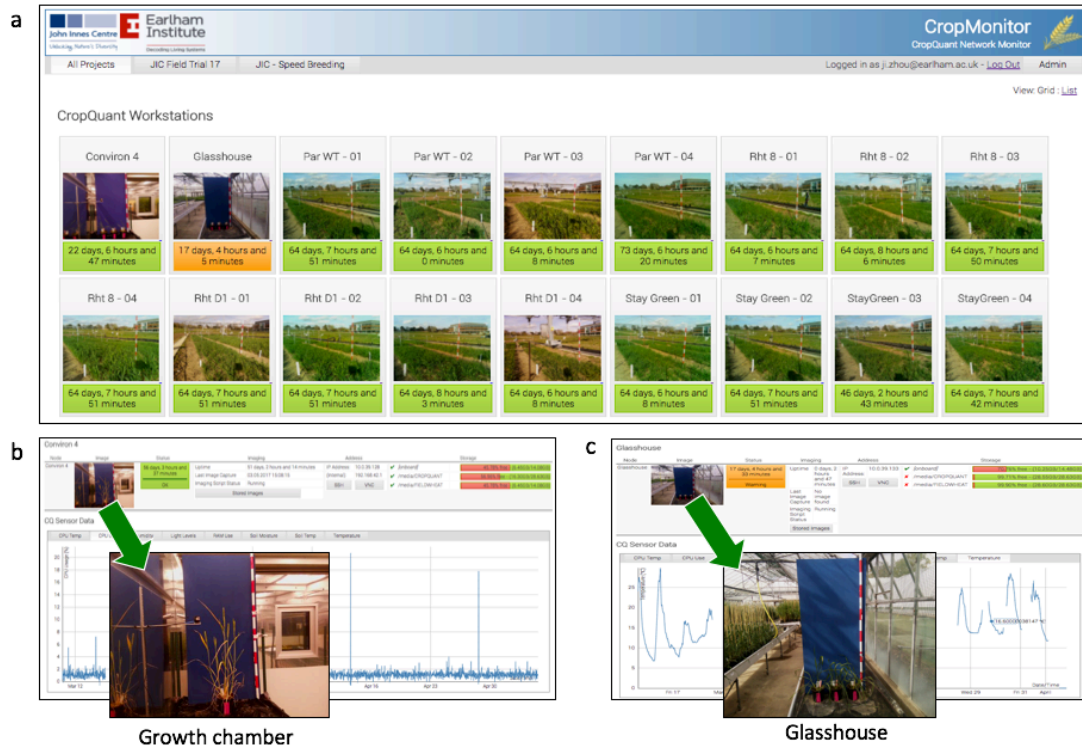
1147 running on each CQ device.

1148

## Tools and Resources

1149 *Supplementary figure 6*

1150



1151

1152

1153

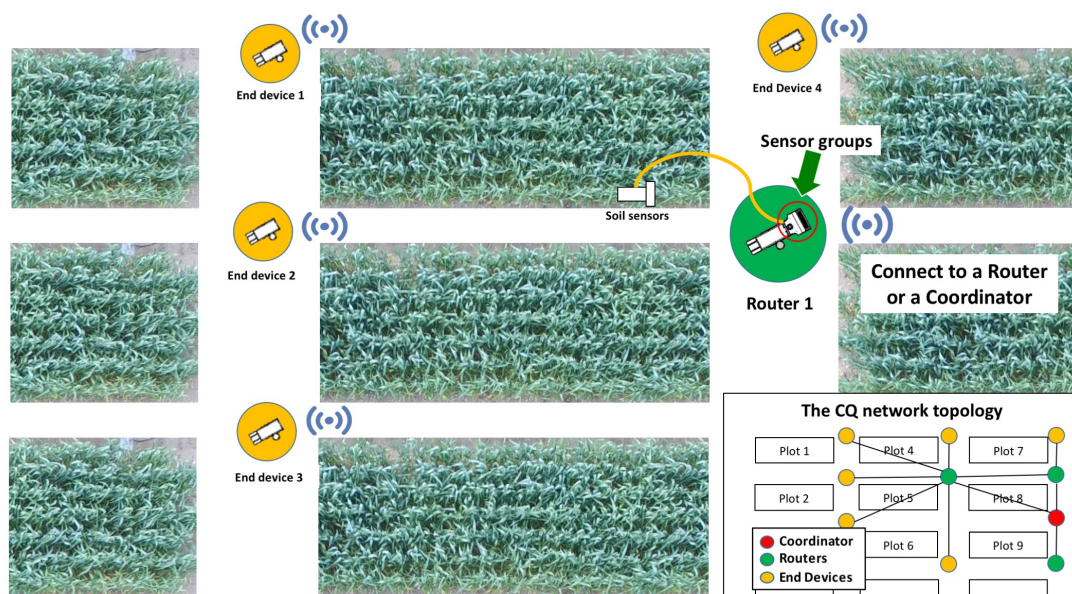
1154

**Supplementary figure 6. Outdoor and indoor wheat experiments (e.g. wheat speed breeding) monitored by the CQ platform in 2017.**

## Tools and Resources

### 1155 *Supplementary figure 7*

1156



1157

1158

1159

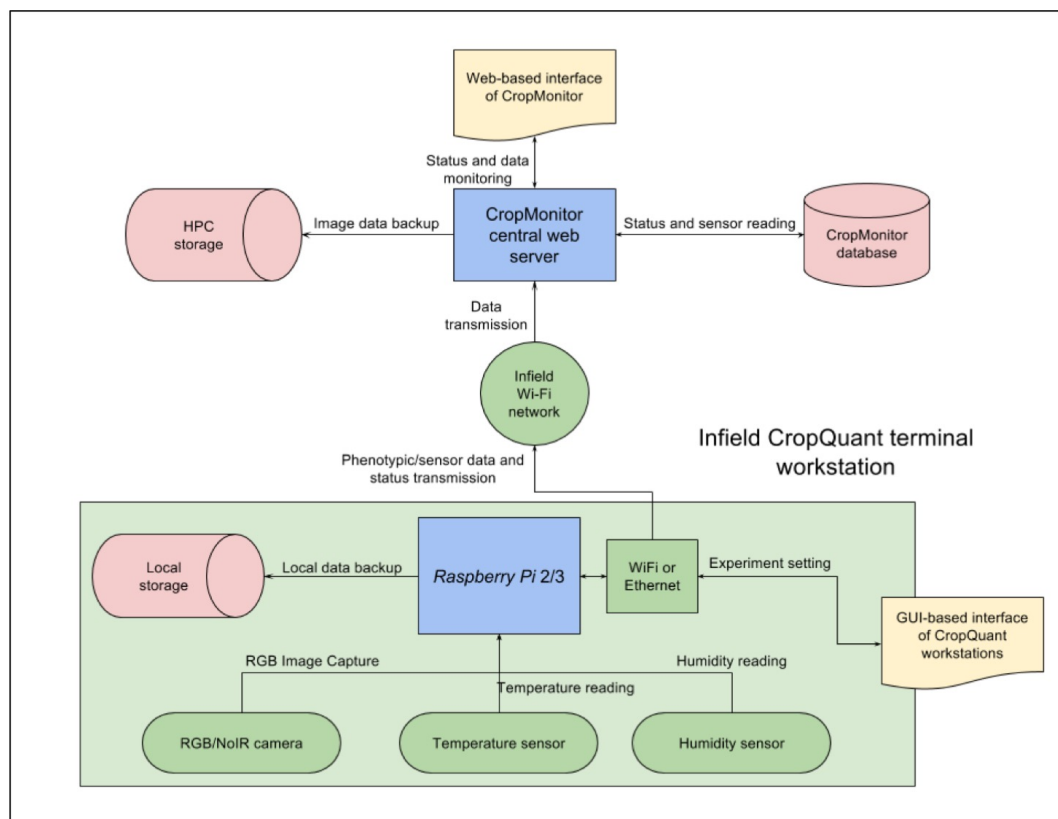
1160

**Supplementary figure 7. The network topology of the mesh network system established by many all-in-one CQs jointly operating in the field.**

## Tools and Resources

1161 *Supplementary figure 8*

1162



1163

1164

1165

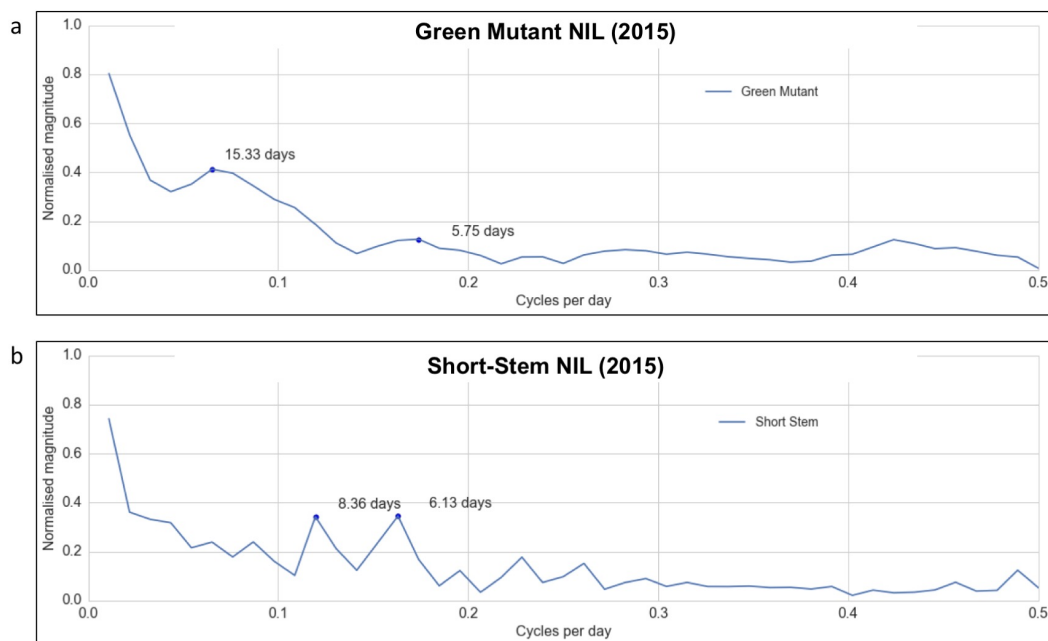
1166

**Supplementary figure 8. The systems architecture diagram of how to network CropQuant terminals, the CropMonitor system, and a HPC cluster.**



1167 *Supplementary figure 9*

1168



1169

1170

1171

1172

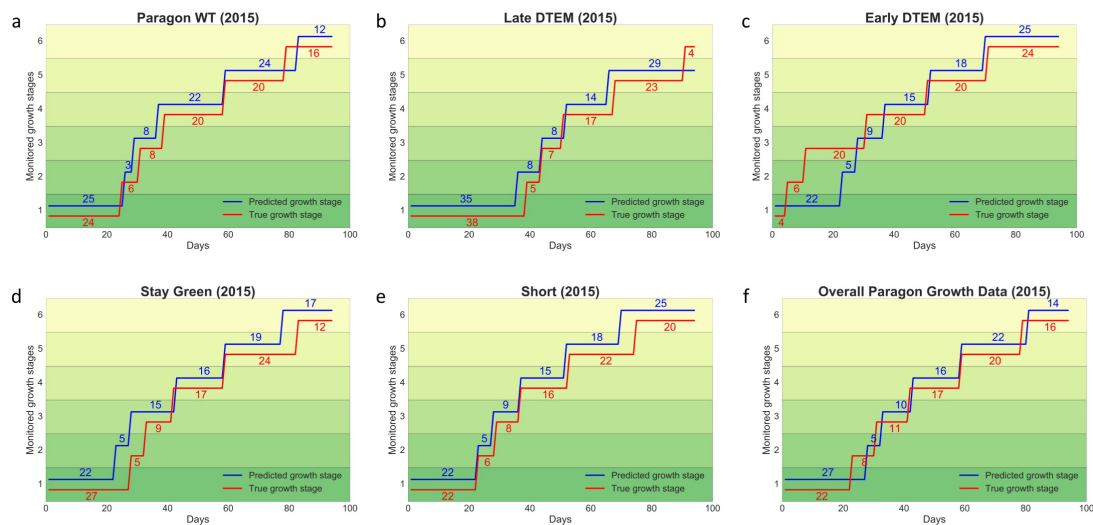
1173

**Supplementary figure 9. The underlying growth patterns for Green Mutant NIL and Short-Stem NIL extracted through a fast Fourier transform (FFT) approach.**

## Tools and Resources

### 1174 *Supplementary figure 10*

1175



1176

1177

### Supplementary figure 10. A predictive model to forecast the timing and duration of key growth stages for genotypes with *Paragon* background.

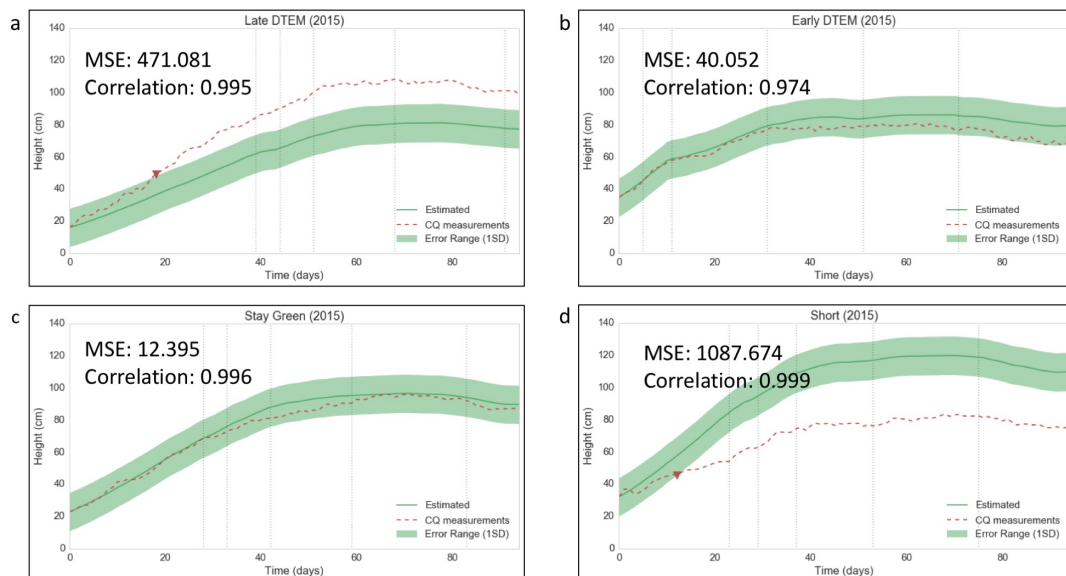
1178

1179 (a-e) The model predicts the growth stages based on all *Paragon* growth data  
1180 (acquired in 2015 and 2016) and environmental factors selected by the GxE  
1181 correlation model. (f) The model forecasts the growth stages based on overall

1182 *Paragon* data (the hypothetical growth dataset) and environmental factors.  
1183

## Tools and Resources

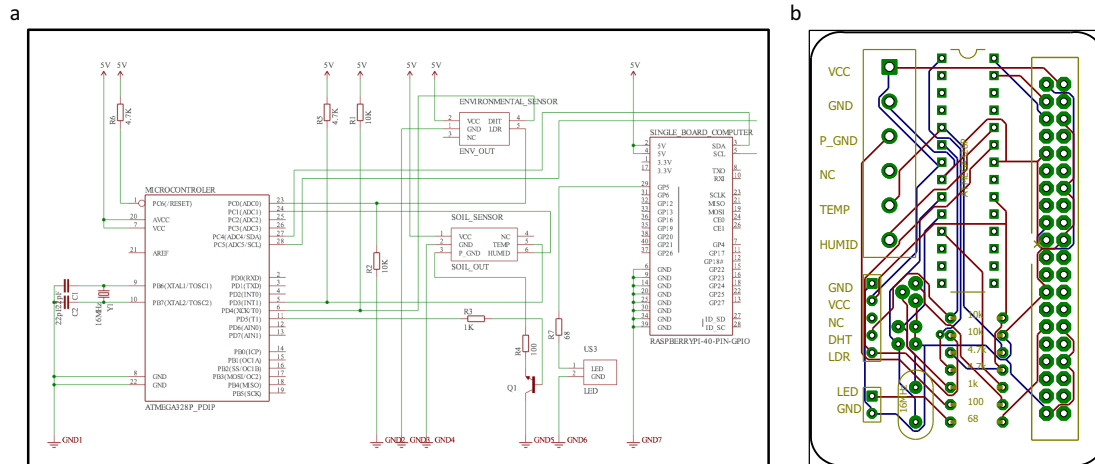
### 1184 *Supplementary figure 11* 1185



1186  
1187 **Supplementary figure 11. A predictive model to forecast the dynamic growth**  
1188 **patterns of wheat genotypes with *Paragon* background in field conditions.**  
1189 (a-b) The growth prediction based on Late-DTEM NIL and Early-DTEM NIL growth  
1190 data. Vertical dash lines indicate the manually segmented growth stages, from Stem  
1191 elongation or jointing (GS32-39) to Ripening (GS91-95). The red dotted lines stand  
1192 for real CQ measurements. If outside the safe bounds of the growth estimates, a  
1193 warning message (a triangle coloured red) will be triggered on the CropMonitor  
1194 control system. For example, the Late-DTEM line was growing too slow. (c) The  
1195 growth prediction based on Stay-Green NIL growth data. (d) The height prediction  
1196 based Short-Stem NIL growth data, which was growing too fast.  
1197

## Tools and Resources

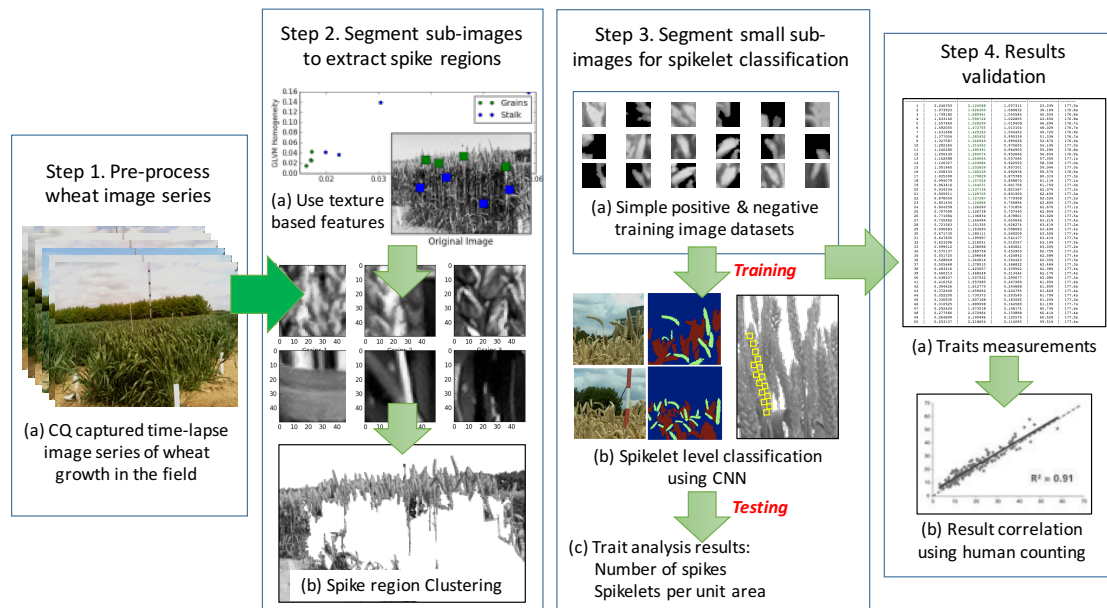
1198 *Supplementary figure 12*  
1199



1200  
1201  
1202  
1203

**Supplementary figure 12. The circuit for the all-in-one CQ workstation.**

1204 **Supplementary figure 13**  
1205



1206 **Supplementary figure 13. Using deep-learning neural network architectures to**  
1207 **train algorithms to quantify yield component traits based on images acquired by**  
1208 **the CQ platform.**

1210 (1) *Step 1*, pre-processing and calibrate image series captured by the CQ platform in  
1211 the field. (2) *Step 2*, based on texture and pattern classification methods such as grey  
1212 level co-occurrence matrices (GLCM) and texture entropy, detecting spike-featured  
1213 regions, which are divided into sub-images as training data for a CNN (Convolutional  
1214 neural network) classifier. (3) *Step 3*, to count spikes per unit area and spike numbers  
1215 on a given image, further smaller sub-images are produced to fit a whole spikelet  
1216 region. CNN is trained to count spikelet-only sub-images. (4) *Step 4*, the machine-  
1217 learning based estimation are correlated with manual measurements, so that the  
1218 estimation model can be improved.

1219