Supplemental materials and methods

Supplemental Method S1. Watering

All plants were watered daily up to their target gravimetric soil water content. Three target gravimetric soil water contents defining the treatments were based on a soil water retention curve calculated for the potting soil (Supplemental Fig. S10), showing the relation between soil water potential and soil gravimetric water content. 2.4 g water/g dry soil was chosen for the Control treatment, because it has a soil water potential of around -10 kPa, which is close to field capacity. 1.4 g water/g dry soil was chosen for the Drought treatment, corresponding to a soil water potential of around -100 kPa. 1.8 g water/g dry soil was chosen for the Mild treatment, leading to an intermediate water potential around -28 kPa. All drought-treated pots started at the well-watered target soil water content and were switched to their drought conditions at either the V5-stage (vegetative stage drought) or the V12-stage (reproductive stage drought). It took several days for the soil to dry down until below the target soil water content and for the pots to receive daily watering again. In the first experiment (March-May 2015) plants were subjected to Control, Mild and Drought treatments. In later experiment only the Control and Drought treatments were used. For the leaf 4 length trial, only plants receiving the Control (2.4g/g) treatment are considered here.

It was necessary to correct the observed weight of the pot for the changing plant weight in order to accurately determine soil water content. In the first experiment, this was done by subtracting the mean plant weight per treatment measured at the most recent destructive sampling time point. For the second to fourth experiment, a Gompertz growth curve (Winsor, 1932) was fitted to the plant fresh weight measurements, of each treatment separately, over time, to predict plant weight as a function of plant age. These predictions were then used to correct the measured weight of the pot.

Supplemental Method S2. Destructive sampling

In the first experiment, 250 plants were destructively sampled throughout development, from the V5-stage until silking (supplemental data S1). The biomass variables fresh and dry weight of roots, leaves, stem and ears were measured for all sampled plants. Starting from the V9-stage, the height of nodes 9, 12, 15, 18 and 20 was determined to measure internode and stem growth. Also starting from the V9-stage, the highest ranked developing ear was sampled and stored in an ethanol:acetic acid (3:1) mixture, to later determine ear traits such as length, spikelet row number and number of spikelets per row.

The three later experiments had a distinct sampling scheme from the first experiment. Different sets of plants were sampled for transcriptome/metabolome on the one hand and destructive phenotyping on the other hand. Only phenotype data from the destructively sampled ears (length, kernel row number and number of kernels per row, supplemental data S2) is used in this study. Results of the transcriptome and metabolome analysis will be reported elsewhere.

Supplemental Method S3. Repeated observations and measurements

In all experiments, plants were visually inspected daily to check their developmental stage. The timing of the V5-stage, leaf 9 appearance, leaf 12 appearance, leaf 15 appearance, leaf 18 appearance, the V12-stage, ear appearance and location, tassel appearance, anthesis and

silking were recorded (supplemental data S3). Additionally, in the first experiment, 100 plants (20 for each of the treatments Control, V5-Drought, V12-Drought, V5-Mild and V12-Mild) were selected for daily measurements of the length of the growing leaves 9, 12, 15 and 18. The growth of leaf 6 was measured for all plants on the platform, including the border plants (14 out of 392 plants had not emerged in the first experiment, resulting in 378 plants divided unevenly over 5 watering treatments). During the leaf 4 length trial, the length of leaf 4 was measured daily for 12 Control treated plants, starting from leaf emergence from the whorl to the end of leaf growth. Leaf length measurements were performed by placing a ruler at soil level at the base of the plant and stretching out the leaf upwards along the ruler.

Supplemental Method S4. Biomass prediction model

Plants subjected to the various watering treatments in the first experiment were destructively sampled throughout development. Plant biomass-related traits, such as stem and leaf fresh and dry weight were measured from the V5-stage onwards until the silking stage, resulting in 236 distinct data points (69 Control, 63 V5-Drought, 61 V5-Mild, 23 V12-Drought, 20 V12-Mild) with both ground truth biomass data and image-derived plant projected area data (supplemental data S5).

Linear models (lm function, R stats package (R Core Team, 2020)) for predicting total or above-ground fresh weight or dry weight were evaluated using leave-one-out-cross-validation (LOOCV), where a prediction is made for each data point using a model that was trained for all other data points (Supplemental table S5). These out-of-bag predictions were then compared to the ground truth values using metrics such as R^2 and the mean-absolute-percentage-error (MAPE) (De Myttenaere et al., 2016) to determine the performance of the model (supplemental table S12). The best performing models were selected using a combination of F tests for nonzero model coefficients, adjusted R^2, LOOCV out-of-bag R^2 and LOOCV out-of-bag MAPE. The final models (eq. 4; eq.5) were trained on all available data points and predicted the total aboveground fresh weight based on the average projected area for all sideview perspectives, the projected area for the top-view camera and the interaction between these two factors and the total aboveground dry weight using a 2nd degree polynomial based on the average projected area for all sideview perspectives:

$$FW_{above} = a_0 + a_1 \times A_{mean.sideview} + a_2 \times A_{topview} + a_3 \times A_{mean.sideview} \times A_{topview} \quad (4)$$

$$DW_{above} = a_0 + a_1 \times {A_{mean.sideview}}^2 + a_2 \times A_{mean.sideview} \quad (5)$$

where FW_{above} and DW_{above} are the aboveground fresh weight and dry weight in g, $A_{mean.sideview}$ is the average projected area for all sideview images in mm^2 and $A_{topview}$ is the projected area for the top-view image in mm^2. R-scripts used for training the linear models, performing the LOOCV and making predictions and accompanying example input datasets for training (experiment 1, supplemental data S5) and predictions (experiment 4) have been made available through Zenodo (https://doi.org/10.5281/zenodo.4323615). Further details regarding the plant biomass modelling can be found in supplemental result S1 and supplemental discussion S1.

The models used to estimate biomass of imaged plants throughout their development were trained on all available imaging and measured biomass data points gathered in the first

experiment. Plant fresh weight and dry weight were predicted for each plant in the second, third and fourth experiment for each day using the data derived from the daily images. As plants were destructively sampled throughout these experiments, the number of replicates for the images decreased over time. Exact numbers of individual plants used for each distinct DAE can be found in supplemental tables S8-S11 and range between 29 and 366 for the Control treatment, 61 and 370 for the V5-Drought treatment and 27 and 137 for the V12-Drought treatment. By 62 DAE the mean silking age of Control plants had passed and the few (15) remaining Control plants were behind developmentally, not accurately representing the population, so biomass estimation results were no longer representative and were not analyzed further.

Fresh and dry weight accumulation rates were calculated for each day in the second, third and fourth experiment by dividing the difference in weight between that day and the previous day over the difference in time between the moment of imaging and the previous moment of imaging. A cubic smoothing spline (smooth.spline function, R stats package (R Core Team, 2020)) was fitted to the fresh and dry weight accumulation rates of each treatment. The values of the smoothing splines for the fresh and dry weight accumulation rates were divided by their maximal value in the Control treatment in order to normalize the data to a scale of 0-1 for comparison with the organ growth rates.

Fresh weight, dry weight and their accumulation rates were analyzed using identical methods. Data were grouped per treatment and per plant age, expressed as days after emergence (DAE). Normality of distribution and equality of variance between treatments were tested for each day. These conditions were not met in many days, so we chose to use the non-parametric Mann-Whitney U test. These tests were performed for each day to test if the distribution for the Control treatment on the one hand and the distributions for the V5-Drought or V12-Drought treatment on the other hand were significantly different. The resulting p-values were corrected for multiple testing using the Holm-Bonferroni method (Holm, 1979).

Supplemental Method S5. Experimental setup

Experiments were conducted in the Phenovision plant phenotyping platform (supplemental Fig. S1). Phenovision is a conveyor belt-based platform, composed of 14 lines with 28 individual positions on each line, leading to a total capacity of 392 pots. The first and final line and the first and final plant on each line were treated as border plants, to eliminate border effects. In total, there were 80 border positions and 312 experimental positions. Plants on the platform were watered daily. Watering was automated through the use of three weighing-watering stations, which supplied water to the pots until they reached a predetermined gravimetric soil water content, allowing us to subject the plant to specific watering/drought treatments. Plants on the platform were imaged daily. The platform was equipped with three imaging systems: a top-view thermal infrared system, an RGB imaging system (combining two sideview cameras with one top-view camera) and a hyperspectral imaging system, composed of two top-view scanners, capable of imaging in the range from 400nm to 2500nm. In this study only the data from the RGB imaging system is considered. Four experiments were conducted in the periods March-May and September-November in the years 2015 and 2016. Additionally, leaf 4 measurements were taken during a small trial occurring in February 2015.

Supplemental Method S6. Environment

All experiments took place in the Phenovision greenhouse, which features automated high-pressure sodium vapor grow lights, heating, humidification and ventilation. In all experiments daytime was from 06:00-22:00. During this period the photosynthetically active radiation depended on solar radiation (mean of 250 µmol photons*m^{-2}* s^{-1}) but was maintained above 150 µmol photons*m^{-2}* s^{-1} to obtain a 16 h day length. In all experiments, the climate was set to a mild 'seedling' scheme from the start of the experiment until the first plants reached the V5-stage, at which point the climate settings were changed to a harsher setting. During the seedling stage of the first experiment daytime temperatures were set to 23°C and nighttime temperatures to 22°C, with daytime relative humidity set to 65% (corresponding to vapor pressure deficit (VPD) ~1 kPa) and nighttime relative humidity to 55% (corresponding to VPD ~1.2 kPa). During the seedling stages in the later experiments, daytime and nighttime temperatures were also 23°C and 22°C respectively but VPD was set to ~1.5 kPa.

After the first plants reached the V5-stage, the system was set to provide a gradual diurnal gradient, with temperature set to range between 22°C and 28°C and VPD set to range between 0.95 kPa and 2kPa, with the highest values for temperature and VPD to be reached in the afternoon. Photosynthetically active radiation, relative humidity, VPD and temperature were continuously monitored above the growth zone of the platform by four environmental monitoring stations containing an SKH 2053 humidity and temperature sensor, and a PAR SKL 2625 sensor (Skye Instruments, UK).

Supplemental Method S7. Leaf growth analysis

During the first experiment, daily leaf length measurements during the growth of leaves 6, 9, 12, 15 and 18 resulted in leaf length profiles over time, which were analyzed to determine LER, LED, FLL and the timing of growth. An issue occurring during the growth period of leaves 12, 15 and 18 was that the stem below the node of these leaves was elongating during the growth period of these leaves. As leaf length was measured from the soil to the tip of the leaf, part of the leaf growth measured this way was in fact stem growth. The leaf length data was corrected for this by subtracting the stem length up to the node concerned from the leaf length and the growth functions were fitted to the corrected leaf length data.
Two growth functions were fitted to the resulting leaf length profiles: a beta-sigmoid growth function (eq. 1) (Yin et al., 2003; Voorend et al., 2014) and a three-piece linear growth function (eq. 2) (supplemental Fig. S3B; supplemental Fig. S4A-C) :

$$\begin{aligned} y &= a * \left(1 + \frac{t_e - t}{t_e - t_m}\right) * \left(\frac{t}{t_e}\right)^{\frac{t_e}{t_e - t_m}} && t < t_e \\ y &= a && t \geq t_e \end{aligned} \quad (1)$$

with y the leaf length, t the plant age, a the final leaf length, t_e the plant age at the end of growth and t_m the plant age at the moment of maximal growth.

$$\begin{aligned} y &= 0 && t \leq t_1 \\ y &= m * \left(\frac{t - t_1}{t_2 - t_1}\right) && t > t_1 \,\&\, t < t_2 \\ y &= m && t \geq t_2 \end{aligned} \quad (2)$$

with y the leaf length, t the plant age, m the final leaf length, t_1 the plant age at the start of growth and t_2 the plant age at the end of growth.

The beta-sigmoid growth function was used to determine the LER over time, the maximal LER and the final leaf length. Mean LER, start and end of leaf growth and LED were determined from the three-piece linear function. The number of independent plants measured per treatment and leaf rank was 12 for leaf 4 under Control conditions, 74 for leaf 6 under Control conditions and 20 for leaves 9, 12, 15 and 18 under Control and V5-Drought conditions and leaves 15 and 18 under V12-Drought conditions. Fitting the growth curves was not successful for all replicates though and the resulting numbers of fitted curves was reduced for leaves 4 and 18 under Control conditions (n=11 and n=18), leaves 15 and 18 under V5-Drought conditions (n=18 and n=19, respectively) and leaf 18 under V12-Drought conditions (n=15). Full details can be found in supplemental table S1. For leaf 9 in the V5-Drought treatment and all observed leaves in the V12-Drought treatment, start of leaf growth occurred before the start of the drought treatment and the resulting values for the Control treatment were used. The mean and standard deviation of the measurements of interest were determined per treatment and leaf. Leaf emergence from the whorl was observed separately for all plants in the experiment, leading to a much higher number of replicates, but was otherwise handled the same as the measurements originating from the growth curves. All variables were tested for normality and equal variance using the Shapiro-Wilk normality test and F-test for equality of variances. If conditions for normality and equal variance were met, the drought treatment was compared to the Control treatment using a two-sample t-test, otherwise the Wilcoxon-Mann-Whitney test was used. The threshold for statistical significance was set at $p<0.05$.

To compare leaf growth to the other organs, the cumulative LER was calculated for each treatment at each time point as the sum of the LER of leaves 4, 6, 9, 12, 15 and 18, which were determined as the first derivatives of the respective fitted beta-sigmoid growth curves. A cubic smoothing spline was fitted to the cumulative LER of each treatment, to compensate for the discontinuous nature of the cumulative LER caused by the lack of data for leaves between the measured leaves. The smoothed curve functioned as a proxy for the true cumulative LER, which would have had a similar shape but much higher values. Finally the values of the smoothing splines for cumulative LER were divided by the maximal value in the Control treatment in order to normalize the data to a scale of 0-1 for comparison with the growth of the other organs and biomass accumulation rate. All analyses were performed in R version 3.6.3 (R Core Team, 2020).

Supplemental Method S8 Stem growth analysis

In the first experiment, the height of nodes 9, 12, 15, 18 and 20 was determined for plants destructively sampled after the V9-stage. In total these node height measurements were taken for 119 plants (48 Control plants, 49 V5-Drought plants and 22 V12-Drought plants, the latter in combination with 15 control treated plants sampled at V12-stage or earlier). Full details can be found in supplemental table S2.

This resulted in measurements of stem length up to a certain node or, by taking the difference between two nodes, measurements of stem fraction length (the length of a group of internodes) over plant age. Both beta-sigmoid (eq.1) (Yin et al., 2003; Voorend et al., 2014) and three-piece linear growth functions (eq.2) (see materials and methods S6: leaf growth analysis) were fitted to these measurements over time (supplemental Fig. S3A; supplemental Fig. S4D-F). The beta-sigmoid growth function was used to determine the elongation rates over time, the maximal elongation rates and the final length. Average elongation rate, start

and end of growth and elongation duration were determined using the three-piece linear growth function. 95% CI were calculated for these measurements using a bootstrap method: data was separated into treatments and for each treatment 10,000 new datasets of the same size as the original dataset were sampled with replacement. For each of these datasets, the beta-sigmoid and three-piece linear growth functions were fitted to the length data and parameters were derived, resulting in a distribution of parameter values for each treatment. As these distributions were often skewed, a bias-corrected and accelerated 95% CI (Zhou et al., 2011) was determined. If the 95% CIs in a comparison do not overlap, this indicates a significant difference at $p<0.05$. If they do overlap, a significant difference at $p<0.05$ is still possible but harder to establish (Cumming and Finch, 2005). In this case, the differences were assessed between the first 1,000 bootstrap parameter values for the Control treatment and the first 1,000 bootstrap parameter values for each of the drought treatments, leading to a distribution of 10^6 difference values. The bias-corrected and accelerated 95% CI was determined for these difference values. If 0 was not included within the CI, the difference was significantly different from 0 at the $p<0.05$ level.

The growth functions were used to estimate the average stem length up to nodes 9, 12, 15, 18 and 20 at each day for each treatment. These estimates were used to correct the 'full leaf length' (from soil to leaf tip, see above). The height of node 20 was used as a proxy for whole stem length and the first derivative of the beta-sigmoid growth function fitted to the height of node 20 over time was used as a measurement of stem elongation rate. In order to compare stem growth to the growth of the other organs, the values of stem elongation rate were divided by the maximal value in the Control treatment in order to normalize the data to a scale of 0-1 for comparison with the growth of the other organs and biomass accumulation rate.

Supplemental Method S9. Ear growth and development

Ear development was analysed by destructive point measurements throughout the four main experiments (289 plants in total, 119 Control, 116 V5-Drought and 54 V12-Drought. The values of 15 Control plants younger than V12-Stage were added to the V12-Drought data to reach 69 data points for V12-Drought (supplemental data S2). The length of these ear samples was measured and the number of ear rows and spikelets per row were determined using a stereomicroscope or by the naked eye, depending on the size of the sample. Not all traits could be measured for each ear, final sample sizes for ear length were 97, 101 and 69 for Control, V5-Drought and V12-Drought conditions, respectively. Sampling continued until silking, at which point leaf and stem growth had concluded and only the ear was still growing. As ear growth occurred hidden within the husk leaves, it was not possible for the RGB imaging system to track ear growth. The experiment was therefore concluded at silking, allowing for other experiments to make full use of the imaging capabilities of the platform. Because the ear data did not comprise the entire growth period, it was not possible to fit a sigmoid growth function to the ear length measurements and an exponential growth function (eq.3) was fitted instead (Supplemental Fig. S4G; Supplemental Fig. S3C). Ear growth rate was determined as the first derivative of this function. Ear growth rate was also normalized to a scale of 0-1 by dividing by the highest value in the Control treatment for comparison with the growth of the other organs and biomass accumulation rate.

$$y = (1 + a)^{t - t_0} - 1 \quad (3)$$

with *y* the ear length, *t* the current plant age, t_0 the plant age at the start of ear growth and *a* the ear length at the start of ear growth.

For each treatment, a 95% CI was calculated for the time point at which the ear elongation rate reached 0.1 mm/h (or 2.4 mm/day) using a bootstrap method. Data were separated into treatments, for each treatment 10,000 new datasets of the same size as the original dataset were sampled with replacement. For each of these datasets the exponential growth function (eq. 3) was fitted to the length data and the time point at which the ear elongation rate reached the threshold was determined from the first derivative, resulting in a distribution of estimates for each treatment. The 2.5% and 97.5% quantile were taken as the limits of the 95% CI. If the 95% CIs of different treatments did not overlap, this indicated a significant difference at $p<0.05$. Full details can be found in supplemental table S4.

Plants older than the average age at which their treatment reached an ear elongation rate of 0.1 mm/h were considered to have reached their final spikelet number. This resulted in a dataset containing 181 plants: 74 Control, 66 V5-Drought and 41 V12-Drought measurements for ear rows and 73 Control, 66 V5-Drought and 41 V12-Drought and for spikelets per row and spikelets per ear (supplemental data S4). Data for husk leaf emergence, silking, tassel emergence and anthesis were observed independently, not only on sampled plants. Full details can be found in supplemental table S3 and supplemental data S3. The mean and standard deviation were determined per treatment for the final ear traits and the developmental timing observations. They were tested for normality and equal variance using the Shapiro-Wilk normality test and F-test for equality of variances. If conditions for normality and equal variance were met, the drought treatment was compared to the Control treatment using a two-sample t-test, otherwise the Wilcoxon-Mann-Whitney test was used.

Supplemental Method S10. RGB imaging setup and plant pixel segmentation

The RGB imaging setup was made up of three RGB cameras, located in a closed-off cabin, to eliminate light from outside the cabin to enter. Plants were positioned on a rotating lift which allowed for imaging at different angles by the three RGB cameras: one top-view camera and two sideview cameras, one at 0° and another at -30°, relative to the sightline perpendicular to the background screen (supplemental Fig. S11A). The top-view image was taken at a lift rotation of 0°, while the sideview images were taken at lift rotations of 0°, 60° and 120° leading to a total of 6 sideview images, at angles -30°, 0°, 30°, 60°, 90° and 120°, capturing an almost hemispheric view of the maize plant. All three cameras were Allied Vision Technologies Prosilica GE4000C (Allied Vision Technologies GmbH, Germany) 11 megapixel cameras equipped with a Canon EF 24 mm f/1.4L II USM lens (Canon Inc., Japan). Images were stored in JPEG format with a resolution of 2673*4009 pixels.

The camera setup was calibrated using chessboard calibration (Gábor, OpenCV: Camera calibration with OpenCV), where a chessboard image with known dimensions was imaged at various positions in order to determine the relation between measurements in pixel values and their real-world values. In applications based on 2D images, these calibrations were determined for the average location of a plant. For the S0 camera (sideview at 0°, 60° and 120°) this value was 0.674mm/pixel, for the S1 camera (sideview at -30°, 30° and 90°) this value was 0.671mm/pixel. For the T0 top-view camera, the resolution was determined at half the plant height, using the formula $resolution = 1.507108 * \frac{distance}{4008\ pixel}$ with *distance* the

distance between the camera and the point at half plant height. This distance was calculated based on the height of the lift and the height of the plant derived from the sideview images.

Plant pixel segmentation was based on a background subtraction followed by support vector machine classification using Mahalanobis distance (Bradski and Kaehler, 2008). The support vector machine classification model was trained with manually segmented data, classifying pixels into classes such as “plant”, “soil”, “background_white” and “background_other”. After plant segmentation the convex hull of the plant was determined and parameters such as number of pixels, amount of edges of the convex hull, the area within the convex hull, the height and the width of the convex hull were determined. These values were converted to real world values using the camera calibration, converting the number of pixels into plant projected area and the convex hull height of the sideview images into plant height.

Literature cited

Bradski GR, Kaehler A (2008) Learning OpenCV: computer vision with the OpenCV library. O’Reilly, Sebastopol, CA, USA

Cumming G, Finch S (2005) Inference by Eye: Confidence Intervals and How to Read Pictures of Data. Am Psychol **60**: 170–180

de Myttenaere A, Golden B, Le Grand B, Rossi F (2016) Mean Absolute Percentage Error for regression models. Neurocomputing **192**: 38–48

Gábor B OpenCV: Camera calibration With OpenCV. https://docs.opencv.org/3.4/d4/d94/tutorial_camera_calibration.html

Holm S (1979) A Simple Sequentially Rejective Multiple Test Procedure. Scand Stat Theory Appl **6**: 65–70

Voorend W, Lootens P, Nelissen H, Roldán-Ruiz I, Inzé D, Muylle H (2014) LEAF-E: a tool to analyze grass leaf growth using function fitting. Plant Methods **10**: 37

Winsor CP (1932) The Gompertz Curve as a Growth Curve. Proc Natl Acad Sci USA **18**: 1–8

Yin X, Goudriaan J, Lantiga EA, Vos J, Spiertz HJ (2003) A Flexible Sigmoid Function of Determinate Growth. Ann Bot **91**: 361–371

Zhou X-H, Obuchowski NA, McClish DK (2011) Appendix B: Jackknife and Bootstrap Methods of Estimating Variances and Confidence Intervals. Statistical Methods in Diagnostic Medicine. John Wiley & Sons, Inc., Hoboken, NJ, USA, pp 477–480

Supplemental results

Supplemental results S1. Treatment specific biomass modeling

Prediction accuracy was treatment dependent. Most noticeably there was an overestimation of the fresh weight of fully grown V5-Drought treated plants and an underestimation of the fresh weight of fully grown Control treated plants. The accuracy for the other treatments was intermediate, ranging from a tendency to overestimate for the V5-Mild treatment, to a tendency to underestimate for the V12-Drought and V12-Mild treatments. This tendency to overestimate the fresh weight of drought-treated plants was caused by a difference in underlying fresh weight density (g/mm^2) between the different treatments (supplemental Fig. S5A). By separately fitting a simplified linear model with only mean sideview projected area as a factor to each treatment, the fresh weight density was estimated as 0.001495 g/mm^2 for V5-Drought and 0.001846g/mm^2 for Control (supplemental table S6). The simplified models removed the issues of treatment specific under- and overestimation, but these models were still further improved by adding the top-view projected area (supplemental Fig. S6; supplemental table S7).When we fitted the full model that also contained top-view projected area and the interaction term to the individual treatments, the fresh weight density estimates for sideview projected plant area became almost equal (0.001513 g/mm^2 and 0.001528 g/mm2 for Control and V5-Drought, respectively). There was a difference in the interaction factor, which is $4.210*10^{-10}$ g/mm^4 for Control and $-7.459*10^{-10}g/mm^4$ for V5-Drought, indicating that the fresh weight density on the sideview image for the V5-Drought treated plants decreased with increasing top-view projected plant area. Plant water content and stem-to-leaf ratio were two factors likely to play a major role in plant density. Here, the lower fresh weight density was not due to differences in plant water content, as plant water content was higher for fully-grown V5-Drought treated plants than for fully-grown Control treated plants (supplemental Fig. S7A). Stem biomass did represent a lower fraction of total shoot biomass in fully grown V5-Drought treated plants compared to fully grown Control plants, both for fresh and dry weights (supplemental Fig. S7B-C). Comparing the density of fresh weight per unit of mean sideview projected plant area to the fraction of stem fresh weight over leaf fresh weight in fully grown plants revealed a significant positive correlation (Pearson's r= 0.62, $p<0.05$), with the V5-Drought treated plants at the low end of the spectrum for both density and stem/leaf ratio (supplemental Fig. S7D). Plant top-view projected area increased during the period of leaf growth, but did not increase during the period of stem growth, as the stem was not visible in the top-view image, while plant sideview projected area kept increasing throughout shoot growth (supplemental Fig. S8A-B). Thus, it was possible that the top-view projected area functioned as a proxy for development, as a plant with higher top-view projected area was more likely to have started stem growth. The relationship between sideview projected plant area, top-view projected plant area, the predicted aboveground fresh weight and the path plants took through this area is shown in supplemental Fig. S8C-E. For plants of all treatments, sideview and top-view projected plant area were strongly correlated (Pearson's r=0.96, p<2.2*10-16 for all points before 40 DAE (n=18638)) at the beginning of growth, as only leaves, which were visible in both views, were growing. Around 40 DAE this correlation was strongly reduced (Pearson's r=0.27, p<2.2*10-16 for all points of 40 DAE or later (n=7921)), as stem growth had accelerated in all treatments. From then the top-view projected area was determined by the leaf area, while sideview projected area was determined by both leaves and stem. In all three treatments, the approximately vertical contour lines indicated the larger role of the sideview projected area in predicting plant fresh weight.

A linear relationship existed between fresh weight and sideview projected plant area, while there was a quadratic relationship between dry weight and sideview projected plant area (supplemental Fig. S5A-B). This meant that in plants with a higher sideview projected area, each mm^2 represented more dry weight than a mm^2 in a plant with a lower sideview projected area. Plants with a higher sideview projected area had, in fact, a higher dry weight density (g/mm^2), while the fresh weight density (g/mm^2) was independent of the sideview projected area. This indicated that an increase in sideview projected area was coupled with an increase in plant dry matter content (dry weight/fresh weight) (supplemental Fig. S5C).

Top-view projected plant area contributed to the fresh weight prediction while it was not retained for the dry weight prediction model (supplemental table S5). For small plants with a top-view projected plant area of less than 0.2 m^2, there was a significant correlation between top-view projected plant area on the one hand and FW or DW on the other hand (Pearson's r value of 0.82 for FW, $p<0.05$ and 0.72 for DW, $p<0.05$) (supplemental Fig. S5D-E). For plants with a top-view projected plant area larger than 0.2 m^2, there was still a significant correlation with FW but not with DW (Pearson's r value of 0.36 for FW, $p<0.05$ and 0.09 for DW, $p=0.29$). Comparing the fraction of stem fresh weight in total aboveground fresh weight to the mean sideview projected plant area revealed that the stem contribution to aboveground fresh weight ranged from less than 5% for plants with mean sideview projected plant areas of 0.05 m^2 (plant age of 20-25 DAE, before the start of stem elongation), to 30%-40% for plants with mean sideview projected plant areas of more than 0.2 m^2 (plant age mostly above 40 DAE, supplemental Fig. S5F). Thus, as expected, stems were a major factor in determining fresh and dry weights of older plants.

Treatment-specific models were evaluated using LOOCV. Treatment-specific datasets for the V5 and V12 treatments contained Control measurements taken before the respective V5- and V12-Stages (16.3 and 39.6 DAE, respectively). This led to a dataset of 69 observations for the Control dataset, 68 observations for the V5-Drought dataset (5 Control treatment, 63 V5-Drought treatment), 66 observations for the V5-Mild dataset (5 Control treatment, 61 V5-Drought treatment), 61 observations for the V12-Drought dataset (38 Control treatment, 23 V12-Drought treatment) and 58 observations for the V12-Mild dataset (38 Control treatment, 20 V12-Mild treatment) (supplemental data S5). Simple models contained only the intercept and the mean sideview projected area term and full models featured mean sideview projected area and top-view projected area, as in eq. 4.

Supplemental discussion

Supplemental discussion S1. Biomass modeling

Our linear 2D image-based plant biomass prediction models reached similar prediction performance as models used in other studies. Klukas et al. (2014) report Pearson's r values of 0.9552 and 0.8370 for correlations between fresh weight or dry weight and digital volume, and these are derived from projected plant area on RGB images for maize plants and a root-mean-squared-relative-error (RMSRE) of 21.7% for fresh weight when testing their prediction model on an independent data set. Ge et al. (2016) find R^2 values for the correlation of fresh weight and dry weight with plant pixel count of 0.993 and 0.952 for young plants up to 26 days after sowing. Cabrera-Bosquet et al. (2016) report an R^2 value of 0.972 for the prediction of shoot fresh weight for plants between 15 and 50 days after emergence. Zhang et al. (2017) report an R^2 of 0.98 and MAPE of 12.31% for fresh weight predictions and an R^2 of 0.97 and MAPE of 15.85% for dry weight predictions on plants from seedling to tasseling stages.

Our fresh-weight prediction model slightly outperformed the dry weight prediction model. This might be related to the simpler relationship between plant fresh weight and the plant projected area on the sideview images compared with that of dry weight. For plant fresh weight this was a linear relationship, while this was a quadratic relationship for plant dry weight. The quadratic relationship of plant dry weight to projected area might be a representation of a developmental factor: as plants grow older, their projected area increased and at the same time, the dry matter as a fraction of total mass increased, thus the dry weight increased more than linearly with the plant projected area. In this case, plant projected area functioned as a proxy for both plant volume and plant age, which determined the dry weight per unit of projected plant area. This can be contrasted to the approach used by Golzarian et al. (2011), who improve a linear dry weight prediction model for wheat by including density (labelled as Plant Specific Weight) as a linear function of plant age.

The fresh weight of fully grown V5-Drought treated plants was consistently overestimated by the biomass prediction model, while there was a tendency to underestimate the fresh weight of the fully grown Control plants. Biases in biomass estimation models due to genotype and treatment variations have been reported before (Golzarian et al., 2011; Ge et al., 2016; Liang et al., 2018) and are attributed to differences in density, i.e., the relationship between plant projected area and biomass. Golzarian et al (2011) consider density as a function of plant age, while Chen et al. (2018) take a machine-learning approach and include traits related to plant physiology, which may explain differences in density, in addition to plant morphology and size. We found a correlation between plant fresh weight density and the fraction of stem fresh weight over leaf fresh weight in fully grown plants. Our hypothesis was that stem biomass was much more compact, in terms of projected plant area per unit of biomass, than leaf biomass, meaning that two plants with the same projected plant area could differ greatly in biomass, depending on the ratio of stem to leaf biomass. A plant with a higher stem/leaf ratio contained more biomass than a plant with a lower stem/leaf ratio with the same projected plant area. In the models trained on the specific treatments, the interaction term accounted for this by reducing the sideview fresh weight density of V5-Drought treated plants with high top-view projected area, a sign they were close to their final leaf area and that stem growth had started. The biomass prediction models could be improved by further separating plant

pixels on the image into pixels attributable to leaves or stem, which would allow stem pixels to carry more weight in both the literal and figurative sense.

Literature cited

Cabrera-Bosquet L, Fournier C, Brichet N, Welcker C, Suard B, Tardieu F (2016) High-throughput estimation of incident light, light interception and radiation-use efficiency of thousands of plants in a phenotyping platform. New Phytol **212**: 269–281

Chen D, Shi R, Pape J-M, Neumann K, Arend D, Graner A, Chen M, Klukas C (2018) Predicting plant biomass accumulation from image-derived parameters. GigaScience **7**: 1–13

Ge Y, Bai G, Stoerger V, Schnable JC (2016) Temporal dynamics of maize plant growth, water use, and leaf water content using automated high throughput RGB and hyperspectral imaging. Comput Electron Agric **127**: 625–632

Golzarian MR, Frick RA, Rajendran K, Berger B, Roy S, Tester M, Lun DS (2011) Accurate inference of shoot biomass from high-throughput images of cereal plants. Plant Methods **7**: 2

Klukas C, Chen D, Pape J-M (2014) Integrated Analysis Platform: An Open-Source Information System for High-Throughput Plant Phenotyping. Plant Physiol **165**: 506–518

Liang Z, Pandey P, Stoerger V, Xu Y, Qiu Y, Ge Y, Schnable JC (2018) Conventional and hyperspectral time-series imaging of maize lines widely used in field trials. GigaScience **7**: 1–11

Zhang X, Huang C, Wu D, Qiao F, Li W, Duan L, Wang K, Xiao Y, Chen G, Liu Q, et al (2017) High-Throughput Phenotyping and QTL Mapping Reveals the Genetic Architecture of Maize Plant Growth. Plant Physiol **173**: 1554–1564