

METHODOLOGY

Open Access



Earbox, an open tool for high-throughput measurement of the spatial organization of maize ears and inference of novel traits

V. Oury¹, T. Leroux¹, O. Turc², R. Chapuis³, C. Palaffre⁴, F. Tardieu², S. Alvarez Prado^{5,6}, C. Welcker² and S. Lacube^{1*} 

Abstract

Background: Characterizing plant genetic resources and their response to the environment through accurate measurement of relevant traits is crucial to genetics and breeding. Spatial organization of the maize ear provides insights into the response of grain yield to environmental conditions. Current automated methods for phenotyping the maize ear do not capture these spatial features.

Results: We developed EARBOX, a low-cost, open-source system for automated phenotyping of maize ears. EARBOX integrates open-source technologies for both software and hardware that facilitate its deployment and improvement for specific research questions. The imaging platform consists of a customized box in which ears are repeatedly imaged as they rotate via motorized rollers. With deep learning based on convolutional neural networks, the image analysis algorithm uses a two-step procedure: ear-specific grain masks are first created and subsequently used to extract a range of trait data per ear, including ear shape and dimensions, the number of grains and their spatial organisation, and the distribution of grain dimensions along the ear. The reliability of each trait was validated against ground-truth data from manual measurements. Moreover, EARBOX derives novel traits, inaccessible through conventional methods, especially the distribution of grain dimensions along grain cohorts, relevant for ear morphogenesis, and the distribution of abortion frequency along the ear, relevant for plant response to stress, especially soil water deficit.

Conclusions: The proposed system provides robust and accurate measurements of maize ear traits including spatial features. Future developments include grain type and colour categorisation. This method opens avenues for high-throughput genetic or functional studies in the context of plant adaptation to a changing environment.

Keywords: Zea mays, Maize ear imaging, CNN-based deep learning, Environmental response, Grain set, Grain abortion, Maize ear spatial organization

Introduction

Characterizing genetic resources and their response to the environment through accurate measurement of relevant traits is crucial to dissect the genetic bases of crop

yield [1], and to tailor genotypes adapted to specific climatic scenarios [2]. In maize, yield results from the number of grains and individual grain size, each of which has higher heritability than overall yield [3, 4], present different genetic architectures [5, 6] and result from environmental conditions during different phases of the crop cycle, namely the vegetative and flowering period for grain number and the post-flowering period for individual grain weight [7]. Sensitivities of grain number to soil

*Correspondence: sebastien.lacube@phymea-systems.com

¹ Phymea Systems, 453 Rue de l'Espinouse, Montpellier, France
Full list of author information is available at the end of the article



© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

water deficit, temperature and light are key parameters in the prevision of grain yield in a wide range of environments [8], thus requiring accurate phenotyping.

Examining the structure of the maize ear provides additional insights into deciphering the response of grain yield to environmental conditions. Indeed, the ear is composed of concentric rings of grains (cohorts) initiated simultaneously within each cohort but sequentially between cohorts [9] (Fig. 1A, B). While the number of grains per cohort is a genetic trait largely independent of environmental conditions, the number of cohorts results from the response to climatic scenarios, with genotype-specific responses. Prior to flowering, suboptimal conditions reduce the number of grains via a reduction in the number of cohorts due to a reduced number of initiated ovaries (Moser et al., 61). Abiotic stresses occurring at flowering result in localized ovary and grain abortion involving cohorts with delayed development [10] located preferentially at the ear apex, with aborted zone increasing with stress intensity (Fig. 1C–E). Stress affecting pollination (pollen availability or viability) results in a wide variety of phenotypes characterised by incomplete cohorts and erratic cohort numbers (Fig. 1F). Stress occurring

beyond two weeks after flowering reduces grain size [11]. Therefore, a fine characterization of the spatial distribution along and around the ear of grain set/abortion and of grain and ear dimensions appears to be a relevant tool to reveal the response of genotypes to environmental scenarios.

Ear phenotyping is still largely manual, time-consuming, costly, and subjective [1]. Several methods have been developed to extract ear and grain characteristics from images [12–14]. They are usually based on manual or non-standardized acquisition involving either isolated grains after shelling [1, 13–16] or one side of the ear [12, 15, 17–20]. Thus, the spatial distributions of grain presence/absence (grain set vs grain abortion) and grain traits along and around the ear is usually not, or only partly, considered.

Several techniques have been used for ear imaging, each providing different advantages and drawbacks. (i) Vertical positioning of the ear on a rotating axis allows imaging different sides of the ear at specific rotation angles [21, 21]. This method is efficient but only considers one ear at a time and requires time-consuming handling for ear positioning before imaging (1–2 min per ear). (ii) Portable imaging systems have been developed, directly threaded around the ear in intact field plants, imaging simultaneously all ear sides, allowing 3D reconstructions of the ear [23]. This technique is affordable and avoids the need to harvest the ears but involves limited throughput because of long ear handling time (husks removing, one ear at a time), while being subjected to various difficulties related to field conditions. Moreover, most of these techniques have been validated with ears from standard commercial hybrids with classical properties (ear and grain shape and colour, regular spatial organization), and therefore fail to provide reliable results for ears with non-regular patterns, a frequent characteristic under non-optimal environmental conditions (heterogeneity of abortion/set zones and grain dimensions, pest, and disease damage).

The aim of this study was to develop a low-cost and open-source system capable of producing automated, standardized, robust and reliable measurements of phenotypic traits of the ear, including the spatial distribution of grain traits along and around the ear. We tested it for genotypes with contrasting ear and grain shape and grain texture (e.g., Dent, flint, pop, waxy, flour). The method of image acquisition consisted of a custom box in which ears are placed horizontally. Motorized rollers are used to rotate the ears. In addition to being easy to setup and implement, this method is easily scalable for multiple ears at once by multiplying the number of rollers and cameras (Fig. 2).

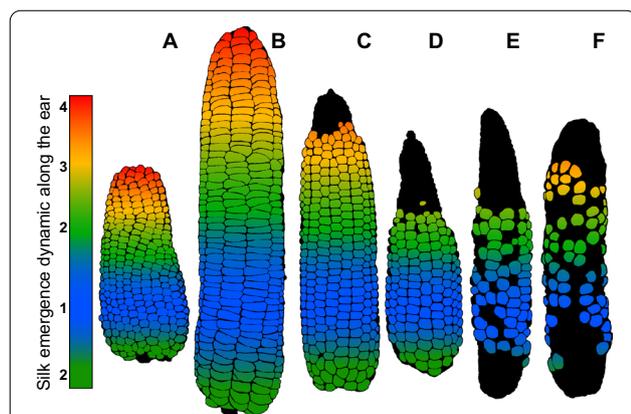
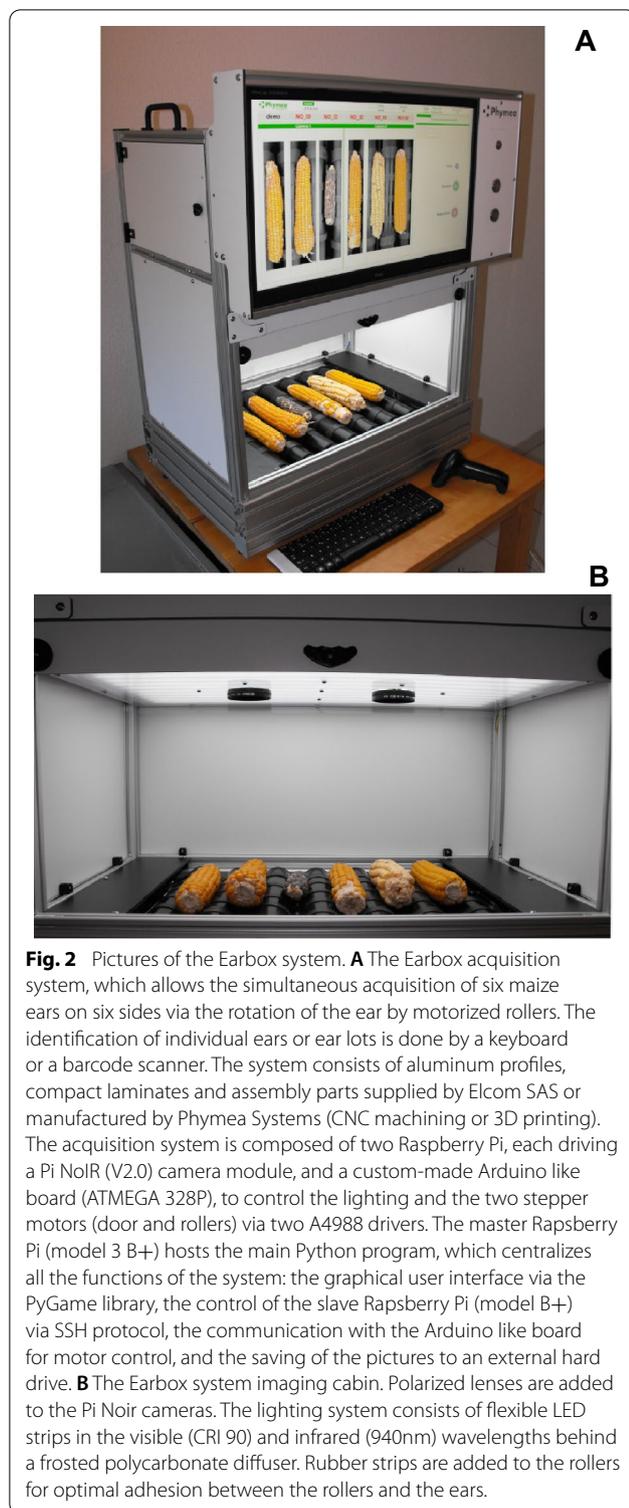


Fig. 1 Spatial organization of grains reflecting the morphogenesis of the ear. The grains are arranged in rings and rows. Each ring corresponds to a cohort of organs with synchronous development, while a developmental gradient exists between cohorts depending on their vertical position along the rows. Floret cohorts are initiated sequentially at the ear apex. The oldest cohorts are located at basal positions and the youngest at apical positions. **A, B** Under optimal conditions, pollination and fertilization follow the order of silk emergence which is illustrated by colors: zone 1 cohorts (blue) are fertilized first, followed by zone 2 (green), zone 3 (yellow) and zone 4 (red). **(C–E)** Abiotic stresses occurring at flowering induce abortion that preferentially affects the youngest apical cohorts in zone 4, followed by the basal cohorts. **E, F** Severe constraints affecting pollination (pollen availability or viability) result in a wide variety of phenotypes characterized by incomplete cohorts and erratic cohort numbers.



We believe that this method will allow measurement of relevant traits in the context of plant adaptation to a changing environment and the enrichment of crop gene bank knowledge base [24].

Methods

A wide phenotypic diversity to test the robustness of the method

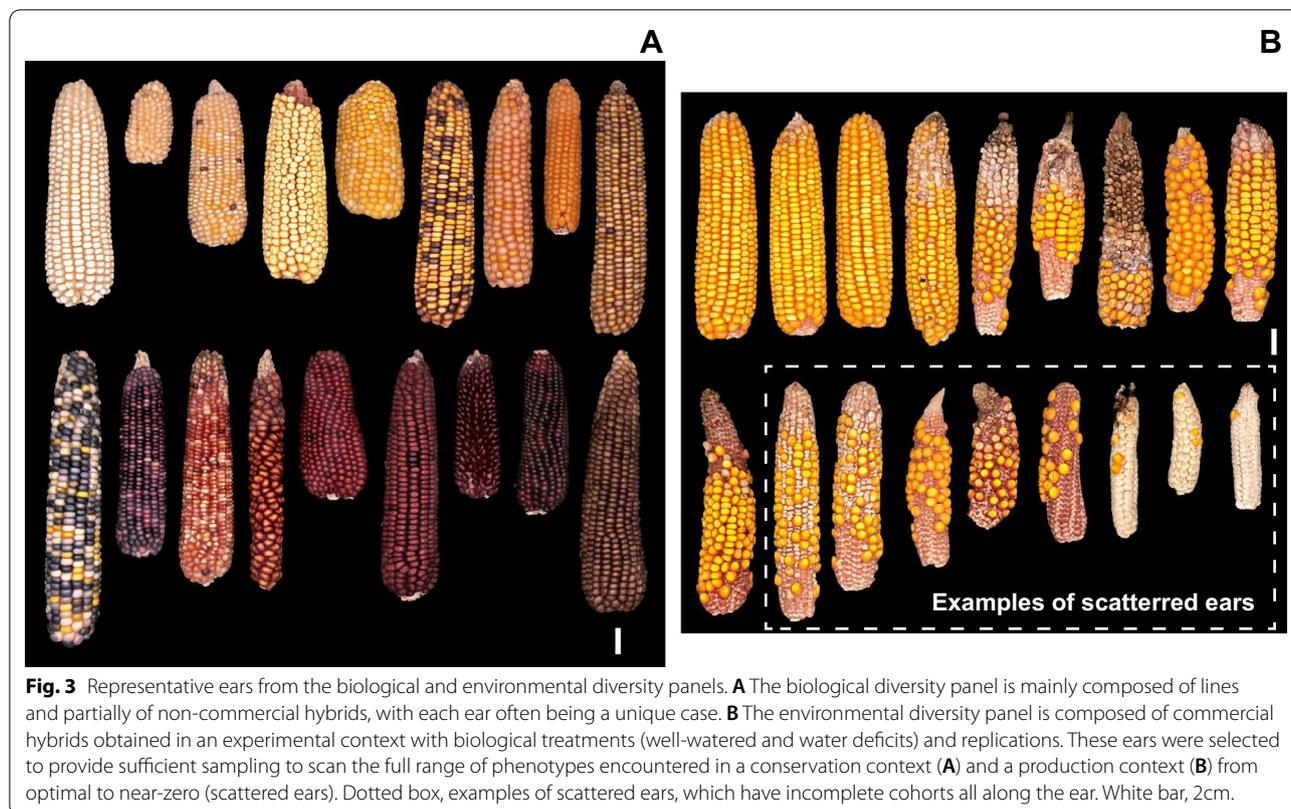
The set of ears used in this study was composed of 796 ears selected from two panels, a ‘biological diversity’ panel chosen to represent the diversity of phenotypes encountered in production contexts (Fig. 3A), and an ‘environmental diversity’ panel, chosen to represent the phenotypes encountered in response to abiotic constraints (soil water deficit) (Fig. 3B).

- The **Biological Diversity panel** represented 16% of the whole set, i.e., 126 ears. Selected ears were of various shapes (length between 3 and 24 cm and diameter between 2 and 5.5 cm). Grain colours were of all existing hues: white, yellow, orange, red, wine, pink, purple, blue, black, and brown, including heterogeneous ears with multiple colours and pearly, opaque, or translucent grains. Grain sizes ranged from 2 mm to 1 cm with variable shapes depending on their position along the ear, from perfectly round to dented or flint grains. Finally, the panel explored a diversity of grain spatial organization, with a range of number of cohorts and number of grains per cohort, and either regular or irregular grain organization along the ear.

- The **Environmental Diversity panel** represented 84% of the set, i.e., 670 ears. First, a set of 431 ears was sampled from a field experiment (INRAE UEDIASCOPE, France) under two water treatments: 321 under well-watered conditions (WW) and 110 under water deficit (WD). The remaining 234 ears were sampled in another experiment under WD treatment. For both experiments, water deficit conditions were triggered by stopping irrigation around 10-leaf stage while continuous irrigation was applied for the WW treatment. The combined variability in plant phenology and water treatments resulted in a wide range of ear phenotypes with various sizes and spatial distribution of fertile and aborted zones.

A simple and low-cost image acquisition system

The ears were imaged with an automaton developed and assembled by Phymea Systems (www.phymea-systems.com)—Montpellier, France). Individual ears are manually positioned in the system (Fig. 2), which acquires images stored in a generic hard drive. Images are uploaded to an independent analysis station where the associated software is installed for output retrieval. The automaton works in independent acquisition sessions to easily separate experiments, genotypes or varieties, and treatments. The ears or ear lots are individually identified by the



keyboard or by a barcode scanner. The analysis software was developed to be flexible (retrieval of one or more phenotypic traits depending on user's needs).

To minimize complexity and cost, the acquisition system was developed to be as simple and robust as possible. It consists of aluminium profiles, compact laminates and assembly parts supplied by Elcom SAS (Bourgoin-Jallieu, France) or manufactured by Phymea Systems (CNC machining or 3D printed) (Fig. 2A). The focus was on developing a flexible system, to be complexified in a second step: adapted for specific use cases, for example, to harvesting machines. Future development could integrate a mirror system to image the whole circumference in less time, combined with automatic ear handling and cleaning systems. Moreover, for an on-board system, software optimisation and recent evolutions of embarked-AI hardware could also be used to enhance the analysis system and propose a direct estimation of production.

The system was designed to take multiple images of the ear via simultaneous rotations of all ears with motorized rollers. The acquisition system used in this study was set with 7 rollers for rotation and imaging of 6 ears (Fig. 2B). Rubber bands were added to the top of the rollers to properly drive the ears without slipping. Because the rollers have fixed dimensions (5.2 cm diameter) and positioning (1 cm spacing), the theoretical ear rotation

angle was calculated from the ear diameter and roller rotation angle and measured in practice by measuring the rotation of ears placed manually on the rollers. The measured and calculated angles fit strongly for the 4 ears tested, representative of the diversity of diameters in the whole ear panel ($R^2 = 0.98$; Additional file 1: Fig. S1). We defined the number of images to be taken for each ear, thus the number of roller rotations, and a fixed roller rotation angle that ensured imaging of the whole ear circumference while minimizing acquisition time. The combination of 6 successive ear images with a roller rotation angle of 58° fulfilled these conditions for the range of 2–6 cm ear diameter (Additional file 1: Fig. S1) which exceeds the range encountered in both panels.

Developing a normalized method for analysing images regardless of ear or grain colour or shape required the use of near-infrared imaging. For this purpose, the system used Pi NoIR Camera v2 (Raspberry.org) driven by a Raspberry Pi to produce two types of images at two wavelengths: visible (RGB) and near infrared (IR, 940 nm). Two sets of cameras and Raspberry Pi were necessary to ensure high resolution images of 6 ears at 6 angles and two wavelengths, for a total of twelve images, in less than 30 s (between 26 and 32 s depending on the images), or 5 s per ear. A custom Arduino-like board was developed to control both the lightning and the two stepper motors

(doors and rollers). A master raspberry Pi was set to centralize all custom functions: host the main python program of the user interface developed with the PyGame python library, control the slave raspberry Pi via SSH protocol, control the Arduino board, and save the images to an external hard drive. The entire system was designed to be affordable and open source, and costs a total of 2500 € in equipment and hardware, excluding labour and development costs.

Ear peduncles at ear base were cut off prior to image acquisition, and husk-free ears were scrubbed and cleaned of silks and/or fungi with a brush so that all grains were accessible to the camera. The ear cleaning procedure appears to be the most variable and time-consuming, up to 10 s or more for an ear covered with fungus. Nevertheless, it is easy to reach a total acquisition rate of 120–220 ears per hour for a team of 2 people, depending on the storage methods, the identification method, and the appearance of the ears. In this study, a

total set of 9492 images (791 ears) were taken from both panels, by one person with a rate of 128 ears per hour.

A combination of empirical segmentation and deep learning to build a robust routine workflow for ear and grain segmentation

RGB and IR images acquired from both panels of ears were first pre-processed (Fig. 4A—Ground for deep learning) using conventional image analysis tools (dilate, open, close, gaussian blur and watershed) and merged to normalize the data for all ear and grain colours (Fig. 4, Step 1), resulting in a pre-processed image set (4746 images: 6 images per ear for 791 ears, hereafter referred to as the ‘dataset’). The dataset images were then empirically segmented (Fig. 4, Step 2) and used to train a Deep Learning Neural Network (Fig. 4, step 3). Finally, ear and grain phenotypic variables were retrieved for both ear panels (Fig. 4B—Routine workflow). The ear masks were retrieved from the RGB images to estimate the ear

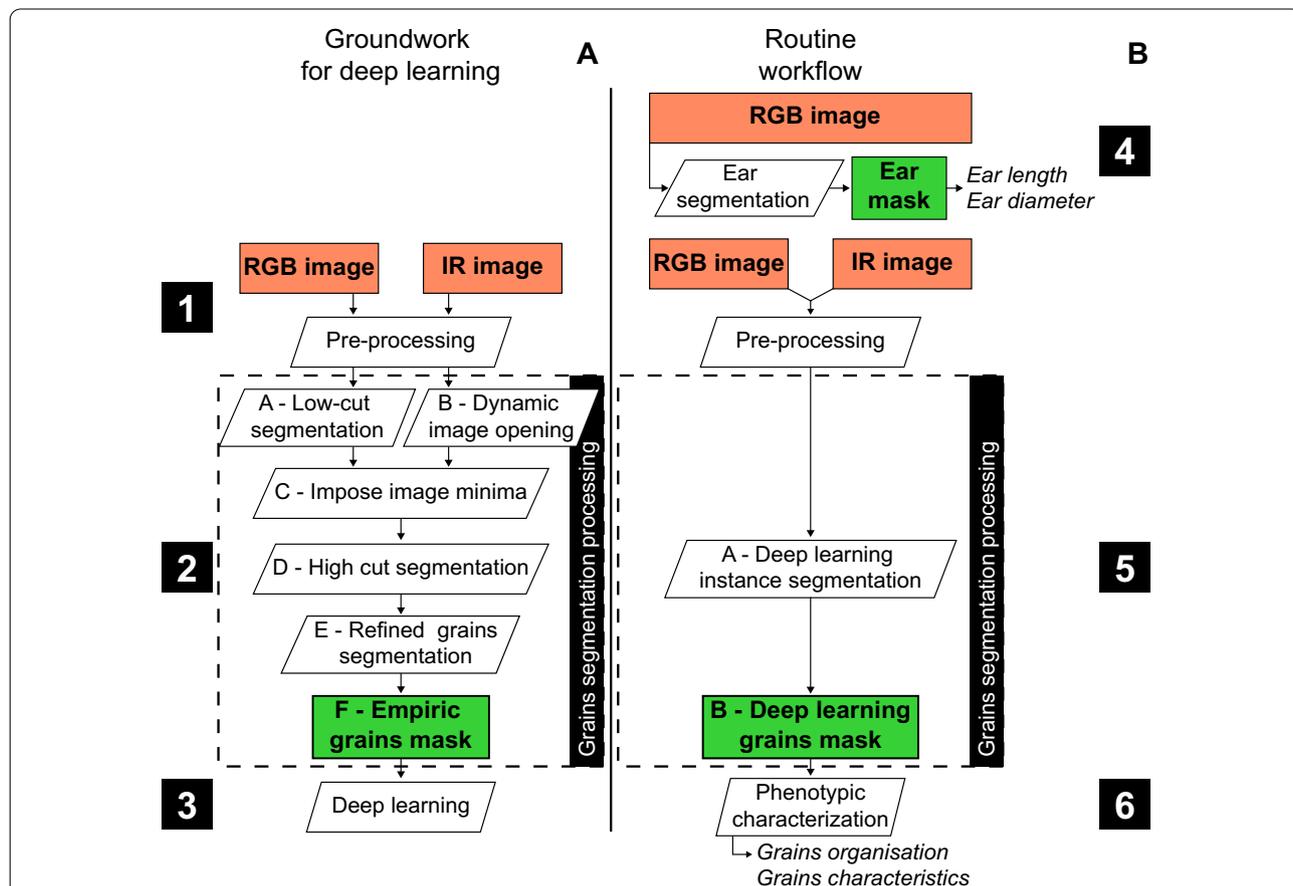


Fig. 4 Workflows for training the neural network and generating phenotypic data. **A** Image processing workflow used to train the learning procedure (Groundwork for deep learning). **B** Image processing workflow used to produce phenotypic data (Routine workflow). Orange box, image acquisition. White box, image, or data processing. Green box processed images or data. Numbers in black boxes, steps of development, from first image acquisition to phenotypic data.

phenotypic data (Fig. 4, Step 4). All images were then processed with the fitted neural network (Fig. 4, step 5) and used to estimate grain phenotypic traits and characterize grain organization on the ear (Fig. 4, Step 6). Data are extracted for each individual ear side pictures (6 sides per ear) and averaged to represent the whole ear.

The whole data extraction pipeline (from treatment of acquired ear images with the EARBOX acquisition system to phenotypic variables extraction) was coded in MATLAB and Python. The code for the analysis using a Graphical User Interface in MATLAB is fully available on a public repository (<https://github.com/Phymea-Systems/Earbox>). It is used in combination with a Python code applying the trained neural network to extract the DL2 images, also available in the same public repository. A benchmark test to assess image analysis capabilities was performed with parallel computing (using MATLAB parallel loop functions, one image per CPU Core) for the whole pipeline for all 791 ears from both panels with medium capacities laptops:

The first step, ‘ear masking’: validating images presence, counting ears on each image, comparing to scanned user-scanned codes, resizing images, and extracting the ear from initial EARBOX images (from 3 ear per image to 1 ear per image on a black background) took approximately 21 secs per ear (Performed on CPU with parallel computing: 11th Gen Intel(R) Core(TM) i7-11800H @ 2.30 GHz—16 Gb Ram).

The second step, ‘Deep Learning’ (DL2): extracting the bounding boxes on grains, correcting for overlapping bounding boxes and resizing took approximately 54 s per ear (Performed on CPU and GPU: AMD Ryzen 7 3700x—NVIDIA GeForce RTX 1080 Desktop GPU, 8 Gb VRAM—32 Gb RAM).

Finally, the extraction of all phenotypic variables, their storage in an organized file system and the computing of a final visualization for user-verification took approximately 62 secs per ear, with a variability depending on the number of grains on the ear and their organisation (Performed on CPU with parallel computing, 11th Gen Intel Core i7-11800H @ 2.30 GHz—16 Gb RAM).

The benchmark showed that, using medium-range CPU and GPU capabilities, the pipeline can check, mask, extract data from raw images in approximately 137 secs per ear. It must be noted that, for all three steps, the approximated computing time represent the calculation and analysis performed 6 times per ear (for the 6 captured sides). The computational capabilities can be greatly increased by using state of the art tools for analysis but increasing the costs quite rapidly. The analysis, as developed, requires relatively little supervision by the user and can thus run overnight or outside of working hours, and is therefore not determinant in the cost

analysis, allowing for a more efficient use of resources. If accounting for data handling and transfer, as a reference one can therefore estimate an approximate time of analysis of 38 h for a set of 1000 ears with medium capacity hardware, highly scalable.

A ground-truth dataset was built prior to any analysis by manually segmenting grains from a set of images from 79 ears (10% of the dataset; not used to train the Neural Network). These ears were randomly selected from each category of grain colour and shape based on its frequency in the entire dataset. Selected ear images were treated manually using adobe photoshop software and a graphic tablet to mark grains by surrounding and filling them on a mask (on top of the image) and converting this mask into a binary image (black and white), similar to the outputs of the processing algorithms used.

A preliminary step of empirical segmentation was performed and used as an automatic annotation to drive deep learning iterations. The architecture of the empirical segmentation was developed to detect grains for a large portion of the dataset, so that all features can be learned and improved in Deep Learning run. RGB and IR images were processed with an algorithm developed by Phymea-Systems (Fig. 4, step 2) using only trial and error (using morphological image processing toolkits) to produce a grain mask that was precise enough to characterise the grains from images with various colours of grains and cob. The images were pre-processed with conventional image analysis tools to enhance their quality and then merged to retrieve complementary contrast and shapes. The cross-checking of the two pieces of information allowed a precise selection of the grains to produce an initial grain segmentation, which was in turn corrected by image analysis to refine the grain shapes and recover the over-segmented grains.

After this step, the processed images were sorted to assess the quality of the output. The output masks were scored by two independent individuals to evaluate the quality of segmentation with a score from 0 (bad segmentation) to 3 (good segmentation). Ears with uniform grain colour, strong colour contrast between cob and grain colour and non-scattered grains were mostly correctly segmented, with only minor problems. Most low scores were encountered for scattered grains (overly segmented), and for ears with similar grain and cob colour, which made it more difficult for the algorithm to distinguish.

A deep learning neural network was trained to segment the grains on ear masks. It is a powerful tool for high-throughput plant phenotyping, yielding valuable results when large datasets are available [25]. More specifically, the Mask-RCNN neural network is commonly used as a framework for instance segmentation [26]. It is a highly flexible, trainable framework that has been

widely validated in many scientific domains, including plant science [12, 27–32]. More specific to maize, Kienbaum et al. [19] have trained mask R-CNN for cob segmentation on images and have highlighted the interest of this type of framework over classical image analysis techniques: its robustness and accuracy.

The entire deep learning (DL) framework was coded in Python 3 using TensorFlow for in-learning visualisation [33]. The baseline architecture of the deep learning model used in this paper for the kernel segmentation is the Mask-RCNN model with a ResNet50 backbone. Mask-RCNN is a two-stage model, the first stage called Region Proposal Network (RPN) generates object bounding box proposals, then a second stage extracts feature from each proposal, performs classification, does bounding box regression, and in parallel predicts object masks. The input layer of this model was adapted to receive 4 channel images (RGB + IR) and no pre-trained weights were used, it was trained from the ground up on the maize dataset. The RPN network was set to use 256 anchors per image with power-of-2 side length from 8 to 128 pixels, and a non-max suppression overlap threshold of 0.9. The smooth-L1 loss function is used for bounding box proposal, and cross-entropy loss is used for both classification and mask prediction. The training was done on a subset containing 2226 images, on random 512*512 pixels crops from original images without resizing, and cross-validated on random samples from a separate subset of the complete dataset. The model was trained for 99 epochs, with 75 cross-validation steps each, and a base learning rate of 0.001 with momentum optimization ($\gamma=0.9$). Detection was done on full-scale images with the non-max suppression overlap threshold of the RPN network set to 0.7.

A set of data augmentation techniques were applied prior to learning. Each image was cropped into a set of 512 by 512 pixels elements on which a random number (between 0 and 2) of augmentation techniques were applied before being introduced into the model. The various augmentation transformations were retrieved from the 'imgaug' package (<https://imgaug.readthedocs.io/en/latest/>): flip up down, flip left right, 90 or 180- or 270-degree rotation, pixel value multiplication and gaussian blur. Deep Learning iterations involved 33 learning epochs with a cross-validation using 75 random images unused in the training dataset.

The neural network training (Fig. 4, step 3) included the following steps (Additional file 2: Fig. S2). First, empirical masks with a mean score equal or greater than 2.5 (2076 images, 43.7% of the dataset, i.e., 346 ears) were used to train the neural network (Additional file 2: Fig. S2, step 3). The resulting DL1 masks outputs were corrected by 'minor' manual corrections (only 'click on grains' to add

or remove mis detected grains) for 1926 images (92,8% of DL1 images, i.e. 321 ears) and 'major' corrections (adding grains and reshaping grains for ears with a large number of missing or mis-segmented grains and/or wrong shapes) for the remaining 150 images (7,2% of DL1, i.e. 25 ears—Additional file 2: Fig. S2, step 4). Second, a set of 72 images (12 ears, i.e., 1.5% of the dataset) from ears incorrectly segmented in the initial empirical segmentation, were manually corrected in the same way as the 'major' corrections seen above. Corrected images from DL1 and initial empirical segmentation were used in a second Deep Learning iteration (DL2 Additional file 2: Fig. S2, step 5) with 2148 images, i.e., 358 ears (45.3% of the dataset).

The 'mean Average Precision' (mAP) was used to estimate the quality of the Deep Learning output [34] and calculated as defined by the latest evaluation's techniques of the COCO dataset [35]. The literature usually considers an algorithm to be highly efficient for mAP values of 0.4.

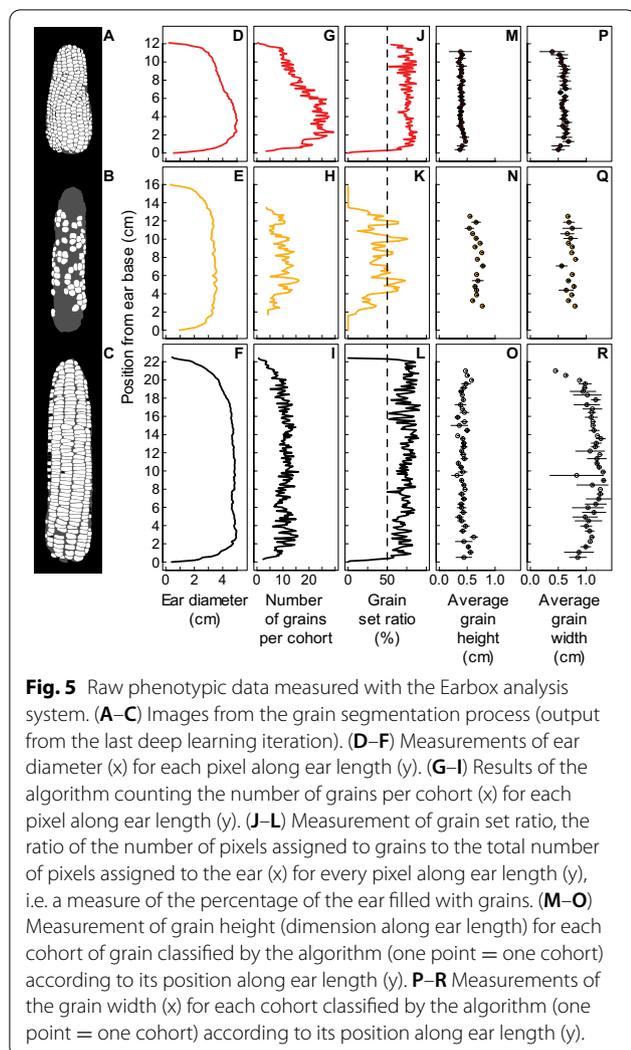
After several steps of learning, small input image correction, re-calibration of the neural network parameters, the resulting network with fitted weights (DL2) was used to extract segmented grains from all acquired images, i.e., the dataset (Fig. 4., step 5).

A routine workflow to access and validate phenotypic traits and their spatial distribution

Image analysis methods were applied on the segmented grains to extract phenotypic data for each trait of interest. To validate this methodology, the set of ears from both panels (791 ears) was also described, for each trait, by a unique observer to generate a set of manual measurements, to be compared to automatic measurements generated by the Earbox system (Additional file 3: Fig. S3). Most of the manual measurements were repeated 4 times around the ear circumference, averaged, and compared to the corresponding automatic measurements. The automatic measurements were repeated on each of the 6 images taken for each ear, and then averaged to produce phenotypic data at the ear scale.

The ear dimensions and form were automatically acquired with the Earbox from the segmented ear in each RGB image (Fig. 5A–C). All measurements were referenced to their spatial position according to the two axes of the image: the principal axis parallel to the ear length (vertical axis, starting from the bottom to the top of the ear) and the perpendicular axis (horizontal axis).

The ear mask was reduced to its centre pixel along the principal axis of the image (Additional file 4: Fig. S4A) to define the central axis of the ear. The *ear length* was calculated as the number of pixels of this central axis running from the bottom to the top of the ear. This method



corrected for the twisting effect of irregular ear shapes. *Ear diameter* was measured at each pixel along the principal axis as the distance between the pixels of the ear contour (Fig. 5D–F).

Automatic measurements were tested by comparison with manual measurements. Ear length and maximum diameter were measured with a ruler and a sliding caliper (gauge), respectively. The measured length ranged from 3.4 cm to 23.8 cm and the maximum diameter from 1.9 cm to 5.5 cm, exploring similar variability for both panels.

Grains were automatically counted with the Earbox from the segmented grains in each image (Fig. 5A–C). Grain objects identified by segmentation were ‘shrunk’ either to a vertical line (one-pixel width) along the principal axis (Additional file 4: Fig. S4B) or to a horizontal line (one pixel height) along the perpendicular axis (Additional file 4: Fig. S4C). Horizontal distances between

objects were corrected by considering each ear section i (one pixel height) along the ear axis as a circle of diameter $Diameter_i$ (Additional file 5: Fig. S5). The mean distance $Dist_i$ between two consecutive vertical lines was calculated at each position i along the ear axis.

The *number of grains per cohort* at position i was estimated by the Earbox as the ratio of the ear perimeter ($\pi * Diameter_i$) to the distance $Dist_i$ between contiguous grains (Fig. 5G–I). The *number of cohorts* was estimated by the Earbox at each horizontal position perpendicular to the ear axis by counting the number of horizontal lines crossed from the bottom to the top of the ear (Additional file 4: Fig. S4C; Additional file 6: Fig. S6). The cohorts were incomplete on ear sides, and we considered the maximum observed value as the number of cohorts in the image (Additional file 6: Fig. S6). The *number of grains per ear* was calculated by the Earbox from the number of cohorts and the number of grains per cohort measured in the 6 ear images and in the basal, median, and apical ear zones. It is derived from a composite calculation performed for each image by averaging: (i) an over-estimator considering the number of grains per cohort in the median zone of the ear and the maximum number of cohorts, and (ii) an under-estimator using information from both the number of grains per cohort and the mean number of cohorts in each third of the ear. The average of these two indicators was identified as the most relevant estimator.

Manual measurements related to grain organisation were performed manually to be tested against automatic measurements. The *number of grains per cohort* was counted at 3 positions along the ear by visually distinguishing a basal zone, a median zone, and an apical zone (Additional file 3: Fig. S3). It was compared to the mean number of grains per cohort averaged over the whole corresponding zone defined automatically i.e., basal third, median third, and apical third of the ear. The *number of cohorts* was counted on 4 sides of the ear and compared to the number of cohorts automatically calculated as an average over the 6 images (Additional file 6: Fig. S6). The *number of grains per ear* was counted with an automatic counting machine (*Contador: Seed counter—Pfeuffer GmbH (Quality control of grain and seeds)*, n. d.) after removing them from ear cob. Because this measurement is destructive, it was only performed on a subset of the Environmental Diversity panel (257 ears), to keep enough ears intact to test and validate future updates and developments of the image analysis algorithm.

Grain dimensions were automatically measured with the Earbox. *Grain height and grain width* were calculated by fitting each segmented grain to a rectangle: grain height was defined as the fitted dimension along the axis of the ear and grain width as the fitted

perpendicular dimension (Fig. 5M–R). As mentioned above, the horizontal distance (grain width) was corrected to consider the circular shape of the ear sections (Additional file 5: Fig. S5).

Automatically measured grain dimensions were confronted to manual measurement. For this purpose, the height and width of 809 grains from the images of 9 reference ears from both panels were manually measured on the images generated by Earbox. We only considered grains located in the centre of the image to avoid distortion of grain widths, and performed a grain-by-grain comparison of dimensions by identifying each grain with its barycentre coordinates on the image.

The spatial arrangement of grains in cohorts was measured with the Earbox system by assigning each segmented grain to a cohort (Additional file 7: Fig. S7). To achieve this, the grains were scanned from the bottom to the top of the ear and classified according to their relative proximity, which depends on the mean size of objects (grains) in the image. More precisely, the algorithm starts with the lowest grain in the ear, checks the centroids of the grains at half the mean ear grain width and assigns the selected grains to the first cohort. Previously classified grains are removed for the remainder of the analysis and subsequent cohorts of grains are iteratively classified in the same way until no grains remain. This allowed grain dimensions to be plotted against cohort ranking (Fig. 5M–R) which is relevant to account for the developmental gradient along the ear due to morphogenesis (Fig. 1).

Automatic measurements of grains spatial arrangement were confronted to manual measurements. Each grain was manually assigned to a cohort by determining its rank along a row, from 1 at the ear bottom to n at the ear apex, at 3 positions around the ear.

Finally, an automatic method for abortion zones characterization was developed. The abortion zone is an area on the ear (inflorescence) for which the ovaries (floret) have not produced a grain: hence, they are considered loss of production areas that are directly visible and measurable on the ear (Fig. 1). Grain masks allowed discriminating grain pixels from pixels outside the grains but within the ear contours. The latter were considered as corresponding to aborted zones. The *grain set ratio* was calculated for each ear section i (one pixel height) along the ear axis as the ratio of the number of grain pixels to the number of pixels of ear diameter at that section. It was smoothed by a running average over 2% of the total vertical pixels to make it less sensitive to high variations (Fig. 5J–L). At the ear level, we considered the zones with grain set ratio greater than 50% as fertile zones and the others as aborted zones. When no

fertile zone was detected, its length was set to 0 and the apical and basal aborted zones were each set to half of the total ear length.

The automatic method of abortion characterization was tested against manual measurements. The dimensions of the aborted and fertile zones were visually positioned and measured manually with a ruler to represent the approximate positions at which the cohort abortion rate was greater (aborted zone) or lower (fertile zone) than 50%.

Results

A high-quality segmentation allowed for reliable trait measurements

The masks computed with the trained neural network provided a standardized and reliable method for ear and grain segmentation on the whole set of acquired images. The mean average precision metric (mAP) used to assess the quality of segmentation with manually segmented images was 0.4 for Empirical Segmentation (ES) masks and 0.52 and 0.55 for Deep Learning DL1 and DL2 results, respectively. The ES masks had an acceptable value above the standard threshold, while both DL1 and DL2 iterations improved the indicator. In addition to validating the quality of the segmentation, these results also show an overall improvement at each step, highlighting their importance in the method. The high-quality segmentation obtained for a wide diversity of ear and grain phenotypes allows the production of comparable, standardized, and automatic data for both studied panels, composed of ears as different as small horny strawberry-shaped, and dented commercial hybrids.

The use of Neural Networks for Deep Learning is a powerful tool for image analysis, commonly used for plant phenotyping, both for morphological measurements and feature classification. The downside of our method is its requirement for annotated images that are difficult to acquire and analyse: they involve time-consuming annotation work, often performed manually. To overcome this, we chose an empirical approach using simple image analysis tools to produce a set of automatically annotated ear images with little time-input (avoiding manual annotation of the whole set of pre-processed images) to be used for deep learning iterations. The resulting empirical masks were used to train a Neural Network. Indeed, this is a direct and efficient mean to synthesize the best information while potentially improving the outputs. Furthermore, it provides a straightforward and efficient way to improve the analysis system if needed, by adding information

through new segmented masks from previously unaccounted-for diversity.

Ear dimensions and shape, fertile and aborted zones

The automatic measurements of ear length and ear diameter were accurate. The linear regressions with the manual measurements closely fit to the bisector for both variables: respectively $R^2=0.99$, $RMSE=0.44$ cm for ear

length; and $R^2=0.97$, $RMSE=0.15$ cm for ear diameter (Fig. 6A, B). The small differences may be due to differences in methodology: the automatic algorithm measured the ear length with the joined line of the central pixels of the ear (usually not a straight line) while the manual measurement was a straight-line measurement. Since most ears are curvilinear, the automatic measurement appears to be more accurate in describing the diversity

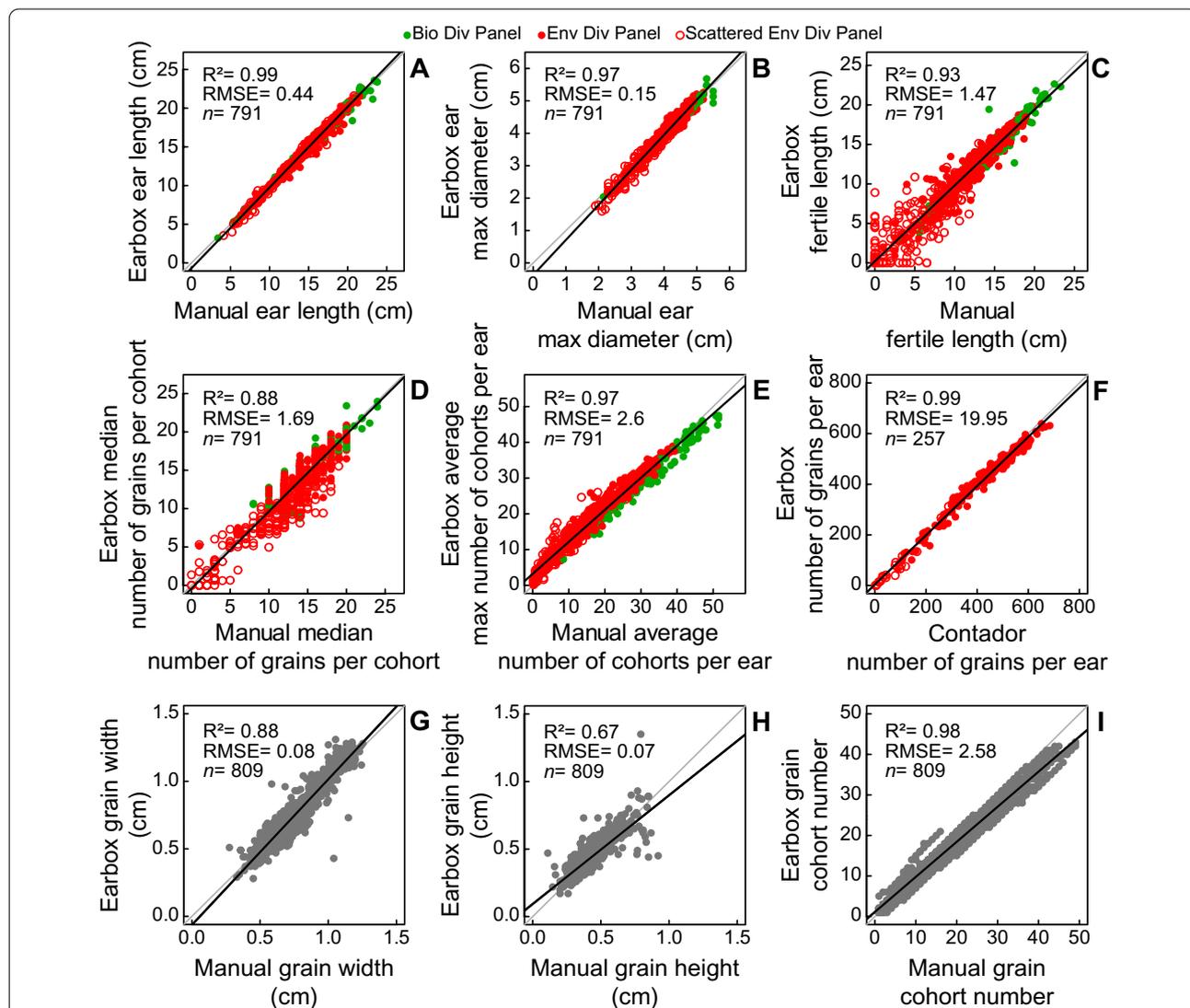


Fig. 6 Comparison of Earbox (y) and reference (x) data. **A** Ear length in centimeters. **B** Maximum ear diameter in centimeters. **C** Length of the fertile zone in centimeters. **D** Number of grains per cohort in the median third of the ear. **E** Average number of cohorts per ear, the mean of 4 observations around the ear (x), the average of the maximum number of cohorts per ear of 6 images of the ear (y). Graphs **A–E** include all 791 ears from both panels. **F** Number of grains per ear measured automatically by the Earbox (y) and manually with an automatic grains counter (x) for 257 ears selected from the environmental diversity panel. **G** Grain width in centimeters, measured along the axis perpendicular to the ear. **H** Grain height in centimeters, measured along the main axis of the ear. **I** Assignment of a cohort number to each grain. Data presented in **G–I** correspond to a set of 809 grains measured and classified automatically by the Earbox (y) and measured manually on the acquisition images (x) for comparison. Green dots: ears from the biological diversity panel; red dots: ears from the environmental diversity panel. Empty red dots: scattered ears of the environmental diversity panel (Fig. 3B). Grey dots: grain dimensions (one dot = one grain). Grey line: bisector line. Black line: linear regression of the data. R^2 : correlation coefficient between x and y values, $RMSE$: root mean square error, n: number of observations in each graph.

of ear shapes. For the same reasons, the algorithm might also be more accurate in determining the maximum ear diameter, which was inferred via an algorithm, rather than visually. Moreover, the accurate measurements of the algorithm at each pixel along the image (Fig. 5D–F) provide access to new variables describing ear shape, such as ear diameter along the ear and centreline curvature, to be investigated and explored for new avenues of comparison between phenotypes and varieties.

Promising results were obtained towards a standardized way of characterising maize ear abortion. Comparison of automatic and manual data for fertile and aborted zones (fertile zone: Fig. 6C; apical and basal abortion zones: Additional file 8: Fig. S8C and S8D) indicates good agreement between them, especially for the fertile zone length ($R^2=0.93$, RMSE=1.47 cm; Fig. 6C). A discrepancy appears in the case of scattered ears (empty dots; Additional file 8: Fig. S8C and 8D) for which a precise visual positioning of the zones is difficult because the grains are randomly scattered along length and circumference of the ear (Additional file 6: Fig. S6). In these cases, the automatic method is probably more relevant, as it uses a common and accurate rule for all ear types. Moreover, characterization at each vertical pixel provides information on the spatial distribution of abortion and grain set along the ear (Fig. 5J–L), inaccessible by conventional methods.

Grain counting

The spatial organisation of grains along the ear measured with the automatic method was both accurate and trustworthy, even for scattered ears. First, for the number of grains per cohort, the manual and automatic estimators were highly correlated in the median ($R^2=0.88$, RMSE=1.69 grains; Fig. 6D) and basal zone ($R^2=0.87$, RMSE=2.07 grains; Additional file 8: Fig. S8A), and less correlated in the apical zone ($R^2=0.68$, RMSE=3.43 grains; Additional file 8: Fig. S8B). Most of the discrepancies are due to scattered ears (empty dots in Fig. 6 and Additional file 8: Fig. S8) *i.e.*, ears with incomplete cohorts, making the cohort identification uncertain and, as a result, counting their grain number difficult (Fig. 3B; Additional file 3: Fig. S3, scattered ear). Earbox data tend to be more objective and closer to the true average number, as they incorporate information from the entire apical zone, whereas manual estimates can be considered more subjective where abortion was high (empty dots in Fig. 6D and Additional file 8: Fig. S8A and B). In addition, the Earbox data provide access to the vertical distribution of this variable (Fig. 5G–I).

Manual and automatic measurements were also highly correlated for the number of cohorts (Fig. 6E, $R^2=0.97$, RMSE=2.6 cohorts), even for scattered ears.

The number of grains per ear was highly variable, ranging from almost 0 to about 700 grains per ear (Fig. 6F). The Earbox estimator of the number of grains per ear, which considers the number of cohorts and the number of grains per cohort, was highly correlated with the Contador measurement (Fig. 6F; $R^2=0.99$, RMSE=19.95 grains). These results validate the potential of the whole system to be used under both optimal and constraining conditions and for the study of the determinism of grain number in maize and its response to the environment.

Grain dimensions and spatial positions: a potential framework for studying developmental gradients

The method adequately estimates grain dimensions (Fig. 6G, H), ranging from 0.3 to 1.3 cm wide (Fig. 6G) and 0.1 to 0.9 cm high (Fig. 6H) in both panels. The correlation between manual and automatic measurements was indeed high for grain width ($R^2=0.88$, RMSE=0.08 cm; Fig. 6G), and slightly lower for grain height ($R^2=0.67$, RMSE=0.07 cm; Fig. 6H). The slight differences may be due to the narrower range of variation observed for grain height versus grain width in the training dataset, which can be easily addressed with further Deep Learning iterations with suitable datasets. Most of the noise comes from isolated grains on scattered ears that tend to have a more circular shape when space is available around them (examples Fig. 3B), which could be easily resolved with an increase in the proportion of data or an individual training for scattered ears. Nevertheless, the results indicate that the method correctly positions the grain barycentre and properly captures shape variations between and within ears.

Automatic and manual measurements were consistent in assigning a cohort number to each grain, *i.e.*, its vertical positioning along ear rows (Fig. 6I). Manual and automatic grain cohort numbers were highly correlated in the 809-grains sample set ($R^2=0.98$, RMSE=2.58 grains), indicating that the grains were properly located in the spatial organisation of the ear.

Thus, the system was able to characterize and discriminate a large variability in grain dimensions (width and height) and shapes (width/height ratio) among the studied ears (Additional file 9: Fig. S9), potentially allowing a reliable characterisation of genotypes based on these traits. In addition, by gathering grains into cohorts with synchronous development, the method gives access to the distribution of grain dimensions along developmental gradients. These distributions differ among ears: they are almost flat, decrease at different rates, or display a

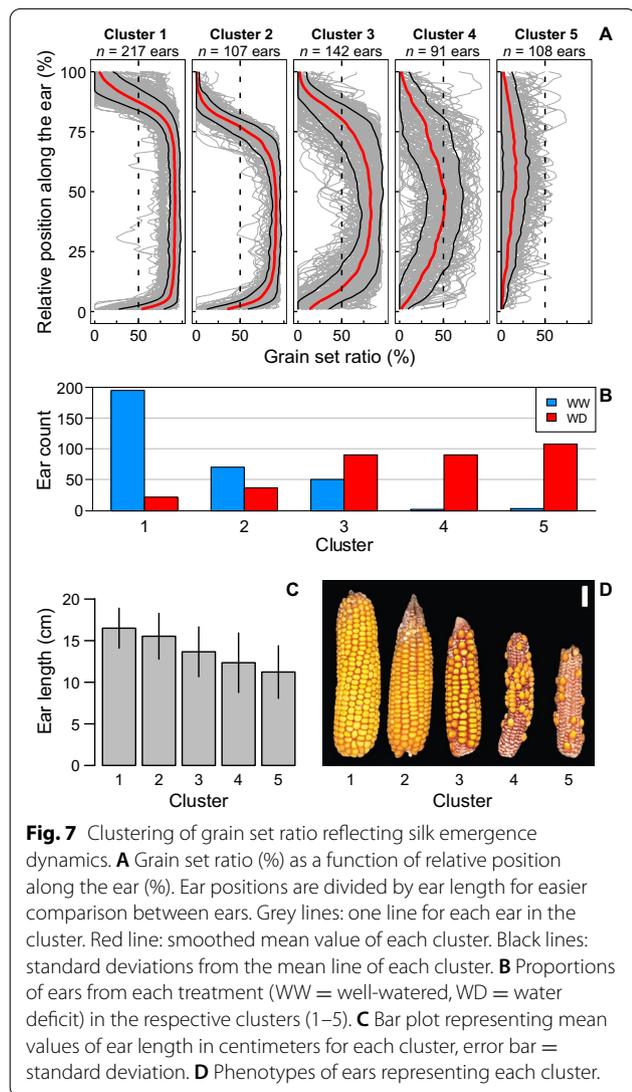
maximum at different vertical positions (Additional file 9: Fig. S9). Since developmental gradients are relevant to ear morphogenesis, they provide a framework for study and analysing the determinism of grain dimensions and, consequently, grain filling and weight.

The spatial distribution of abortion reveals plant response to stress

The grain set ratio (GSR) emerged as a synthetic trait characterizing the response to environmental scenarios (Fig. 7). GSR profiles, i.e., grain set ratio as a function of vertical position along the ear axis, were established for the environmental diversity panel grown under contrasting soil water availability during flowering (665 ears). We performed an a priori-free analysis of all these profiles, using a clustering algorithm (k-means), only based on these GSR profiles, independent of other ear variables.

Precisely, for the analysis to be independent of absolute ear length variations between ears, vertical positions along ear of GSR values were normalized by ear length (dividing absolute values in cm by ear length to use values between 0 and 1 for all ears), and averaged for all sides of each ear.

All curves were processed using Euclidean distance matrix and ward method to generate a cluster tree synthesizing the similarities between ears and finally discriminating ears into 5 clusters with increasing intensity of abortion from cluster 1 to cluster 5 (Fig. 7A, D). Cluster 1 coincided with ears with no or limited abortion. Abortion was limited to the apical zone of ears in cluster 2 and extended to the basal zone in cluster 3. The apical and basal aborted zones were wider in cluster 4, whereas they extended to the entire vertical profile of the ears from cluster 5. The progress of abortion from cluster 1 to cluster 5 followed the reverse order of silk emergence, which is reported in the literature as the main predictor of ovary/grain abortion frequency in response to constraints during flowering (Fig. 1 and [10]). The distribution of well-watered and water-stressed plants among clusters also indicated an increasing impact of stress from cluster 1 to cluster 5. Well-watered plants mainly belonged to cluster 1, and barely to clusters 2 and 3, whereas water-stressed plants mainly belong to clusters 3 to 5 (Fig. 7B). Moreover, the average ear length for each cluster decreased from cluster 1 to cluster 5 (Highly significant, p-value < 0.001 for One-Way ANOVA, detailed in Additional file 10: Table S1) whereas it was not considered as a factor for cluster calculation, not explicitly included in the k-means sorting algorithm (Fig. 7C).



Discussion

A robust phenotyping pipeline to evaluate biological resources, complementary to existing procedures

The phenotyping pipeline developed and presented in this study (hardware and software) was able to accurately characterize, independently of the colour, shape, or transparency of grains and ears: the shape and dimensions of the ear, the number of grains and their spatial organisation, and the dimensions of the grains along the ear. The data were very similar to conventional manual data, with a much lower acquisition time.

Compared to existing systems for maize ear handling and phenotyping [21, 21], the EARBOX system has the advantage of imaging multiple ears at a time while greatly reducing data acquisition time per ear and being scalable to more ears by increasing the number of cameras and rollers (but increasing costs). Considering an acquisition time of 15 s per ear (cleaning and imaging), it is greatly enhanced from comparable systems for which information is available in the literature: one minute per ear [21]

and Warman, 2021). The choice of imaging 6 sides of the ear has been tested and proven trustworthy in this study for a precise measurement but could be reconsidered in case where faster acquisition time and less precision can be needed, that would greatly increase the throughput, comparable to single imaging systems [17, 19]. Nonetheless, IR images being a basis of the analysis to normalize ear and grain colors, the pipeline developed and presented in this study cannot be used as is to extract phenotypic variables from simpler RGB imaging systems alone (ex: smartphone pictures taken in the field with common RGB cameras). For the analysis and variable extractions, most pipelines do not extract as much information as the EARBOX system from non-destructive analysis of ears and yield better results, but with non-comparable hardware (A few seconds for both [15] and [19]. The benchmark done in this study shows that affordable laptop hardware (~1500 euros) can be used to extract masks and phenotypic variables from data acquired with the EARBOX with reasonable computing time (~2 min per ear). This can be greatly enhanced if ran on GPU clusters, but those usually require more technical expertise, stable internet access and recurring payments, not necessarily possible and accessible in all phenotyping laboratories.

The system developed here can also be distinguished by the phenotypic variability used to develop, train, and test its robustness. Most studies focus on specific colors and types of ears and grains [1, 21] while a few explore a variability on commercial hybrids and various ear, grain and cob colors [13, 15, 19, 22], or abortion phenotypes [12], but none investigate the whole range of these possibilities, specifically from water deficits at flowering coupled with cases of biotic stresses with one model training. The development of methods that allow to treat both healthy ears and ears suffering from drought or diseases is an important step forward to produce efficient selection and research tools for field application to tackle the image analysis bottleneck of phenomics [36].

In addition to this, the system provides new traits, inaccessible by conventional methods, especially grain dimensions as a function of the grain cohort number, relevant to ear morphogenesis, and the distribution of abortion frequency along the ear, relevant to plant response to stress. Analysis of the genetic bases of these traits could enlighten the role and regulation of crucial genes determining ear phenotype and grain organisation, as well as their response to the environment. This could lead to new breeding traits responding to climatic challenges, which could be used in marker-assisted selection. Adding to our system an automated and standardized calculation of standard qualitative descriptors of maize cultivars, such as the conicity of the ear or, for the grains, their shape, type, colour, or organization on the ear (GEVES

or UPOV technical documentation) would only require simple general statistical classification methods based on machine learning or deep learning. Our methodology is also complementary to other methods used for varietal description, such as cross-sections or ear deseeding, to characterize the cob, or the morphology of the grains and their physiological characteristics [37, 38], involving robotics for ear and grain handling [39]. Its relative simplicity and flexibility allow easy adaptation of the ear processing line to integrate the Earbox phenotyping solution before ear deseeding and grain phenotyping.

Finally, while the panel used in this study was specifically defined to capture much of the existing phenotypic variability encountered for maize, the ability of the pipeline to handle unseen or new phenotypes could be discussed. This kind of limitation can be easily overcome by first testing the system as it is and reapplying the learning logic if the results are not satisfactory. In such a case, great attention should be applied to reworking the ratio of the various ear types and grains (colors and shapes) so as not to unbalance the learning specifically towards one overrepresented type. It would be necessary to bring both annotated image data and hand-measured phenotypic variables to enrich the existing data and test the outputs under appropriate experimental conditions. The first developed 'empirical segmentation' presented in this study could be of great use for a first iteration of ground truth to be corrected afterwards, to avoid time-consuming annotations and focus more efforts on model training.

Contribution to the analysis of adaptation/tolerance to environmental scenarios in combination with crop models

The new features obtained by the phenotyping pipeline open new avenues in the characterisation of maize grain yield formation in response to genetic and/or environmental factors. In particular, the spatial distribution of the grain set ratio appears to be a marker of the dynamics of silk emergence and ovary/grain abortion, a major component of the plant's response to environmental scenarios [2, 10]. Consistent with the literature [10, 40–42] our results (Fig. 7) suggest that the measurement of this variable potentially provides a high-throughput proxy for the complex processes involved in ear morphogenesis (rate and number of ovary initiations, growth rates of silks and pollen tube, development of the husks). This would greatly facilitate ecophysiological studies of the mechanisms determining yield components and their response to the environment.

Yield losses in maize are most pronounced when stress occurs around flowering [43, 44] affecting grain number determination. Reproductive failure has different faces

and can manifest as ear barrenness, incomplete ear pollination due to lack of pollen, and grain abortion [10, 45–47]. The effects of the timing of stressful conditions and the pattern of zygote development along the ear row (successive cohorts) determine the nature of the ear phenotype associated with reproductive failure [9]. In this sense, the ear phenotype can tell whether the crop experienced a stressful scenario and give some clues about the timing of the stressful condition. However, its implementation in plant genetics is difficult or impossible due to the size of studied panels of genotypes and/or to the number of traits resulting from stress x phenology combinations. The Earbox system presented in this study answers this bottleneck by providing standardized and high-throughput traits measurement. Furthermore, such a development paves the way towards new analyses to understand the interaction between genotype and environment in the context of water deficit for maize. While many efforts have been made to study the dynamics of water deficit related to phenology, [48, 49], less progress has been made to study its genotypic variability under drought scenarios [9]. As such, additional investigations provide insights to link the observed phenotype at harvest with events occurring at flowering, thus helping in the identification of varieties best adapted to a specific water deficit scenario.

Conclusion

The system developed and presented in this study is a scalable system providing an accurate, robust, and reliable way to extract precise measurements from maize ear images, including spatial features of grain organization. This work illustrates, like many others [50–54], the possibilities and the efficiency that open-source technologies and low-cost electronics now offer to plant science. They make accurate phenotyping accessible to everyone. In the case of the Earbox, even research structures with limited resources, farmer cooperatives, or multi-site research projects (limited by multiple observers and non-standardized methodologies), can claim reliable and reproducible ear phenotyping data with a system that can be easily modified to be integrated into complete ear and grain processing chains. For example, cameras can be replaced for higher resolutions or multispectral acquisition for characterization of grain physiology [31, 32, 55–59]. Additional steps of deep learning would probably be sufficient to develop a method for the recognizing and classifying of maize diseases [60, 31, 32], or for characterising early grain development, by processing immature ears and grains a few days after flowering. Finally, the results of this work pave the way for future development of tools for inflorescence phenotyping of other crops, such as

wheat and sunflower, for which the present system will be adapted.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13007-022-00925-8>.

Additional file 1 Fig. S1. Choice of ear rotation angle based on Earbox specifications. (A) Heatmap of the theoretical cumulative ear percentage seen in all 6 images for roller rotation angles between 0° and 360° and diameters between 2cm and 6cm. Values are calculated assuming 120° of the ear circumferences is captured in each image. The dotted line represents the selected roller rotation angle, for which no critical diameter is encountered, capturing 100% of the information from each ear. (B) Match between theoretical and measured rotation angle for a set of ears chosen to represent diameter diversity in maize. Dots: measured and theoretical rotation angles. Colors: ear diameters. Dotted line: $y = x$. Solid line: linear regression. R^2 : correlation coefficient between x and y values, RMSE: root mean square error.

Additional file 2 Fig. S2. Steps of the deep learning iteration workflow. (A) Steps of the deep learning (DL) workflow and (B,C,D) example images used for 3 typical ears representing the problems encountered and the manual correction performed. (B) Typical ear with good empirical grain masks used directly for training the first iteration of deep learning (B2) and requiring minimum manual corrections (B3 to B4) to produce the routine segmentation method (B5). (C, D) Typical ears with false detections and erroneous grain segmentations from the empirical segmentation (C2, D2) and requiring major manual corrections (step 3 to 4) used for training the second deep learning iteration to produce the routine segmentation method (C5, D5). Orange boxes, image acquisition. White boxes, image, or data processing. Green boxes processed images or data. Green areas, manual grain mask corrections.

Additional file 3 Fig. S3. Methodology for manual phenotyping of maize ears. (A) Sample ear from the biological diversity panel. (B) Sample ear from the environmental diversity panel grown under water deficit conditions showing partial pollination and/or ovary and grain abortion. Red areas: example of grains used for manual phenotyping of basal (BGC), median (MGC) and apical (AGC) number of grains per cohort, particularly difficult to characterize on "scattered" ears. To standardize the measurement, the method consisted in counting the number of rows (= lines of grains along the ear perpendicular to the cohorts) with at least one grain per row for each third of the ear. Green areas: example of grains used for manual phenotyping of the average number of cohorts per ear (CN). The measurement was repeated 4 times around the ear. White zone: fertile zone of the ear (FZ), where the surface occupied by grains visually represents more than 50% of the visible surface of the ear. Grey zones: basal (BA) and apical (AA) abortion zones, where the surface occupied by grains visually represents less than 50% of the visible surface of the ear. The lengths of the FZ, BA, and AA zones were measured manually along the main axis of the ear.

Additional file 4 Fig. S4. (A) Illustration for 3 ears of the processing used to extract ear length from ear masks. The black line represents the centerline of pixels along the main axis of the ear (vertical axis, starting from the bottom to the top of the ear), the measured ear length is the number of pixels in this line. (B) Illustration of the processing procedure to calculate the number of grains per cohort. Grain objects were reduced to a one-pixel wide vertical line along the principal axis. (C) Illustration of the processing procedure to calculate the number of grain cohorts. Grain objects were reduced to a one-pixel wide horizontal line along the ear axis perpendicular to the principal axis.

Additional file 5 Fig. S5. Illustration of the image correction applied to project the distances and positions of the reference points onto a hypothetical circular section of the ear. Orange circle, boundaries of the ear. The reference image measurement (dref) is corrected using the horizontal distances measured between its extreme reference points (black dots) and the center of the ear (dmax and dmin). The final corrected measured

is an estimate of the length of the arc resulting from the projection of dref onto a hypothetical perfect circle of radius (Rear).

Additional file 6 Fig. S6. Output of the estimation of the number of cohorts along the perpendicular ear axis. For each position along the ear diameter (x), the number of cohorts (y) is calculated for each ear shown in Fig. 5 with the methodology presented in Supplementary Fig. S3. Black line, ear with white grains; red line, ear with vine grains; golden line, aborted ear with yellow grains. Dotted lines represent the maximum number of cohorts for each curve. The maximum number of cohorts is used as the output of the routine workflow and its average over the 6 ear sides is used for the correlation in Fig. 6. Each curve represents data from a single image/ear side.

Additional file 7 Fig. S7. Illustration of the cohort classification algorithm. Grains are classified in cohorts (each identified by a color) by scanning the ear from bottom to top starting with the lowest grain (bold black dots on the barycenter). The classification is done sequentially, one cohort at a time. For each cohort, grains whose barycenter lie within a specific range of pixels along the main axis of the ear are classified into a common cohort, and removed for the rest of the classification process. Black dots: grain barycenter's. Colors: grains classified in the same cohort.

Additional file 8 Fig. S8. Comparison of Earbox (y) and reference (x) data. (A) Number of grains per cohort in the basal third. (B) Number of grains per cohort in the apical third. (C) Length of the basal aborted zone in centimeters. (D) Length of the apical aborted zone in centimeters. Green dots: ears from biological diversity panel; red dots: ears from environmental diversity panel. Empty red dots: scattered ears of the environmental diversity panel (Fig.3B). Grey line: bisector line. Black line: linear regression of data. R^2 : correlation coefficient between x and y values, RMSE: root mean square error, n: number of observations in each graph.

Additional file 9 Fig. 9. Examples of grain dimensions as a function of cohort and position along the ear across 9 contrasting ears. Each point represents the average dimension of all grains classified in the same cohort (Fig. 5P, 5Q, 5R) across the 6 sides (images) of the ear. Red dots, average grain height in centimeters. Green dots, average grain width in centimeters. Error bars, standard deviation. Errors for position from ear base are shown but are smaller than the dots.

Additional file 10 Table S1. Results of ANOVA testing the effect of Clusters of hydric conditions on mean ear length. The table was calculated using SPSS Software's function '1-factor ANOVA'. Columns contain the values for Sum of squares, degrees of freedom (DF), Observed Fischer coefficient (Fobs) and the calculated significance, for between Groups (clusters) and inside groups (clusters) tests.

Acknowledgements

We thank all the staff of the Station Expérimentale de Mauguio & Saint-Martin-de-Hinx for their special contribution to grow, harvest and provide access to the biological diversity used in the dataset of this study. We particularly thank the French INRAE agents, especially Serge Malavieille, for their patience during the testing of our prototypes, their feedback and their valuable knowledge that allowed us to successfully complete this project. A special thanks to Christian Fournier (INRAE LEPSE) for his helpful advice to put in place the first steps of the empirical segmentation. Special thanks to Romain Chapuis and Olivier Turc for investing time in the project from beginning to end. Finally, we thank Joubin Ezami for his patience and his valuable contribution for the 3D modeling and the production of the imaging system.

Author contributions

Conceptualization: VO, TL, OT, SL, CP, RC, CW; Project supervision: VO. Project administration: VO, SL, TL, OT; Biological resources and experimentation: CP, RC, CW; Data measurements and gathering: VO, RC, CP, TL, SL; Hardware-development: VO, TL. Software-development: VO, SL, TL; Deep Learning implementation and training: TL. Formal analysis and investigation: SL, VO, TL, OT; Data visualization: VO, SL, OT. Writing—review and editing: SL, VO, OT, FT, SAP, TL, CW; *Authors are listed in the order of contribution for each item.

Funding

This study is a joined effort from Phymea-Systems and the INRAE. For Phymea-Systems, the work was largely financed by the company's own resources with a participation from Région Occitanie funds. For the INRAE participation, this work was supported by the EU project FP7-244374 (DROPS), the projects ANR-10-BTBR-01 (Amaizing) and ANR-11-INBS-0012 (Phenome) and the EU project H2020 731013 (EPPN2020).

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare they have no competing interest.

Author details

¹Phymea Systems, 453 Rue de l'Espinouse, Montpellier, France. ²LEPSE, Univ Montpellier, INRAE, Institut Agro, Montpellier, France. ³MELGUEIL, Univ Montpellier, INRAE, Montpellier, France. ⁴UE Maïs, INRAE, Univ. Bordeaux, Bordeaux, Saint Martin de Hinx, France. ⁵Facultad de Agronomía, IFEVA-CONICET, Universidad de Buenos Aires, Av. San Martín 4453 (C1417DSE), Ciudad de Buenos Aires, Argentina. ⁶Cátedra de Sistemas de Cultivos Extensivos-GIMUCE, Facultad de Ciencias Agrarias, Universidad Nacional de Rosario, Campo Experimental Villarino S/N, S21125ZAA Zavalla, Prov. de Santa Fe, Argentina.

Received: 18 January 2022 Accepted: 13 July 2022

Published online: 28 July 2022

References

- Liang X, Wang K, Huang C, Zhang X, Yan J, Yang W. A high-throughput maize kernel traits scorer based on line-scan imaging. *Measurement*. 2016;90:453–60. <https://doi.org/10.1016/j.measurement.2016.05.015>.
- Tardieu F, Simonneau T, Muller B. The physiological basis of drought tolerance in crop plants : a scenario-dependent probabilistic approach. *Annu Rev Plant Biol*. 2018;69(1):733–59. <https://doi.org/10.1146/annurev-arplant-042817-040218>.
- Messmer R, Fracheboud Y, Bänziger M, Vargas M, Stamp P, Ribaut J-M. Drought stress and tropical maize: QTL-by-environment interactions and stability of QTLs across environments for yield components and secondary traits. *TAG Theor Appl Genetics*. 2009;119(5):913–30. <https://doi.org/10.1007/s00122-009-1099-x>.
- Peng B, Li Y, Wang Y, Liu C, Liu Z, Tan W, Zhang Y, Wang D, Shi Y, Sun B, Song Y, Wang T, Li Y. QTL analysis for yield components and kernel-related traits in maize across multi-environments. *TAG Theor Appl Genet*. 2011;122(7):1305–20. <https://doi.org/10.1007/s00122-011-1532-9>.
- Alvarez Prado S, López CG, Senior ML, Borrás L. The Genetic architecture of maize (*Zea mays* L.) Kernel weight determination. *G3 Genes Genomes Genetics*. 2014;4(9):1611. <https://doi.org/10.1534/g3.114.013243>.
- Amelong A, Gambín BL, Severini AD, Borrás L. Predicting maize kernel number using QTL information. *Field Crop Res*. 2015;172:119–31. <https://doi.org/10.1016/j.fcr.2014.11.014>.
- Gambín BL, Borrás L. Resource distribution and the trade-off between seed number and seed weight : a comparison across crop species. *Ann Appl Biol*. 2010;156(1):91–102. <https://doi.org/10.1111/j.1744-7348.2009.00367.x>.
- Millet EJ, Kruijer W, Coupel-Ledru A, Alvarez Prado S, Cabrera-Bosquet L, Lacube S, Charcosset A, Welcker C, van Eeuwijk F, Tardieu F. Genomic prediction of maize yield across European environmental conditions. *Nat Genet*. 2019;51(6):952–6. <https://doi.org/10.1038/s41588-019-0414-y>.

9. Messina CD, Hammer GL, McLean G, Cooper M, van Oosterom EJ, Tardieu F, Chapman SC, Doherty A, Gho C. On the dynamic determinants of reproductive failure under drought in maize. *in silico Plants*. 2019. <https://doi.org/10.1093/insilicoplants/diz003>.
10. Oury V, Tardieu F, Turc O. Ovary apical abortion under water deficit is caused by changes in sequential development of ovaries and in silk growth rate in maize. *Plant Physiol*. 2016;171(2):986–96. <https://doi.org/10.1104/pp.15.00268>.
11. Saini HS, Westgate ME. Reproductive development in grain crops during drought. In: Sparks DL, editor. *Advances in agronomy*. Cambridge: Academic Press; 1999. [https://doi.org/10.1016/S0065-2113\(08\)60843-3](https://doi.org/10.1016/S0065-2113(08)60843-3).
12. Chipindu L, Mupangwa W, Mtsilizah J, Nyagumbo I, Zaman-Allah M. Maize kernel abortion recognition and classification using binary classification machine learning algorithms and deep convolutional neural networks. *AI*. 2020;1(3):361–75. <https://doi.org/10.3390/ai1030024>.
13. Makanza R, Zaman-Allah M, Cairns JE, Eyre J, Burgueño J, Pacheco Á, Diepenbrock C, Magorokosho C, Tarekegne A, Olsen M, Prasanna BM. High-throughput method for ear phenotyping and kernel weight estimation in maize using ear digital imaging. *Plant Methods*. 2018;14(1):49. <https://doi.org/10.1186/s13007-018-0317-4>.
14. Severini AD, Borrás L, Cirilo AG. Counting maize kernels through digital image analysis. *Crop Sci*. 2011;51(6):2796–800. <https://doi.org/10.2135/cropsci2011.03.0147>.
15. Miller ND, Haase NJ, Lee J, Kaeppeler SM, de Leon N, Spalding EP. A robust, high-throughput method for computing maize ear, cob, and kernel attributes automatically from images. *Plant J*. 2016;89(1):169–78. <https://doi.org/10.1111/tpj.13320>.
16. Ni C, Wang D, Vinson R, Holmes M, Tao Y. Automatic inspection machine for maize kernels based on deep convolutional neural networks. *Biosyst Eng*. 2019;178:131–44. <https://doi.org/10.1016/j.biosysteng.2018.11.010>.
17. Khaki S, Pham H, Han Y, Kuhl A, Kent W, Wang L. Convolutional neural networks for image-based corn kernel detection and counting. *Sensors*. 2020;20(9):2721. <https://doi.org/10.3390/s20092721>.
18. Khaki S, Pham H, Han Y, Kuhl A, Kent W, Wang L. DeepCorn : a semi-supervised deep learning method for high-throughput image-based corn kernel counting and yield estimation. *Knowl-Based Syst*. 2021;218:106874. <https://doi.org/10.1016/j.knsys.2021.106874>.
19. Kienbaum L, Correa Abondano M, Blas R, Schmid K. DeepCob : Precise and high-throughput analysis of maize cob geometry using deep learning with an application in genebank phenomics. *Plant Methods*. 2021;17(1):91. <https://doi.org/10.1186/s13007-021-00787-6>.
20. Wu D, Cai Z, Han J, Qin H. Automatic kernel counting on maize ear using RGB images. *Plant Methods*. 2020;16(1):79. <https://doi.org/10.1186/s13007-020-00619-z>.
21. Grift TE, Zhao W, Momin MA, Zhang Y, Bohn MO. Semi-automated, machine vision based maize kernel counting on the ear. *Biosyst Eng*. 2017;164:171–80. <https://doi.org/10.1016/j.biosysteng.2017.10.010>.
22. Warman, C., Sullivan, C. M., Preece, J., Buchanan, M. E., Vejlupekova, Z., Jaiswal, P., & Fowler, J. E. (2021). A cost-effective maize ear phenotyping platform enables rapid categorization and quantification of kernels. *The Plant Journal*, 106(2), 566–579. <https://doi.org/10.1111/tpj.15166>.
23. Arvalis (2018). KMScan Documentation. <https://kmscan.fr/index.php/page-d-exemple>. Accessed 20 Dec 2021
24. Law JR, Anderson SR, Jones ES, Nelson BK, Mu E. (2011). *Characterization of maize germplasm : Comparison of morphological datasets compiled using different approaches to data recording*. 13.
25. Warman C, Fowler JE. Deep learning-based high-throughput phenotyping can drive future discoveries in plant reproductive biology. *Plant Reprod*. 2021;34:81–9. <https://doi.org/10.1007/s00497-021-00407-2>.
26. He K, Gkioxari G, Dollár P, Girshick R. Mask R-CNN. 2018. ArXiv:1703.06870. <http://arxiv.org/abs/1703.06870>
27. Davis CC, Champ J, Park DS, Breckheimer I, Lyra GM, Xie J, Joly A, Tarapore D, Ellison AM, Bonnet P. A new method for counting reproductive structures in digitized herbarium specimens using mask R-CNN. *Front Plant Sci*. 2020. <https://doi.org/10.3389/fpls.2020.01129>.
28. Ganesh P, Volle K, Burks TF, Mehta SS. Deep orange : mask R-CNN based orange detection and segmentation. IFAC-PapersOnLine. 2019;52(30):70–5. <https://doi.org/10.1016/j.ifacol.2019.12.499>.
29. Machefer M, Lemarchand F, Bonnefond V, Hitchins A, Sidiropoulos P. Mask R-CNN refitting strategy for plant counting and sizing in UAV imagery. *Remote Sensing*. 2020;12(18):3015. <https://doi.org/10.3390/rs12183015>.
30. Wang X, Xuan H, Evers B, Shrestha S, Pless R, Poland J. High-throughput phenotyping with deep learning gives insight into the genetic architecture of flowering time in wheat. *GigaScience*. 2019. <https://doi.org/10.1093/gigascience/giz120>.
31. Zhang C, Zhao Y, Yan T, Bai X, Xiao Q, Gao P, Li M, Huang W, Bao Y, He Y, Liu F. Application of near-infrared hyperspectral imaging for variety identification of coated maize kernels with deep learning. *Infrared Phys Technol*. 2020;111: 103550. <https://doi.org/10.1016/j.infrared.2020.103550>.
32. Zhang J, Dai L, Cheng F. Corn seed variety classification based on hyperspectral reflectance imaging and deep convolutional neural network. *J Food Meas Charact*. 2020. <https://doi.org/10.1007/s11694-020-00646-3>.
33. Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, Corrado GS, Davis A, Dean J, Devin M, Ghemawat S, Goodfellow I, Harp A, Irving G, Isard M, Jia Y, Jozefowicz R, Kaiser L, Kudlur M, Zheng X. TensorFlow : Large-Scale Machine Learning on Heterogeneous Distributed Systems. arXiv:1603.04467 [cs]. 2016. <http://arxiv.org/abs/1603.04467>
34. Henderson P, Ferrari V. End-to-end training of object class detectors for mean average precision. Cham: Springer; 2017.
35. Lin T-Y, Maire M, Belongie S, Bourdev L, Girshick R, Hays J, Perona P, Ramanan D, Zitnick CL, Dollár P. Microsoft COCO : common objects in context. Cham: Springer; 2015.
36. Tardieu F, Cabrera-Bosquet L, Primrose T, Bennett M. Plant phenomics, from sensors to knowledge. *Curr Biol*. 2017;27:R770–83. <https://doi.org/10.1016/j.cub.2017.05.055>.
37. Baye TM, Pearson TC, Settles AM. Development of a calibration to predict maize seed composition using single kernel near infrared spectroscopy. *J Cereal Sci*. 2006;43(2):236–43. <https://doi.org/10.1016/j.jcs.2005.11.003>.
38. Spielbauer G, Armstrong P, Baier JW, Allen WB, Richardson K, Shen B, Settles AM. High-throughput near-infrared reflectance spectroscopy for predicting quantitative and qualitative composition phenotypes of individual maize kernels. *Cereal Chem*. 2009;86(5):556–64. <https://doi.org/10.1094/CCHEM-86-5-0556>.
39. Seed Analysis Automation. (n. d.). *Optomachines*. <https://optomachines.fr/home/agronomics/seed-analysis-automation/>. Accessed 28 Jan 2021
40. DeBruin JL, Hemphill B, Schussler JR. Silk development and kernel set in maize as related to nitrogen stress. *Crop Sci*. 2018;58(6):2581–92. <https://doi.org/10.2135/cropsci2018.03.0160>.
41. Liu X, Wang X, Wang X, Gao J, Luo N, Meng Q, Wang P. Dissecting the critical stage in the response of maize kernel set to individual and combined drought and heat stress around flowering. *Environ Exp Bot*. 2020;179: 104213. <https://doi.org/10.1016/j.envexpbot.2020.104213>.
42. Shen S, Zhang L, Liang X-G, Zhao X, Lin S, Qu L-H, Liu Y-P, Gao Z, Ruan Y-L, Zhou S-L. Delayed pollination and low availability of assimilates are major factors causing maize kernel abortion. *J Exp Bot*. 2018;69(7):1599–613. <https://doi.org/10.1093/jxb/ery013>.
43. Campos H, Cooper M, Edmeades GO, Löffler C, Schussler JR, Ibanez M. Changes in drought tolerance in maize associated with fifty years of breeding for yield in the US Corn Belt (*Zea mays* L). *Maydica*. 2006;51(2):369–81.
44. Claassen MM, Shaw RH. Water deficit effects on corn. II. Grain components I. *Agron J*. 1970;62(5):652–5. <https://doi.org/10.2134/agronj1970.00021962006200050032x>.
45. Cárcova J, Andrieu B, Otegui ME. Silk elongation in maize. *Crop Sci*. 2003;43(3):914–20. <https://doi.org/10.2135/cropsci2003.9140>.
46. Cárcova J, Otegui ME. Ear temperature and pollination timing effects on maize kernel set. *Crop Sci*. 2001;41(6):1809–15. <https://doi.org/10.2135/cropsci2001.1809>.
47. Moss GI, Downey LA. Influence of drought stress on female gametophyte development in corn (*Zea mays* L) and subsequent grain yield1. *Crop Sci*. 1971;11(3):368–72. <https://doi.org/10.2135/cropsci1971.0011183X001100030017x>.
48. Chenu K, Deihimfard R, Chapman SC. Large-scale characterization of drought pattern : a continent-wide modelling approach applied to the Australian wheatbelt—spatial and temporal trends. *New Phytol*. 2013;198(3):801–20. <https://doi.org/10.1111/nph.12192>.
49. Harrison MT, Tardieu F, Dong Z, Messina CD, Hammer GL. Characterizing drought stress and trait influence on maize yield under current and

- future conditions. *Glob Change Biol.* 2014;20(3):867–78. <https://doi.org/10.1111/gcb.12381>.
50. Bagley SA, Atkinson JA, Hunt H, Wilson MH, Pridmore TP, Wells DM. Low-cost automated vectors and modular environmental sensors for plant phenotyping. *Sensors.* 2020;20(11):3319. <https://doi.org/10.3390/s20113319>.
 51. Czedik-Eysenberg A, Seitner S, Güldener U, Koemeda S, Jez J, Colombini M, Djamei A. The 'PhenoBox', a flexible, automated, open-source plant phenotyping solution. *New Phytol.* 2018;219(2):808–23. <https://doi.org/10.1111/nph.15129>.
 52. Gaggion N, Ariel F, Daric V, Lambert É, Legendre S, Roulé T, Camoirano A, Milone DH, Crespi M, Blein T, Ferrante E. ChronoRoot : high-throughput phenotyping by deep segmentation networks reveals novel temporal parameters of plant root system architecture. *BioRxiv.* 2020. <https://doi.org/10.1101/2020.10.27.350553>.
 53. Pearce JM. Economic savings for scientific free and open source technology : a review. *HardwareX.* 2020;8: e00139. <https://doi.org/10.1016/j.ohx.2020.e00139>.
 54. Salter WT, Shrestha A, Barbour MM. Open source 3D phenotyping of chickpea plant architecture across plant development. *BioRxiv.* 2020. <https://doi.org/10.1101/2020.09.08.288530>.
 55. Caporaso N, Whitworth MB, Fisk ID. Near-Infrared spectroscopy and hyperspectral imaging for non-destructive quality assessment of cereal grains. *Appl Spectrosc Rev.* 2018;53(8):667–87. <https://doi.org/10.1080/05704928.2018.1425214>.
 56. Chu X, Wang W, Ni X, Li C, Li Y. Classifying maize kernels naturally infected by fungi using near-infrared hyperspectral imaging. *Infrared Phys Technol.* 2020;105: 103242. <https://doi.org/10.1016/j.infrared.2020.103242>.
 57. Pang L, Men S, Yan L, Xiao J. Rapid vitality estimation and prediction of corn seeds based on spectra and images using deep learning and hyperspectral imaging techniques. *IEEE Access.* 2020. <https://doi.org/10.1109/ACCESS.2020.3006495>.
 58. Türker-Kaya S, Huck CW. A review of mid-infrared and near-infrared imaging : principles, concepts and applications in plant tissue analysis. *Molecules.* 2017;22(1):168. <https://doi.org/10.3390/molecules22010168>.
 59. Zhou Q, Huang W, Tian X, Yang Y, Liang D. Identification of the variety of maize seeds based on hyperspectral images coupled with convolutional neural networks and subregional voting. *J Sci Food Agric.* 2021. <https://doi.org/10.1002/jsfa.11095>.
 60. Hobbs J, Khachatryan V, Anandan BS, Hovhannisyan H, Wilson D. Broad dataset and methods for counting and localization of on-ear corn kernels. *Front Robot AI.* 2021;8: 627009. <https://doi.org/10.3389/frobt.2021.627009>.
 61. Moser, S. B., Feil, B., Jampatong, S., & Stamp, P. (2006). Effects of pre-anthesis drought, nitrogen fertilizer rate, and variety on grain yield, yield components, and harvest index of tropical maize. *Agricultural Water Management*, 81(1-2), 41-58. Chicago. <https://doi.org/10.1016/j.agwat.2005.04.005>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

